Research Article

A Novel Time-Incremental End-to-End Shared Neural Network with Attention-Based Feature Fusion for Multiclass Motor Imagery Recognition

Shidong Lian,[1,2,3] **Jialin Xu** [iD],[2,3] **Guokun Zuo,**[2,3] **Xia Wei** [iD],[1] **and Huilin Zhou**[2,3,4]

[1]*College of Electrical Engineering, Xinjiang University, Urumqi 830047, China*
[2]*Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering,*
 Chinese Academy of Sciences, Ningbo, Zhejiang 315201, China
[3]*Zhejiang Engineering Research Center for Biomedical Materials, Ningbo Institute of Materials Technology and Engineering,*
 Chinese Academy of Sciences, Ningbo, Zhejiang 315300, China
[4]*University of Chinese Academy of Sciences, Beijing 100049, China*

Correspondence should be addressed to Jialin Xu; xujialin@nimte.ac.cn and Xia Wei; 30462111@qq.com

In the research of motor imagery brain-computer interface (MI-BCI), traditional electroencephalogram (EEG) signal recognition algorithms appear to be inefficient in extracting EEG signal features and improving classification accuracy. In this paper, we discuss a solution to this problem based on a novel step-by-step method of feature extraction and pattern classification for multiclass MI-EEG signals. First, the training data from all subjects is merged and enlarged through autoencoder to meet the need for massive amounts of data while reducing the bad effect on signal recognition because of randomness, instability, and individual variability of EEG data. Second, an end-to-end sharing structure with attention-based time-incremental shallow convolution neural network is proposed. Shallow convolution neural network (SCNN) and bidirectional long short-term memory (BiLSTM) network are used to extract frequency-spatial domain features and time-series features of EEG signals, respectively. Then, the attention model is introduced into the feature fusion layer to dynamically weight these extracted temporal-frequency-spatial domain features, which greatly contributes to the reduction of feature redundancy and the improvement of classification accuracy. At last, validation tests using BCI Competition IV 2a data sets show that classification accuracy and kappa coefficient have reached $82.7 \pm 5.57\%$ and 0.78 ± 0.074, which can strongly prove its advantages in improving classification accuracy and reducing individual difference among different subjects from the same network.

1. Introduction

The brain-computer interface (BCI) is a communication control system established between the brain and the external devices through the signals generated by brain activity. Creating direct communication between the brain and the external device, the system does not rely on muscles or peripheral nerves but the central nervous system [1]. Motor imagery (MI) is a psychological process in which an individual simulates the body movements. During the process of performing different MI tasks, when a certain area of the cerebral cortex is activated, the metabolism and blood flow of this area increase. Meanwhile, a simultaneous information processing will lead to an amplitude decrease or even block of EEG in its mu and beta spectrum oscillation. This electrophysiologic concept is called event-related desynchronization (ERD). In contrast, the phenomenon of a manifest amplitude increase of mu and beta oscillation, which appears in resting or inert states, is called event-related synchronization (ERS) [2].

The purpose of MI-BCI is to identify the imagined movements by classifying the electroencephalogram (EEG) characteristics of the brain, to control the external devices, such as robots [3, 4]. On the one hand, MI-BCI can help

patients with severe dysfunction and establish communication channels with the outside world. On the other hand, to some extent, it can activate the brain region to promote the remodeling of the patient's central nervous system [5]. In contrast to the traditional rehabilitation training, it can improve the patient's subjective initiative to achieve the rehabilitation effect, which overcomes the defect of the passive and single means of traditional rehabilitation [6]. Therefore, MI-BCI has a growing potential value in the fields of motor function assist and motor neurorehabilitation. However, the high complexity and instability of the EEG signals make the feature extraction and pattern classification of signals very challenging.

The very important part of the MI-BCI system is how to classify the EEG characteristics of MI task correctly and convert it into external control instructions [4]. At present, the traditional MI-EEG signal feature extraction is mainly based on ERD/ERS in the μ band (8–12 Hz) and the β band (16–31 Hz), including signal bandpass filtering [7], autoregressive model [8], frequency domain statistics [9], phase-locking value (PLV) [10], wavelet transformation and wavelet-packet transformation [11, 12], information entropy [13], and common spatial pattern (CSP) [14]. Based on the above methods, Li et al. [15] used wavelet-packet transform (WPT) to analyze and rebuild the MI-EEG signals and extract the energy characteristics of the μ band and the β band. Zhang et al. [16] analyzed MI-EEG signals and extracted temporal and spatial features by using a one-versus-rest filter. However, traditional feature extraction relies on manual selection of specific frequency bands, and features are very limited, which may lose part of the EEG information. In addition, the pattern classification methods of MI-EEG signals include linear discrimination analysis (LDA) [17], bayesian linear discrimination analysis (BLDA) [18], logic regression (LR) [19], support vector machine (SVM) [20], and neural network (NN) [21]. The classification performance of these methods depends on the quality of feature extraction.

In more recent years, deep learning has made excellent achievements in the fields of speech recognition, image recognition, and natural language processing [22, 23]; it has been used as a good machine learning method in these fields for its advantages on self-learning of features [24–26]. Therefore, deep learning is also gradually used in the feature extraction and pattern classification of EEG signals, in some cases not only improve the accuracy, but also provide a new method to learn features from EEG data [27, 28]. For example, Tang et al. [29] investigated how convolution neural network (CNN) displayed spectral features of the series of MI-EEG samples. Yang et al. [30] used augmented CSP to extract the spatial features and CNN to learn deep structural features for MI-EEG classification on the BCI Competition IV data sets, which revealed that feature extraction no longer relied on manual ways. Tabar and Halici [31] proposed using a kind of deep learning method to classify MI-EEG signal patterns, CNN was used to extract features, and then stacked autoencoder (SAE) was used to classify the extracted features. However, MI-EEG signals are time series with strong time-varying characteristics, and CNN is not completely suitable for learning time-series features. Therefore, Lee and

Choi [32] used continuous wavelet transform to extract the temporal-frequency features of MI-EEG signals and classified them by CNN. Zhou et al. [33] adopted a way based on wavelet-packet and long short-term memory (LSTM) neural network, which divided MI-EEG signals into several categories through the amplitude features and time-series information. An et al. [34] did some research on deep belief network (DBN) based on the restricted Boltzmann machine (RBM) linked up with fast Fourier transform (FFT) for MI-BCI pattern recognition, and the results were significantly better than those of traditional SVM-based algorithm. However, these methods simply extracted the temporal domain, frequency domain, or temporal-frequency domain features and did not fully extract the EEG signal features. Many other methods of deep learning have also been used in the recognition of MI-EEG signals, but the network structure is overly complex.

To sum up, all the above deep learning methods used in the recognition of MI-EEG signals do not take full advantages on self-learning of features, which still manually select features of specific frequency bands before pattern classification. Because the features selected manually are very limited and the objective function of feature extraction is different from that of pattern classification, it is easy to get information loss. What is more, multiclass MI-BCI classification mainly adopts splitting strategy. The whole process is extremely cumbersome and the classification accuracy is not high. Besides, the signal-to-noise ratio of MI-EEG signals is relatively low, and the data of the same person in the same task has randomness, instability, and individual variability, which makes the network trained with small-sample data sets have limitations. To reduce these limitations, a multimodal neural network is designed to form a novel end-to-end shared neural network in this paper. The main contributions are as follows:

(1) For the sample size of BCI Competition IV 2a data sets is small, 1s time window is used to intercept the training data, and then autoencoder (AE) network is used to enlarge the sample size of all subjects' training data which is intercepted and merged in advance. It meets the requirement of a large amount of training data for neural network and effectively reduces the bad effect on signal recognition because of randomness, instability, and individual variability of EEG data.

(2) To ensure the classification results of MI-EEG signals, a novel convolution neural network structure named the shallow convolution neural network (SCNN) is proposed to extract the different-dimension frequency-spatial domain features. Because of its simple structure and fewer parameters, the training model is not easy to overfitting. Furthermore, the EEG signals processed in the frequency-spatial domain are input into bidirectional long short-term memory (BiLSTM) network to extract the time-series features, so that the features in the MI-EEG signals are fully extracted. Finally, to reduce the redundancy of the fusion features and improve the

classification accuracy, the attention model is introduced into the feature fusion layer to dynamically weight the extracted temporal-frequency-spatial domain features.

(3) Through the proposed multimodal neural network, the training data of all subjects is used to train the end-to-end shared neural network, and it is tested by the test data of each subject and compared with the state-of-the-art methods in the MI-EEG recognition field to prove its higher classification accuracy and the minimum individual difference.

The structure of this paper is as follows: Section 1 is the introduction. Section 2 describes the data sets and the details of the neural network method that we proposed. The experiment and its results are presented in Section 3. The discussion is presented in Section 4. Finally, Section 5 is the conclusion.

2. Materials and Methods

Different from images and videos, MI-EEG signals are time series with strong time-varying characteristics, which own a mass of data information while the amount of data is not large. In addition, its signal-to-noise ratio is low, and randomness, instability, and individual variability still exist in the process of signal acquisition. What is more, with the increase in the number of MI-EEG classifications, the strategy that multiclass task was split and the method that patterns were classified after feature extraction were both introduced in the past, but it is still difficult to improve the classification accuracy. To solve these problems, we put forward a method of an attention-based time-incremental end-to-end shared neural network, as shown in Figure 1. With a combination of SCNN network and BiLSTM network, and an attention model introduced into the feature fusion process, it is practicable for feature extraction and pattern classification in a step-by-step way for the temporal-frequency-spatial domain features of multiclass MI-EEG signals. This method is simply called the method of SCNN-BiLSTM network based on attention.

Before feature extraction and pattern classification are carried out in a step-by-step way, the training data of all subjects is expanded using AE network. Then, the different-dimensional frequency-spatial domain features which were abstract are extracted by different convolutional kernels of SCNN network, and the time-series features are extracted through BiLSTM network with time increments; after that, all the temporal-frequency-spatial domain features are combined with the attention mechanism. Finally, the above fusion features are input to output layer of the network for classifying. During the training process of this attention-based time-incremental end-to-end shared neural network, the convolution layers and the recurrent layers can receive the reverse propagation error of the output layer at the same time, and the gradient drop caused by the error will gradually spread to the front of the network. So, after many iterations, the network parameters are gradually updated, and the error will become smaller and smaller.
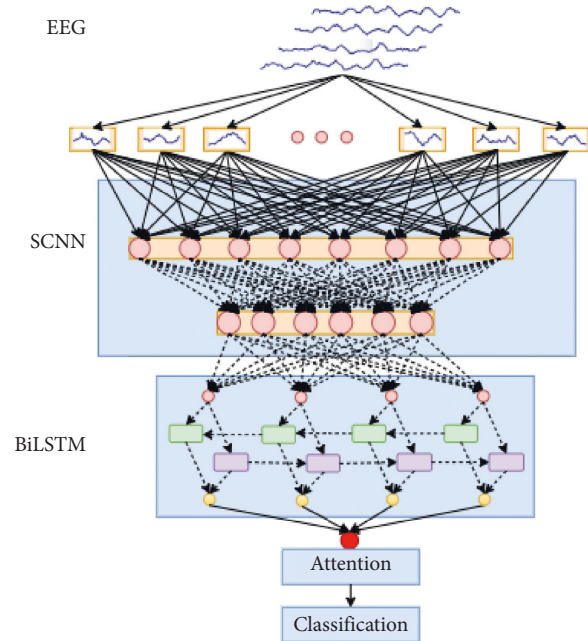

Figure 1: Our method of SCNN-BiLSTM network based on attention. The EEG signals are input to convolution layers and therefore preprocessed time series are generated. Moreover, the time series are input to BiLSTM cells for the exchange of information among different time points. And the attention mechanism module receives the output of BiLSTM cells, calculates weights for different time points, and outputs the ultimate result.

2.1. Data Description and Processing. In this paper, the data sets are taken from four-class MI-EEG data of left hand, right hand, foot, and tongue in BCI Competition IV 2a in 2008 [35]. In the data sets, the EEG data of 9 subjects was recorded with 22 Ag/AgCl electrodes and labeled as A01–A09. The sample frequency is 250 Hz, band-pass filtering is set between 0.5 Hz and 100 Hz, and line noise is suppressed by a 50 Hz notch filter. Each experiment consists of two sessions. The first session is training and the second session is testing. One run contains 48 trials (12 for each of the four possible classes), resulting in 288 trials per session. The timing scheme of experimental data acquisition is shown in Figure 2.

Firstly, we select 2 s–6 s data from the training data sets T and intercept it with a time window of 1 s. After processing, the training data sets T of 9 subjects are merged and then enlarged with AE network. Secondly, to accelerate the convergence speed of the network, prevent interference caused by abnormal EEG data, and avoid unnecessary numerical problems, the segmented training data is standardized. What is more, to increase the stability of the network, the training data sets T with training labels are reordered randomly.

Particularly, in the process of data standardization, we standardize the EEG data based on the mean and standard deviation of the raw data, so as to avoid the influence of outliers and extreme values in the data through centralization. The processed EEG data conforms to the standard normal distribution with mean 0 and standard deviation 1.
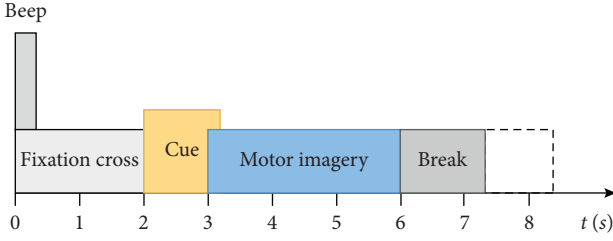
FIGURE 2: The paradigm timing scheme for MI-EEG data of BCI Competition IV 2a.

The training sets and testing sets are both standardized before being input into the network as follows:

$$s_{in}(i) = \frac{s_{in}(i) - \text{mean}(s_{in})}{\text{std}(s_{in})}, \tag{1}$$

$$p_{in}(i) = \frac{p_{in}(i) - \text{mean}(p_{in})}{\text{std}(p_{in})}, \tag{2}$$

where s_{in} is the segmented training data and p_{in} is the segmented testing data.

And, during a single trial of BCI Competition IV 2a data, when $t = 2$ s, the prompt arrow appeared and lasted for 1.25 s; the subjects observed and imagined the corresponding action. When $t = 3.25$ s~6 s, the subjects imagined the corresponding action. Because the EEG signals are instantaneous and susceptible to interference, in the process of motor imagery, the ERD/ERS characteristics of subjects' motor imagery EEG are uncertain during the transition from the preparation stage to the imagination stage, which is easy to cause invalid edge data. So, the method proposed in this paper is verified by selecting the motor imagery data of 4 s~5 s with 0.75 s interval from $t = 3.25$ s in the final.

2.2. Data Expansion Based on Autoencoder. The deep learning methods need a large amount of data to train the network models. However, the data samples of BCI Competition IV 2a are small. In the meantime, the different periods and the size of the electrode caps during the data acquisition process make each subject's EEG signals have randomness and instability. Therefore, for the training sets, we select the data of 2 s~6 s and use the 1 s time window to intercept and then use the autoencoder (AE) network to enlarge the data by generating the reconstructed data from the real training data in this paper, which is a three-layer neural network composed of an input layer, a hidden layer, and an output layer [36], as shown in Figure 3. After processing, it satisfies the need for a mass of training data of neural networks and improves the robustness of the network model, while effectively reducing the bad effect on signal recognition because of randomness, instability, and individual variability of EEG data.

In AE network, the output layer y has the same size as the input layer x, so y can be considered as an approximation of x. f and g represent the encoding and decoding functions, respectively. The encoding and decoding procedures are as follows:

$$h = f(x) = s_f(W_1 x + a), \tag{3}$$

$$y = g(h) = s_g(W_2 h + b), \tag{4}$$

where h denotes the hidden layer information, W_1 denotes weight of the input layer to the hidden layer, and W_2 denotes weight of the hidden layer to the output layer; a and b are biases of the hidden layer and the output layer, respectively; s_f and s_g are activation function of the encoding and decoding procedures, respectively. Here, both s_f and s_g adopt the sigmoid function. And, to simplify the calculation, let $W_2 = W_1^T = W$.

In this paper, the AE network firstly encodes the real training data to reduce the data dimension, and the important features of the data are retained through unsupervised learning. Then, the encoded data is decoded to obtain the reconstructed data. And, finally, the average value of the error reconstruction function between the reconstructed data and the real training data, namely, the loss cost function, is calculated to measure the similarity between them. The smaller the loss cost function is, the more similar the reconstructed data is to the real training data. However, in the process of network learning to obtain the parameter $\theta = \{W, a, b\}$, the value of the loss cost function will become smaller and smaller, which may result in overfitting. Therefore, we adopt cross-entropy in reconstruction error function to suppress overfitting to obtain an AE network with strong generalization ability. The reconstruction error function $R(x, y)$ is defined as follows:

$$R(x, y) = -\sum_{i=1}^{n} [x_i \log(y_i) + (1 - x_i)\log(1 - y_i)]. \tag{5}$$

For the entire training sets $S = \{x_1, x_2, \ldots, x_m\}$, the overall loss cost function $J(\theta)$ is

$$J(\theta) = \frac{1}{m} \sum_{x \in s} R(x, y). \tag{6}$$

The function $J(\theta)$ is minimized to obtain the parameter θ by the gradient descent method.

2.3. Shallow Convolution Neural Network. The structure of convolution neural network (CNN) is different from that of traditional hierarchical connections. The connections between neurons in CNN are not fully connected; what is more, the sharing weight of convolution kernel can reduce the complexity of the network model and reduce the weight parameters of network training, making it easier to train than the previous neural network [37].

Nevertheless, compared with the information volume of images and videos, that of EEG signals is very small. Besides, it is a kind of nonstationary, random, very weak, and low signal-to-noise ratios signal with unstable waveform. When classifying EEG signals, we found that too many convolutional layers of CNN can easily lead to overfitting of the training model. Therefore, it is very crucial to structure a suitable CNN model. In this paper, the different-dimensional frequency-spatial domain features which were
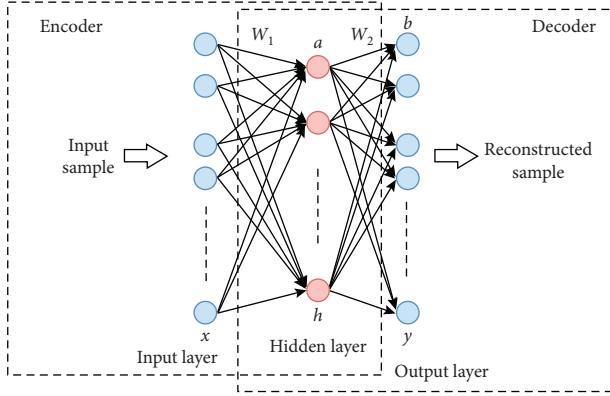
FIGURE 3: The structure of autoencoder neural network.

abstract are extracted by different convolutional kernels of shallow CNN network. The design of SCNN adopts the principle of the Visual Geometry Group (VGG) network [38]. It is a special CNN with a simple structure and few parameters. The training model is not easy to overfit and can directly extract the frequency-spatial domain features from the EEG data. The structure of SCNN is shown in Figure 4. The details of the structure used in this paper are shown in Table 1 in Section 3.

The input is one-dimensional feature vector $X = [x_1, x_2, x_3, \ldots, x_N]$ with a length of N corresponding to EEG signals of N channels; the convolution layer is composed of K convolution kernels, the size of each convolution kernel is $1 * S$, the coefficient of the convolution kernel is $w_k \in R^s$, $k = 1, 2, \ldots, K$, and the output is $h = [h_1, h_2, h_3, \ldots, h_k] \in R^{(N-S+1)K}$, where

$$h_k = R\left(\text{conv}\left(w_k * x + b_k\right)\right), \tag{7}$$

where b_k denotes the bias of the convolution kernel and R denotes the nonlinear activation function that adopts the Leaky ReLU function [39, 40].

$$f(x) = \begin{cases} x, & x > 0, \\ 0.01 * x, & x \leq 0. \end{cases} \tag{8}$$

In a single SCNN network structure, the network connects one or more full connection layers and a Softmax output layer after multiple convolution, pooling, and dropout layers. Supposing that the network has a total L layers, where the L_M layer is the full connection layer, the L_L layer is the final output layer, and the output number of cells is the number of classification categories n, the entire calculation process is as follows:

$$h^l = f\left(w^l X + b^l\right), \quad L_M \leq l < L, \tag{9}$$

$$a^L = w^L h^l + b^L, \tag{10}$$

$$P(t \mid x) = \text{Soft max}\left(a^L\right) = \frac{\exp\left(a_t^L\right)}{\sum_{t'=1}^n \exp\left(a_{t'}^L\right)}, \quad t = 1, 2, \ldots, n, \tag{11}$$

where h^l denotes the output of the convolution network's hidden layer information, w^l and b^l are the learning parameters of the network, a^L is the value which is not activated before the last output layer, and $P(t \mid x)$ is a posterior probability judging whether the input x belongs to the category t. The label for each input's EEG signal category is $T = [1, 2, \ldots, n]$. For all samples in the training sets, cross-entropy Loss $= (1/n) \sum_{i=1}^n \log p(t = i \mid X)$ is taken as the objective function to optimize.

2.4. Bidirectional Long Short-Term Memory Network. EEG signals are not images in the traditional sense, but time series with a strong correlation in time. The SCNN network is not fully suitable for learning time-series features of EEG signals; however, recurrent neural network has certain advantages on that [41, 42]. Therefore, in this paper, BiLSTM network with time increments which is a kind of recurrent neural network is connected in series before the full connection layer and after multiple convolution, pooling, and dropout layers of the SCNN network. Different from the traditional unidirectional LSTM network, the BiLSTM network improves on network structure so as to solve the gradient disappearance well and more fully extracts the information of each time point, which is suitable for EEG processing in temporal domain. The input at each moment of BiLSTM network comes from the information transmitted by the hidden layers in the forward and backward directions, and then the network combines the output of the forward and backward hidden layers to obtain its final output of each moment.

In this paper, to reduce the local convergence caused by fewer layers and the gradient disappearance caused by too many layers, the two-layer BiLSTM network is designed to converge more quickly and effectively reduce the gradient disappearance caused by too deep propagation between layers. The network's structure and principle are shown in Figure 5.

BiLSTM network is a unidirectional LSTM network when it is performing forward calculation, and the forward calculation requires the input data before the current time. The forward calculation of the network is as follows:

$$\overrightarrow{h_t} = W_{x_h^{\rightarrow}} x_t + W_{\overrightarrow{h}\overrightarrow{h}} \overrightarrow{h_{t-1}} + b_{\overrightarrow{h}}, \tag{12}$$

When it is performing backward calculation, it is associated with the future input data after the current time. The backward calculation of the network is as follows:

$$\overleftarrow{h_t} = W_{x_h^{\leftarrow}} x_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h_{t-1}} + b_{\overleftarrow{h}}. \tag{13}$$

The LSTM networks in the forward and backward directions maintain the state information of their own network, respectively, there is no connection between them, and the unfolded diagram of the network is not a circular structure. Having superimposed the state information coming from both directions simultaneously, then the output layer can be calculated. The overall calculation of the network is as follows:
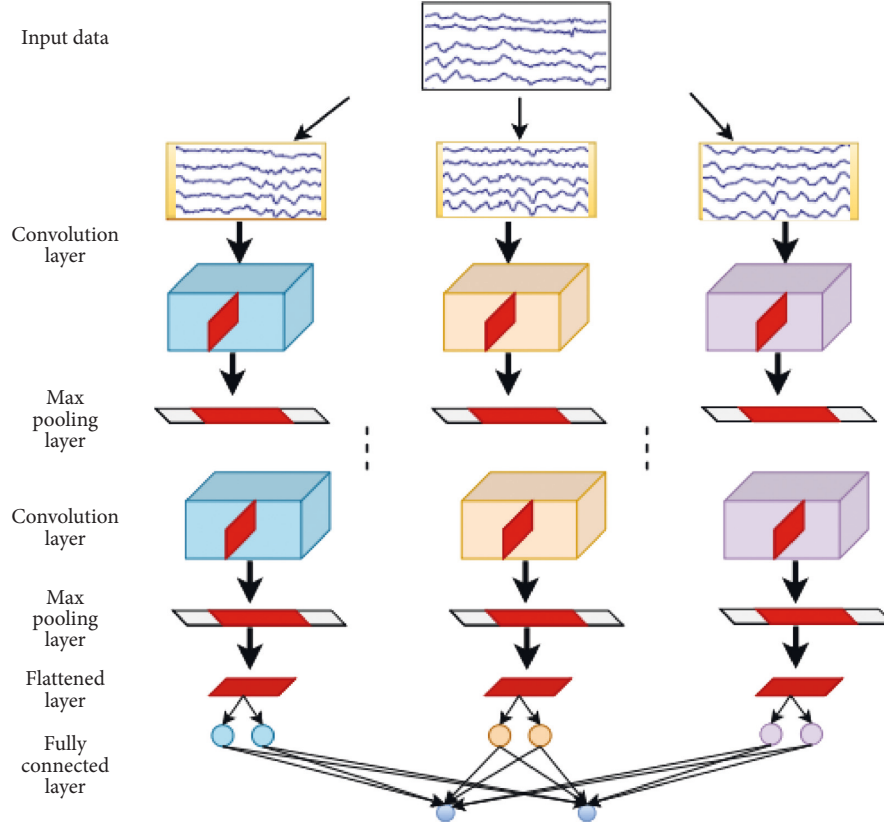
FIGURE 4: The structure of shallow convolution neural network. During the training, the input EEG data is $N * f$ (channel $*$ sample). Then, the maximum pooling layer is used to adjust the size of the convolution, which has kernels of $1 * S$ with a stride of $1 * S$.

TABLE 1: The details of the network model.

Layers	Type
Conv 1	Convolution 1×5
Norm 1	Batch normalization
Pool 1	Max pooling 1×2
Conv 2	Convolution 1×3
Norm 2	Batch normalization
Pool 2	Max pooling 1×2
Drop 2	Dropout layer
Conv 3	Convolution 1×3
Norm 3	Batch normalization
Drop 3	Dropout layer
Conv 4	Convolution 1×3
Norm 4	Batch normalization
Flattened	Flattened layer
LSTM 1	BiLSTM layer 32
LSTM 2	BiLSTM layer 20
Fc 5	Fully connected layer
Drop 5	Dropout layer
Attention	Attention layer
Fc 6	Fully connected layer
Classification	Softmax layer

$$y_t = W_{\overrightarrow{h} y} \overrightarrow{h}_t + W_{\overleftarrow{h} y} \overleftarrow{h}_t + b_y, \tag{14}$$

where x_t and y_t denote the input and the output layers, respectively; W and b represent the network's weights and biases, respectively.

When BiLSTM network is combined in series with SCNN network, the calculation method for the whole network is as follows:

$$h^l_{scnn} = f\left(w^l X + b^l\right), \quad L_M \leq l < L, \tag{15}$$

$$a^L_{SCNN} = w^L h^l_{scnn} + b^L, \tag{16}$$

$$a^t_{BiLSTM} = W_{\overrightarrow{h} a} \overrightarrow{h}_t + W_{\overleftarrow{h} a} \overleftarrow{h}_t + b_a n, \tag{17}$$

$$a_{BiLSTM} = U_t a^t_{BiLSTM}, \tag{18}$$

$$h_{FFL} = W_{FFL}\left(a^L_{SCNN} + a_{BiLSTM}\right) + b_{FFL}, \tag{19}$$

$$P(t|x) = \text{Soft max}\left(h_{FFL}\right) = \frac{\exp\left(h_{FFLt}\right)}{\sum_{t'=1}^{n} \exp\left(h_{FFLt'}\right)}, \quad t = 1, 2, \ldots, n, \tag{20}$$

where h^l_{scnn} is the hidden layer information that is output from the whole SCNN network. After a linear transformation, it is found that a^L_{SCNN} is a set of effective EEG features for different categories extracted by the network from the input EEG data. Having the hidden feature of EEG signals in temporal dimension, a_{BiLSTM} is the synthesis of
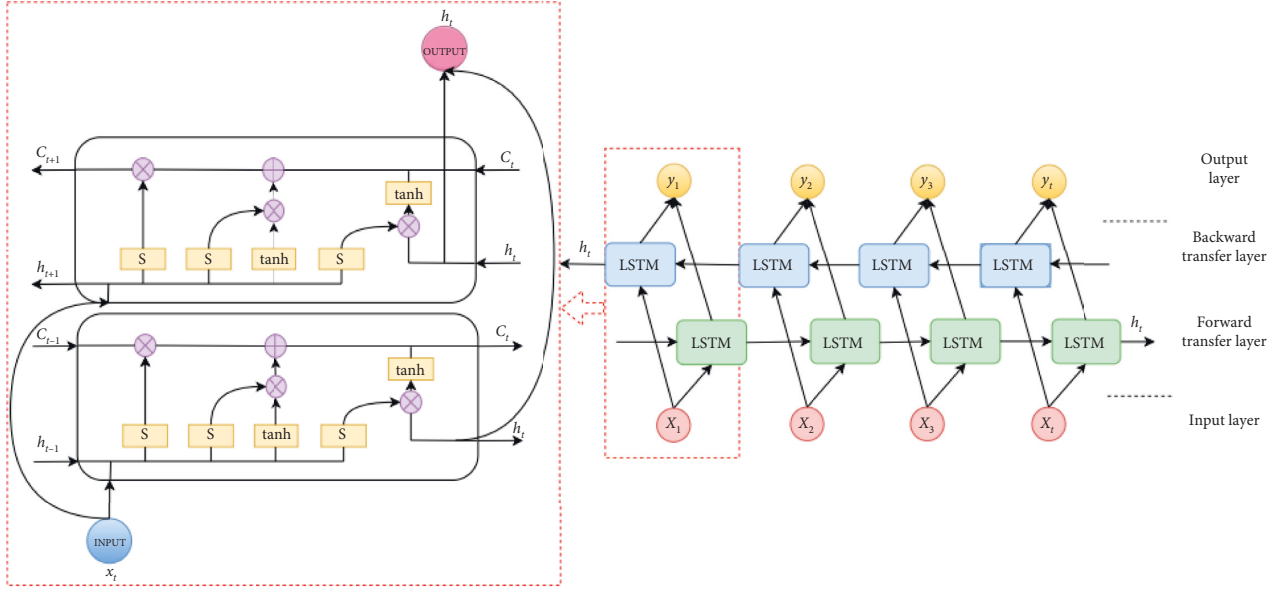
FIGURE 5: The structure and principle of bidirectional long short-term memory network with time increments. "*S*" represents sigmoid operator; "tanh" signifies hyperbolic tangent operator. "C_t" represents the state of BiLSTM cell at t moment.

a_{BiLSTM}^t output from BiLSTM network of all time nodes, and it can largely reflect the abnormal changes of EEG signals in the temporal dimension. Then, the frequency-spatial domain features extracted by SCNN network and the time-series features extracted by BiLSTM network are synthesized in the feature fusion layer and the fusion features of the temporal-frequency-spatial domain are obtained. Finally, the highly abstract features that have undergone multiple convolutions and cycles will be fused after a linear transformation. The relative proportions of the "good" and "bad" features are adjusted by learning weights from the training data; then, the proportions are sent into the output layer for the probability calculation of each category.

The above feature fusion is only synthesized in the direction of one-dimensional vector. The frequency-spatial domain features extracted by SCNN network and the time-series features extracted by BiLSTM network will have some redundancy. The fusion features, mechanically synthesized, will be redundant which will slow down the network training speed and then spoil final classification effect. Therefore, in this paper, attention mechanism is added to process the fusion features.

2.5. Attention Mechanism.
In cognitive science, to reasonably use the finite resources of visual information processing, humans usually choose to ignore part of the information and pay attention to the more critical part of all the information; that is to say, the brain's attention is focused on the specific visual area; this mechanism is called attention mechanism [43]. In this paper, the feature fusion process is optimized through the attention mechanism. The frequency-spatial domain features extracted by SCNN network and the time-series features extracted by BiLSTM network are fused and the important degree of the fusion feature is calculated to obtain the effective attention, so as to realize the automatic

classification of MI-EEG signals by more effectively fusing temporal-frequency-spatial domain features.

In traditional sequence-to-sequence learning, the encoder-decoder structure is often used for learning, as shown in Figure 6.

The encoder encodes the input sequence to get the intermediate state information C and then uses the intermediate vector as the input of the decoder to get the output of each sequence at the decoding end. The overall process is as follows:

$$P(y_t|\{y_1, y_2, \ldots, y_{t-1}\}, c) = g(y_{t-1}, s_t, c). \tag{21}$$

The output at each moment uses the same context semantic vector C, but, in the process of sequence encoding and decoding, we hope that the context semantic vector for each moment's output is an appropriate vector, so the attention mechanism is introduced to select the appropriate context semantic vector according to the output of different moments. The attention model is shown in Figure 7.

The decoding process for the attention model is as follows:

$$P(y_i|\{y_1, y_2, \ldots, y_{i-1}\}, x) = g(y_{i-1}, s_i, c_i), \tag{22}$$

$$s_i = f(s_{i-1}, y_{i-1}, c_i), \tag{23}$$

where c_i is the added attention; its role is to associate the output with the relevant input and to calculate the correlation a_{ij} between the current output and all inputs; then,

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j, \tag{24}$$

where h_j is the hidden layer information at the position j of the encoder's input. In this paper, the SCNN outputs EEG signal with frequency-spatial domain features as the input of
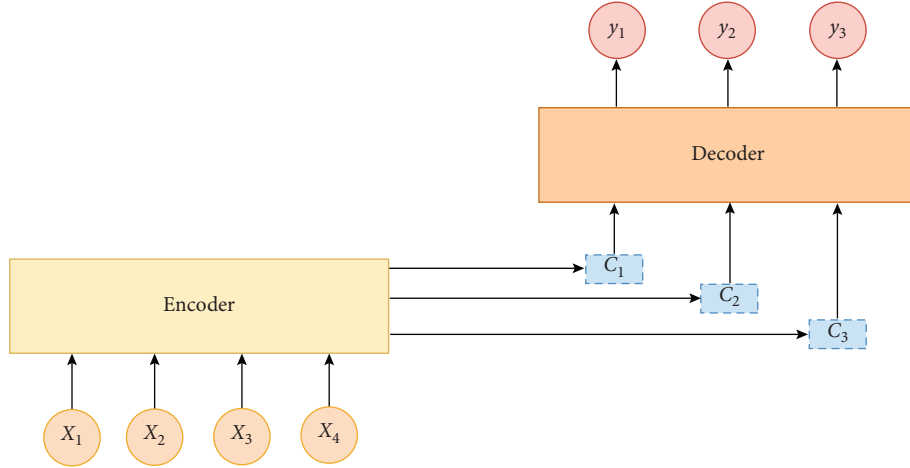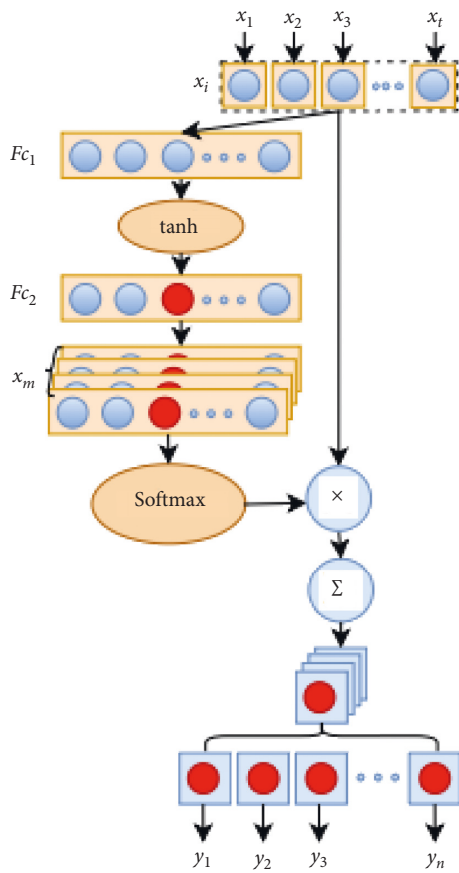
FIGURE 6: Encoder-decoder process.

FIGURE 7: The structure of attention model. In the attention model, there are two fully connected layers F_{c1} and F_{c2} and a hyperbolic tangent function tanh. The first fully connected layer calculates $x_i' = \sum_{k=1}^{t} w_{ik}x_k + b_i$ from each input feature vector $x_i (i = 1, \ldots, t)$, where b_i is the bias. The hyperbolic tangent function performs a nonlinear transformation on x_i' to obtain $x_i'' = \tanh(x_i')$. And, after the second full connected layer, $x_{mi} = \sum_{k=1}^{t} w_{mik}x_k'' + b_{mi}, (m = 1, \ldots, n; i = 1, \ldots, t)$ is obtained. Then, Softmax calculates the weight coefficient α_{mi} of each feature in x_m sequences. Therefore, the attention model outputs important features for each input feature vector x_i, which is $y_m = \sum_{i=1}^{t} \alpha_{mi}x_i, (m = 1, \ldots, n)$.

the encoder BiLSTM; that is, $h_j = [\overrightarrow{h_j}, \overleftarrow{h_j}]$, where the forward and backward of the hidden layer information are synthesized. Weight a_{ij} identifies the relevancy of the input sequence to the current output sequence, which is a normalized probability value, meaning the probability of the relationship between item j of the input and the output at the current moment. And weight a_{ij} is defined as

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}, \quad (25)$$

$$e_{ij} = a(s_{i-1}, h_j). \quad (26)$$

The definition of a_{ij} introduces the symbol e_{ij} as a feedforward neural network, which is jointly determined by the state information s_{i-1} of the hidden layer at the decoding end and h_j of the hidden layer at the encoding end. In the SCNN-BiLSTM network based on attention that we designed above, the attention module is an additional neural network, which can give different weights to each part of the fusion features and is more sensitive to the classification target; it can effectively enhance the performance of the whole neural network in a natural way.

3. Experiment and Results

We refine the model of SCNN-BiLSTM network based on attention that is designed in Section 2 and then train and test the model to verify its superiority in MI-EEG multiclass recognition. The model is trained and tested on the Intel 3.6 GHz Core i7-10700F CPU and 16 GB RAM NVIDIA GeFore RTX 2060 GPU.

The details of the network model are shown in Table 1.

For the above deep neural network model, the minibatch gradient descent method is used for network training. To accelerate the attenuation of the network, the Adam optimizer is used for the network model, so that the model converges to the optimal value [44]. In the training of the model, the setting of the learning rate and the selection of the minibatch size affect the model's final accuracy and training speed, so in this paper we fix the other parameters of the

model, constantly change the size and attention frequency of the learning rate, set the minibatch size in different sizes, and compare the final accuracy of the model. After multiple comparative experiments, the minibatch size takes 200 and the initial value of the learning rate is set to 0.001.

While training the neural network, random dropout and padding strategies are used. Among them, the random dropout strategy for the SCNN network can prevent the network model from overfitting the training data, while the padding strategy makes the output size of the convolution layer equal to the input size to prevent the loss of feature size [45]. In this paper, the random dropout parameter P is 0.2.

Figure 8 shows the training loss rate and accuracy curve of the neural network model after 500-time repeated training. It can be seen that, after 240 iterations, the training accuracy curve converges to 0.9 and the training loss rate is about 0.1.

We test the trained network model by the test data sets E; each point of the high-dimensional data of four-class MI-EEG features is assigned on the low-dimensional map and is avoided to concentrate in the center of the map, so as to form a scatter plot of T-distributed Stochastic Neighbor Embedding (T-SNE) [46], as shown in Figure 9. In the T-SNE scatter plot, the classification categories are represented by different colors, and it can be seen that all categories are clearly separated, but there is also some data hard to identify, which may be caused by interference during data acquisition.

Further, the network model which has been trained is measured by indicators such as accuracy, precision, sensitivity and specificity, and these indicators are calculated as follows:

$$ACC = \frac{TP_n + TN_n}{TP_n + TN_n + FP_n + FN_n}, \quad (27)$$

$$PPV = \frac{TP_n}{TP_n + FP_n}, \quad (28)$$

$$TPR = \frac{TP_n}{TP_n + FN_n}, \quad (29)$$

$$TNR = \frac{TN_n}{TN_n + FP_n}, \quad (30)$$

where TP represents the number of testing samples whose real value and model predicted value of classification category are both true, TN represents the number of testing samples whose real value and model predicted value of classification category are both negative, FP represents the number of testing samples whose real value of classification category is negative but their model predicted value is positive, and FN represents the number of testing samples whose real value of classification category is positive but their model predicted value is negative. Accuracy (ACC) is the proportion of the total number of model's correct judgments in the total model prediction results of testing samples. Precision (PPV) is the proportion of the number of model's correct judgments in the model prediction results of

testing samples whose predicted value of classification category is positive. Sensitivity (TPR) is the proportion of the number of model's correct judgments in the model prediction results of testing samples whose true value of classification category is positive. Specificity (TNR) is the proportion of the number of model's correct judgments in the model prediction results of testing samples whose true value of the classification category is negative. n denotes the classification categories. The trained model classifies the test data of subjects in BCI Competition IV 2a; the classification accuracy rate of each subject and average classification accuracy rate of all subjects are shown in Figure 10.

4. Discussion

Our method of SCNN-BiLSTM network based on attention in this paper is compared with the methods in the literature [16, 31, 47–52] and the classification accuracy of each method is measured by kappa coefficient. In the classification problem, the higher the kappa coefficient [53], the higher the classification accuracy. The kappa coefficient is calculated as follows:

$$Kappa = \frac{ACC - (1/C)}{1 - (1/C)}, \quad (31)$$

where C denotes the number of known categories and ACC is the average classification accuracy.

For analyzing the literature methods, see Table 2. The literature in [47] proposed the Filter Bank Common Spatial Pattern (FBCSP) method to extract features of MI-EEG signals and adopted the "one-versus-rest" multi-classification mechanism, which won the 2008 International Brain-Computer Interface Competition. The literature in [48] proposed an automatic method for the classification of general artifactual source components, which was a kind of Independent Component Analysis (ICA) for artifact removal in MI-EEG signals, and the classification accuracy was $69.7 \pm 14.2\%$. The literature in [49] proposed a method which spectral regression kernel discriminant analysis (SRKDA), with a classification accuracy of $78.4 \pm 14.0\%$. The literature in [50] proposed a method which combined CSP and Local features-scale Decomposition (LCD) to extract features of MI-EEG signals, with classification accuracy of $80.2 \pm 8.10\%$. The literature in [51] proposed a method of adaptive Stacked Regularized Linear Discriminant Analysis (SRLDA) to analyze the temporal, spatial, and spectral information of MI-EEG signals. The results showed that the adaptive SRLDA method was superior to the method of Data Space Adaptation (DSA) based on Kullback-Leibler divergence. However, the above-mentioned literature methods completely rely on human's current cognition of EEG signals and require relevant professional knowledge in the process of feature extraction, which makes the feature extraction too complicated and the classification effect poor. The literature in [52] proposed a method based on the combination of wavelet transformation and 2-layer CNN network, with classification accuracy of $81.2 \pm 28.5\%$. The literature in [16] proposed using "one-versus-rest" Filter
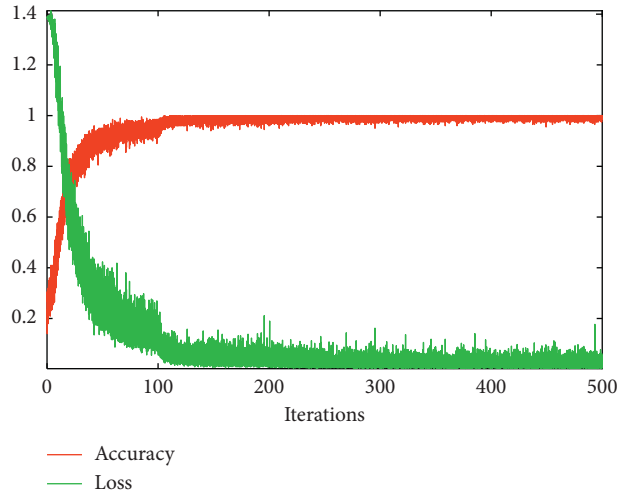
FIGURE 8: The relationship between the training loss rate, accuracy curve, and iterations. We repeated the training process of the network model for 500 times.
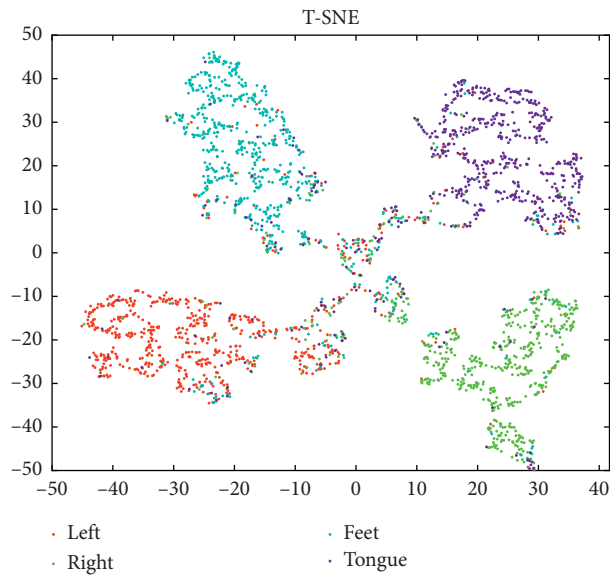

FIGURE 9: Visualizations of the high-dimensional data of four-class MI-EEG features.

Bank Common Spatial (OVR-FBCSP) mode to extract features of MI-EEG signals primarily; then, CNN and LSTM networks were applied to reextract and classify those primary processing features. The classification accuracy was $83.0 \pm 8.34\%$. Although these methods have achieved some accomplishments, they did not fully utilize the advantage of deep learning's self-learning characteristics and still followed the idea of manually extracting features first and then classifying patterns. Tabar and Halici [31] and Amin et al. [37] proposed new deep learning methods of feature extraction and pattern classification for MI-EEG signals, but the classification accuracy was not high, which was $66.2 \pm 11.2\%$ and $74.5 \pm 10.1\%$, respectively.

In this paper, we propose a method of an attention-based time-incremental end-to-end shared neural network. After extracting the frequency-spatial domain features by SCNN network and extracting the time-series features by BiLSTM network with time increments, the method effectively learns the temporal-frequency-spatial domain features of MI-EEG signals. Finally, an attention mechanism is added to the network feature fusion layer, and the extracted temporal-frequency-spatial domain features are dynamically weighted to reduce the redundancy of the fusion features and improve the classification accuracy rate to $82.7 \pm 5.57\%$. The results of comparison between our method and the literature methods are shown in Figure 11. It is obvious from (a) and (b) that, compared with the nondeep learning methods and deep
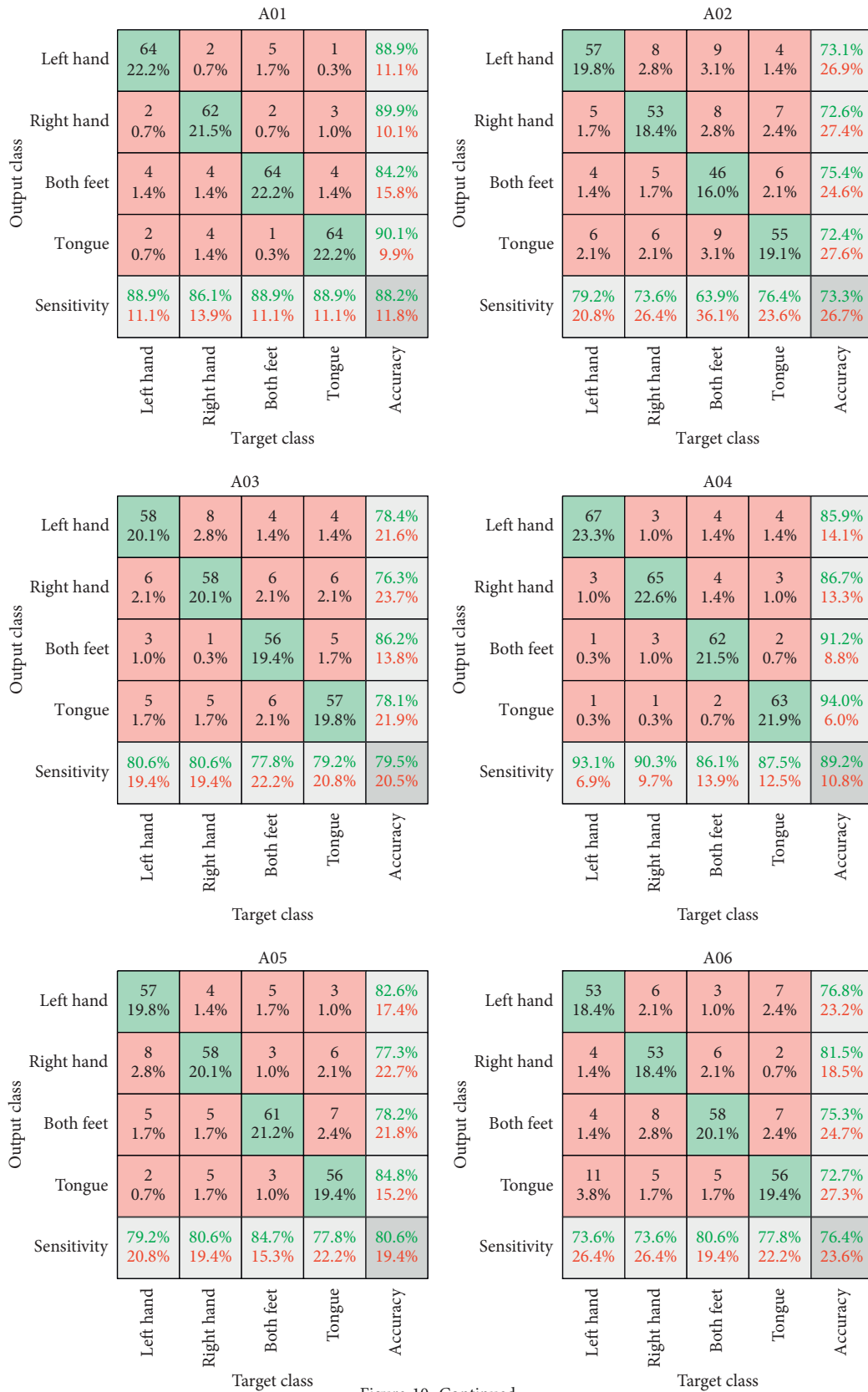
A01

Output class	Left hand	Right hand	Both feet	Tongue	Accuracy
Left hand	64 / 22.2%	2 / 0.7%	5 / 1.7%	1 / 0.3%	88.9% / 11.1%
Right hand	2 / 0.7%	62 / 21.5%	2 / 0.7%	3 / 1.0%	89.9% / 10.1%
Both feet	4 / 1.4%	4 / 1.4%	64 / 22.2%	4 / 1.4%	84.2% / 15.8%
Tongue	2 / 0.7%	4 / 1.4%	1 / 0.3%	64 / 22.2%	90.1% / 9.9%
Sensitivity	88.9% / 11.1%	86.1% / 13.9%	88.9% / 11.1%	88.9% / 11.1%	88.2% / 11.8%

Target class

A02

Output class	Left hand	Right hand	Both feet	Tongue	Accuracy
Left hand	57 / 19.8%	8 / 2.8%	9 / 3.1%	4 / 1.4%	73.1% / 26.9%
Right hand	5 / 1.7%	53 / 18.4%	8 / 2.8%	7 / 2.4%	72.6% / 27.4%
Both feet	4 / 1.4%	5 / 1.7%	46 / 16.0%	6 / 2.1%	75.4% / 24.6%
Tongue	6 / 2.1%	6 / 2.1%	9 / 3.1%	55 / 19.1%	72.4% / 27.6%
Sensitivity	79.2% / 20.8%	73.6% / 26.4%	63.9% / 36.1%	76.4% / 23.6%	73.3% / 26.7%

Target class

A03

Output class	Left hand	Right hand	Both feet	Tongue	Accuracy
Left hand	58 / 20.1%	8 / 2.8%	4 / 1.4%	4 / 1.4%	78.4% / 21.6%
Right hand	6 / 2.1%	58 / 20.1%	6 / 2.1%	6 / 2.1%	76.3% / 23.7%
Both feet	3 / 1.0%	1 / 0.3%	56 / 19.4%	5 / 1.7%	86.2% / 13.8%
Tongue	5 / 1.7%	5 / 1.7%	6 / 2.1%	57 / 19.8%	78.1% / 21.9%
Sensitivity	80.6% / 19.4%	80.6% / 19.4%	77.8% / 22.2%	79.2% / 20.8%	79.5% / 20.5%

Target class

A04

Output class	Left hand	Right hand	Both feet	Tongue	Accuracy
Left hand	67 / 23.3%	3 / 1.0%	4 / 1.4%	4 / 1.4%	85.9% / 14.1%
Right hand	3 / 1.0%	65 / 22.6%	4 / 1.4%	3 / 1.0%	86.7% / 13.3%
Both feet	1 / 0.3%	3 / 1.0%	62 / 21.5%	2 / 0.7%	91.2% / 8.8%
Tongue	1 / 0.3%	1 / 0.3%	2 / 0.7%	63 / 21.9%	94.0% / 6.0%
Sensitivity	93.1% / 6.9%	90.3% / 9.7%	86.1% / 13.9%	87.5% / 12.5%	89.2% / 10.8%

Target class

A05

Output class	Left hand	Right hand	Both feet	Tongue	Accuracy
Left hand	57 / 19.8%	4 / 1.4%	5 / 1.7%	3 / 1.0%	82.6% / 17.4%
Right hand	8 / 2.8%	58 / 20.1%	3 / 1.0%	6 / 2.1%	77.3% / 22.7%
Both feet	5 / 1.7%	5 / 1.7%	61 / 21.2%	7 / 2.4%	78.2% / 21.8%
Tongue	2 / 0.7%	5 / 1.7%	3 / 1.0%	56 / 19.4%	84.8% / 15.2%
Sensitivity	79.2% / 20.8%	80.6% / 19.4%	84.7% / 15.3%	77.8% / 22.2%	80.6% / 19.4%

Target class

A06

Output class	Left hand	Right hand	Both feet	Tongue	Accuracy
Left hand	53 / 18.4%	6 / 2.1%	3 / 1.0%	7 / 2.4%	76.8% / 23.2%
Right hand	4 / 1.4%	53 / 18.4%	6 / 2.1%	2 / 0.7%	81.5% / 18.5%
Both feet	4 / 1.4%	8 / 2.8%	58 / 20.1%	7 / 2.4%	75.3% / 24.7%
Tongue	11 / 3.8%	5 / 1.7%	5 / 1.7%	56 / 19.4%	72.7% / 27.3%
Sensitivity	73.6% / 26.4%	73.6% / 26.4%	80.6% / 19.4%	77.8% / 22.2%	76.4% / 23.6%

Target class

Figure 10: Continued.

FIGURE 10: Confusion matrix of the trained network model. The classification accuracy rate of each subject in BCI Competition IV 2a (A01–A09). The average classification accuracy rate of all subjects (average). Take the confusion matrix of A01 as an example. The first column indicates that, in a total of 72 testing samples from 72 left-hand trials, the number of testing samples is 64 whose model predicted value of classification category is left hand, and the number of testing samples is 8 whose model predicted value of classification category covers right hand, feet, and tongue; that is, TP = 64 and FN = 8; then, TPR = 64/72 = 88.9%. The second to fourth columns are the same. The first row indicates that, in a total of 72 testing samples from 64 left hand, 2 right hand, 5 feet, and 1 tongue trials, all the testing samples' model predicted values of classification category are left hand; that is, TP = 64 and FP = 8; then, PPV = 64/72 = 88.9%. The second to fourth rows are the same. The main diagonal indicates that the total number of model's correct judgments is 64 + 62+64 + 64 = 254 times, while the total model prediction results of testing samples are 288, so, ACC = 254/288 = 88.2%.

TABLE 2: Comparison of the accuracy and kappa coefficient between the method in this paper and the methods in the literature.

| Methods	Statistics	Subjects									Mean ± sd
		A01	A02	A03	A04	A05	A06	A07	A08	A09	
Tabar and Halici [31]		64.0%	49.7%	49.7%	92.5%	74.5%	68.5%	61.7%	62.5%	59.5%	66.2 ± 11.2%
		0.520	0.330	0.330	0.910	0.660	0.580	0.490	0.500	0.460	0.550 ± 0.152
Ang et al. [47]		76.0%	56.5%	81.3%	61.0%	55.0%	45.3%	82.7%	81.3%	70.7%	67.8 ± 12.9%
		0.680	0.420	0.750	0.480	0.400	0.270	0.770	0.750	0.610	0.570 ± 0.173
Winkler et al. [48]		69.3%	55.7%	86.5%	67.7%	53.5%	46.7%	84.2%	85.0%	79.0%	69.7 ± 14.2%
		0.590	0.410	0.820	0.570	0.380	0.290	0.790	0.800	0.700	0.600 ± 0.189
Nicolas-Alonso et al. [49]		87.2%	63.3%	91.0%	76.0%	67.0%	51.2%	92.5%	88.0%	89.5%	78.4 ± 14.0%
		0.830	0.510	0.880	0.680	0.560	0.350	0.900	0.840	0.750	0.700 ± 0.180
Ai et al. [50]	Accuracy kappa	82.7%	65.5%	88.0%	77.5%	72.2%	70.7%	82.7%	88.0%	89.5%	80.2 ± 8.10%
		0.770	0.540	0.840	0.700	0.630	0.610	0.770	0.840	0.860	0.730 ± 0.108
Nicolas-Alonso et al. [51]		88.0%	66.2%	92.5%	78.2%	74.5%	58.0%	**95.5%**	**88.7%**	82.0%	80.4 ± 11.8%
		0.840	0.550	0.900	0.710	0.660	0.440	**0.940**	**0.850**	0.760	0.740 ± 0.157
Xu et al. [52]		76.7%	68.5%	**100%**	**100%**	**87.2%**	**85.0%**	78.2%	71.5%	73.0%	81.2 ± 28.5%
		0.690	0.580	**1.000**	**1.000**	**0.830**	**0.800**	0.710	0.620	0.640	0.750 ± 0.147
Zhang et al. [16]		87.7%	65.5%	88.2%	83.5%	82.7%	74.5%	93.2%	87.2%	**91.5%**	83.0 ± 8.34%
		0.850	0.540	0.870	0.780	0.770	0.660	0.920	0.830	**0.900**	0.800 ± 0.120
SCNN-BiLSTM network based on attention		**88.2%**	**73.3%**	79.7%	89.2%	80.6%	76.4%	90.3%	81.6%	85.4%	**82.7 ± 5.57%**
		0.840	**0.650**	0.730	0.870	0.740	0.690	0.870	0.760	0.810	**0.780 ± 0.074**

(a)

(b)
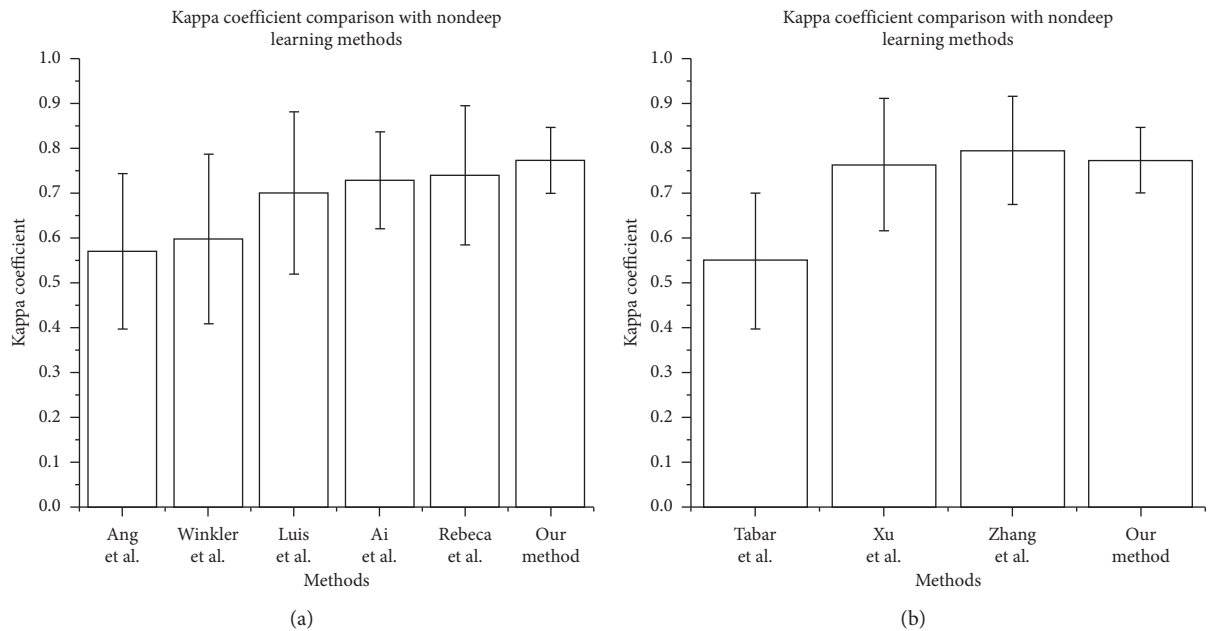
FIGURE 11: Continued.
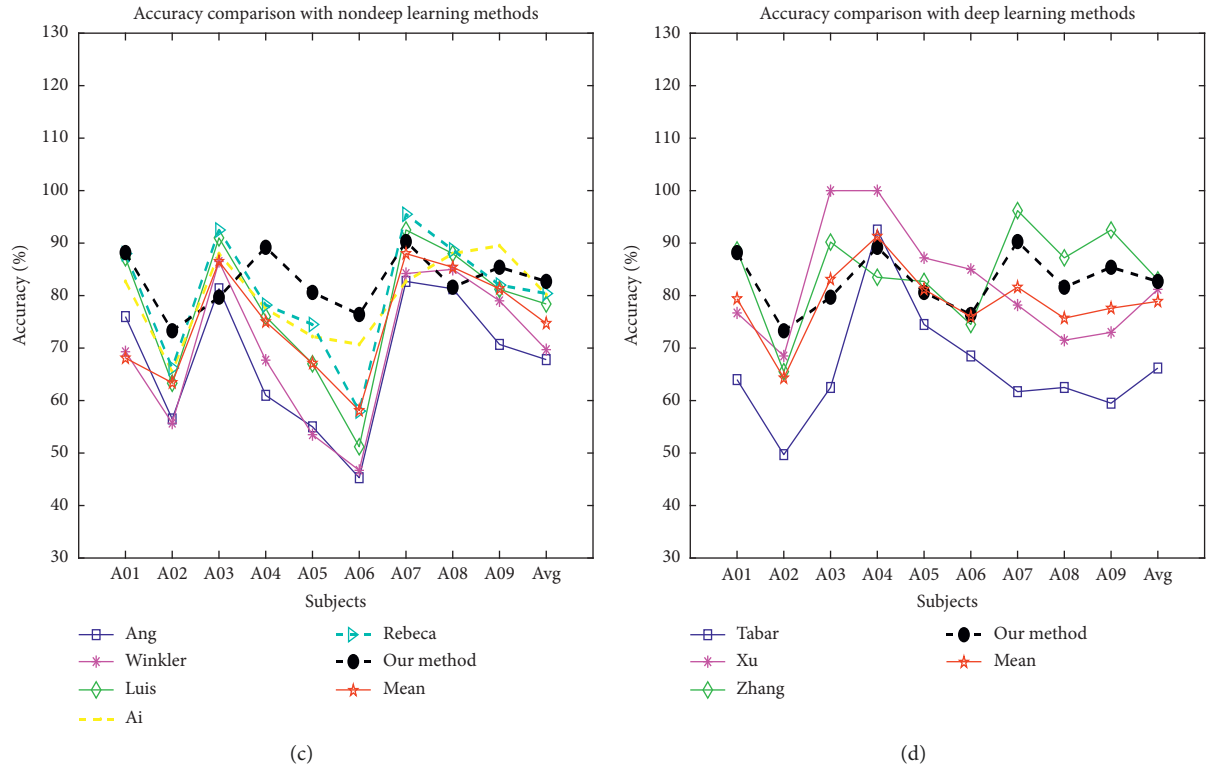
FIGURE 11: Comparison results of the accuracy and kappa coefficient. (a, c) The comparison results between the traditional nondeep learning methods and our method in this paper. (b, d) The comparison results between the deep learning methods and our method in this paper. It is obvious from (a, b) that our method has the minimum individual difference among different subjects, and also as can be seen from (c, d) our method has greatly improved the overall accuracy of all 9 subjects.

learning methods, our method has the minimum individual difference among different subjects, and also as can be seen from (c) and (d) the classification accuracy of each subject is greater than 73.3%. In other words, our method has greatly improved the overall accuracy of all 9 subjects. Therefore, our method is more suitable for the multiclass recognition of MI-EEG signals that are short-time series and is very effective in improving the overall classification and recognition.

5. Conclusions

In this paper, we propose an attention-based time-incremental end-to-end shared neural network which is essentially an end-to-end trainable model formed through the unification of SCNN network, BiLSTM network, and attention mechanism. With this end-to-end shared neural network, the feature extraction and pattern classification of MI-EEG signals are performed in a step-by-step way, which effectively improves the accuracy and robustness of EEG recognition.

In much research of deep neural network methods for EEG signals, the amount of training data is not enough for the network's training, and EEG signals are treated as images in usual, which may lead to the loss of information about time. Moreover, a simple network stacking can cause redundancy of features. So, to solve these issues, the method in

this paper is divided into the following steps: first, a combination of all the sample training data followed by an increase in sample number using autoencoder meets the needs of a mass of training data for deep learning while effectively reducing the bad effect on signal recognition because of randomness, instability, and individual variability of EEG data. Second, BiLSTM network with time increments is connected in series with SCNN network, so that feature extraction of MI-EEG in its frequency-spatial domain and temporal domain successively uses SCNN network and BiLSTM network, which can make the features of MI-EEG in its temporal-frequency-spatial domain fully learned and ensure the classification results of MI-EEG signal. Third, the attention mechanism introduced into the dynamic weighted feature fusion of MI-EEG reduces the redundancy of the fusion features and improves the classification accuracy.

The results of comparison with the traditional nondeep methods and deep learning methods have shown the effectiveness of this end-to-end shared neural network that we proposed. The method is more suitable for the multiclass recognition of MI-EEG signals that are short-time series. It has the minimum individual difference among different subjects and is very effective in improving the overall classification and recognition of subjects.

In the near future, we will continue keeping focus on the research of information in raw MI-EEG data. Through the analysis of the irregularities of both distribution structure of

EEG channels and EEG data, we will try to learn more about the algorithms of feature extraction, feature fusion, and pattern classification to recognize motor imagery tasks more compared to four classes and improve the classification accuracy while reducing the individual difference of the same network model in different subjects. What is more, we will also try to deploy the combination of brain-computer-interface and limb rehabilitation robot in an online system.

Data Availability

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] R. Abiri, S. Borhani, E. W. Sellers et al., "A comprehensive review of EEG-based brain-computer interface paradigms," *Journal of Neural Engineering*, vol. 16, no. 1, Article ID 011001, 2019.

[2] Y. Aoh, H.-J. Hsiao, M.-K. Lu et al., "Event-related desynchronization/synchronization in spinocerebellar ataxia type 3," *Frontiers in Neurology*, vol. 10, p. 822, 2019.

[3] S. Aggarwal and N. Chugh, "Signal processing techniques for motor imagery brain computer interface: a review," *Array*, vol. 1-2, Article ID 100003, 2019.

[4] M. A. Ramírez-Moreno and D. Gutiérrez, "Evaluating a semiautonomous brain-computer interface based on conformal geometric algebra and artificial vision," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 9374802, 19 pages, 2019.

[5] A. Malekmohammadi, H. Mohammadzade, A. Chamanzar et al., "An efficient hardware implementation for a motor imagery brain computer interface system," *Scientia Iranica*, vol. 26, pp. 72–94, 2019.

[6] X. Mao, M. Li, W. Li et al., "Progress in EEG-based brain robot interaction systems," *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 1742862, 25 pages, 2017.

[7] V. Vimala, K. Ramar, and M. Ettappan, "An intelligent sleep apnea classification system based on EEG signals," *Journal of Medical Systems*, vol. 43, no. 2, p. 36, 2019.

[8] R. Salazar-Varas and R. A. Vazquez, "Evaluating the effect of the cutoff frequencies during the pre-processing stage of motor imagery EEG signals classification," *Biomedical Signal Processing and Control*, vol. 54, Article ID 101592, 2019.

[9] S. Tarai, R. Mukherjee, Q. A. Qurratul, B. K. Singh, and A. Bit, "Use of prosocial word enhances the processing of language: frequency domain analysis of human EEG," *Journal of Psycholinguistic Research*, vol. 48, no. 1, pp. 145–161, 2019.

[10] M. J. Hülsemann, E. Naumann, and B. Rasch, "Quantification of phase-amplitude coupling in neuronal oscillations: comparison of phase-locking value, mean vector length, modulation index, and generalized linear modeling cross-frequency coupling," *Frontiers in Neuroscience*, vol. 13, p. 573, 2019.

[11] Y. Liu, J. Gao, W. Cao et al., "A hybrid double-density dual-tree discrete wavelet transformation and marginal Fisher analysis for scoring sleep stages from unprocessed single-channel electroencephalogram," *Quantitative Imaging in Medicine and Surgery*, vol. 10, no. 3, p. 766, 2020.

[12] S. Taran and V. Bajaj, "Motor imagery tasks-based EEG signals classification using tunable-Q wavelet transform," *Neural Computing and Applications*, vol. 31, no. 11, pp. 6925–6932, 2019.

[13] Y. Lu, M. Wang, W. Wu, Y. Han, Q. Zhang, and S. Chen, "Dynamic entropy-based pattern learning to identify emotions from EEG signals across individuals," *Measurement*, vol. 150, Article ID 107003, 2020.

[14] R. Fu, M. Han, Y. Tian, and P. Shi, "Improvement motor imagery EEG classification based on sparse common spatial pattern and regularized discriminant analysis," *Journal of Neuroscience Methods*, vol. 343, Article ID 108833, 2020.

[15] M.-A. Li, M. Zhang, and Y.-J. Sun, "A novel motor imagery EEG recognition method based on deep learning," in *2016 International Forum on Management, Education and Information Technology Application*, Guangzhou, China, 2016.

[16] R. Zhang, Q. Zong, L. Dou et al., "A novel hybrid deep learning scheme for four-class motor imagery classification," *Journal of Neural Engineering*, vol. 16, no. 6, Article ID 066004, 2019.

[17] J. Long, J. Wang, and T. Yu, "An efficient framework for EEG analysis with application to hybrid brain computer interfaces based on motor imagery and P300," *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 9528097, 6 pages, 2017.

[18] H. Rajaguru and S. K. Prabhakar, "Correlation dimension and bayesian linear discriminant analysis for alcohol risk level detection," in *Proceedings of the 2019 International Conference on Computational Vision and Bio Inspired Computing*, Coimbatore, India, September 2019.

[19] A. Quintero-Rincón, M. Flugelman, J. Prendes et al., "Study on epileptic seizure detection in EEG signals using largest lyapunov exponents and logistic regression," *Revista Argentina de Bioingeniería*, vol. 23, no. 2, pp. 17–24, 2019.

[20] Y. Xu, J. Hua, H. Zhang et al., "Improved transductive support vector machine for a small labelled set in motor imagery-based brain-computer interface," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 2087132, 16 pages, 2019.

[21] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, "EEG based multi-class seizure type classification using convolutional neural network and transfer learning," *Neural Networks*, vol. 124, pp. 202–212, 2020.

[22] R. Haeb-Umbach, S. Watanabe, T. Nakatani et al., "Speech processing for digital home assistants: combining signal processing with deep-learning techniques," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 111–124, 2019.

[23] C. Tian, L. Fei, W. Zheng et al., "Deep learning on image denoising: an overview," *Neural Networks*, vol. 131, pp. 251–275, 2020.

[24] S. Kurkin, A. Hramov, P. Chholak et al., "Localizing oscillatory sources in a brain by MEG data during cognitive activity," in *Proceedings of the 2020 4th International Conference*

on *Computational Intelligence and Networks (CINE)*, IEEE, Kolkata, India, February 2020.

[25] G. Niu, X. Yi, C. Chen et al., "A novel effluent quality predicting model based on genetic-deep belief network algorithm for cleaner production in a full-scale paper-making wastewater treatment," *Journal of Cleaner Production*, vol. 265, Article ID 121787, 2020.

[26] S. L. Oh, J. Vicnesh, E. J. Ciaccio, R. Yuvaraj, and U. R. Acharya, "Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals," *Applied Sciences*, vol. 9, no. 14, p. 2870, 2019.

[27] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[28] M. Li, R. Wang, J. Yang et al., "An improved refined composite multivariate multiscale fuzzy entropy method for MI-EEG feature extraction," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 7529572, 12 pages, 2019.

[29] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, 2017.

[30] H. Yang, S. Sakhavi, K. K. Ang et al., "On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification," in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, Milano, Italy, 2015.

[31] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of Neural Engineering*, vol. 14, no. 1, Article ID 016003, 2016.

[32] H. K. Lee and Y.-S. Choi, "A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequecy image," in *Proceedings of the 2018 International Conference on Information Networking (ICOIN)*, IEEE, Chiang Mai, Thailand, 2018.

[33] J. Zhou, M. Meng, Y. Gao et al., "Classification of motor imagery EEG using wavelet envelope analysis and LSTM networks," in *Proceedings of the 2018 Chinese Control and Decision Conference (CCDC)*, IEEE, Shenyang, China, 2018.

[34] X. An, D. Kuang, X. Guo et al., "A deep learning method for classification of EEG data based on motor imagery," in *Proceedings of the 2014 International Conference on Intelligent Computing*, Taiyuan, China, 2014.

[35] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, *BCI Competition 2008—Graz Data Set A*, Vol. 16, Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, Graz, Austria, 2008.

[36] R. Shams, M. Masihi, R. B. Boozarjomehry, and M. J. Blunt, "Coupled generative adversarial and auto-encoder neural networks to reconstruct three-dimensional multi-scale porous media," *Journal of Petroleum Science and Engineering*, vol. 186, Article ID 106794, 2020.

[37] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18940–18950, 2019.

[38] T. Urakawa, Y. Tanaka, S. Goto, H. Matsuzawa, K. Watanabe, and N. Endo, "Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network," *Skeletal Radiology*, vol. 48, no. 2, pp. 239–244, 2019.

[39] A. K. Dubey and V. Jain, "Comparative study of convolution neural network's ReLu and leaky-ReLu activation functions," in *Applications of Computing, Automation and Wireless Systems in Electrical Engineering*, pp. 873–880, Springer, Berlin, Germany, 2019.

[40] C. Liu, W. Ding, X. Xia et al., "Circulant binary convolutional networks: enhancing the performance of 1-bit dcnns with circulant back propagation," in *Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, June 2019.

[41] X. Chen, J. He, X. Wu, W. Yan, and W. Wei, "Sleep staging by bidirectional long short-term memory convolution neural network," *Future Generation Computer Systems*, vol. 109, pp. 188–196, 2020.

[42] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

[43] H. Fukui, T. Hirakawa, T. Yamashita et al., "Attention branch network: learning of attention mechanism for visual explanation," in *Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019.

[44] S. Bock and M. Weiß, "A proof of local convergence for the Adam optimizer," in *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 2019.

[45] M. Giménez, J. Palanca, and V. Botti, "Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. A case of study in sentiment analysis," *Neurocomputing*, vol. 378, pp. 315–323, 2020.

[46] A. C. Belkina, C. O. Ciccolella, R. Anno et al., "Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets," *Nature Communications*, vol. 10, no. 1, pp. 1–12, 2019.

[47] K. K. Ang, Z. Y. Chin, C. Wang et al., "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, p. 39, 2012.

[48] I. Winkler, S. Haufe, and M. Tangermann, "Automatic classification of artifactual ICA-components for artifact removal in EEG signals," *Behavioral and Brain Functions*, vol. 7, no. 1, p. 30, 2011.

[49] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Alvarez, and R. Hornero, "Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain -computer interfaces," *Neurocomputing*, vol. 159, pp. 186–196, 2015.

[50] Q. Ai, A. Chen, K. Chen et al., "Feature extraction of four-class motor imagery EEG signals based on functional brain network," *Journal of Neural Engineering*, vol. 16, no. 2, Article ID 026032, 2019.

[51] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Alvarez, and R. Hornero, "Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 4, pp. 702–712, 2015.

[52] B. Xu, L. Zhang, A. Song et al., "Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification," *IEEE Access*, vol. 7, pp. 6084–6093, 2018.

[53] N. Bagh and M. R. Reddy, "Improving the performance of motor imagery based brain-computer interface using phase space reconstruction," in *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, Berlin, Germany, 2019.