Motor Imagery EEG Decoding Based on Multi-Scale Hybrid Networks and Feature Enhancement

Xianlun Tang, *Member, IEEE*, Caiquan Yang, Xia Sun, Mi Zou, *Member, IEEE*, and Huiming Wang, *Member, IEEE*

***Abstract*— Motor Imagery (MI) based on Electroencephalography (EEG), a typical Brain-Computer Interface (BCI) paradigm, can communicate with external devices according to the brain's intentions. Convolutional Neural Networks (CNN) are gradually used for EEG classification tasks and have achieved satisfactory performance. However, most CNN-based methods employ a single convolution mode and a convolution kernel size, which cannot extract multi-scale advanced temporal and spatial features efficiently. What's more, they hinder the further improvement of the classification accuracy of MI-EEG signals. This paper proposes a novel Multi-Scale Hybrid Convolutional Neural Network (MSHCNN) for MI-EEG signal decoding to improve classification performance. The two-dimensional convolution is used to extract temporal and spatial features of EEG signals and the one-dimensional convolution is used to extract advanced temporal features of EEG signals. In addition, a channel coding method is proposed to improve the expression capacity of the spatiotemporal characteristics of EEG signals. We evaluate the performance of the proposed method on the dataset collected in the laboratory and BCI competition IV 2b, 2a, and the average accuracy is at 96.87%, 85.25%, and 84.86%, respectively. Compared with other advanced methods, our proposed method achieves higher classification accuracy. Then we use the proposed method for an online experiment and design an intelligent artificial limb control system. The proposed method effectively extracts EEG signals' advanced temporal and spatial features. Additionally, we design an online recognition system, which contributes to the further development of the BCI system.**

***Index Terms*— Brain–computer interface, EEG decoding, feature enhancement, multi-scale hybrid network, artificial limb control.**

Manuscript received 13 April 2022; revised 17 August 2022, 12 December 2022, and 24 January 2023; accepted 31 January 2023. Date of publication 3 February 2023; date of current version 9 February 2023. This work was supported in part by the National Nature Science Foundation of China under Grant 61673079, in part by the Natural Science Foundation Project of Chongqing under Grant CSTB2022NSCQ-MSX0380, and in part by the Major projects of Chongqing Municipal Education Commission under Grant KJZDM202001901 and Grant KJZD-M202200603. *(Corresponding author: Xianlun Tang.)*

Xianlun Tang is with the Chongqing Key Laboratory of Complex Systems and Bionic Control, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with the Guangyang Bay Laboratory, Chongqing Institute for Brain and Intelligence, Chongqing 400064, China (e-mail: tangxl@cqupt.edu.cn).

Caiquan Yang, Mi Zou, and Huiming Wang are with the Chongqing Key Laboratory of Complex Systems and Bionic Control, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: ycqwer@qq.com).

Xia Sun is with the Chongqing Institute of Engineering, Chongqing 400056, China (e-mail: sunxia@cqie.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2023.3242280, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3242280

I. Introduction

BRAIN-COMPUTER Interface (BCI), a technology for information interaction between the nervous system and external devices, establishes a direct connection between the brain and external devices [1]. BCI technology collects brain nerve activity signals through sensors, e.g., electrodes placed on the scalp or in the skull. Through signal processing, feature extraction, and pattern recognition, the BCI system can predict human control intention, cognitive or mental states, and neurological disease states. Besides, it offers new communication channels or rehabilitation methods for patients with difficulty in body or language [2], [3] and provides more information output channels for healthy people. At present, there is a large body of research in many fields on BCI systems, e.g., sports rehabilitation [4], smart home [5], and entertainment [6].

Commonly used BCI paradigms are Steady-State Visual Evoked Potentials (SSVEP), P300, and Motor Imagery BCI (MI-BCI) [7]. MI-BCI is one of the most valuable paradigms. When the subject imagines the movement of the left or the right hand (there is no movement of the left and right hands), the cerebral cortex will produce two salient rhythm signals. The EEG rhythm energy drops significantly in the motor-sensory area on the contralateral side of the cerebral cortex. In contrast, the EEG rhythm energy of the ipsilateral motor-sensory area increases. This phenomenon is called Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS) [8]. EEG signals are classified by extracting the features of this phenomenon, enabling direct communication and control between the human brain and external devices. In most research [9], [10], feature extraction is designed based on people's knowledge and experience, which usually demands sophisticated experiments and close observation. Designing an effective feature extractor consumes a lot of human resources and the generalization of feature extractors designed through experience is poor. The convolutional neural network shows great promise in Computer

Vision (CV) and Natural Language Processing (NLP). Many researchers have begun to apply CNN to nonlinear EEG signal classification to improve decoding ability and implement a BCI system with more robust generalization performance [11], [12], [13].

At present, many CNN-based classification methods apply the one-dimensional (1D) convolution or the two-dimensional (2D) convolution and use a single-scale convolution kernel, which limits the CNN network's adaptability to the extraction of different temporal and spatial features. For example, some classical networks DeepNet [11] and EEGNet [12] used for EEG signal decoding use 2D convolution with a single-scale convolution kernel, which cannot effectively extract deep temporal features and do not take into account inter-individual differences, since the optimal kernel size of each individual varies from person to person. In recent years, MSCNN [14] incorporates a 1D convolution and a multi-scale strategy which can effectively extract the temporal features of EEG signals and balance the differences between individuals to some extent, but cannot extract spatial features well. In addition, [15] proposed an interesting serial multiscale network, but the multiscale features were not characterized from the original data because it is a deeper serial network. To take into account the differences between different individuals and the extraction of spatio-temporal features, we propose a novel parallel end-to-end network model-Multi-Scale Hybrid Convolutional Neural Network (MSHCNN), which decodes dichotomous MI-EEG signals to improve classification performance. In addition, considering that 1D convolutional networks can only effectively extract temporal features, a coding method of EEG signals is proposed to enhance the expression of temporal and spatial features.

We highlight the contributions of this paper as follows:

1) A method for enhancing EEG signal features is proposed, which is more suitable for encoding between EEG signal channels in motor imagery.
2) An end-to-end network called MSHCNN is built, which can achieve good classification performance on EEG signals with less preprocessing.
3) An intelligent artificial limb control system is designed based on our proposed method. Experiments in section IV show that the BCI system is feasible.

The rest of this paper is organized as follows: The second section reviews the work related to the classification of MI-EEG signals. The third section describes the proposed MSHCNN and feature enhancement method. The fourth section presents the experimental results and the related analysis. The fifth section summarizes our work.

II. Related Work

The BCI system mainly comprises signal acquisition, signal processing and conversion, control object, and feedback. The most crucial part is signal processing and transformation, which involves feature extraction and classification. We focus on time-frequency features and spatial features for EEG feature extraction. The classification of EEG signals is primarily studied within the framework of traditional and deep learning methods.

A. EEG Feature Extraction

Common Spatial Pattern (CSP) and improved methods based on CSP are mainly used for spatial feature extraction of EEG signals. CSP uses the diagonalization of the matrix to find a set of optimal spatial filters for projection, which maximizes the variance of the two types of signals but does not consider the local temporal information. Wang et al. propose a new optimal spatiotemporal filter-Local Temporal Common Space Patterns (LTCSP) for robust single-experiment EEG classification. This method takes local temporal information into account [16]. Ang et al. apply an FBCSP method to classify MI-EEG signals and optimize the subject-specific frequency band of CSP [17]. According to the literature, Fourier Transform and Wavelet Transform are mainly adopted for time-frequency feature extraction of EEG signals. For example, Lu et al. use Fast Fourier Transform (FFT) and Wavelet Packet Decomposition (WPD) to obtain frequency domain features to classify MI-EEG signals [18]. Ji et al. apply a feature extraction method based on Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD), and approximate entropy for MI-EEG signal classification [19]. In addition, some researchers use Power Spectral Density (PSD) to extract frequency domain features for EEG signal classification [20], [21].

B. EEG Pattern Classification

Traditional EEG signal classification methods mainly include K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM). Vidaurre et al. propose an unsupervised adaptive method based on a LDA classifier, and this unsupervised classifier is applied to online experiments [22]. Siuly and Li apply a feature extractor based on cross-correlation, where a least square support vector machine (LS-SVM) is used to classify MI-EEG signals [23].

Given the superiority of deep learning in CV and NLP, many researchers use CNN to decode EEG signals. Schirrmeister et al. propose three CNN architectures with different frameworks to decode MI-EEG from the original EEG, such as ShallowNet, DeepNet, and HybridNet [11]. Lawhern et al. propose a compact EEG feature extraction model based on depth separable convolution to classify EEG signals of different paradigms [12]. Tang et al. employ a novel method based on conditional empirical mode decomposition (CEMD) and 1D multi-scale convolutional neural network (1DMSCNN) to decode MI-EEG signals and for the control of intelligent wheelchairs [14]. Jia et al. apply a novel end-to-end model, Multi-branch Multi-Scale Convolutional Neural Network (MMCNN), to determine the optimal convolution scale [24]. In addition, many researchers have introduced the attention mechanism into the classification of EEG signals. Liu et al. propose a convolutional neural network based on parallel spatial-temporal self-attention, which is used to classify four types of MI-EEG signals and apply the proposed method to control drones [25]. In order to apply state-of-the-art methods in other fields to BCI systems, Song et al. use the transformer for the extraction of temporal and spatial features of EEG signals for the first time [26].

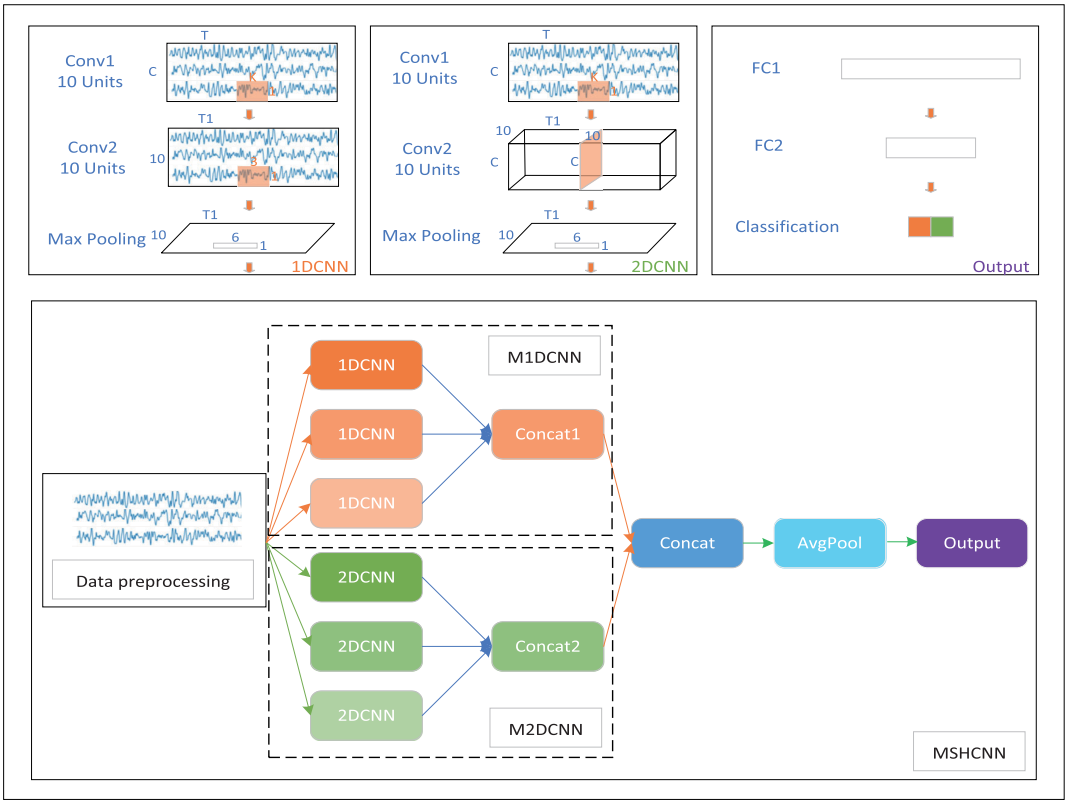

Fig. 1. The structure of MSHCNN. The 1DCNN block is composed of two one-dimensional convolutions, the 2DCNN block is composed of two two-dimensional convolutions, and the Output block is two fully connected layers. The colors of 1DCNN and 2DCNN in the MSHCNN structure represent different convolution kernel sizes.

In general, CNN can not only extract and classify the features of complex EEG signals simultaneously, but also extract features from multiple dimensions, such as the temporal domain, spatial domain, and frequency domain. Many researchers have made headway in BCI using CNN-based methods. However most of the current deep learning methods use a single 1D convolution or 2D convolution and use a single convolution kernel scale. Due to the individual differences of EEG signals, the optimal scale may vary from subject to the subject [24]. A single convolution kernel and a single convolution method cannot fully extract the features of EEG signals [14], [27]. To fully extract EEG signals' temporal and spatial features, we have designed a novel hybrid network combining multi-scale 1D convolution with 2D convolution to classify EEG signals. In addition, 1D convolution cannot extract the correlation between channels well, so an encoding method suitable for EEG data feature enhancement is proposed.

III. Method

In response to the above problems, this paper proposes a Multi-Scale Hybrid Convolutional Neural Network, extracting deep temporal and spatial features on multiple scales to improve classification performance. The structure of the proposed MSHCNN is shown in Fig. 1. This paper also presents a data preprocessing method to ameliorate the properties of the MI-EEG signal to improve classification accuracy. The following is a detailed description of our proposed method.

A. Proposed MSHCNN Structure

CNN is first applied to the handwriting digit recognition system in the paper [28]. It is inspired by the human visual nervous system, which uses a convolution kernel to replace the field of vision of human eyes. CNN generally consists of three parts, including a convolutional layer, a pooling layer, and a fully connected layer. The convolutional layer and pooling layer are used to extract features, and the fully connected layer is used for classification, and the convolution formula is shown in (1). Due to its powerful adaptive feature extraction applications, it has gained great popularity in machine vision and is applied to image classification [29], object detection [30], semantic segmentation [31], and style transfer [32], etc.

$$x_j^d = f\Big(\sum_{i \in M_j} x_i^{d-1} * w_{ij}^d + b_j^d\Big). \tag{1}$$

where x_j^d is the j^{th} feature map of the d^{th} layer convolution, x_i^{d-1} is the i^{th} feature map of the previous convolutional layer, M_j is the set of input feature maps, w_{ij}^d is the connection weight between the j^{th} feature map of the d^{th} layer convolution and the i^{th} feature map of the previous layer of convolution, $*$ represents the convolution operation, b_j^d is the bias of the j^{th} feature map of the d^{th} layer convolution, $f(\bullet)$ is the activation function, and the commonly used activation function is Sigmoid($f(x) = \frac{1}{1+e^{-x}}$), Relu($max(0, x)$), etc.

In object detection, YOLO proposes a multi-scale detection strategy to be compatible with the detection accuracy of large

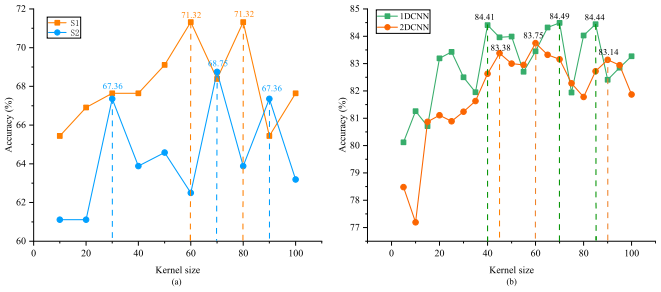

Fig. 2. Effects of convolution kernel size on classification performance. (a) classification accuracy of two different subjects in Dataset A of different convolution kernel sizes. (b) the average classification accuracy of 1D convolution and 2D convolution models of different convolution kernel sizes in Dataset A.

and small objects. It takes an image with a resolution of 416 as input, and generates 3 different scale feature maps (52×52, 26×26, 13×13), then performs object detection of different scales [30]. For the classification of MI-EEG signals, different subjects also have different optimal receptive fields. In order to explore the effect of the convolution kernel size on the classification accuracy of MI-EEG signals, we have designed a two-dimensional CNN similar to ShallowNet [11]. Fig. 2 (a) presents the classification accuracy of two random subjects on Dataset A with different scale convolution kernels. It can be concluded that subject 1 achieves better classification results with convolution kernels of 60×1 and 80×1, while subject 2 achieves better classification results with convolution kernels of 30×1, 70×1, and 90×1. Inspired by the above discoveries, we note that the size of the receptive field is closely related to feature extraction, and that 1D convolution and 2D convolution can effectively extract temporal and spatial features, respectively. Therefore, an MSHCNN is proposed to improve the classification of MI-EEG signals.

As shown in Fig. 1, it consists of four parts: data input block, one-dimensional multi-scale convolutional neural network (M1DCNN) and two-dimensional multi-scale convolutional neural network (M2DCNN) feature extraction block, feature splicing block, and feature classification. The input data shape of the M1DCNN block is (B, N, T), and the input data shape of the M2DCNN block is (B, 1, T, N), where B represents the batch of the input network, T represents the length of the EEG signal, and N indicates the number of channels to select EEG signals. We extract deep temporal features through multi-scale 1D convolution while extracting spatio-temporal features in parallel using multi-scale the 2D convolution, as described below. The M1DCNN block extracts the shallow and deep temporal features of the EEG signals on multiple scales, which consists of three 1DCNN blocks and feature splicing layers. The shades of the colors of the 1DCNN block represent different convolution kernel sizes. In the 1DCNN block, the EEG signal first passes through 10 one-dimensional filters with a kernel size of K to extract shallow temporal features. Then we use 10 one-dimensional filters with a kernel size of 3 to extract deep temporal features. M2DCNN blocks are used for multi-scale extraction of shallow temporal and spatial features of EEG signals. Similar to M1DCNN, its color shades represent different convolution kernel scales, and 10 filters with a kernel size of K×1 are used to extract temporal features. Then, 10 filters with the same kernel size as the number of EEG signal channels are used to extract the spatial features. From the analysis above, we can see that the optimal convolution kernel size for each subject varies from person to person. Therefore, when choosing the size of the convolution kernel, we use 1DCNN and 2DCNN to implement a series of experiments on Dataset A to explore the influence of the size of the convolution kernel on the average accuracy of all subjects. The result is shown in Fig. 2 (b). The average classification accuracy varies with the size of the convolution kernel. According to the experimental results, the 1DCNN structure selects three different convolution kernel sizes (40, 70, 85), and the 2DCNN structure selects three different convolution kernels of 45×1, 60×1, and 90×1. For feature splicing, the three splicing blocks all perform feature fusion in the time dimension, and the process can be described as:

$$R_1^{(b,10,t_1^1)} + R_1^{(b,10,t_1^2)} + R_1^{(b,10,t_1^3)} \rightarrow R_1^{(b,10,t_1^*)}. \quad (2)$$

$$R_2^{(b,10,t_2^1)} + R_2^{(b,10,t_2^2)} + R_2^{(b,10,t_2^3)} \rightarrow R_2^{(b,10,t_2^*)}. \quad (3)$$

$$R_1^{(b,10,t_1^*)} + R_2^{(b,10,t_2^*)} \rightarrow R^{(b,10,t)}. \quad (4)$$

R_1, R_2, R respectively denote the size of the feature map in the M1DCNN block, the size of the feature map in the M2DCNN block, the size of the feature map after the M1DCNN block and the M2DCNN block are joined; b represents the batch size, and t represents the size of the time dimension, where $t_1^1 + t_1^2 + t_1^3 = t_1^*$, $t_2^1 + t_2^2 + t_2^3 = t_2^*$, $t_1^* + t_2^* = t$.

The spliced temporal feature and spatial feature are subject to average pooling and then mapped to the 1D feature as the input of the Output block. The Output block is composed of two fully connected layers, with the hidden layer set to 100 neurons, and the output to 2 neurons, and then are classified by the Softmax. In the experiment, we use the Rectified linear unit (Relu) [33] as the activation function, which alleviates the problem of vanishing gradient and speeds up the learning of the network. To prevent network overfitting, we introduce L2 regularization, BatchNorm, and Dropout methods to reduce the risk of overfitting. Table I shows the detailed parameters of the basic blocks 1DCNN, 2DCNN, and Output blocks to build the MSHCNN structure. Since the basic network structure is identical, only the parameters of 1DCNN and 2DCNN with a single kernel size are provided in the table. It should be noted that each convolutional layer is followed by a BatchNorm layer, a Dropout layer, and a Relu layer.

B. Feature Enhancement

In the 1DCNN block, an one-dimensional convolution cannot extract the correlation between channels, thus Tang et al. propose an EEG signal combination method to encode ERS/ERD information, which improves the classification accuracy of MI-EEG signals [14]. However, only the difference between the left channel and the right channel is considered, and the similarity between the channel has not been taken into account. Therefore, the following methods are proposed to enhance the features of EEG signals. Suppose

TABLE I
DETAILED PARAMETERS OF THE MAIN BLOCKS OF THE PROPOSED MSHCNN

Block type Input layer	Filters	Feature map $1 \times 1000 \times 3$	Kernel	Stride	Parameters
1DCNN		3×1000			
Conv1 layer	10	10×321	40	3	1210
Conv2 layer	10	10×319	3	1	310
MaxPool layer	10	10×53	6	6	
2DCNN		$1 \times 1000 \times 3$			
Conv1 layer	10	$10 \times 319 \times 3$	45×1	3×1	460
Conv2 layer	10	$10 \times 319 \times 1$	1×3	1×1	310
MaxPool layer	10	10×53	6×1	6×1	
Concat		10×309			
AvgPool layer	10	10×38	8	8	
Flatten		380			
FC1+Dropout		100			38100
FC2+Softmax		2			202

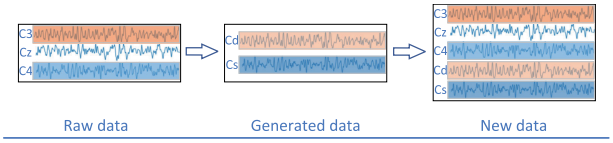

Fig. 3. EEG signal feature enhancement method. C3, Cz, and C4 represent the original EEG data, and Cd and Cs represent the EEG data after feature enhancement.

the data of the C3 channel on the left side of the brain is represented by C_3^T, where T represents the data length of the EEG signal, and the data of the C4 channel symmetrical to the C3 channel is represented by C_4^T, the difference and similarity of the symmetric channel data are expressed as follows:

$$C_s = C_3^T - C_4^T. \tag{5}$$

$$C_d = C_3^T + C_4^T. \tag{6}$$

Take the EEG signal channels C3, C4, and Cz in Dataset A as an example. The EEG signal feature enhancement method is presented in Fig. 3. The steps of the EEG signal feature enhancement method are as follows:

1) Determine the symmetrical channel of the EEG signal.
2) Use formulas (5) and (6) to process symmetric channel data to obtain a new pair of data.
3) Add the obtained new EEG data to the original data in parallel.

IV. EXPERIMENT

A. Dataset and Experimental Method

To evaluate the effectiveness of our proposed method, we have conducted related experiments on BCI Competition IV 2b [34] (Dataset A), BCI Competition IV 2a [35] (Dataset B), and laboratory data [36] (Dataset C). The following is the detailed description of each dataset:

Dataset A: It is based on visually evoked left-hand and right-hand motor imagery and contains data from three channels C3, C4, and Cz. The dataset collects the EEG signals of 9 normal subjects. The EEG data of each subject includes 5 sessions. There are 240 trials in the first 2 sessions, and 120 trials in each session (60 for the left hand and 60 for the right hand). The last 3 sessions have 480 trials, and each session has 160 trials (80 for the left hand and 80 for the right hand). All data have been processed with a 0.5-100Hz bandpass filter and a 50Hz notch filter, the sampling frequency is 250Hz, and the amplitude range of the EEG data is $\pm 50\mu$V.

Dataset B: It is composed of EEG data from 9 normal subjects, including four different motor imagery tasks, involving the left hand, the right hand, the feet, and the tongue. Each subject has two sessions on different days, each session has 6 cycles, and each cycle has 48 trials (There are 12 of each of the four motor images), and a total of 288 trials have been conducted for each session. The data collect information on 25 channels, including 22 EEG channels and 3 EOG channels, with a sampling frequency of 250Hz.

Dataset C: It is an EEG dataset of left-hand and right-hand motor imagery. The EEG data of 7 subjects are collected by the Emotiv EEG acquisition instrument developed by Emotiv System in the United States(in the experiment we use the first 6 subjects). Each subject has performed 240 trials, 120 times for the left and the right hand respectively. There are a total of 14 electrodes in the EEG acquisition equipment. This dataset selects 6 channels F3, F4, FC5, FC6, T7, T8 located in the motion perception area to identify the EEG signals of left and right motor imagery. The sampling frequency is 128 Hz, in the dataset, we retain only the 3-4 seconds of each channel.

When subjects imagine the movement of the left or the right hand, the ERD/ERS phenomenon of μ rhythm (8-13Hz) and β rhythm (13-30Hz) is significant [8]. To simplify preprocessing, we have performed 6-order Butterworth bandpass filtering and Z-Score normalization on the original data. To preserve the complete information of μ and β rhythms, the filtered frequency bands are extended to 0.5-40 Hz. In addition, the standardized formula we adopt is expressed as below:

$$\dot{X}_{T\times C} = \frac{X_{T\times C} - \bar{X}_{1\times C}}{\delta_{1\times C}}. \tag{7}$$

where $X_{T\times C}$ is the original data of a sample, T represents the length of the time dimension of the data, C represents the number of channels, $\bar{X}_{1\times C}$ represents the average in the time dimension, and $\delta_{1\times C}$ represents the standard deviation in the time dimension.

In reference to Dataset A, we select the corresponding three channels of C3, C4, and Cz in Dataset B, and select the samples of left-hand and right-hand motor imagery to do the two classifications.

In the experiment, the data are divided into a training set and a test set at the ratio of 4 to 1. Pytorch1.8.0 is used to build our proposed MSHCNN network. The loss function uses cross-entropy. The dropout probability is set at 0.25. L2 regularization parameter is set at 0.1 and the momentum is set at 0.9. Stochastic Gradient Descent (SGD) method is used to optimize our network, the learning rate is set to 0.001, the batch size is set to 20, and 100 epochs are trained.

B. Experiments on Dataset A and B

1) Performance of MSHCNN: A series of experiments are conducted on Dataset A using a network with a single convolution kernel and MSHCNN to verify the performance of our

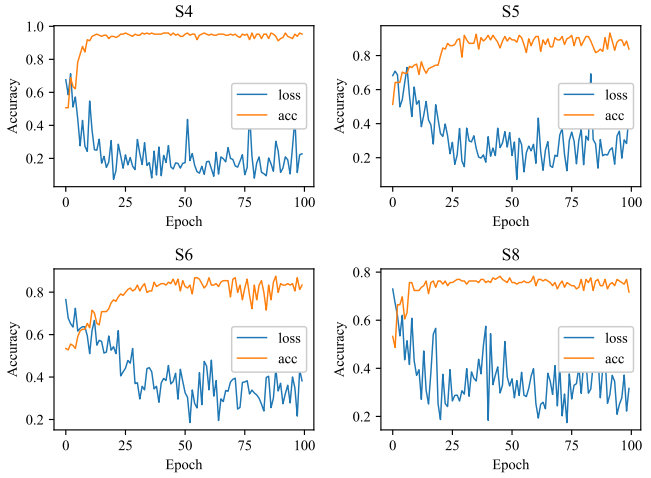

Fig. 4. Training loss and validation accuracy curves for subjects S4, S5, S6 and S8 in Dataset A.

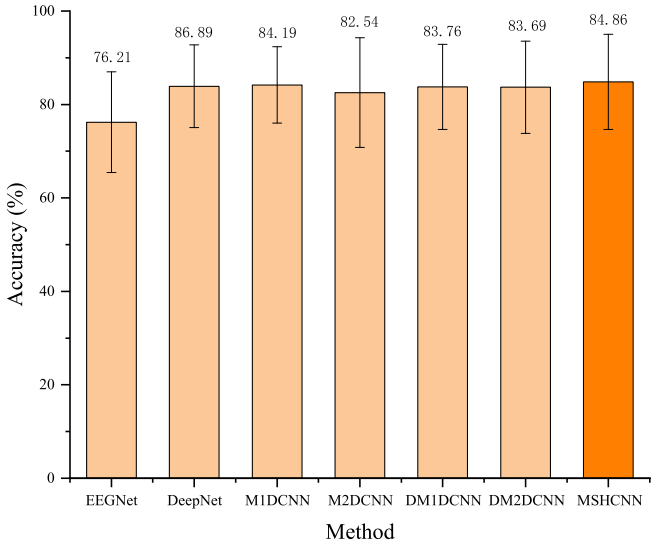

Fig. 5. Experimental results of ablation experiment on Dataset A.

proposed method in multi-scale and spatial-temporal feature extraction. The one-dimensional convolutional network with a single convolution kernel uses the combination of the 1DCNN block and the Output block in Fig. 1 (denoted as 1DCNN). A two-dimensional convolutional network with a single convolution kernel uses a combination of a 2DCNN block and an Output block (denoted as 2DCNN). From the analysis in Fig. 2 (b), the convolution kernel size of the 1DCNN block in MSHCNN is set at 40, 70, and 85, and the convolution kernel size of the 2DCNN block is set at 45, 60, and 90. The average accuracy of the MSHCNN network on Dataset A is 84.86%. The results obtained by the proposed method are compared with the results of the separate 1D convolution and 2D convolution models in Fig. 2 (b), it is concluded that the multi-scale hybrid network, which combines the advantages of one- and two dimensional convolution in both temporal and spatial feature extraction, outperforms the single convolutional kernel network. At the same time, we find that in Dataset A the one-dimensional convolution is generally slightly better than the two-dimensional convolution. To demonstrate the reliability of our proposed network, Fig. 4 shows the training loss and validation accuracy curves for subjects S4, S5, S6, and S8 in Dataset A. From the accuracy curves, we can see that the model achieves decent classification performance in about 10-25 epochs.

2) Comparing With Baselines: We choose the widely used network EEGNet [12] and DeepNet [11] as the baseline. In addition, we choose the combination of blocks in our proposed network for ablation experiments. The networks include M1DCNN, M2DCNN, DM1DCNN, and DM2DCNN. Among them, M1DCNN is a combination of M1DCNN block and Output block, M2DCNN integrates M2DCNN block with Output block, DM1DCNN combines 2 parallel M1DCNN blocks and Output block, and DM2DCNN is an integration of 2 parallel M2DCNN blocks and Output block. We conduct experiments on Dataset A. In the experiment, the size of the convolution kernel of other networks is consistent with the size of the convolution kernel in MSHCNN, and the hyperparameters are the same as those given in the experimental method. The comparative results are shown in Fig. 5. MSHCNN achieved the best classification results, with an average classification accuracy rate of 84.86%. Statistically, our proposed method is significantly different from EEGNet (0.018, 0.825) and M2DCNN (0.036, 0.211); it is not significantly different from DeepNet (0.441, 0.102), M1DCNN (0.498, 0.073), DM1DCNN (0.484, 0.114) and DM2DCNN (0.233, 0.117), but the average classification results it achieves are better. The first number in the brackets is the p-value, which helps determine statistical significance. Generally, A p-value less than 0.05 is statistically significant. The second is the Cohen's d-value, which characterizes the effect size by relating the mean difference to variability, and a value less than 0.2 means that the difference is very small; a value between [0.2, 0.5) indicates a small difference; a value between [0.5, 0.8) indicates a medium difference; a value greater than 0.8 indicates a very large difference.

3) Performance of Feature Enhancement Method: We use EEGNet and MSHCNN to evaluate our proposed feature enhancement method on Dataset A. Fig. 6 shows the experimental results. EEGNet and MSHCNN represent the results of filtering and standardization of the original data. S_EEGNet and S_MSHCNN are the results obtained using the subtractive encoding method. A_EEGNet and A_MSHCNN are the results obtained by using the additive encoding method. SA_EEGNet and SA_MSHCNN use the encoding methods proposed in this paper. The analysis of the experimental results shows that the classification accuracy of the three encoding methods has been improved in the EEGNet network. Compared with unencoded data, the classification accuracy of a single encoding method has decreased by about 0.1% in the MSHCNN network, whereas there is an improvement in the encoding method we proposed. Therefore, our proposed encoding method is more adaptive. Statistically, SA_EEGNet is significantly different from A_EEGNet (0.022, 0.653) and not significantly different from S_EEGNet (0.139, 0.201). SA_MSHCNN is not significantly different from S_MSHCNN (0.441, 0.064) or A_MSHCNN (0.285, 0.067), but our proposed method is more stable. In addition, we have verified the feature enhancement

TABLE II
COMPARISON OF THE AVERAGE PERCENTAGE CLASSIFICATION ACCURACY OF DIFFERENT MODELS ON DATASET A

Subject	FBCSP 2012 [17]	MSNN 2021 [15]	DeepNet 2017 [11]	S3T 2021 [26]	MANN 2021 [37]	MSCNN 2020 [14]	MMCNN 2021 [24]	Proposed
S1	70.00	84.72	86.11	81.67	82.81	80.56	84.90	86.80
S2	60.36	69.11	72.79	68.33	60.36	65.44	70.40	77.94
S3	60.94	62.50	68.05	66.67	59.06	65.97	75.50	65.97
S4	97.50	97.97	96.62	98.33	97.50	99.32	96.30	97.97
S5	93.12	91.21	92.56	88.33	91.88	89.19	92.40	93.24
S6	80.63	86.11	85.41	90.00	86.38	86.11	86.30	88.88
S7	78.13	79.16	83.33	85.00	84.06	81.25	87.60	86.80
S8	92.50	87.50	83.35	93.33	93.44	88.82	84.20	82.89
S9	86.88	88.88	86.80	86.67	86.88	86.81	81.80	86.80
Avg	80.00*	83.01	83.89*	84.26	82.54	82.61	84.37	85.25
Std	13.85	11.10	8.85	10.64	13.73	11.00	7.91	9.19
p-value	0.08	0.18	0.09	0.59	0.11	0.21	0.26	-

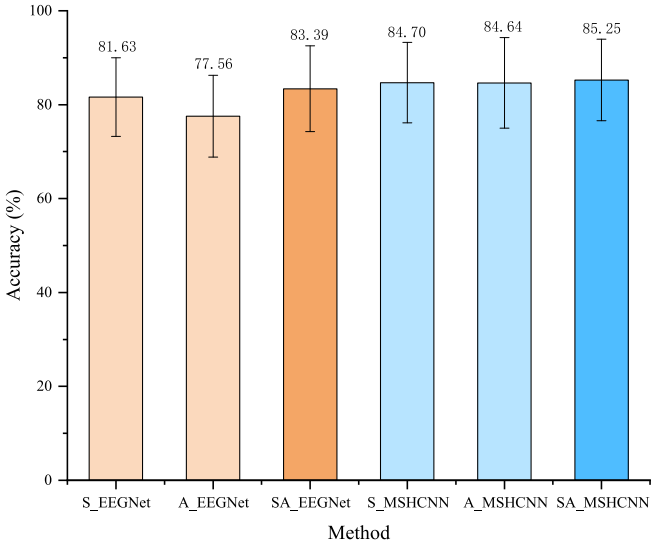

Fig. 6. Experimental results of feature enhancement on Dataset A.

method on Dataset B. The average accuracy rate can reach 84.87%, with the data processed by our feature enhancement method.

4) Comparison With Other Methods: We compare our proposed method with open-source methods. The following is a brief introduction to these methods:

- FBCSP [17]: It is a feature extraction method based on CSP, which is achieved through frequency band grouping and feature selection algorithms.
- MSNN [15]: It is a novel and serial deep CNN that classifies multi-paradigm EEG by representing multi-scale spatio-temporal features.
- DeepNet [11]: It consists of 5 parts. The first block has two convolutional layers to extract spatial and temporal features. Then there are three standard convolutional layers and finally a fully connected layer for classification.
- S3T [26]: It is a Transformer-based network structure that includes a spatial transformer and a temporal transformer using the attention mechanism.
- MAAN [37]: It is a new multi-attention adaptive network that integrates attention with transfer learning for the classification of EEG signals.
- MSCNN [14]: It uses an improved empirical mode decomposition data preprocessing method and a multi-scale one-dimensional convolution network to classify EEG signals.
- MMCNN [24]: It is a novel end-to-end EEG signal classification model. Without filtering, it can effectively decode the original EEG signal with a multi-scale and attention mechanism.

Table II compares the average accuracy of our proposed method with several state-of-the-art methods on Dataset A. From the table, we can draw the conclusion that the average accuracy of our proposed method is the highest. Compared with the traditional FBCSP method, our proposed method improves the average classification accuracy improved by 5.25%, and only the accuracy achieved by subjects 8 and 9 is lower than that achieved in the traditional method. Compared with the highly cited convolutional neural network EEGNet, our proposed method improves the average accuracy by 9.04%. Relative to the new methods proposed in recent years, the classification results of our proposed method are also very competitive. We use the Wilcoxon signed-rank test to perform statistical analysis on the classification results. In the table for the average results, * indicates that there is a significant difference at 10%, and ** indicates that there is a significant difference at %5. The annotation applies to subsequent tables.

5) Visualization Analysis: To demonstrate the learning patterns of our proposed network, we use EEG activation patterns and t-sne to visualize and analyze the learning patterns of the network. The t-sne is an embedding model that can map data in a high-dimensional space to a low-dimensional space and preserve the local characteristics of the data set. It is mainly used for dimensionality reduction and visualization of high-dimensional data [38]. The visualization patterns are generated based on information from the fourth subject in Dataset A.

As Fig. 7 indicates, we map the learning weights of the spatial convolution in the MSHCNN and visualize them as a topological map based on the activation pattern. We use a 22-channel EEG mapping to facilitate the observation of activation patterns. We normalize the learning weights for channels C3, Cz and C4, then fill the remaining 19 channels with zeros. In this investigation, we have found that the

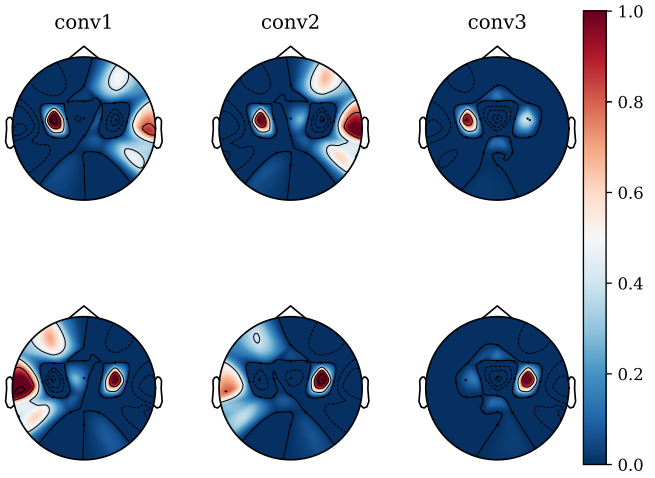

Fig. 7. Topological visualization of activation pattern maps for MSHCNN spatial convolution. The visualization depicts the fourth subject of Dataset A. The first row is the learning weight activation map of left-hand motor imagery and the second row is the learning weight activation map of right-hand motor imagery, where conv1 denotes the first branch, conv2 denotes the second branch and conv3 denotes the third branch.

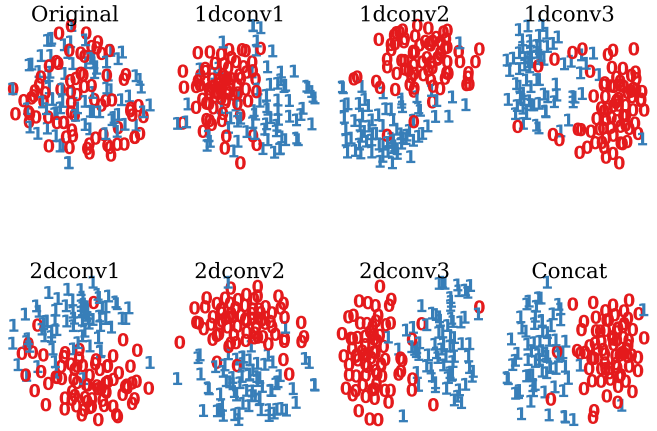

Fig. 8. The t-sne algorithm is used to visualize the features of the MSHCNN at different convolutional layers.The visualization uses the fourth subject of Dataset A. Original represents the visualization of the original data. 1dconv1-3 are the features of the first layer of convolution of the three branches of the 1D convolution and 2dconv1-3 are the features of the first layer of convolution of the three branches of the 2D convolution. Concat is a concatenation of six branch features.

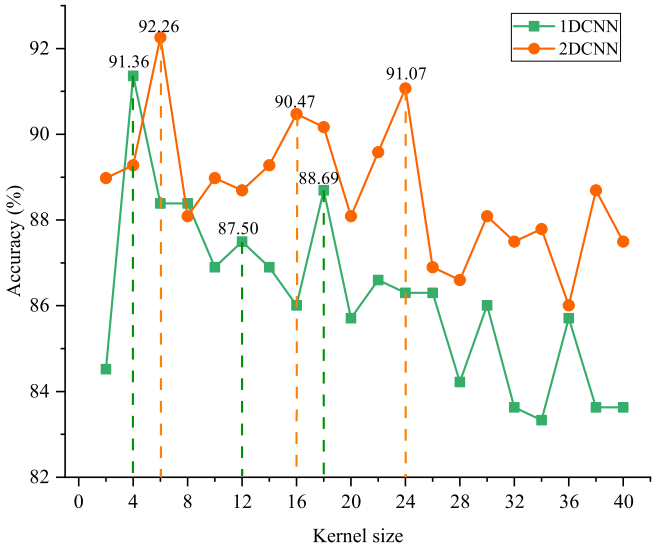

Fig. 9. The average accuracy of different convolution kernel sizes on subjects in Dataset C.

weights of spatial convolution represent the different degrees of activation in the left and right sides of the brain on left- and right-hand motor imagery. Thus, our proposed network is capable of spatial feature extraction of EEG signals from multiple temporal scales.

In addition, we visualize the feature map after the first layer of multi-scale temporal convolution and concatenation, and the visualization results are shown in Fig. 8. We visualize the original input features, the three feature maps from the 1D multiscale convolution and the three feature maps from the 2D multiscale convolution, where the convolution kernel size increases sequentially. Finally, the features are visualized with all branches converged into one. We found that in the one-dimensional convolution, the features of the second and third branches are more distinct from those of the first branch. In two-dimensional convolution, the features of the second branch are more prominent than those of the first and third branches. After feature concatenation, the features are more clearly differentiated, especially in the middle part, where only a few samples are not differentiated. We conclude that our proposed network is better at extracting temporal features on different scales relative to other methods.

C. Experiment on Dataset C

We do relevant experiments on the data collected in the laboratory to verify the adaptability of our proposed method to other datasets. Since the length of the input data is different from Dataset A, and the input data shape of Dataset C is 128×6, we have modified some parameters in the network. We use 8 filters for the first layer of convolution, with the stride size set at 1, and the number of filters for the second layer of convolution is set at 16. When choosing the size of the convolution kernel, we repeat the experiment, and the result is shown in Fig. 9. We conclude that as the convolution kernel increases, the average classification accuracy witnesses a downward trend. In the end, we choose 4, 12, and 18 as the convolution kernels of 1DCNN, 6, 16, and 24 as the convolution kernels of 2DCNN, and the remaining hyperparameters remain unchanged. On Dataset C, we found that two-dimensional convolutions generally obtain better results than one-dimensional, which might is impacted by the sampling frequency, duration, and number of channels.

1) Method Validation: We evaluate the performances of EEGNet, DeepNet, M1DCNN, M2DCNN, DM1DCNN, DM2DCNN, and MSHCNN on Dataset C. The experimental results are shown in Fig. 10. It can be clearly seen that the method we proposed reports a higher average accuracy. In addition, the multi-scale convolutional networks M1DCNN and M2DCNN obtain better classification results than the single convolution kernel EEGNet and DeepNet networks, and we also find that two parallel network structures (DM1DCNN and DM2DCNN) can improve the classification accuracy. As Fig. 11 indicates, the training loss and accuracy curves

TABLE III
COMPARISON OF THE AVERAGE PERCENTAGE CLASSIFICATION ACCURACY OF DIFFERENT MODELS ON DATASET C

Subject	MKELM	LSTM	k-SAE	DeepNet	MSNN	Proposed
S1	92.72	89.50	94.17	89.58	93.75	97.91
S2	92.83	94.83	99.83	100.00	100.00	100.00
S3	88.33	85.00	89.67	66.66	81.25	95.83
S4	98.74	96.50	100.00	91.66	100.00	97.91
S5	92.53	87.00	98.83	100.00	97.91	97.91
S6	82.03	91.50	89.47	91.66	83.33	91.66
Avg	91.20**	90.72**	95.33	89.92	92.70	96.87
Std	5.09	4.06	4.51	11.18	8.41	2.62
p-value	0.04	0.02	0.24	0.13	0.14	-

for subjects S1, S2, S4, and S5 in Dataset C on the validation set are shown. It can be seen that the convergence speed of subjects S2 and S4 is fast, and the convergence speed of subjects S1 and S5 is slightly slower, probably due to the mental state or environmental factors, which cause the data distribution of the collected EEG signals to be more complex, making subjects S1 and S5 converge more slowly. We use MSHCNN to evaluate the effectiveness of feature enhancement. The average results of the experiments are S_MSHCNN (96.52±3.88), A_MSHCNN (95.48±5.80), SA_MSHCNN (96.87±2.87). We found that using one encoding method may be ineffective or even counterproductive, while using our proposed encoding method has boosting effects to some extent. It can be concluded that our proposed data encoding method is more robust.

2) Comparison With Other Methods: We compare our proposed method with other methods. Table III provides the comparative results, among which the experimental results of MKELM [39], LSTM [40] and, k-SAE [36] are all from [36], the experimental results of DeepNet and MSNN are from our implementation. According to the analysis in Table III, our proposed method has achieved the best results, and the minimum standard deviation is 2.62, indicating that our proposed method is the most robust.

D. Cross-Subject Experiments

We use our proposed network structure to conduct cross-subject experiments on three datasets to explore the adaptability of MSHCNN to different subjects. There are 9 subjects in Dataset A and B respectively. We use the data of the first subject for testing, and the data of the remaining 8 subjects for training. Then we select the second subject as the test set until the ninth subject is selected as the test set. There are 6 subjects in Dataset C. In the same way, we select one of the subjects as the test set, and the rest as the train set. The experimental results are shown in Table IV, where we perform cross-subject experiments using four models, MSHCNN, EEGNet, DeepNet, and MSNN, where A, B, and C denote the datasets. The MSHCNN has achieved competitive results on three datasets. The average classification accuracy rate of 9 subjects on Dataset A is 76.03%, the average classification accuracy rate on Dataset B reaches 72.60%, and the average accuracy rate of 6 subjects on Dataset C is 72.28%. Although we found that the MSNN outperforms our method on Dataset C, it did not perform well on the other two datasets. In general, our proposed method fares better in cross-subject experiments compared to the other three networks.

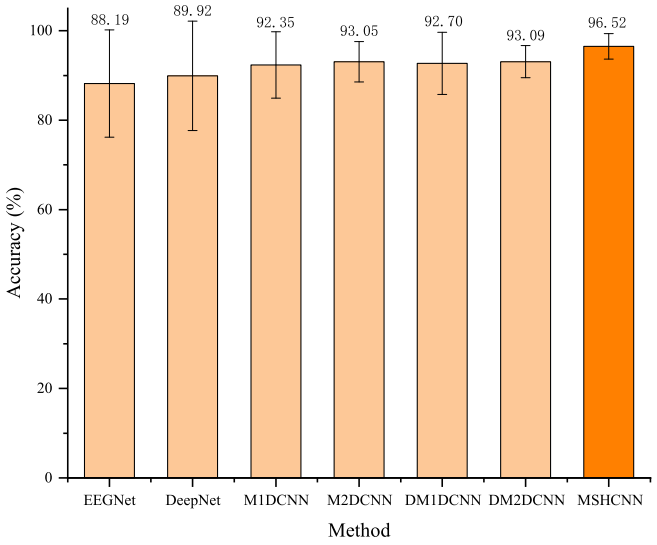

Fig. 10. Experimental results of ablation experiment on Dataset C.

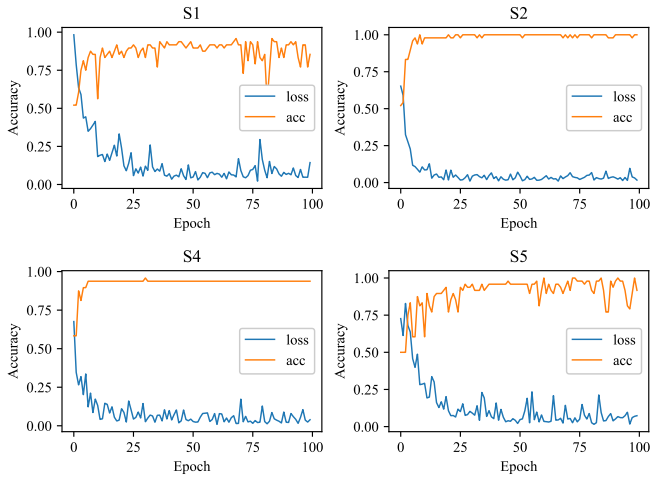

Fig. 11. Training loss and validation accuracy curves for subjects S1, S2, S4 and S5 in dataset C.

E. Online Experiments

We apply the proposed algorithm to the actual control system and design a BCI-based online control system for intelligent artificial limb. The intelligent artificial limb system mainly includes EEG signal acquisition equipment, signal processing equipment, microprocessor, and artificial limb. As shown in Fig. 12. It is the intelligent artificial limb control system. The real-time EEG data is obtained through the TCP/IP protocol, then preprocessed before sent to the model to get the classification results. The results are fed back to the subject, and the results are converted into control instructions. The instructions are sent to the STM32 microprocessor through Bluetooth to control the grip of the artificial limb.

Different from dataset C, we use new equipment produced by Brain Products in Germany to collect EEG signals and have conducted an online experiment on three subjects aged 23 to 27. The device consists of an actiCHamp amplifier,

TABLE IV
THE AVERAGE PERCENTAGE CLASSIFICATION ACCURACY OF CROSS-SUBJECT EXPERIMENTS ON THREE DATASETS

Method	S1	S2	S3	S4	S5	S6	S7	S8	S9	Avg	Std
MSHCNN_A	76.80	66.32	57.36	91.75	79.59	82.63	74.16	80.13	75.55	76.03	9.79
MSHCNN_B	69.44	59.37	88.88	75.34	64.58	69.09	59.02	80.22	87.50	72.60	11.17
MSHCNN_C	74.58	77.50	62.91	78.33	87.91	52.50	-	-	-	72.80	12.58
EEGNet_A	78.61	66.47	55.13	91.08	78.64	71.25	69.72	60.26	58.05	69.91	11.54
EEGNet_B	61.11	54.86	76.04	67.36	71.52	63.19	60.06	70.48	83.68	67.59	8.90
EEGNet_C	63.33	56.25	58.33	62.08	51.66	53.75	-	-	-	57.56	4.59
DeepNet_A	75.41	63.82	57.50	92.43	78.37	81.52	75.83	80.00	75.83	75.63	10.07
DeepNet_B	71.18	56.25	86.11	71.35	73.61	61.45	66.66	75.69	64.58	69.65	8.73
DeepNet_C	67.08	78.75	62.08	82.91	86.25	52.91	-	-	-	71.66	13.06
MSNN_A	74.72	65.29	57.63	91.21	74.72	85.55	72.91	76.57	76.66	75.02	9.87
MSNN_B	58.68	52.77	68.05	61.80	55.20	54.16	54.51	57.98	50.00	57.01	5.40
MSNN_C	65.83	77.91	63.75	87.91	88.33	51.25	-	-	-	72.49	14.76

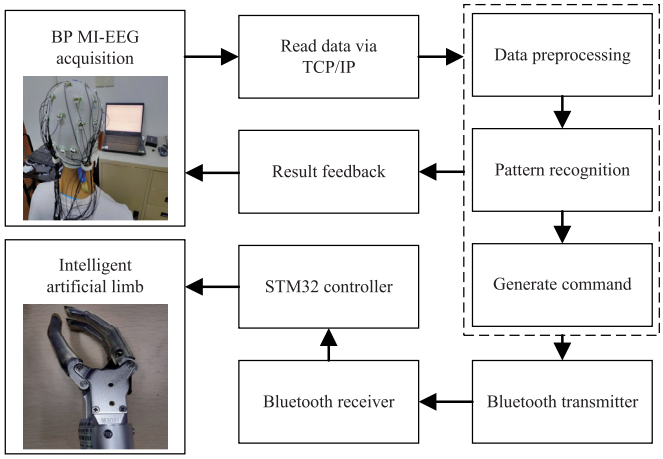

Fig. 12. The structure of the intelligent manipulator system.

electrode cap, signal recording software, and analysis software. The data collection paradigm consists of three parts. The first part is the preparation phase of 2s, where a white plus sign is displayed on the screen; the second part is the motor imagery phase of 4s, where the left and right white arrows alternately appear on the screen; the third part is the rest phase of 4s, and the screen is black. We use a combination of EMG and EEG control, using the clenching of teeth as the start signal of the experiment. After detecting the clenching signal, the subjects start motor imagery for four seconds and use the left-hand and right-hand motor imagery to control the grasping and releasing of the manipulator. After many experimental observations, we select the FT9 channel as the EMG signal detection of teeth clenching and use the variance within 0.2s to detect whether the teeth are clenched, and the success rate is 100%. C3, Cz, and C4 are selected as MI-EEG data channels. To better understand our experimental procedure, we provide a video of a successful live demonstration in our supplementary material. In addition, the EEG signal is classified by the method we proposed, and the EEGNet method is utilized for comparative experiments.

For each subject, we collect 600 samples, 300 left-hand and right-hand motor imagery, respectively. Each session collects 100 samples and rests 5-10 minutes in between. It is divided into training set and validation set at a ratio of 5:1. A few days later, we conducted an online control experiment, in which

TABLE V
PERCENTAGE CLASSIFICATION ACCURACY OF ONLINE EXPERIMENTS

Subject	SA_EEGNet		SA_MSHCNN	
	validation	online	validation	online
S1	84	73	88	81
S2	82	74	84	79
S3	91	79	93	85
Avg	85.67	75.33	88.33	81.67
Std	4.72	3.21	4.51	3.05

each subject performed 100 motor imagery tasks, alternating the left and the right hand. The accuracy of the validation set and online experimental results are shown in Table V.

From the experimental results in Table V, it can be concluded that the proposed method obtains a higher average accuracy than the EEGNet model. The EEGNet model achieves an average online control accuracy rate of 75.33%, while our proposed method achieves 81.67%. In addition, results of the EEGNet model differ from the average accuracy of the online experiments by 10.34% on the validation set. The experimental results of our proposed method differ by 6.66%, indicating that our proposed method is more adaptable.

V. CONCLUSION

Based on the theory of deep learning, this paper proposes a multi-scale hybrid convolutional neural network, which extracts the depth temporal and spatial features of EEG signals from multiple scales. In addition, a more robust coding method for EEG signals is proposed. We use BCI Competition IV 2b, BCI Competition IV 2a, and Laboratory data datasets to verify the effectiveness of our proposed method. Compared with traditional methods and deep learning methods, our proposed network achieves higher average accuracy rates of 85.25%, 84.86%, and 96.87%, respectively. Competitive results are also obtained in cross-subject experiments. Experiments show that our method can effectively extract the temporal and spatial features of EEG signals, and can be used in brain-computer interface systems. In addition, we apply our method to the online artificial limb control system, and the classification accuracy in real-time control reaches 81.67%.

For future avenues for research on brain-computer interface in the field of athletic rehabilitation, we believe it is worthwhile to: (1) Increase the categories of motor imagery

classification to provide more control commands; (2) Build 3D EEG data, introduce 3D convolutional neural networks, and extract EEG signals features in three-dimensional space; (3) Introduce other bioelectrical signals (Electrooculogram signals, electromyographic signals), and combine them with MI-EEG signals to study a hybrid brain-computer interface system.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] D.-Y. Lee, M. Lee, and S.-W. Lee, "Decoding imagined speech based on deep metric learning for intuitive BCI communication," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1363–1374, 2021.

[3] A. Cruz, G. Pires, A. Lopes, C. Carona, and U. J. Nunes, "A self-paced BCI with a collaborative controller for highly reliable wheelchair driving: Experimental tests with physically disabled individuals," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 2, pp. 109–119, Apr. 2021.

[4] R. Mane, T. Chouhan, and C. Guan, "BCI for stroke rehabilitation: Motor and beyond," *J. Neural Eng.*, vol. 17, no. 4, Aug. 2020, Art. no. 041001.

[5] W. Zhao, X. Zhang, J. Qu, J. Xiao, and Y. Huang, "A virtual smart home based on eeg control," in *Proc. IEEE 9th Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC)*, Jul. 2019, pp. 85–89.

[6] D. Szajerman, M. Warycha, A. Antonik, and A. Wojciechowski, "Popu lar brain computer interfaces for game mechanics control," in *Multi media and Network Information Systems*. Cham, Switzerland: Springer, 2017, pp. 123–134. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-43982-2_11

[7] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of eeg-based brain-computer interface paradigms," *J. Neural Eng.*, vol. 16, no. 1, pp. 1–21, 2019.

[8] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 42, no. 6, pp. 817–826, Jun. 1977.

[9] Q. Xiong, X. Zhang, W.-F. Wang, and Y. Gu, "A parallel algorithm framework for feature extraction of EEG signals on MPI," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–10, May 2020.

[10] C. Kim, J. Sun, D. Liu, Q. Wang, and S. Paek, "An effective feature extraction method by power spectral density of EEG signal for 2-class motor imagery-based BCI," *Med. Biol. Eng. Comput.*, vol. 56, no. 9, pp. 1645–1658, Sep. 2018.

[11] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for eeg decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[12] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Jul. 2018, Art. no. 056013.

[13] J.-H. Jeong, K.-H. Shim, D.-J. Kim, and S.-W. Lee, "Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1226–1238, May 2020.

[14] X. Tang, W. Li, X. Li, W. Ma, and X. Dang, "Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113285.

[15] W. Ko, E. Jeon, S. Jeong, and H.-I. Suk, "Multi-scale neural network for EEG representation learning in BCI," *IEEE Comput. Intell. Mag.*, vol. 16, no. 2, pp. 31–45, May 2021.

[16] H. Wang and W. Zheng, "Local temporal common spatial patterns for robust single-trial EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 2, pp. 131–139, Apr. 2008.

[17] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, no. 1, p. 39, 2012.

[18] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 566–576, Jun. 2016.

[19] Ji, Ma, Dong, and Zhang, "EEG signals feature extraction based on DWT and EMD combined with approximate entropy," *Brain Sci.*, vol. 9, no. 8, p. 201, Aug. 2019.

[20] W. B. Ng, A. Saidatul, Y. Chong, and Z. Ibrahim, "PSD-based features extraction for EEG signal during typing task," in *Proc. IOP Conf. Mater. Sci. Eng.*, vol. 557, no. 1, Jun. 2019, Art. no. 012032.

[21] M. Demuru, S. M. La Cava, S. M. Pani, and M. Fraschini, "A comparison between power spectral density and network metrics: An EEG study," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101760.

[22] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K.-R. Müller, "Toward unsupervised adaptation of LDA for brain–computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 587–597, Mar. 2010.

[23] S. Siuly and Y. Li, "Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 526–538, Jul. 2012.

[24] Z. Jia, Y. Lin, J. Wang, K. Yang, T. Liu, and X. Zhang, "Mmcnn: A multi-branch multi-scale convolutional neural network for motor imagery classification," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2020, pp. 736–751. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-67664-3_44

[25] X. Liu, Y. Shen, J. Liu, J. Yang, P. Xiong, and F. Lin, "Parallel spatial–temporal self-attention CNN-based motor imagery classification for BCI," *Frontiers Neurosci.*, vol. 14, p. 1157, Dec. 2020.

[26] Y. Song, X. Jia, L. Yang, and L. Xie, "Transformer-based spatial–temporal feature learning for EEG decoding," 2021, *arXiv:2106.11170*.

[27] Y. Han, B. Wang, J. Luo, L. Li, and X. Li, "A classification method for EEG motor imagery signals based on parallel convolutional neural network," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103190.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[30] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[31] Y. Zhang, Z. Qiu, T. Yao, D. Liu, and T. Mei, "Fully convolutional adaptation networks for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6810–6818.

[32] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2015, *arXiv:1508.06576*.

[33] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.

[34] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set B," Graz Univ. Technol., Graz, Austria, Tech. Rep., 2008, pp. 1–6.

[35] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set A," Inst. Knowl. Discovery (Lab. Brain-Comput. Interfaces), Graz Univ. Technol., Graz, Austria, 2008, pp. 1–6, vol. 16.

[36] X. Tang, T. Wang, Y. Du, and Y. Dai, "Motor imagery EEG recognition with KNN-based smooth auto-encoder," *Artif. Intell. Med.*, vol. 101, Nov. 2019, Art. no. 101747.

[37] P. Chen, Z. Gao, M. Yin, J. Wu, K. Ma, and C. Grebogi, "Multiattention adaptation network for motor imagery recognition," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 8, pp. 5127–5139, Aug. 2022.

[38] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[39] Y. Zhang et al., "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces," *Expert Syst. Appl.*, vol. 96, pp. 302–310, Apr. 2018.

[40] A. Zhao, L. Qi, J. Dong, and H. Yu, "Dual channel LSTM based multi-feature extraction in gait for diagnosis of neurodegenerative diseases," *Knowl. Based Syst.*, vol. 145, pp. 91–97, Apr. 2018.