

# A Model Combining Multi Branch Spectral-Temporal CNN, Efficient Channel Attention, and LightGBM for MI-BCI Classification

Hai Jia<sup>ID</sup>, Shiqi Yu, Shunjie Yin, Lanxin Liu, Chanlin Yi<sup>ID</sup>, Kaiqing Xue, Fali Li<sup>ID</sup>,  
Dezhong Yao<sup>ID</sup>, Peng Xu<sup>ID</sup>, and Tao Zhang<sup>ID</sup>

**Abstract**—Accurately decoding motor imagery (MI) brain-computer interface (BCI) tasks has remained a challenge for both neuroscience research and clinical diagnosis. Unfortunately, less subject information and low signal-to-noise ratio of MI electroencephalography (EEG) signals make it difficult to decode the movement intentions of users. In this study, we proposed an end-to-end deep learning model, a multi-branch spectral-temporal convolutional neural network with channel attention and LightGBM model (MBSTCNN-ECA-LightGBM), to decode MI-EEG tasks. We first constructed a multi branch CNN module to learn spectral-temporal domain features. Subsequently, we added an efficient channel attention mechanism module to obtain more discriminative features. Finally, LightGBM was applied to decode the MI multi-classification tasks. The within-subject cross-session training strategy was used to validate classification results. The experimental results showed that the model achieved an average accuracy of 86% on the two-class MI-BCI data and an average accuracy of 74% on the four-class MI-BCI data, which outperformed current state-of-the-art methods. The proposed MBSTCNN-ECA-LightGBM can efficiently decode the spectral and temporal domain information of EEG, improving the performance of MI-based BCIs.

**Index Terms**—Motor imagery, deep learning, spectral-temporal, attention mechanism, LightGBM.

Manuscript received 15 November 2022; revised 23 December 2022 and 30 January 2023; accepted 6 February 2023. Date of publication 10 February 2023; date of current version 21 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62006197, Grant 61961160705, and Grant U19A2082; and in part by the Project of Science and Technology Department of Sichuan Province under Grant 2018JY0526, Grant 2021YFSY0040, and Grant 2020ZYD013. (Corresponding authors: Peng Xu; Tao Zhang.)

Hai Jia, Shiqi Yu, Shunjie Yin, Lanxin Liu, and Tao Zhang are with the Mental Health Education Center and School of Science, Xihua University, Chengdu 610039, China (e-mail: zhangtao1698@mail.xhu.edu.cn).

Chanlin Yi, Fali Li, Dezhong Yao, and Peng Xu are with MOE Key Laboratory for Neuroinformation, Clinical Hospital of Chengdu Brain Science Institute, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: xupeng@uestc.edu.cn).

Kaiqing Xue is with the School of Computer and Software Engineering, Xihua University, Chengdu 610039, China.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNSRE.2023.3243992>, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3243992

## I. INTRODUCTION

**B**RAIN-COMPUTER interface (BCI) system can directly convert individual brain cognitive activity (i.e., motor intentions) into actual instructions to help people to communicate with external devices [1], [2]. For this to happen, several BCIs-based electroencephalography (EEG) signals were widely used, for instance, steady-state visual evoked potentials (SSVEP), P300 evoked potentials, and motor imagery (MI) BCIs [3]. Among these BCI systems, only the MI-BCI system depended on the individual's spontaneous brain potential without external stimulation. It is increased the flexibility of participants and enriched the application scenarios of MI-BCI. However, it also brings new challenges to decode EEG signals to understand the movement intentions of users.

MI usually refers to the mental activity of imagining movement in the brain without producing any actual motor action [4]. For different MI tasks, such as left hand, right hand, and both feet, the oscillatory activities of movement-related brain regions and the brain network patterns among regions were different [5]. For example, in the early EEG study, Pfurtscheller et al. [6], [7] found that the event-related synchronization (ERS) and event-related desynchronization (ERD) phenomenon during the subject performed the MI task. ERD occurs in the motor area of the right cortex (left cortex) several hundred milliseconds before the start of the left-hand (right hand) MI task, the amplitude and energy of alpha (8-13 Hz) and beta (13-30 Hz) bands in the related areas decreased; ERS occurs with a rapid increase in the energy of beta band within 1-2s after the end of MI [8]. The MI-BCI system can effectively identify MI-EEG signals based on the differences in ERD and ERS generated by imagining the movement so as to know the movement intention of the subjects [9]. Although the MI task activities are different, decoding of MI based on the EEG signals is still difficult, and there are greater individual differences in MI ability [10]. Thus, accurately recognizing these MI activities is one of the main goals of MI-BCI studies.

In the MI-BCI system, feature learning and classification are two important parts of MI task recognition. For the feature learning stage, a lot of researches have focused on feature extraction methods. Common spatial patterns (CSP) [11], for example, have been widely used in MI classification studies. Ang et al. [12] proposed the filter band common spatial pattern (FBCSP) algorithm to extract multi band CSP features, a feature selection algorithm was used to select discriminative band pairs and corresponding CSP features automatically. Yang et al. [13] used pairwise projection matrices to generate augmented common spatial pattern (ACSP) features, and then proposed a frequency band feature selection strategy to screen features for classification. In addition, other well-known feature extraction methods are also used in MI recognition studies, such as Wavelet Transform, Wavelet Packet Decomposition, and Short-time Fourier Transform [14], [15], [16]. However, these feature learning methods still rely on hand-designed features based on human experience, which may limit the improvement of classification accuracy and cost a lot of time [17].

In recent years, with the successful application of deep learning in the field of computer vision (CV) [18] and natural language processing (NLP) [19], many researchers [20], [21] have applied this technology to the field of BCI because of its powerful automatic feature learning capability during network training. Among these deep learning models, the convolutional neural network (CNN) has become one of the most popular models used in EEG classification. Compared to other deep learning models, CNN could capture the intrinsic features in EEG signals with fewer parameters based on the weight sharing technique [22]. Tabar and Halici [5] used CNN to extract a new form of features with time-frequency-location information, and a stacked autoencoder (SAE) was used for MI-EEG two classification. Sakhavi et al. [23] proposed a new temporal representation of EEG data and utilized CNN to learn MI task features. The recently proposed EEGNet [24] takes a different approach, exploiting an end-to-end approach to extract spectral and temporal features, and achieves good results on several BCI paradigms. To obtain the contribution of EEG channels to MI classification, Zhang et al. [25] used a squeeze-and-excitation attention module to automatically learn channel weights based on the contribution of EEG channels to MI classification, a CNN model was used to further learn the time-frequency features. Overall, in the feature learning model building stage, machine learning utilized a set of predefined and interpretable algorithmic steps applied to input data, while deep learning models automatically learn parameters directly from the input data. In our study, a CNN network with attention module was constructed to adaptively acquire feature information representing motor intentions for each EEG sample.

In addition, the methods mentioned above are all single branch CNN results, while multi-branch CNN results have been widely used in MI classification tasks. For example, Dai et al. [17] performed the MI-EEG classification tasks by feeding EEG signals of three different frequency bands into three different CNN branches. In order to preserve more

temporal features and spatial features, Zhao et al. [26] proposed a multi branch 3D-CNN model to decode MI signals, which achieved better robustness without changing the spatial distribution of electrodes. Jia et al. [27] proposed to capture MI-EEG features through five parallel CNNs with different convolution scales, and reduce the temporal variance to a certain extent. Although these multi branch methods achieved a good performance compared with single-branch CNN, additional preprocessing and model parameters further increase the computational cost.

For the classification part, traditional algorithms such as support vector machine (SVM) [28], random forests [29], and Bayesian classifiers [30] were used in MI studies. However, most of the current MI-EEG classification studies focus on two classification tasks or convert multi classification tasks into two classification tasks, which increases the computational cost. In 2017, a boosting algorithm based on the Gradient Boosting Decision Tree (GBDT) model, Light Gradient Boosting Machine (LightGBM), was proposed by Microsoft [31]. With fast training speed, lower memory consumption, and fast processing of massive data [22], it has been well applied in biomedical [32] and financial forecasting [33]. More importantly, LightGBM can directly implement multi classification tasks. The superiority of LightGBM has also been gradually applied to EEG classification. Zeng et al. [34] used the CSP algorithm to learn EEG features and used an improved LightGBM classifier to classify mental states in fatigued driving and achieved better performance than traditional classifiers such as SVM, CNN, and gated recurrent unit (GRU). Aggarwal et al. [35] applied LightGBM and XGBoost algorithms to emotion recognition, which obtained higher accuracy and faster training speed. However, the application of LightGBM on MI classification is still few.

To address these issues, in our study, a model that combining multi branch spectral-temporal CNN, attention mechanism, and LightGBM classifier is used for feature learning and classification tasks of MI-BCI. The proposed method is evaluated on BCI Competition Dataset IV-2a achieve good performance in two-class and four-class MI classification tasks. The main contributions of this paper are summarized as follows:

- (1) We propose a novel MBSTCNN model for MI-EEG decoding, which integrates both spectral and temporal feature extraction into the deep learning model. Particularly, the proposed MBSTCNN captures MI-EEG features more accurately than widely used single branch CNN.
- (2) We introduce an efficient channel attention module, which extracts more discriminative features from spectral-temporal features in an adaptive way.
- (3) An efficient classifier LightGBM, which can directly classify high-dimensional features, is used to complete the final MI-BCI multi classification tasks.

## II. METHODS

### A. Dataset

The public MI-BCI competition EEG data set (BCI Competition Dataset IV-2a) is used to train and test our model in the current study, which can be download

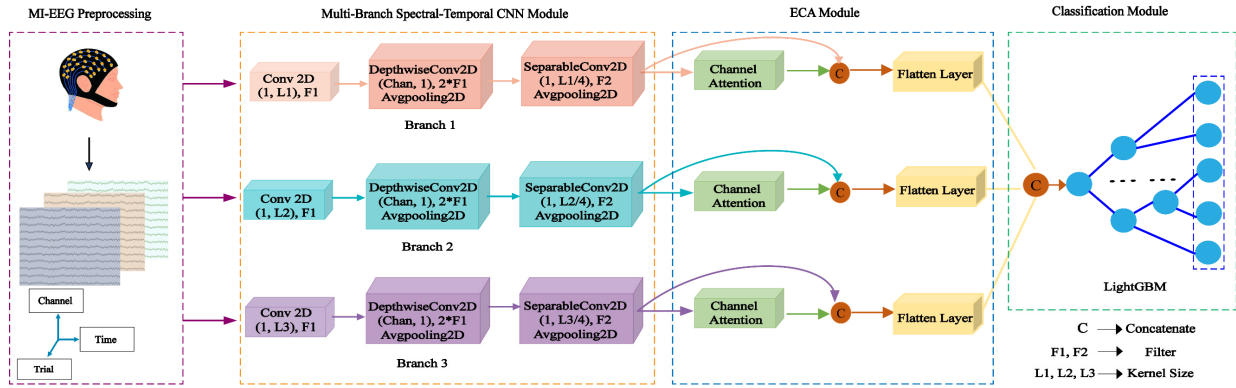


Fig. 1. The overall process of MBSTCNN-ECA-LightGBM construction. The main parameters  $L1$ ,  $L2$ ,  $L3$ , and  $F1$ ,  $F2$  in the MBSTCNN module represent the convolution kernel size and the number of filters respectively. Parameter  $C$  represents the concatenate layer. The “Channel Attention” represents the efficient channel attention mechanism module we introduced.

at <https://www.bbc.de/competition/iv/#dataset2a> [36]. These MI-EEG signals were recorded by twenty-two Ag/AgCl electrodes (10-20 international standard lead system) and were sampled with 250 Hz and bandpass filter between 0.5 Hz and 100 Hz (with the 50Hz notch filter enabled). The dataset is provided by the Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology [36]. Nine healthy subjects were recruited in this dataset. This work involves human subjects in its research. We confirm all human subject research procedures and protocols are exempt from review board approval.

## B. MI-BCI Experiment

The cue-based MI-BCI paradigm contains four experiment tasks, including imagining the movement of the left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Two sessions are called “T” session and “E” session were performed for each subject on different days. Each session consists of 6 runs separated by short breaks. One run contains 48 trials (each class for 12 trials), resulting in 288 trials for each session. In our study, the EEG data from “T” session is used to train our proposed model, and the “E” session is used to test our model (within-subject cross-session validation). During the experiment, a cue in the form of an arrow appeared and stayed on the screen at 2s, and all subjects started imagining the movement from 3 seconds to the end of 6 seconds according to the prompting on the screen, then there was a 1.2s rest before the next trial began. For more detailed information, please reference [36].

## C. EEG Data Preprocessing and Segment

Pre-processing is used in MI-EEG to remove noise and artifacts generated in the experiments. In order to learn the raw EEG signals with our proposed deep learning model, we did not perform any EEG artifact removal and bandpass filtering. Therefore, the raw MI signal time frame  $X_i \in R^{C_{EEG} \times T}$  ( $i = 1, 2, \dots, N$ ) from the trial is extracted, where  $N$  represents the number of trials,  $C_{EEG}$  represents the number of EEG channels, and  $T$  represents the time points.

After that, channel-wise normalization along the channel level is used for  $X_i$ , calculated as

$$X_i = \frac{X_i - \text{mean}(X_i)}{\text{std}(X_i)} \quad (1)$$

where  $\text{mean}$  represents the channel-wise average value and  $\text{std}$  represents the channel-wise standard deviation value. In our work, the time segments we extracted consisted of 4.5s (from 0.5s before the start of the cue to the end of MI).

## D. MBSTCNN-ECA Framework

In order to efficiently extract spectral and temporal domain features from raw EEG signals, we propose an end-to-end multi branch spectral-temporal convolutional neural network with efficient channel attention and LightGBM (MBSTCNN-ECA-LightGBM) framework. This framework has three main modules for decoding MI-EEG classification, including the multi branch spectral-temporal convolutional neural network (MBSTCNN) module, the efficient channel attention (ECA) module, and the classification module. In the multi branch spectral-temporal CNN module, the raw EEG signals are directly input into the MBSTCNN module. Then the different spectral and temporal information are extracted through 2D-CNN. In the ECA module, we use a neural network to assign attention weights to the extracted spectral-temporal features, and optimize the network parameters adaptively according to the importance of the features. The extracted new features are input into the flatten layer, which is used to flatten 2D features into 1D. In the classification module, the extracted multi branch features are connected by the concatenation layer after passing through the flatten layer. For the MBSTCNN-ECA-LightGBM module, the features are directly input into the LightGBM classifier to get the classification results. The activation function  $\text{softmax}$  is used to classify the MI tasks, which is defined as:

$$S(x)_i = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}} \quad (2)$$

where  $x$  is the input vector to the  $\text{softmax}$  function  $S$ , it contains elements for  $n$  outcomes,  $x_i$  is the  $i$ th element in the input vector  $x$ , and  $N$  is the number of classes. The overall framework of our proposed model is shown in Fig. 1.

### E. MBSTCNN Module Spectral-Temporal Wise Feature Learning

The basic idea of CNN is to extract different features through convolution operations. Previous studies have shown that for different MI tasks, the optimal number of convolution kernels varies for different subjects and the same subject at different times [17]. Motivated by the EEG feature extraction method EEGNet [24] and HS-CNN [17], the raw MI-EEG samples are directly input into the MBSTCNN module. To capture more frequency and temporal features, we construct a multi branch CNN module by adjusting the network convolution size.

Different convolution kernel sizes can extract features from various fields of view. Therefore, we capture information in different frequency domains of EEG signals by setting multiple different convolution kernels. For example, in the first branch, the convolution kernel size is  $(1, 32)$ , and for EEG signals with a sampling rate of 250Hz, 2D convolution can capture information above about 8Hz as demonstrated in [24]. To speed up model convergence, we apply batch normalization along the feature map dimension. Then, an exponential linear unit (ELU) activation function (3) is used.

$$f(x) = \begin{cases} x & x > 0 \\ \alpha(\exp(x) - 1) & x \leq 0 \end{cases} \quad (3)$$

Subsequently, depthwise convolution is used to learn spatial filters for specific frequency bands [37]. An averaging pooling layer was used to reduce the sampling rate to 64Hz as depicted in EEGNet (the original sample rate is 250Hz). To help regularize our model, we use the Dropout technique [38] and set the dropout rate to 0.5 in our work. Further, we use a separable convolution to learn the temporal summation of each frequency feature map separately, and then mix these feature maps. Finally, we obtain the spectral-temporal features extracted by all branches, respectively. To define the optimal number of branches, we performed comparative experiments with different branch parameters (For more details see the supplementary materials *section B* and *section C* STABLE I). According to the classification accuracy (see *section RESULTS B*), a three-branch module was used in our study.

### F. ECA Module

Inspired by the successful use of the attention mechanism in a recent study [25] for channel selection in the MI-BCI, the attention mechanism is introduced to recalibrate spatial and spectral features. In our proposed model, a novel attention mechanism by avoiding dimensionality reduction and cross-channel interaction is used to rescale spectral-temporal features [39]. The architecture of the ECA module in each branch are shown in the Fig. 2. The main parameters are shown in STABLE II of supplementary materials *section C*, respectively. The specific steps of spectral-temporal features re-calibration are divided into the following steps.

(a) The spectral-temporal features extracted by each branch in the MBSTCNN module are integrated into 1D feature vectors by using global average pooling (GAP) [40] without

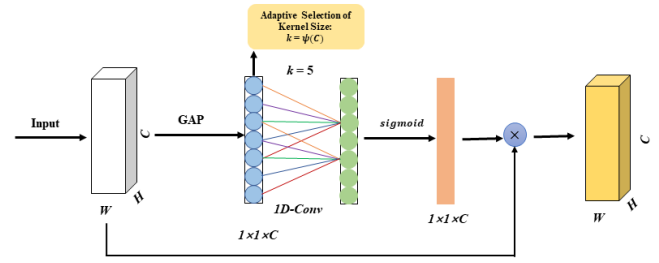


Fig. 2. The architecture of ECA module for recalibrate spectral-temporal features. The parameter  $C$ ,  $W$ , and  $H$  represent the feature maps dimension. The GAP represents the global average pooling layer,  $k$  represents the 1D convolution size.

dimensionality reduction. Suppose the input dimension of each branch in the module is  $(N, 1, 35, C_{features})$  ( $C_{features} = 8, 16, 32$ ) where  $C_{features}$  represents the feature maps dimension extracted by the MBSTCNN module. The idea of GAP is to use global average pooling to replace the fully connected layer, and more importantly, it retains the efficient information extracted by the previous MBSTCNN module. After the GAP layer, the feature maps become  $(N, C_{features})$ .

(b) After the GAP layer, new feature maps are learned by using an adaptive 1D convolution. In fact, the kernel size of the 1D convolution can affect the number of feature channels in the attention mechanism when calculating the weights, i.e., the coverage of cross-channel interaction [39]. Therefore, local cross-channel exchange information is performed by considering each channel and its  $k$  neighbors (i.e., the coverage of the local cross-channel interaction with kernel size  $k$  in 1D-convolution). An adaptive method is proposed to determine the value of  $k$ , where  $k$  is proportional to the feature channel dimension. The mathematical expression is shown in Equation (4). In order to extract nonlinear features, nonlinear function (3) is used.

(c) After step  $b$  computes the coverage of cross-channel interactions by using 1D convolution, a sigmoid activation function (2) is used to map the attention weights between 0-1.

(d) Inspired by the ResNet proposed by He et al. [41], we innovatively fuse the features extracted by the MBSTCNN module with the new features reconstructed by the ECA module to avoid overfitting. Flatten layers are used in subsequent steps.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

$$2^{\phi(k)} = 2^{\gamma * k - b} \quad (5)$$

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{odd} \quad (6)$$

### G. Classification Module

LightGBM was an improved gradient boosting decision tree (GBDT) machine learning algorithm proposed by Ke and colleagues in 2017 [31]. Its distributed and efficient characteristics could handle large data volume tasks. Recently, it has been widely used in machine learning tasks such as multi classification [42] and regression [43]. LightGBM algorithm



mainly contains two key techniques, including the Gradient-based One-side Sampling (GOSS) algorithm and the Exclusive Feature Bundling (EFB) algorithm. The detailed procedure of LightGBM is listed as follows.

First, we defined a supervised training set  $\{x_1, x_2, \dots, x_n\}$ , LightGBM aimed to find an approximation  $f^\wedge(x)$  of the function  $f(x)$  to minimize the expected value of a specific loss function  $L(y, f(x))$  as follows:

$$f^\wedge(x) = \operatorname{argmin}_{y, X} E_{y, X} L(y, f(x)). \quad (7)$$

Specifically, LightGBM minimized the objective function by fitting the gradient of the loss function in the decision tree growth stage so that a better decision tree was obtained. The objective function of decision tree growth was as follows:

$$\operatorname{obj} = \sum_{i=1}^n \left[ g_i f(x_i) + \frac{1}{2} h_i f^2(x_i) \right] + \Omega(f) \quad (8)$$

where  $g_i$  and  $h_i$  were the first and second derivatives of the loss function, respectively,  $\Omega(f)$  was the regularization terms for all samples.

Subsequently, the GOSS algorithm was proposed in LightGBM to achieve a balance between the number of samples and the accuracy of the decision tree. As LightGBM have demonstrated that if a sample is associated with a small gradient, the training error for this sample is small and it is already well-trained. Therefore, GOSS saved large gradient samples and randomly discarded small gradient samples. The principal of GOSS is shown as follows:

$$G = \sum_{x_i \in A} g_i + \frac{1-a}{b} \sum_{x_i \in B} g_i \quad (9)$$

where  $A$  was the large gradient sample and  $B$  was the small gradient sample. For example, first, GOSS sorted the absolute value of the sample gradients and selected the top  $a \times 100\%$  samples. Then,  $b$  samples were randomly selected from the remaining samples. Finally, we multiplied the  $b$  samples by the coefficient  $\frac{1-a}{b}$ , to focus on the under-trained samples (large gradient sample) without changing the original data distribution. Furthermore, LightGBM used the EFB algorithm to bound mutually exclusive features into a single feature in a high-dimensional sparse feature space. It made the number of features for training is reduced.

The original code for LightGBM was in <https://github.com/microsoft/LightGBM>. In our model, the new features obtained from the ECA module in each branch are concatenated through a flattening layer to obtain a 1D feature vector with a dimension  $(N, C_{new})$ , where  $C_{new}$  represents the final feature maps dimension. Final feature maps are input into the LightGBM classifier. The main parameters of LightGBM we use as shown in STABLE III of the supplementary materials section C.

#### H. Model Training

For the model training, we selected ‘‘T’’ session as the training set, and the ‘‘E’’ session as the testing set for each subject, we called it ‘‘within-subject cross-session’’. We randomly select 20% of samples in the training set as the validation set, and the remaining datasets as the training set. The parameter

update of the CNN feature learning network is based on the results of the validation set. Finally, the testing set is used to test our proposed model. During the feature learning network training phase, the Adam optimizer [44] is used, the learning rate is set to 0.001, and the total number of training epochs is 1000. In addition, an early stopping mechanism is used in the model training process, and the training process stops when the accuracy of the validation set does not increase within 200 epochs.

For the MI classification part, the LightGBM classifier was used to classify the feature learned by MBSTCNN-ECA module on the public dataset to complete the MI-EEG tasks of four-classification and two-classification. Finally, the testing set results are calculated by our proposed model. Different from previous methods [25], [45], in our work, in order to compare the results of the classification process of deep learning with the deep learning feature learning network plus LightGBM classifier, we used MBSTCNN-ECA-Net and MBSTCNN-ECA-LightGBM, respectively.

#### I. Model Evaluation

To evaluate the performance of the MI-BCI classification model, two commonly used measurements are selected, including accuracy and kappa [5], [46]. They are defined as,

$$\operatorname{accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$

$$\operatorname{Kappa} = \frac{P_0 - P_e}{1 - P_e} \quad (11)$$

$$P_0 = \frac{\sum_{i=1}^n a_{ii}}{N} \quad (12)$$

$$P_e = \frac{\sum_{i=1}^n a_{i+} * a_{+i}}{N^2}. \quad (13)$$

where  $TP$  (true positive) is the number of samples in which the original data is positive and still positive after classification;  $TN$  (true negative) is the number of samples in which original data is negative and still negative after classification;  $FN$  (false negative) is the number of samples in which original data is positive but negative after classification;  $FP$  (false positive) is the number of samples in which original data is negative but positive after classification; for kappa value, in the confusion matrix,  $a_{i+} = \sum_j a_{ij}$ ,  $a_{+j} = \sum_i a_{ij}$ , and  $n$  is the number of categories,  $N$  is the total number of samples.

### III. RESULTS

This project uses the programming language is Python3.7. The *MNE* library was used for EEG signals preprocessing. The python library *LightGBM* was used for classifier construction. The *Scikit-learn* library was used for splitting data and model evaluation. The feature learning network was designed by *Tensorflow 2.5*, which is a concise, efficient and fast computing deep learning framework [47]. Our frame runs in Windows10 on a Core i9 PC with 32GB RAM, and the model is trained on an NVIDIA GTX3070Ti GPU.

TABLE I  
TWO-CLASS CLASSIFICATION RESULTS OF MBSTCNN-ECA-LIGHTGBM MODEL

Pairs	Statistics	Subject									Avg
		1	2	3	4	5	6	7	8	9	
L vs. R	<i>Acc</i>	0.88	0.78	0.87	0.76	0.93	0.77	0.87	<b>0.94</b>	0.78	0.84
	<i>Kappa</i>	0.76	0.57	0.74	0.51	0.86	0.54	0.74	<b>0.88</b>	0.79	0.71
L vs. F	<i>Acc</i>	<b>0.98</b>	0.71	0.90	0.86	0.83	0.78	0.82	0.83	0.85	0.84
	<i>Kappa</i>	<b>0.96</b>	0.42	0.81	0.72	0.65	0.56	0.64	0.67	0.71	0.68
L vs. T	<i>Acc</i>	0.95	0.78	0.95	0.84	0.83	0.74	0.76	<b>0.97</b>	0.94	<b>0.86</b>
	<i>Kappa</i>	0.90	0.57	0.92	0.68	0.67	0.47	0.51	<b>0.93</b>	0.89	<b>0.73</b>
R vs. F	<i>Acc</i>	0.92	0.80	<b>0.96</b>	0.74	0.80	0.70	0.76	0.72	0.82	0.80
	<i>Kappa</i>	0.83	0.60	<b>0.92</b>	0.49	0.60	0.40	0.53	0.43	0.64	0.60
R vs. T	<i>Acc</i>	<b>0.97</b>	0.71	0.99	0.76	0.80	0.80	0.78	0.88	0.85	0.84
	<i>Kappa</i>	<b>0.94</b>	0.42	0.97	0.53	0.60	0.60	0.57	0.76	0.71	0.68
F vs. T	<i>Acc</i>	0.74	0.85	0.83	0.78	0.81	0.78	0.83	<b>0.86</b>	0.61	0.79
	<i>Kappa</i>	0.49	0.71	0.81	0.57	0.63	0.56	0.65	<b>0.72</b>	0.22	0.60

*Acc* represents the accuracy, L represents the left hand, R represents the right hand, F represents the feet, and T represents the tongue.

TABLE II  
FOUR-CLASS CLASSIFICATION RESULTS OF MBSTCNN-ECA-NET AND MBSTCNN-ECA-LIGHTGBM MODEL

Proposed Method	Statics	Subject									Avg
		S1	S2	S3	S4	S5	S6	S7	S8	S9	
MBSTCNN-ECA-Net	<i>Acc</i>	0.77	0.60	0.87	0.61	0.66	0.57	0.64	0.77	0.86	0.71
	<i>Kappa</i>	0.69	0.47	0.83	0.48	0.55	0.42	0.51	0.70	0.81	0.61
MBSTCNN-ECA-LightGBM	<i>Acc</i>	0.82	0.61	0.89	0.63	0.71	0.64	0.72	0.79	0.84	0.74
	<i>Kappa</i>	0.76	0.48	0.85	0.50	0.62	0.51	0.63	0.72	0.78	0.65

### A. Two-Class and Four-Class MI-BCI Tasks Classification

To explore possible effects between different MI tasks (i.e., left hand vs. right hand, left hand vs. tongue, left hand vs. both feet, right hand vs. tongue, right hand vs. both feet, tongue vs. both feet) and test the performance of the MBSTCNN-ECA-LightGBM model for two-classification, we conduct comparative experiments by “within-subject cross-session”. As listed in TABLE I, the accuracy and kappa values of various pairs of MI tasks are achieved by our proposed model for all subjects on BCI Competition Dataset IV-2a. The highest two-classification decoding accuracy is achieved in Subject 1 “L vs. F” task. Subject 8 achieves the highest accuracy in multiple two-classification tasks. The overall two-classification and four-classification accuracy of subject 2 is lower compared with other subjects. This finding is in line with previous reports [17], [48]. The highest 98% and average 86% accuracy are obtained by our model.

For the four-classification, the same “within-subject cross-session” strategy is used to calculate the classification accuracy and kappa. Finally, we obtain the testing set accuracy and kappa values as our results. Furthermore, a fully connected layer and a *softmax* activation layer are used instead of LightGBM to be used as the baseline model for this part which is called MBSTCNN-ECA-Net. TABLE II shows the classification results of MBSTCNN-ECA-Net and MBSTCNN-ECA-LightGBM. The highest 89% and average 74% accuracy are obtained by our model. Furthermore, we also display the confusion matrix to illustrate the classification results for each of the four categories for each subject (please see SFig. 1 and 2 in the supplementary materials section A). In the confusion matrix, the diagonal elements represent the decoding accuracy of each class.

For the overall results, the t-SNE algorithm [49] is used to visualize the results of our proposed model MBSTCNN-ECA-LightGBM in four-class and various two-class pairs. The t-SNE results for the highest subject accuracy in each classification task are shown in Fig. 3.

### B. Different Numbers of Branches Results and Comparison

The same “within-subject cross-session” training method is used to compare the effect of different branch numbers on the model results. We selected one-branch, two-branch, three-branch and four-branch model to determine the optimal number of branches. As shown in Fig. 4, the four-classification accuracy of three-branch model exhibits the best results for our proposed model. The results of the four-branch model are presented in the supplementary materials section B. However, as the number of branches increases to 4, the performance of the model does not improve more than three-branch while increasing the amount of network parameters (SFig. 3 in the supplementary materials section B). Therefore, after many experiments in our study, three branches deep neural network frameworks are used in our model. The parameters of one-branch and two-branch are the same as those in STABLE I in the supplementary materials section C. The accuracy of each subject under different numbers of branches is shown in Fig. 4 (a).

To compare the difference in results without the ECA or without the LightGBM module, we use ablation experiments in the proposed MBSTCNN model. In this part, our proposed MBSTCNN-ECA-LightGBM is used as the baseline model. Two compared methods are used in the ablation study. The first model is the MBSTCNN-ECA-Net model without using LightGBM, while the second model is not using the ECA

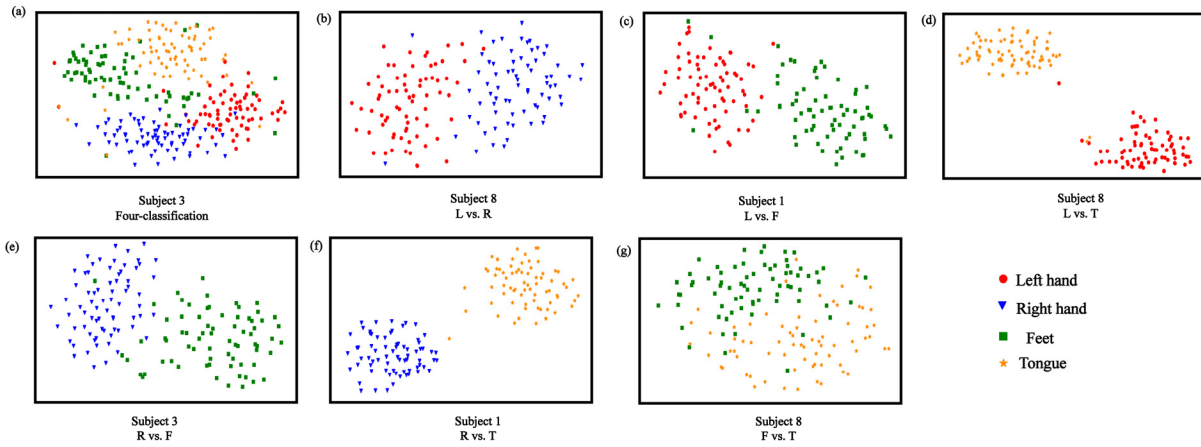


Fig. 3. Results of the t-SNE for the subjects with the highest accuracy in each classification task based on MBSTCNN-ECA-LightGBM. (a) Four-class result of subject 3, (b) L vs. R result of subject 8, (c) L vs. F result of subject 1, (d) L vs. T result of subject 8, (e) R vs. F result of subject 3, (f) R vs. T result of subject 1, (g) T vs. F result of subject 8.

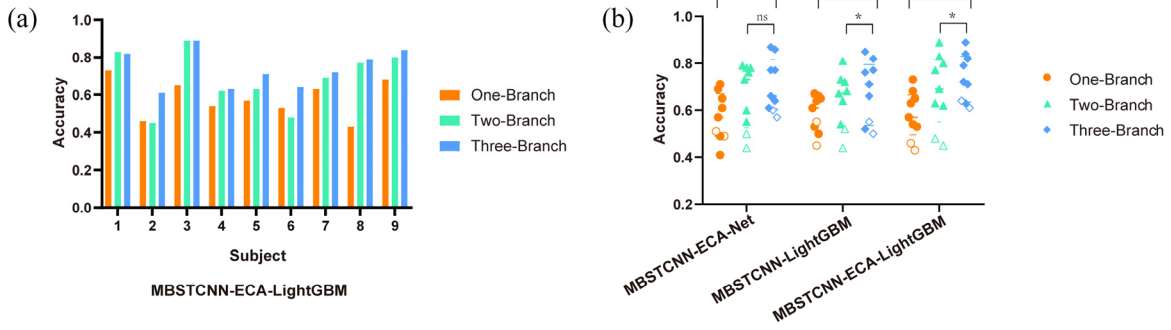


Fig. 4. Comparison of different numbers of branches on the two proposed methods. (a) The MBSTCNN-ECA-LightGBM results under the different number of branches. (b) Paired t-tests are used to test the differences of models with three branches compared to one and two branches, respectively.  $*p < 0.05$ ,  $**p < 0.01$ , *n.s.* means no significant difference. The hollow points in Fig. 4 (b) indicate that the accuracy of the corresponding subjects (subject 2 and subject 6) is significantly improved under our proposed three-branch model (MBSTCNN-ECA-LightGBM).

module (MBSTCNN-LightGBM). The comparison of the effect of different numbers of branches on model performance is also used in ablation experiments. Paired t-tests are used to statistically test the differences between the baseline model and the two ablation models, respectively. The experimental results are shown in Fig. 4 (b). The results showed that for the MBSTCNN-ECA-Net model, three branches could significantly improve the model accuracy compared to one branch ( $p = 0.0040$ ), and there is no significant difference between the two branches and three branches ( $p = 0.1052$ ). For the MBSTCNN-LightGBM model, three branches can significantly improve the model accuracy compared to one branch ( $p = 0.0129$ ) and two branches ( $p = 0.0155$ ). For the MBSTCNN-ECA-LightGBM model, three branches can significantly improve the model accuracy compared to one branch ( $p = 0.0007$ ) and two branches ( $p = 0.0367$ ). Further, we compare the differences of our model under the three branches with the other two models, respectively. The results of the comparison show that there are significant differences between the MBSTCNN-ECA-LightGBM model and the MBSTCNN-ECA-Net model ( $p = 0.0067$ ) as well as the MBSTCNN-LightGBM model ( $p = 0.0008$ ). Most importantly, our three-branch model achieves

improvements on subjects 2 and 6, compared to the literature [12], [23].

### C. Overall Classification Results and Comparison

The MI multi classification results validated the efficiency of our proposed MBSTCNN-ECA-LightGBM on the MI-EEG dataset. Subsequently, we compared our method with several state-of-the-art methods for MI four-class classification. As shown in the TABLE III, the classification accuracies reveal that our method outperformed the previous studies by the FBCSP [12], SVM+FBCSP [23], ShallowNet [45], EEGNet [24], TSFCNN [50] and SPCNN [51]. The experiments show that the performance of MBSTCNN-ECA-LightGBM is improved by 9% (average accuracy) compared with the winning algorithm FBCSP on BCI competition dataset IV-2a. Moreover, for 6 of 9 subjects, the MBSTCNN-ECA-LightGBM achieves the highest accuracy.

## IV. DISCUSSION

In this work, we propose an end-to-end deep learning model, MBSTCNN-ECA-LightGBM, for decoding multi classification MI-BCI tasks using raw EEG signals. The MBSTCNN-ECA-LightGBM mainly consists of three modules: (1) A multi

TABLE III  
FOUR-CLASSIFICATION COMPARISON RESULTS OF OUR PROPOSED MODEL AND STATE-OF-THE-ART MODELS

Method	Subject									Avg
	1	2	3	4	5	6	7	8	9	
FBCSP	0.76	0.57	0.81	0.61	0.55	0.45	0.83	0.81	0.71	0.68
SVM+FBCSP	<b>0.82</b>	0.60	0.83	<b>0.73</b>	0.61	0.44	<b>0.86</b>	0.77	0.75	0.71
ShallowNet	0.63	0.44	0.76	0.53	0.48	0.43	0.65	0.75	0.73	0.60
EEGNet	0.73	0.45	0.82	0.55	0.65	0.49	0.64	0.62	0.82	0.64
TSFCNN	0.72	0.76	0.82	0.68	0.52	0.55	<b>0.86</b>	<b>0.87</b>	0.83	0.73
SPCNN	0.79	0.51	<b>0.89</b>	0.68	0.64	0.57	0.82	0.81	0.79	0.72
<b>MBSTCNN-ECA-Net (Proposed)</b>	0.77	0.60	0.87	0.61	0.66	0.57	0.64	0.77	<b>0.86</b>	0.71
<b>MBSTCNN-ECA- LightGBM (Proposed)</b>	<b>0.82</b>	<b>0.61</b>	<b>0.89</b>	0.63	<b>0.71</b>	<b>0.64</b>	0.72	<b>0.79</b>	0.84	<b>0.74</b>

*Bold values indicate the highest accuracy for each column.*

branch spectral-temporal CNN module to learn frequency and temporal domain features; (2) An efficient channel attention module to recalibrate the spectral-temporal features; (3) A GBDT-based classifier to classify the recalibrated spectral-temporal features. Our results show that the MBSTCNN-ECA-LightGBM method can successfully handle cross-session MI-EEG signals and achieve relatively high results on both two and four-classification tasks.

In general, the non-stationary and individual variability of EEG signals makes the model difficult to decode the cross-session MI tasks [10], [52]. Our current work uses the MBSTCNN-ECA-LightGBM method does overcome the low accuracy problem on the cross-session MI-EEG classification tasks. As listed in TABLE I and TABLE II, the highest accuracy 98% and average accuracy 86% are obtained on the two-classification tasks by our model, while for the four-class task, the highest 89% accuracy and average accuracy 74% are obtained. In a recent MI-EEG classification study, Dai et al. [17] found that the performance of the model is related to the convolution scale in CNN. In our experiments, similar results are obtained with a multi branch CNN network model inspired by a single EEGNet, which obtains more spectral-temporal information by setting different convolution kernels. As shown in TABLE III, compared with the results of single-branch EEGNet, our multi branch CNNs can improve MI decoding performance by capturing different spectral-temporal information. Furthermore, decoding MI-EEG signals with traditional CNN models usually needs to use some data augmentation methods [17], [53], while the multi branch CNNs model can directly learn features from a small number of samples. The t-SNE results (Fig. 3) also indicate that the proposed model is able to learn different MI states from complex EEG signals on the MI-EEG dataset. Thus, our proposed MBSTCNN module can effectively capture the spectral-temporal information in MI-EEG signals and has the potential to solve the problem of less EEG data. To further improve the ability of the model to decode MI tasks, the attention mechanism method is used to adaptively assign higher weights on the spectral-temporal features. Essentially, the attention mechanism enables the model to focus on the key information, providing an important way to capture more reliable features [54]. Recently, as the attention mechanism has been widely used in the MI-EEG classification tasks

[25], [55], for example, Liu et al. [55] used an attention mechanism on the extracted CSP-wise features and class-wise features to further improve the MI classification accuracy. In our work, the function of the ECA module is to recalibrate features by explicitly modeling the interdependencies between feature channels, and it can be easily applied in the existed CNN model [39], [55]. Therefore, ECA assign higher weights adaptively on the spectral-temporal features to obtain more discriminative features. That is, the useful spectral-temporal in each branch can be enhanced and redundant features can be weakened. Furthermore, the architecture of the ECA module (i.e., the adaptive selection of the 1D convolution kernels) enables the model parameters and time consumption to be reduced compared to the attention mechanism used in [56]. Finally, our ablation experiments (Fig. 4 (b)) show a 9% improvement in average classification accuracy (with ECA vs. without ECA).

In terms of classifiers, different from other studies [12], [24], we combine the powerful adaptive learning feature capability of deep learning with the excellent classification effect of LightGBM. The recalibrated spectral-temporal features after MBSTCNN-ECA module as the input of the LightGBM classifier. MI-BCI classification tasks often employ a One-VS-Rest (OVR) strategy, in which one category is regarded as one class, and all the remaining categories are regarded as the other class [12], [57]. Unlike previous studies [12], [58], the LightGBM classifier could directly complete multi classification based on the decision tree model of GBDT. It is worth mentioning that our model achieves better performance on both two and four-classification. In the current study, we trained the model by using a within-subject cross-session strategy. Good results were obtained for each subject in the experiment, which indicated that our model could solve the problem of individual differences to a certain extent. Moreover, our results further illustrate the LightGBM classifier can improve MI classification accuracy (compared to a fully connected layer). Subsequently, as shown in Fig. 4 (b), the result proves that as the model complexity increases, the impact of LightGBM on the decoding accuracy decreases.

Furthermore, different numbers of branches are used in the three models (i.e., MBSTCNN-ECA-Net, MBSTCNN-LightGBM, and MBSTCNN-ECA-LightGBM) to investigate the effect of the number of branches on the MI-EEG



classification accuracy. As described earlier, the purpose of the convolution layers in our MBSTCNN module is mainly to extract spectral and time-domain features under specific frequency bands. Fig. 4 (a) indicates that our three-branch network structure can more effectively learn important frequency band information for MI-EEG [21], [24]. In addition, the ablation experiment results show that the three-branch model achieves the highest accuracy compared to one and two branches (Fig. 4 (b)). As shown in Fig. 4 (b), subject 2 and subject 6 achieve the highest accuracy under the three-branch model. This shows that the multi branch spectral and temporal features have a better effect on the subjects with low classification accuracy in [12] and [23]. Therefore, a certain number of branches have a positive effect on the results of the MI-EEG classification tasks.

In addition, the use of activation functions is also critical for the training of deep neural networks. In order to verify the influence of different activation functions on our proposed MBSTCNN-ECA-Net model, we selected multiple activation functions to perform comparative experiments. The results show that the *ELU* activation function has achieved the best accuracy in 7 of the 9 subjects. (For details, please see supplementary materials section E). Similarly, in order to verify the influence of the selection of different channels on the model, we selected multiple channel combinations for ablation experiments. The results show that for the BCI competition dataset, the best accuracy can be obtained using all channels. This may be the loss of important information caused by random deletion of inappropriate channels, which is consistent with [59]. (Please see supplementary materials section F for details)

Overall, our proposed MBSTCNN-ECA-LightGBM model achieved an improvement in performance compared with some state-of-the-art methods. As shown in TABLE III, traditional machine learning and deep learning are separately used for our model comparison. The four-classification accuracy results show a 9% improvement over FBCSP and a 15% improvement over EEGNet. In addition, our proposed MBSTCNN model does not require much preprocessing and manual feature extraction, which can greatly save time cost. We use ablation experiments to verify whether data preprocessing has an impact on classification performance (More details see supplementary materials section D). The results show that data preprocessing have a negative effect on deep neural networks, which is also consistent with the results in [60] and [61].

However, the learning ability of deep learning relies on large amounts of data, such as in CV [62] and NLP [63]. In fact, our work built a multi branch CNN feature learning network architecture for a relatively small sample size. In future work, the data augmentation work deserves to be added to solve the small sample problem. In addition, although our work achieved good performance, the interpretability of deep learning is poor, and the abstract features learned by CNN are difficult to interpret physiological meanings. In future work, we will further explore the interpretability research of deep learning in MI-BCI. Moreover, only the within-subject cross-session strategy is used in our experiments. For the cross-subject task

on the BCI Dataset, the accuracy obtained by our model is limited. The stability and robustness of cross-subject tasks should be enriched in follow-up studies.

## V. CONCLUSION

In our study, we propose an end-to-end MBSTCNN-ECA-LightGBM framework based on a multi branch spectral-temporal CNN module, an attention mechanism module, and the LightGBM to decode MI-BCIs. By the multi branch CNN, we could extract spectral-temporal features of EEG. We added the ECA module to assign weight values for the features, improving the discrimination of features. Finally, we used the LightGBM classifier to classify the MI tasks. The experiment results demonstrated that our proposed model could effectively capture the spectral-temporal information from the original EEG signal, achieving a better classification performance.

## REFERENCES

- [1] B. Hamner, R. Leeb, M. Tavella, and J. D. R. Millán, "Phase-based features for motor imagery brain-computer interfaces," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 2578-2581.
- [2] K. K. Ang et al., "A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke," *Clin. EEG Neurosci.*, vol. 46, no. 4, pp. 310-320, 2015.
- [3] L.-W. Ko, S. S. K. Ranga, O. Komarov, and C.-C. Chen, "Development of single-channel hybrid BCI system using motor imagery and SSVEP," *J. Healthcare Eng.*, vol. 2017, pp. 1-7, Aug. 2017.
- [4] T. Zhang et al., "Structural and functional correlates of motor imagery BCI performance: Insights from the patterns of fronto-parietal attention network," *Neuroimage*, vol. 134, pp. 475-485, Jul. 2016.
- [5] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, 2016, Art. no. 016003.
- [6] G. Pfurtscheller and A. Aranibar, "Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 46, no. 2, pp. 138-146, 1979.
- [7] G. Pfurtscheller, "Event-related synchronization (ERS): An electrophysiological correlate of cortical areas at rest," *Electroencephalogr. Clin. Neurophysiol.*, vol. 83, no. 1, pp. 62-69, Jul. 1992.
- [8] Y. Hou, L. Zhou, S. Jia, and X. Lun, "A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN," *J. Neural Eng.*, vol. 17, no. 1, Feb. 2020, Art. no. 016048.
- [9] B. Blankertz et al., "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303-1309, 2010.
- [10] X. Zhao, J. Zhao, C. Liu, and W. Cai, "Deep neural network with joint distribution matching for cross-subject motor imagery brain-computer interfaces," *BioMed Res. Int.*, vol. 2020, pp. 1-15, Feb. 2020.
- [11] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 4, pp. 441-446, Dec. 2000.
- [12] 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.
- [13] H. Yang, S. Sakhavi, K. K. Ang, and C. Guan, "On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2620-2623.
- [14] T. Gandhi, B. K. Panigrahi, and S. Anand, "A comparative study of wavelet families for EEG signal classification," *Neurocomputing*, vol. 74, no. 17, pp. 3051-3057, 2011.
- [15] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 4, pp. 317-326, Aug. 2008.
- [16] H. Baali, A. Khorshidtalab, and M. Mesbah, "A transform-based feature extraction approach for motor imagery tasks classification," *IEEE J. Transl. Eng. Health Med.*, vol. 3, 2015, Art. no. 2100108.

- [17] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016025.
- [18] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*.
- [19] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8599–8603.
- [20] X. Deng, B. Zhang, N. Yu, K. Liu, and K. Sun, "Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces," *IEEE Access*, vol. 9, pp. 25118–25130, 2021.
- [21] X. Zhu, P. Li, C. Li, D. Yao, R. Zhang, and P. Xu, "Separated channel convolutional neural network to realize the training free motor imagery BCI systems," *Biomed. Signal Process. Control*, vol. 49, pp. 396–403, Mar. 2019.
- [22] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, "A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting," *IEEE Access*, vol. 7, pp. 28309–28318, 2019.
- [23] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.
- [24] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
- [25] H. Zhang, X. Zhao, Z. Wu, B. Sun, and T. Li, "Motor imagery recognition with automatic EEG channel selection and deep learning," *J. Neural Eng.*, vol. 18, no. 1, Nov. 2020, Art. no. 016004.
- [26] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3D convolutional neural network for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2164–2177, Oct. 2019.
- [27] 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*, 2021, pp. 736–751.
- [28] W. S. Noble, "What is a support vector machine?" *Nature Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [29] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [30] F. V. Jensen, "Causal and Bayesian networks," in *Bayesian Networks and Decision Graphs*. Cham, Switzerland: Springer, 2001, pp. 3–34.
- [31] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3146–3154.
- [32] C. Chen, Q. M. Zhang, Q. Ma, and B. Yu, "LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion," *Chemometrics Intell. Lab. Syst.*, vol. 191, pp. 54–64, Aug. 2019.
- [33] X. Sun, M. Liu, and Z. Sima, "A novel cryptocurrency price trend forecasting model based on LightGBM," *Finance Res. Lett.*, vol. 32, Jan. 2020, Art. no. 101084.
- [34] H. Zeng et al., "A LightGBM-based EEG analysis method for driver mental states classification," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–11, Sep. 2019.
- [35] S. Aggarwal, L. Aggarwal, M. S. Rihal, and S. Aggarwal, "EEG based participant independent emotion classification using gradient boosting machines," in *Proc. IEEE 8th Int. Advance Comput. Conf. (IACC)*, Dec. 2018, pp. 266–271.
- [36] 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., 2008, pp. 1–6, vol. 16.
- [37] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
- [38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.
- [39] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.
- [40] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [42] P. Li, "Robust LogitBoost and adaptive base class (ABC) LogitBoost," 2012, *arXiv:1203.3491*.
- [43] M. Richardson, E. Dominowska, and R. Ragno, "Predicting clicks: Estimating the click-through rate for new ads," in *Proc. 16th Int. World Wide Web Conf. (WWW)*, 2007, pp. 521–530.
- [44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [45] 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.
- [46] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "EEG classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, no. 3, p. 551, 2019.
- [47] J. V. Dillon et al., "TensorFlow distributions," 2017, *arXiv:1711.10604*.
- [48] F. Shamsi, A. Haddad, and L. Najafzadeh, "Early classification of motor tasks using dynamic functional connectivity graphs from EEG," *J. Neural Eng.*, vol. 18, no. 1, Feb. 2021, Art. no. 016015.
- [49] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.
- [50] X. Jia, Y. Song, L. Yang, and L. Xie, "Joint spatial and temporal features extraction for multi-classification of motor imagery EEG," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103247.
- [51] X. Zhao et al., "Deep CNN model based on serial-parallel structure optimization for four-class motor imagery EEG classification," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103338.
- [52] R. Srinivasan, "Methods to improve the spatial resolution of EEG," *Int. J. Bioelectromagnetism*, vol. 1, no. 1, pp. 102–111, 1999.
- [53] K. Zhang et al., "Data augmentation for motor imagery signal classification based on a hybrid neural network," *Sensors*, vol. 20, no. 16, p. 4485, Aug. 2020.
- [54] T. Zhang et al., "Separated channel attention convolutional neural network (SC-CNN-attention) to identify ADHD in multi-site RS-fMRI dataset," *Entropy*, vol. 22, no. 8, p. 893, Aug. 2020.
- [55] C. Liu et al., "Distinguishable spatial-spectral feature learning neural network framework for motor imagery-based brain-computer interface," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 0460e4.
- [56] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7132–7141.
- [57] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *J. Neural Eng.*, vol. 16, no. 6, Oct. 2019, Art. no. 066004.
- [58] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A CSP\AM-BA-SVM approach for motor imagery BCI system," *IEEE Access*, vol. 6, pp. 49192–49208, 2018.
- [59] X. Lun, Z. Yu, T. Chen, F. Wang, and Y. Hou, "A simplified CNN classification method for MI-EEG via the electrode pairs signals," *Frontiers Human Neurosci.*, vol. 14, p. 338, Sep. 2020.
- [60] A. Hassanpour, M. Moradikia, H. Adeli, S. R. Khayami, and P. Shamsinejadbabaki, "A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals," *Exp. Syst.*, vol. 36, no. 6, Dec. 2019, Art. no. e12494.
- [61] H. Altaheri et al., "Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review," *Neural Comput. Appl.*, 2021, doi: [10.1007/s00521-021-06352-5](https://doi.org/10.1007/s00521-021-06352-5).
- [62] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 558–567.
- [63] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.