Motor Imagery (MI)-Electroencephalogram (EEG) Decoding Method Based on Multi-modal Temporal Fusion and Spatial Asymmetry

Zhikang YIN[1], Chunjiang SHUAI[2]*

1. School of Physics and Telecommunication Engineering, Shaanxi University of Technology, Hanzhong 723000, China; 2. Trine Engineering Institute, Shaanxi University of Technology, Hanzhong 723000, China

Abstract Deep learning methods have been widely applied in motor imagery (MI)-based brain-computer interfaces (BCI) for decoding electroencephalogram (EEG) signals. High temporal resolution and asymmetric spatial activation are fundamental properties of EEG during MI processes. However, due to the limited receptive field of convolutional kernels, traditional convolutional neural networks (CNNs) often focus only on local features, and are insufficient to cover neural processes across different frequency bands and duration scales. This limitation hinders the effective characterization of rhythmic activity changes in MI-EEG signals over time. Additionally, MI-EEG signals exhibit significant asymmetric activation between the left and right hemispheres. Traditional spatial feature extraction methods overlook the interaction between global and local regions at the spatial scale of EEG signals, resulting in inadequate spatial representation and ultimately limiting decoding accuracy. To address these limitations, in this study, a novel deep learning network that integrates multi-modal temporal features with spatially asymmetric feature modeling was proposed. The network first extracts multi-modal temporal information from EEG data channels, and then captures global and hemispheric spatial features in the spatial dimension and fuses them through an advanced fusion layer. Global dependencies are captured using a self-attention module, and a multi-scale convolutional fusion module is introduced to explore the relationships between the two types of temporal features. The fused features are classified through a classification layer to accomplish motor imagery task classification. To mitigate the issue of limited sample size, a data augmentation strategy based on signal segmentation and recombination is designed. Experimental results on the BCI Competition IV-2a (bbic-IV-2a) and BCI Competition IV-2b (bbic-IV-2a) datasets demonstrated that the proposed method achieved superior accuracy in multi-class motor imagery classification compared with existing models. On the BCI-IV-2a dataset, it attained an average classification accuracy of 84.36%, while also showing strong performance on the binary classification BCI-IV-2b dataset. These outcomes validate the capability of the proposed network to enhance MI-EEG classification accuracy.

Key words Deep learning; Brain-computer interface (BCI); Convolutional neural network (CNN); Electroencephalogram (EEG); Motor imagery (MI)

DOI: 10.19759/j.cnki.2164 –4993.2025.06.018

Brain-computer interface (BCI), an emerging technology that establishes direct communication between the human brain and external devices through electroencephalographic signals, has demonstrated broad application prospects in recent years in fields such as rehabilitation, prosthetic control, and human-computer interaction[1]. The motor imagery (MI) paradigm based on non-invasive electroencephalography (EEG) has become a crucial control strategy in BCI systems, as it can elicit distinguishable neural activity in the motor cortex without requiring external stimuli[2]. Furthermore, it has emerged as a promising technology in non-medical fields, such as virtual reality, gaming[3], and robotic arm control[4 –5]. Despite the significant potential of MI-EEG in clinical and engineering applications, its decoding process still faces several challenges: the inherently low signal-to-noise ratio of EEG signals, strong inter-individual differences, and non-stationarity influenced by factors such as channel layout and electrode contact. These issues make reliable, robust and real-time MI-EEG decoding an unresolved problem.

Traditional MI-EEG decoding methods predominantly rely on manually designed feature extraction and classical classifiers. For instance, Common Spatial Pattern (CSP)[6], one of the most widely used features in brain-computer interface (BCI) systems, aims to identify an optimal spatial filter that maximizes the difference between two classes of EEG signals. Building upon CSP, numerous variants of the CSP method have been proposed to enhance decoding performance[7 –11]. Within the Riemannian geometry-based classification framework, the covariance structure of EEG data is directly utilized as features[12 –14]. Various methods based on wavelets have been applied to extract time-frequency features from EEG[15 –16]. Following feature extraction, classifiers such as Support Vector Machine (SVM) or Linear Discriminant Analysis (LDA) are commonly applied to obtain decoding results. Although these methods have achieved satisfactory performance under specific experimental settings, they heavily rely on prior feature engineering and typically separate feature extraction and classification into two distinct stages, making end-to-end joint optimization challenging. To address this, researchers have recently widely adopted deep learning approaches, particularly Convolutional Neural Networks (CNNs), which possess the capability to learn directly from EEG, enabling automatic extraction of discriminative spatiotemporal features from raw EEG[17 –19]. Architectures such as DeepConvNet and ShallowConvNet employ a two-stage spatial and temporal convolutional input layer structure to integrate

Received: August 25, 2025 Accepted: Ocotober 25, 2025

Zhikang YIN (1999 –), male, P. R. China, master, devoted to research about brain-computer interfaces and deep learning.

* Corresponding author.

feature extraction and classification for processing of EEG data[20]. Lawhern *et al.* [21] proposed EEGNet, which uses the channel size as a deepwise convolution kernel to extract spatial information, achieving notable progress in MI-EEG decoding and demonstrating the advantages of data-driven representation learning[22].

However, the limitations of traditional CNNs are increasingly evident. The receptive fields of convolutional kernels are typically confined to local spatiotemporal regions, making it difficult to adequately capture long-range temporal dependencies in EEG signals. Since EEG serves as a kind of highly dynamic time-series signal, time convolution on a single scale is often insufficient to cover neural processes in different frequency bands and duration scales. To address these shortcomings, attention mechanism-based models, known for their ability to capture long-term dependencies and widely applied in image processing[23] and natural language processing[24], have been increasingly introduced into the EEG field to enhance decoding performance. The Attention-based Temporal Convolutional Network (ATCNet) integrates multi-head self-attention with temporal convolutional networks[25] to highlight the most critical features. The EEG Convolutional Decoder (Conformer) employs convolutional modules for feature extraction and subsequently passes the features to self-attention modules to capture global dependencies[26]. While these studies have improved decoding performance, they have not yet fully combined attention mechanisms and temporal features. On the other hand, numerous neuroscience studies have demonstrated that the left and right cerebral hemispheres exhibit varying degrees of activation differences during motor imagery and other tasks[27]. However, existing MI-EEG decoding methods overlook the complementary role between local hemispheric representations and global spatial representations. Therefore, this paper argues that spatial feature extraction should simultaneously account for the complementarity between local hemispheric representations and global representations.

Based on the aforementioned challenges, we argue that an effective MI-EEG decoder must meet two key requirements. First, it must adequately capture multimodal temporal information and capture global dependencies across time segments with the help of self-attention mechanism. Second, it should incorporate targeted spatial designs to explicitly learn asymmetric patterns between the left and right hemispheres and overall spatial activation, thereby obtaining more physiologically meaningful and discriminative spatial representations. To this end, in this study, a novel end-to-end deep learning network was proposed to capture multimodal temporal features and asymmetric spatial features in MI-EEG decoding tasks. Design idea of the network architecture: In the temporal dimension, it retains dual-modal temporal extraction based on median pooling and standard deviation pooling, while employing a shared self-attention module to learn global dependencies of the two types of temporal features. In the spatial dimension, it introduces a neurophysiologically inspired asymmetric spatial module (comprising parallel designs of hemispheric convolutional kernels and global convolutional kernels) to explicitly extract hemispheric and global spatial representations. Subsequently, a multi-scale convolutional fusion module integrates temporal and spatial features to obtain more discriminative representations. In addition, to enhance the model's generalization capability, a data augmentation strategy based on signal segmentation and recombination is adopted, aiming to mitigate overfitting issues caused by limited sample size and signal non-stationarity.

Network Overview

This section provides a detailed introduction to the proposed network and data augmentation method. The architecture of the proposed network is illustrated in Fig. 1.

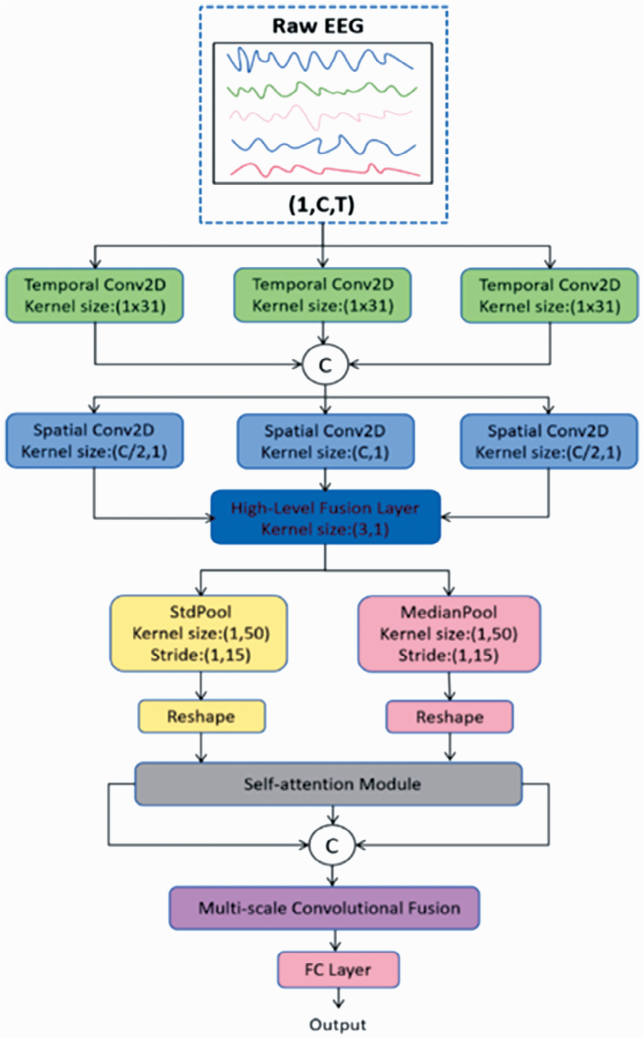
Raw EEG
(1,C,T)
Temporal Conv2D Kernel size:(1x31)
Temporal Conv2D Kernel size:(1x31)
Temporal Conv2D Kernel size:(1x31)
C
Spatial Conv2D Kernel size:(C/2,1)
Spatial Conv2D Kernel size:(C,1)
Spatial Conv2D Kernel size:(C/2,1)
High-Level Fusion Layer Kernel size:(3,1)
StdPool Kernel size:(1,50) Stride:(1,15)
MedianPool Kernel size:(1,50) Stride:(1,15)
Reshape
Reshape
Self-attention Module
C
Multi-scale Convolutional Fusion
FC Layer
Output

Fig. 1 Overall architecture of the proposed network

Proposed network architecture

The network consists of four components in an end-to-end process: a feature extraction module, a self-attention module, a multi-scale convolutional fusion module, and a classification module.

Feature extraction module The feature extraction module is designed to extract discriminative features from raw EEG signals. First, the raw EEG data is expanded along the channel dimension in one dimension. Subsequently, a multi-scale temporal convolutional filter bank is employed to perform local modeling of the signal in time series, capturing dynamic features across different time ranges. Compared with a single convolutional kernel, the multi-scale temporal convolutional filter bank enables more comprehensive extraction of temporal features and enhances the model's sensitivity to different frequency bands and rhythmic components[28]. In the architecture propsed in this study, a filter bank composed of three two-dimensional convolutional layers with different kernel sizes is selected, with kernel sizes set to (1,15), (1,31), and (1,63), respectively, while the output size remains unchanged along the temporal dimension. The three outputs of the temporal filter bank are concatenated along the convolutional channel dimension. They are further normalized through a batch normalization layer to adjust the weight values[29]. Following batch normalization, spatial dimension features in the EEG signals are further considered. Existing studies have demonstrated that motor imagery tasks elicit activation patterns in specific regions of the cerebral cortex, and these activations often exhibit significant inter-hemispheric differences. Based on this neuroscientific evidence, the design concept of spatial asymmetry is introduced during spatial modeling in the architecture propsed in this study. In specific, unlike traditional methods that directly apply global convolution across all channels, the electrode channels are divided into left and right hemispheres and perform convolutional operations separately, aiming to capture inter-hemispheric differences and improve classification performance. Spatial convolution is applied to the electrode channels of the left and right hemispheres respectively, with the kernel size set to half of the global convolution kernel, ensuring coverage of the channel range of each hemisphere. After feature extraction from the left and right hemispheres, a spatial convolutional layer with a kernel size of (C, 1) is used to learn representations of interactions between different electrode channels. This approach not only captures local asymmetry, but also integrates overall inter-hemispheric connections, thereby forming more discriminative spatial feature representations. Here, C denotes the number of electrode channels in the EEG data. To fuse spatial information from global and hemispheric sources, an advanced fusion layer is employed to integrate the three types of spatial information. In specific implementation, a one-dimensional convolution (with a kernel size of 3 × 1) is applied to the output of the asymmetric spatial layer to fuse information along the spatial dimension. Subsequently, a batch normalization layer is employed to enhance the training process and mitigate overfitting. Furthermore, the Exponential Linear Unit (ELU)[30] is adopted as the activation function.

To aggregate temporal information and reduce computational complexity, traditional methods often employ average pooling to reduce computational complexity and compress feature dimensions[20-21,26,31]. However, average pooling only reflects the central tendency of the signal and is insufficient for fully characterizing the complex dynamic features of non-stationary EEG signals, making it inadequate for discriminative feature extraction in MI-EEG decoding[32-34]. Standard deviation pooling, on the other hand, characterizes the intensity of signal fluctuations and can capture the variation amplitude of rhythmic activities in different motor imagery tasks, which is of significant importance for enhancing classification discriminability[35-36]. Therefore, both median pooling and standard deviation pooling are performed along the temporal dimension with a kernel size of (1,50) and a stride of (1,15). The features obtained from median pooling and standard deviation pooling can be treated as representations of different time segments. Finally, the electrode channel dimensions are compressed, and the convolutional channel dimensions are transposed with the time dimension before output. Subsequently, the output feature maps need to be reshaped. This reshaping operation helps map the feature maps of each time segment into sequence tokens, facilitating their input into the self-attention module to capture global dependencies.

Self-attention module In MI-EEG decoding, significant temporal dependencies often exist between different time segments, which are closely related to the rhythmic activities of the brain during task execution. Therefore, capturing the global correlations among time-series features is crucial for improving classification performance. In the architecture proposed in this study, a self-attention mechanism is introduced after the feature extraction module to enhance the model's ability to selectively focus on key temporal segments, thereby highlighting discriminative features and suppressing redundant information[37]. The feature extraction module outputs two types of feature representations: median-pooled features and standard deviation-pooled features. Both are subsequently fed into the self-attention module to capture dependencies along the temporal dimension. The self-attention module primarily consists of two components.

The first layer is a multi-head attention mechanism. Self-attention calculates the correlations between features using Query (Q), Key (K), and Value (V). The input feature matrix is linearly projected into Q, K, and V, and the attention weights are computed by scaling the dot product.

$$Attention\ (Q,\ K,\ V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

In formula (1), d_k represents the dimension of the key vector, used for ensuring numerical stability. Unlike single-head attention, multi-head attention (MHA) enables the model to jointly attend to information from multiple subspaces of the representation at various positions[23]. Therefore, the input features are mapped into Q, K, and V matrices through linear transformations. Different learnable weight matrices are used to project Q, K, and V into spaces of dimensions d_q, d_k, and d_v, respectively. Subsequently, attention computation is performed in parallel in each projected space, yielding output representations from multiple attention heads. These outputs are concatenated along the channel dimension and then projected back to the original feature dimension, thereby forming the final attention representation.

$$\begin{aligned} &MHA\ (Q,\ K,\ V) = Concat\ (head_1,\ \cdots,\ head_h)\ W^O \\ &head_i = Attention\ (QW_i^Q,\ KW_i^K,\ VW_i^V) \end{aligned} \tag{2}$$

In formula (2), $W_i^Q \in R^{d_m \times d_q}$, $W_i^K \in R^{d_m \times d_k}$, $W_i^V \in R^{d_m \times d_v}$, and

$W^{O} \in R^{hd_v \times d_m}$. The input feature dimension is evenly distributed to each head: $d_q = d_k = d_v = d_m / h$.

The second layer is a fully connected feed-forward neural network (FFN). Features at each time step are independently fed into the FFN to further enhance nonlinear representation capabilities. The FFN consists of two linear transformations with a GELU activation therebetween, which can be expressed as:

$$GELU\ (x) = \frac{x}{2}\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right) \qquad (3)$$
$$FFN\ (x) = GELU\ (xW_1 + b_1)\ W_2 + b_2)$$

In formula (3), $erf\ (x)$ denotes the Gaussian error function. To ensure training stability and prevent gradient vanishing, layer normalization is applied before the attention layer and the FFN[37-38], and residual connections are employed in both layers[39]. The entire computational process is repeated N times within the self-attention module, where N represents the depth of the self-attention module.

Multi-scale convolutional fusion module To more comprehensively explore the correlations between temporal features of different modalities, a Multi-scale Convolutional Fusion (MCF) module is introduced after the self-attention module. Compared with using a single convolutional kernel to learn the connections between two different features, the MCF achieves multi-scale fusion of cross-modal features by parallelly configuring multiple convolutional kernels. Following the convolutional layers, batch normalization layers and ELU activation are further applied. In this way, the median-pooled features and standard deviation-pooled features are fused together, generating more discriminative features for final classification.

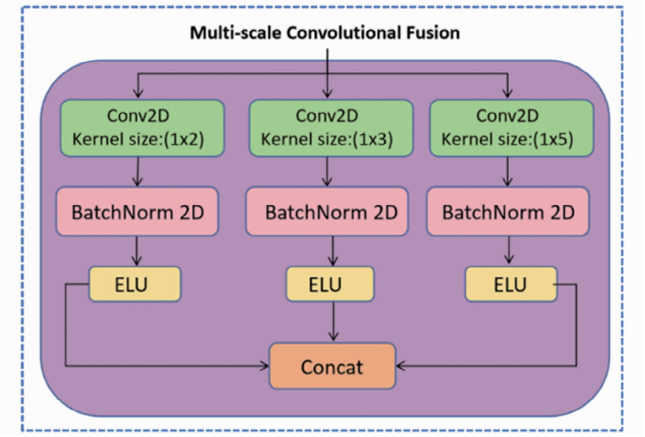

Fig. 2 Architecture of the multi-scale convolutional fusion module

Classification Based on the features extracted above, a classifier is designed to provide the final classification results. A fully connected (FC) layer is used to classify the aforementioned features, and the output is passed through a softmax function to generate prediction probabilities. The label with the highest probability is considered the final result.

Data augmentation

For deep neural network-based MI-EEG decoding, model performance is highly dependent on the quantity and diversity of training samples. However, EEG data collection is time-consuming and resource-constrained, resulting in a relative scarcity of real samples available for training, which can easily lead to overfitting. To address this, a common practice is to generate additional samples from the original training set through data augmentation to enrich the training distribution[20,26,31-32,40]. For the architecture proposed in this study, a data augmentation strategy that combines signal partitioning and smooth transition splicing is adoped. Specifically, each EEG trial belonging to the same class is uniformly divided into "N" _"s" segments along the temporal dimension to ensure each sub-segment has a consistent length. Subsequently, corresponding segments from different trials are randomly selected and sequentially reassembled in chronological order to form new EEG sequences. Unlike the direct splicing methods used in previous approaches, a smooth transition splicing mechanism is introduced at the segment junctions, which better avoids artifacts caused by abrupt changes at segment boundaries, making the generated augmented data more closely resemble real EEG signals in terms of temporal continuity.

Experimental Methods

Datasets

To validate the effectiveness of the proposed model in motor imagery electroencephalogram (MI-EEG) decoding, this study utilized publicly available datasets from the 4th Brain-Computer Interface Competition (BCI Competition IV) organized by Technische Universität Berlin (TU Berlin): Dataset 2a (BCI-IV-2a) and Dataset 2b (BCI-IV-2b)[41]. The datasets feature rigorous experimental design and publicly available protocols, have been widely adopted internationally, and are recognized as a standard benchmark. Their use ensures the comparability of experimental results and the generalizability of research findings, and they have been extensively applied in motor imagery-related brain-computer interface studies. Therefore, in this study, a comprehensive evaluation of the proposed network was conducted on these datasets.

bbic-iv-2a The BCI-IV-2a dataset contains EEG data from 9 healthy subjects. Each subject performed four different motor imagery tasks: imagination of left hand, right hand, both feet, and tongue movements. EEG signals were recorded using 22 Ag/AgCl electrodes at a sampling rate of 250 Hz. During the experiment, data were collected from each subject on two separate dates for training and testing purposes, respectively. Each experimental session consists of 288 trials, with each of the four different motor imagery tasks comprising 72 trials. In this study, the training data were sourced from the first session, while the test data were recordings from the second session.

bbic-iv-2b The BCI-IV-2b dataset also includes EEG data from 9 healthy subjects. Participants were required to perform two different motor imagery tasks: left-hand and right-hand imagination. EEG signals were recorded using three bipolar electrodes (C3, Cz, C4) at a sampling rate of 250 Hz. For each subject, a total of five experimental sessions were conducted. The first two sessions without feedback contained 120 trials each, while the subsequent three sessions with feedback contained 160 trials each. During the experiment, the first three sessions were used for training, while the last two sessions were used for testing. For all datasets, the EEG data for each trial were extracted using the same time window [0, 4] seconds relative to the cue onset. Each trial was treated as a sample, and each sample was represented as a two-dimensional matrix of channel × sample.

Comparative methods

To evaluate the performance of the proposed network, this study compared it with four leading deep learning networks. In experiments, all comparative models were retested and validated on both datasets. Brief descriptions of the comparative models are provided below.

Deep convolutional neural network The Deep Convolutional Neural Network is a deep learning model based on the classical Convolutional Neural Network architecture, which has been demonstrated to perform well in MI-EEG decoding tasks[20].

EEGNet-8, 2 EEGNet-8, 2 is a lightweight neural network specifically designed for EEG signals, particularly suitable for brain-computer interface tasks at a 128 Hz sampling rate[21]. To ensure fairness, all raw EEG data in this study were resampled to 128 Hz.

FBCNet FBCNet extracts discriminative features from multiple frequency bands of EEG signals and utilizes a variance pooling layer to reduce feature dimensions, thereby achieving efficient feature extraction[35]. This model has demonstrated outstanding performance on multiple public MI-EEG datasets.

EEG decoder EEG Conformer is a compact convolutional transformer that integrates convolutional modules with self-attention modules to extract both local and global features from EEG data[26]. This model has demonstrated state-of-the-art performance in MI-EEG decoding tasks.

Experimental details

Default network configuration used in this study: Each 2D convolutional layer had an output channel count of 8, and the kernel sizes of the temporal filter bank were set to different values. The depth (N) of the self-attention module and the attention dimension per head were set to 4 and 8, respectively. Considering the trial duration of 4 s across all datasets and a sampling rate of 250 Hz, the time segment was set to 0.5 s for data augmentation, meaning each time segment consisted of 0.5 s of trials.

This study implemented all experiments using PyTorch[42] and performed training on two NVIDIA RTX 3080Ti GPUs. The models were trained with a cross-entropy loss function, with a maximum of 1 800 training epochs. The Adam optimizer[43] was employed with an initial learning rate set to 0.000 2. To evaluate the decoding performance of different networks, classification accuracy and the Kappa coefficient were used as evaluation indicators[37]. The Kappa coefficient was calculated using following formula:

$$k = \frac{P_o - P_e}{1 - P_e} \tag{4}$$

In the experiments, classification accuracy and the kappa coefficient were used to evaluate the decoding performance of different networks. In formula (4), P_o represents the average classification accuracy, and P_e denotes the classification consistency based on chance.

Results and Analysis

Overall decoding performance comparison

The proposed method was systematically evaluated on two public motor imagery EEG datasets: BCI Competition IV 2a and BCI Competition IV 2b. MI-EEG data represent typical multi-channel time-series physiological signals, which identify subjects' motor imagination categories from EEG signals according to their different motor imaginations during the experiment. The BCI-IV-2a dataset includes four-class motor imagery tasks (left hand, right hand, both feet, and tongue), while the BCI-IV-2b dataset contains two-class motor imagery tasks (left hand and right hand). In this study, the performance of the proposed method was compared with four representative deep learning methods. The overall results are shown in Fig. 3 and Fig. 4. It can be observed that the proposed method achieves the best performance in average classification accuracy on both datasets, demonstrating its strong decoding capability.

On the BCIC-IV-2a dataset, the proposed method achieved an average accuracy of 84.36% and a kappa coefficient of 0.791 4 on 9 subjects, both of which were the highest among all comparative methods. The kappa coefficient is shown in Fig. 5. Compared with Deep ConvNet, EEGNet-8, 2, and FBCNet, the accuracy was improved by 13.38%, 11.87%, and 7.58%, respectively. Even when compared with EEG Conformer, the proposed method maintained an advantage with an improvement of 5.75%. For high-performance subjects such as A03, A07, and A09, the proposed method achieved accuracy rates close to or exceeding 95%, significantly outperforming the comparative methods. These results demonstrate that the introduced multi-scale convolutional fusion module can effectively integrate multi-modal temporal features. By combining spatially asymmetric convolutions with self-attention mechanisms, it exhibits higher classification accuracy in cross-temporal dependencies and spatial asymmetry.

Similarly, on the BCIC-IV-2b dataset, the proposed method also demonstrated leading performance, achieving an average accuracy of 89.98% and an average kappa coefficient of 0.853 0. Compared with EEGNet-8, 2 and FBCNet, the accuracy was improved by 3.73% and 7.23%, respectively. Moreover, it achieved optimal results in 7 out of the 9 subjects. Since this dataset contains only three channels (C3, Cz, C4), the spatial information is relatively limited. As a result, the performance gap between different methods was smaller compared with that observed on BCIC-IV-2a. Nevertheless, under these conditions, the proposed method still maintained a leading position.

From the overall trend, the performance improvement on BCIC-IV-2a was significantly greater than that on BCIC-IV-2b. The proposed method significantly enhanced MI-EEG accuracy by effectively integrating multi-modal temporal and spatial features through the multi-scale convolutional fusion module and capturing cross-temporal dependencies with the self-attention module. Even under conditions of limited channel numbers, the proposed approach maintained stable performance and surpassed existing methods in most subjects, further validating the superiority of the proposed network in multi-subject motor imagery decoding tasks.

The above results demonstrate that the design combining multi-scale convolutional fusion, self-attention mechanisms, and multi-modal feature aggregation can better identify classification patterns in MI-EEG signals, thereby enhancing the reliability of brain-computer interface systems in practical applications.

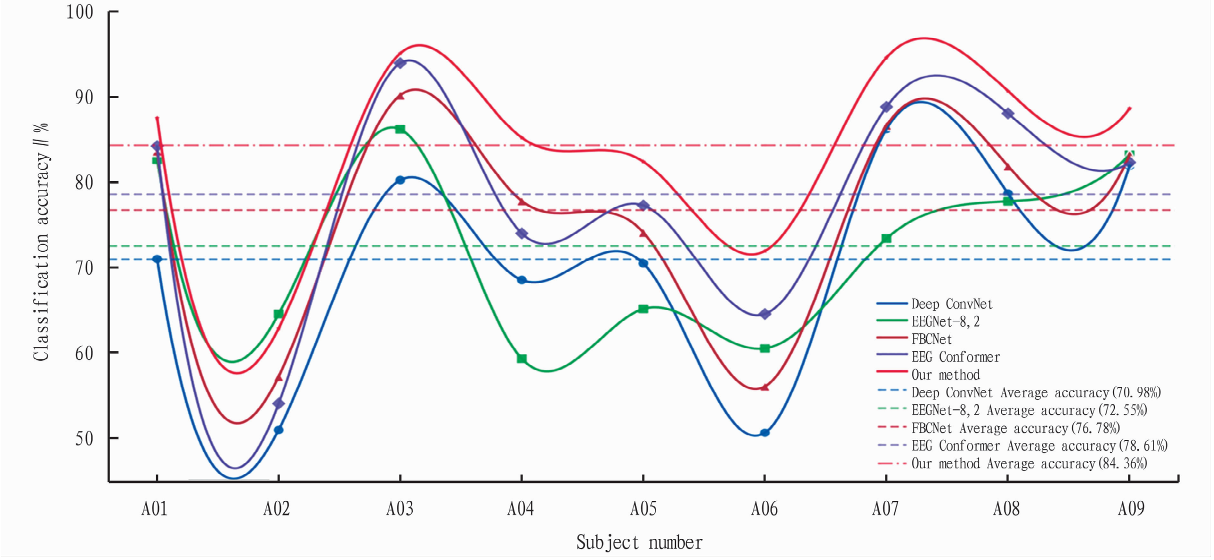

Fig. 3 Comparation of decoding performance on bbic-iv-2a

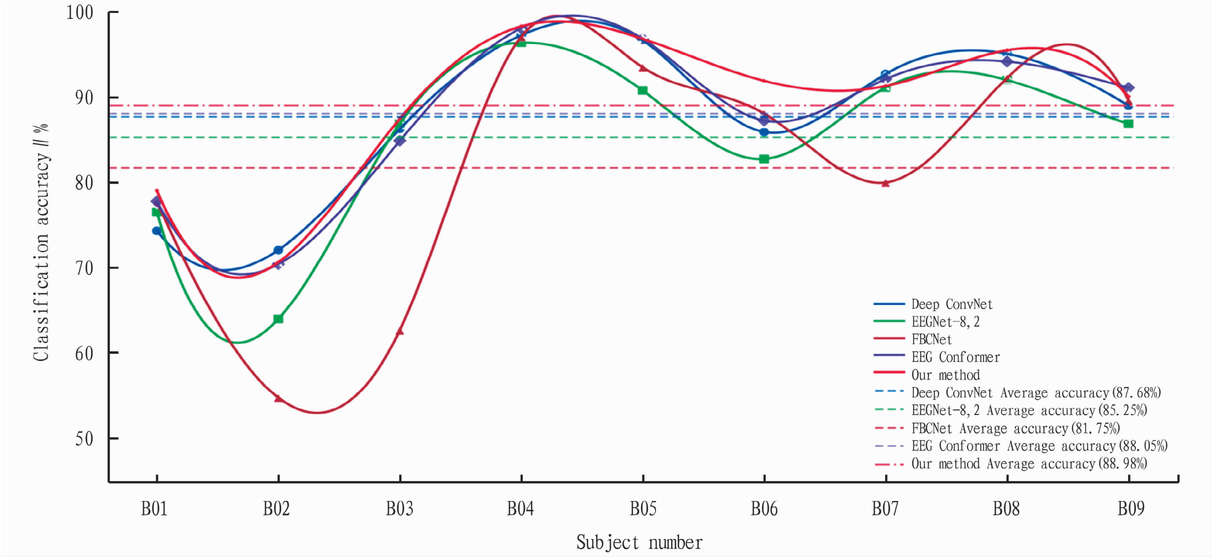

Fig. 4 Comparation of decoding performance on bbic-iv-2b

Ablation experiment analysis

In the proposed network, data augmentation, the self-attention module, and the Multi-scale Convolutional Fusion module are key components. To validate their contributions to the overall classification performance, ablation experiments were conducted on the BCIC-IV-2a and BCIC-IV-2b datasets. Specifically, each module was individually removed, and the resulting changes in classification accuracy on the two public datasets were observed. The results are summarized in Table 1. From the perspective of data augmentation's effect, when the data augmentation module was removed, the classification accuracy on both datasets showed a significant decline. On BCIC-IV-2a, the accuracy dropped from 84.36% to 77.76%, and the kappa coefficient decreased from 0.791 4 to 0.703 5. On BCIC-IV-2b, the accuracy fell from 89.64% to 86.21%, and the kappa coefficient was reduced from 0.853 0 to 0.816 1. These results indicate that the data augmentation strategy can effectively mitigate overfitting issues caused by limited and non-stationary EEG data, thereby significantly enhancing the model's generalization capability. The study further investigated the role of the self-attention module. When the self-attention module was removed, the accuracy on BCIC-IV-2a decreased to 82.85%, and the kappa coefficient dropped to 0.771 3. On BCIC-IV-2b, the accuracy fell to 88.54%, and the kappa coefficient was reduced to 0.847 2. The results demonstrate that the self-attention mechanism effectively captures global dependencies across different time segments, thereby helping the model maintain stable decoding performance in cross-subject and cross-session tasks. When the MCF module was removed, the model's accuracy on BCIC-IV-2a and BCIC-IV-2b decreased to 82.86% and 88.32%, respectively, while the kappa coefficients dropped to 0.771 5 and 0.844 3, respectively. The MCF effectively integrates the complementary information from average-pooled features and variance-pooled features, enhancing the network's ability to capture spatio-temporal characteristics, thereby significantly improving the classification accuracy for MI-EEG. In summary, the purpose of the ablation experiments is to analyze the individual contributions and synergistic effects of each module within the network. Both the self-attention module and the multi-scale convolutional fusion module help enhance the discriminative capability of the learned features, while data augmentation during training improves the model's generalization. The synergistic integration of these three components significantly boosts the classification accuracy of MI-EEG.

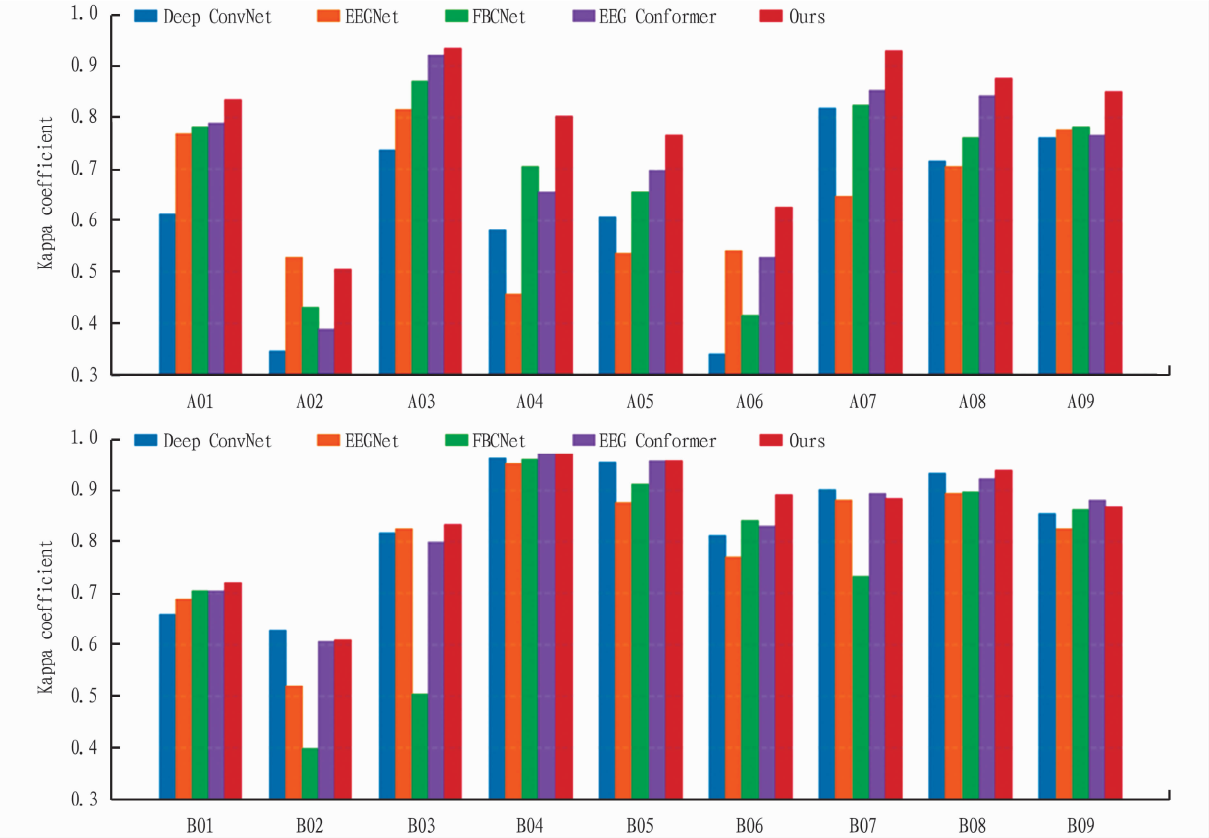

Fig. 5 Comparison of model kappa coefficients on bbic-iv-2a and bbic-iv-2b

Table 1 Ablation study on bbic-iv-2a and bbic-iv-2b

Dataset	Method	Accuracy//%	Kappa
bcic-iv-2a	Our method-w/o data augmentation	77.76	0.703 5
	Our method-w/o self-attention	82.85	0.771 3
	Our method-w/o mcf	82.86	0.771 5
	Our method	84.36	0.791 4
bcic-iv-2b	Our method-w/o data augmentation	86.21	0.816 1
	Our method-w/o self-attention	88.54	0.847 2
	Our method-w/o mcf	88.32	0.844 3
	Our method	88.98	0.853 0

Conclusions

In this study, a deep learning network based on multi-modal temporal fusion and spatial asymmetry was proposed for MI-EEG decoding. The model comprehensively captures temporal information in EEG signals by employing multi-scale temporal convolutional filters to extract dynamic features across different frequency bands and time domains. The introduction of a spatially asymmetric convolutional structure separately captures spatial information from electrodes in the left hemisphere, right hemisphere, and global brain regions, thereby better characterizing the asymmetric activation patterns of the cerebral cortex during motor imagery tasks. The integration of global median pooling and standard deviation pooling into the self-attention module enables the learning of global dependencies. Furthermore, a multi-scale convolutional fusion module is designed to explore the correlations among temporal features from different modalities. The data augmentation strategy effectively mitigates overfitting issues caused by limited EEG data. Experimental results on two public MI-EEG datasets demonstrated that the proposed network achieved significantly higher classification accuracy than existing methods. These outcomes fully validate the effectiveness of multi-modal temporal information capture, spatially asymmetric feature extraction, and self-attention mechanisms in

EEG decoding. The main contribution of this work lies in substantially improving the classification accuracy of MI-EEG decoding. Future plans include extending this network to real-time online decoding tasks to further enhance its practicality and generalization capability.

References

[1] WOLPAW JR, BIRBAUMER N, MCFARLAND DJ, *et al.* Brain-computer interfaces for communication and control[J]. Clin. Neurophysiol., 2002, 113(6): 767 – 791.

[2] MCFARLAND DJ, WOLPAW JR. EEG-based brain-computer interfaces [J]. Curr. Opin. Biomed. Eng., 2017, 4: 194 – 200.

[3] WU JA, WANG ZY, CHEN CC, *et al.* Attention state detection based on brain-computer interface and network endogenous intelligence with simulated experiments[J]. Mobile Communications, 2025, 49(3): 107 – 116. (in Chinese).

[4] BOUSSETA R, EL OUAKOUAK I, GHARBI M, *et al.* EEG based brain computer interface for controlling a robot arm movement through thought [J]. IRBM, 2018, 39(2): 129 – 135, .

[5] HE XY, FANG HJ, LUO JL. Research on brain controlled manipulator based on Petri Net and Hybrid Brain-Computer Interface[J]. Computer Simulation, 2025, 42(04): 417 – 423. (in Chinese).

[6] RAMOSER H, MULLER-GERKING J, PFURTSCHELLER G. Optimal spatial filtering of single trial EEG during imagined hand movement[J]. IEEE Trans. Rehabil. Eng., 2000, 8(4): 441 – 446.

[7] LI JB, XIANG CL, YAO XZ. MI-EEG Classification Based on Common Time-Frequency Spatial Patterns[J]. Communication Technology, 2024, 57(4): 331 – 337. (in Chinese).

[8] Ang K K, Chin Z Y, Zhang H, *et al.* Filter bank common spatial pattern (FBCSP) in brain-computer interface[C]//2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). IEEE, 2008: 2390 – 2397.

[9] LOTTE F, GUAN C. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms[J]. IEEE Trans. Biomed. Eng., 2010, 58(2): 355 – 362.

[10] ZHANG Y, ZHOU G, JIN J, *et al.* Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface[J]. J. Neurosci. Methods, 2015, 255: 85 – 91.

[11] PARK Y, CHUNG W. Frequency-optimized local region common spatial pattern approach for motor imagery classification[J]. IEEE Trans. Neural Syst. Rehabil. Eng., 2019, 27(7): 1378 – 1388.

[12] LIAN XQ, LIU CQ, GAO C, *et al.* Research on motor imagery EEG classification based on RSCM and Riemann space[J]. Electronic Measurement Technology, 2025, 48(9): 84 – 93. (in Chinese).

[13] NGUYEN CH, ARTEMIADIS P. EEG feature descriptors and discriminant analysis under Riemannian manifold perspective[J]. Neurocomputing, 2018, 275: 1871 – 1883.

[14] LI X. Classification of multi-class motor imagery EEG signals based on the combination of EEGNet and FBCSP[D]. Nanchang: Nanchang University, 2024. (in Chinese).

[15] FAN Y, KUANG SL, XU ZB, *et al.* A convolutional neural network algorithm for simultaneously extracting time-frequency-spatial features of motor imagery signals[J]. Journal of Nanjing University: Natural Science, 2021, 57(6): 1064 – 1074. (in Chinese).

[16] TARAN S, BAJAJ V. Motor imagery tasks-based EEG signals classification using tunable-Q wavelet transform[J]. Neural Comput. Appl., 2019, 31: 6925 – 6932.

[17] LAWHERN VJ, SOLON AJ, WAYTOWICH NR, *et al.* EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces[J]. J. Neural Eng., 2018, 15(5): 056013.

[18] ROBINSON N, LEE S, GUAN C. EEG representation in deep convolutional neural networks for classification of motor image-ry[C]// Proc. IEEE Int. Conf. Syst., Man Cybern., 2019: 1322 – 1326.

[19] WU Y, MAN JZ, SONG Y, *et al.* Research on MI-EEG classification based on channel combination-data alignment-multiscale global CNN [J]. Journal of Chongqing University of Technology, 2024, 38(5): 102 – 112. (in Chinese).

[20] SCHIRRMEISTER RT, SPRINGENBERG JT, FIEDERER LDJ, *et al.* Deep learning with convolutional neural networks for EEG decoding and visualization[J]. Hum. Brain Mappi., 2017, 38(11): 5391 – 5420.

[21] LAWHERN VJ, SOLON AJ, WAYTOWICH NR, *et al.* EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces[J]. J. Neural Eng., 2018, 15(5): 056013.

[22] ZHANG C, KIM YK, ESKANDARIAN A. EEG-inception: An accurate and robust end- to-end neural network for EEG-based motor imagery classification[J]. J. Neural Eng., 2021, 18(4): 046014.

[23] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[24] DOSOVITSKIY A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arxiv preprint arxiv: 2010.11929, 2020.

[25] ALTAHERI H, MUHAMMAD G, ALSULAIMAN M. Physics-informed attention temporal convolutional network for EEG-based motor imagery classification[J]. IEEE Trans. Ind. Inform., 2022, 19(2): 2249 – 2258.

[26] SONG Y, ZHENG Q, LIU B, *et al.* EEG conformer: Convolutional transformer for EEG decoding and visualization[J]. IEEE Trans. Neural Syst. Rehabil. Eng., 2022, 31: 710 – 719.

[27] AHN M, JUN SC. Performance variation in motor imagery brain-computer interface: A brief review[J]. Journal of Neuroscience Methods, 2015, 243: 103 – 110.

[28] HUANG B, CHEN W, LIN CL, *et al.* MLP-BP: A novel framework for cuffless blood pressure measurement with PPG and ECG signals based on MLP-Mixer neural networks[J]. Biomed. Signal Process. Control, 2022, 73: 103404.

[29]IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. pmlr, 2015: 448 – 456.

[30] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (elus)[J]. arxiv preprint arxiv:1511.07289, 2015, 4(5): 11.

[31] CHEN J, YI W, WANG D, *et al.* FB-CGANet: Filter bank channel group attention network for multi-class motor imagery classification[J]. J. Neural Eng., 2022, 19(1): 016011.

[32] XU KK. EEG data augmentation method based on generative adversarial networks[D]. Hangzhou: Hangzhou Dianzi University, 2025. (in Chinese).

[33] WOO S, PARK J, LEE JY, *et al.* Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision, ECCV, 2018: 3 – 19.

[34] WANG CL, LI JX, GAO YX, *et al.* Research on MI-EEG classification based on channel combination-data alignment-multiscale global CNN [J]. Journal of Electronics & Information Technology, 2025, 47(3): 814 – 824. (in Chinese).

(Continued on page 99)

for the Guangdong Province Continuing Education Quality Improvement Project, which corroborates the academic and practical value of this reform exploration from another perspective.

Conclusions and Prospects

The reform and exploration of the experimental and practical teaching component in the "Teochew Gongfu Tea" course have successfully integrated ICH transmission, technical skill development, and innovation-oriented education. By constructing and implementing the hybrid "Three-Dimensional Synergy, Four-Competency Progression, and Five-Integration" teaching model, it has effectively addressed core challenges in intangible heritage pedagogy, including the disconnection between theory and practice, limited teaching contexts, and rigid evaluation systems, thereby enhancing students' practical competence, innovative thinking, and holistic professional development.

The laboratory serves not only as a venue for verifying scientific principles, but also as a crucial platform for cultural inheritance and the cultivation of innovative spirit and practical ability. In the future, the course team will continue to deepen the reform. On one hand, plans are underway to introduce artificial intelligence and big data technologies to achieve more precise analysis and personalized guidance for students' experimental and practical teaching processes and skill mastery. On the other hand, collaboration with industry in "industry-university-research-application" integration will be further strengthened. The establishment of "ICH creative workshops" will be explored to expose student works directly to market evaluation. This approach aims to realize the creative transformation and innovative development of ICH through deeper industry-education integration.

References

[1] HU S, CHANG Z. Innovative construction and practice of the practical teaching system for application-oriented preschool education majors[J]. Xueqian Jiaoyu Yanjiu (Early Childhood Education Research), 2025(8): 61–68.

[2] HU S. Exploration and application of the "Three-Dimensional, Four-Competency, and Five-Integration" practical teaching system[J]. Xueqian Jiaoyu Yanjiu (Early Childhood Education Research), 2025(8): 62–68.

[3] JONASSEN DH. Thinking technology: Toward a constructivist design model[J]. Educational Technology, 1994, 34(4): 34–37.

[4] SPADY WG. Outcome-based education: Critical issues and answers[M]. Arlington, VA: American Association of School Administrators, 1994.

[5] LAN S, FENG D, WANG J. Application of intangible cultural heritage in tourism management education in Chinese universities[J]. Lvyou Gailan (Tourism Overview), 2025(30): 28–30.

[6] WU, Y. Promoting ideological and political education through the "Teochew Gongfu Tea" course[N]. Zhongguo Jiaoyu Bao (China Education Daily), 2024-12-27(07).

[7] XU K, YU L, ZHAO G. What constitutes practical teaching in vocational education: Characteristics and implications of Japan's professional practice specialized courses[J]. Xiandai Daxue Jiaoyu (Modern University Education), 2025(5): 95–103.

[8] WU X, HUA Z, NIU M, *et al.* Reform of the practical teaching curriculum system for the Traditional Chinese Medicine major based on innovative talent cultivation[J]. Huaxue Jiaoyu (Chemical Education, Chinese and English Edition), 2025, 46(18): 78–86.

Editor: Yingzhi GUANG Proofreader: Xinxiu ZHU

(Continued from page 95)

[35] MANE R, CHEW E, CHUA K, *et al.* FBCNet: A multi-view convolutional neural network for brain-computer interface[J]. arxiv preprint arxiv: 2104.01233, 2021.

[36] ZHANG BX, CHAI CL, YIN Z, *et al.* Learning style recognition with multiscale EEG features[J]. Journal of Chinese Mini-Micro Computer Systems, 2021, 42(12): 2479–2484. (in Chinese).

[37] MA X, CHEN W, PEI Z, *et al.* Attention-based convolutional neural network with multi-modal temporal information fusion for motor imagery EEG decoding[J]. Computers in Biology and Medicine, 2024, 175: 108504.

[38] BA J L, KIROS J R, HINTON G E. Layer normalization[J]. arxiv preprint arxiv: 1607.06450, 2016.

[39] HE K, ZHANG X, REN S, *et al.* Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.

[40] LI HL, LIU HY, CHEN HY, *et al.* Multi-scale feature extraction and classification of motor imagery EEG based on time-series data augmentation[J]. Journal of Biomedical Engineering, 2023, 40(3): 418–425. (in Chinese).

[41] TANGERMANN M, MÜLLER K R, AERTSEN A, *et al.* Review of the BCI competition IV[J]. Frontiers in neuroscience, 2012, 6: 55.

[42] PASZKE A, GROSS S, MASSA F, *et al.* Pytorch: An imperative style, high-performance deep learning library[J]. Advances in Neural Information Processing Systems, 2019, 32.

[43] KINGMA D P. Adam: A method for stochastic optimization[J]. arxiv preprint arxiv: 1412.6980, 2014.

Editor: Yingzhi GUANG Proofreader: Xinxiu ZHU