IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 29, 2019, accepted December 17, 2019, date of publication December 27, 2019,
date of current version January 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962740

A Novel MI-EEG Imaging With the Location Information of Electrodes

MING-AI LI[1,2], JIAN-FU HAN[1], AND LI-JUAN DUAN[1]
[1]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[2]Beijing Key Laboratory of Computational Intelligence and Intelligent Systems, Beijing 100124, China

Corresponding author: Ming-Ai Li (limingai@bjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11882003, Grant 81471770, and Grant 61672070, and in part by the Natural Science Foundation of Beijing under Grant 4182009.

ABSTRACT Combination of the Motor Imagery EEG (MI-EEG) imaging and Deep Convolutional Neural Network is a prospective recognition method in brain computer interface. Nowadays, the frequency or time-frequency analysis has been applied to each channel of MI-EEG signal to obtain a spatio-frequency or time-frequency image, and even the images from several channels are infused to generate a combined image. However, the real position information of channels or electrodes is lost in these MI-EEG images, and this is contradictory to the activation area of MI-tasks. In this paper, the MI period and the frequency band covered by μ and β rhythms are divided into ten time windows and three sub-bands, respectively. Then, for each electrode, Fast Fourier Transform (FFT) is employed to transform each time window to spectrum, and its inverse FFT is calculated for each sub-band. The time-domain powers of ten time windows are averaged for the same sub-band. So, three average powers are generated as the time-frequency features of each electrode of MI-EEG. They are further arranged to the electrode coordinate figure by using Clough-Tocher interpolation algorithm, and a complicated image, in which the time-frequency features are correctly located at the real position of each electrode, is obtained to embody the MI-EEG in detail. Furthermore, a VGG network is modified to perform effective recognition for MI-EEG image, and it is called mVGG. Extensive experiments are conducted on three publicly available datasets, and the 10-folds cross validation accuracies of 88.62%, 92.28% and 96.86% are achieved respectively, and they are higher than that of the state-of-the-art imaging methods. Kappa values and ROC curves demonstrate our method has lower class skew and error costs. The experimental results show that the effectiveness of proposed MI-EEG imaging method, and it is well-matched with mVGG.

INDEX TERMS Brain computer interface, convolutional neural network, interpolation method, machine learning, MI-EEG imaging method.

I. INTRODUCTION

Electroencephalography (EEG)-based Brain-Computer interface (BCI) technology is an efficient pathway of communication between patients and peripheral equipment. BCI aims at transforming the mental activity of the human brain into EEG signals that can be processed by a computer. Due to the high temporal resolution and easy operation, this non-invasive method is frequently used in some applications, e.g., text input systems [1], wheelchairs [2], and rehabilitation devices [3], [4] etc., which has practical implications for disabled patients.

MI-EEG refers to the EEG signal of imagining body movement without actual movement. The power of MI-EEG signals will decrease or increase in the μ and β rhythm when a subject imagines body parts moving [5]; they are called event-related desynchronization (ERD) or event-related synchronization (ERS) phenomena [6]. Therefore, the MI-EEG signal has significant time-frequency characteristics. In addition, differences exist in the activated brain regions between different MI tasks. Thus, the MI-EEG signal also has spatial distribution characteristics. So, the key of feature extraction and recognition lies in the effective use of the time, frequency and spatial information of MI-EEG.

The convolutional neural network (CNN) has been applied to EEG recognition. CNN, which is inspired by the sensory field mechanism of the biological visual nerve [7],

The associate editor coordinating the review of this manuscript and approving it for publication was Benyun Shi.

can learn local non-linear features (through convolutions and nonlinearities) and represent higher-level features as compositions of lower level features (through multiple layers of processing) [8]. The combination of CNN and MI-EEG recognition can be distinguished as two classes, non-imaging measure and imaging measure. The non-imaging measure directly uses the raw or the spatially filtered MI-EEG signal as the CNN's input. The raw MI-EEG without prior time-frequency domain knowledge and handcrafted features was directly handled, even the feature fusion was considered in CNN framework to retain the information in both abstract and global levels [9], [10]. Furthermore, the raw signals were filtered by common spatial pattern (CSP) or its extended algorithms, the CSP features from multi-level decomposition of the frequency ranges were calculated, and they were stacked to input to a sequential CNN, or they were directly input to a separated CNN to independently encode temporal and/or frequency information [11]–[14]. To make full use of the temporal, frequency and spatial domains characteristics of MI-EEG and the superiority of CNN in image recognition, some imaging methods were also developed. Based on Fast Fourier Transform (FFT), Short Time Fourier Transform (STFT) and Wavelet Transform (WT) etc., each electrode of MI-EEG sequence was converted to form a spatio-frequency or time-frequency image. The images from multi-electrodes were fed to a CNN separately or merged into one image as the representation of a trail [15]–[23].

In summary, the spatio-frequency or time-frequency characteristics are infused into the MI-EEG image, and the recognition accuracies of imaging methods are relatively better than non-imaging ones. However, the real location information of electrodes is not reflected in the MI-EEG image, and it is very relevant to the activation of MI-tasks. In this paper, based on Clough-Tocher interpolation algorithm and FFT, a novel MI-EEG imaging method is proposed to integrate the time-frequency features with the exact location information of electrodes, and the power change of each electrode MI-EEG signal with time and frequency is simultaneously displayed in a MI-EEG image. In addition, the Deep CNN (DCNN) structure is modified to deal with the exploding information. The extensive experimental results show the effectiveness of the proposed imaging, the superiority of the modified DCNN for MI-EEG images, and the good degree of matching between the imaging method and DCNN.

The rest of this paper will be organized as follows. Section II details the related work about imaging methods. Section III relays the feature extraction, imaging methods and CNN design. Section IV introduces comparative experiments and the results on three public datasets. Section V discusses internal relations of experiments and a summary of the results. Section VI concludes the study and summarizes the shortcomings and future directions.

II. RELATED WORK

In this section, the related FFT, STFT and WT based imaging methods of MI-EEG will be described in detail.

A. FFT-BASED IMAGING METHOD

Uktveris and Jusas [15] employed FFT for computing the ith MI-EEG channel signal energy estimation in the frequency domain, forming a single row in the FFT energy map (FFTEM). A full signal window was used to gain a global energy view. Each element was calculated using formula (1).

$$H(i) = |FFT(x(i))|, \quad i = 1, 2, \cdots, N \tag{1}$$

where N is the number of electrodes. The resulting spatio-frequency image was generated, as shown in Fig. 1.

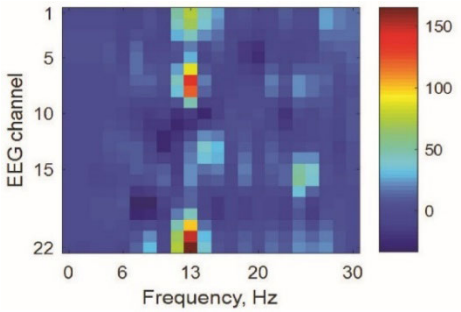


FIGURE 1. FFT energy map.

B. STFT-BASED IMAGING METHOD

Based on STFT, each MI-EEG time-series can be converted into a time-frequency representation by computing spectrograms, and the images generated from multi-electrodes are even integrated, yielding a combined image.

In [21], each channel of MI-EEG in the selected time period (4s) was calculated by STFT to obtain a time-frequency image with a frequency band in the range from 0∼120 Hz, Fig. 2 shows the time-frequency image of the C3 electrode. Three images from the C3, Cz and C4 channels were input into CNN. Similarly, Wang *et al.* [16] used STFT to generate a time-frequency image for each channel of MI-EEG. The frequency components from 8∼30 Hz were extracted, which included the rhythms of μ and β. In contrast, the images of C3, Cz and C4 were merged into one image, as shown in Fig. 3, and further input into CNN.

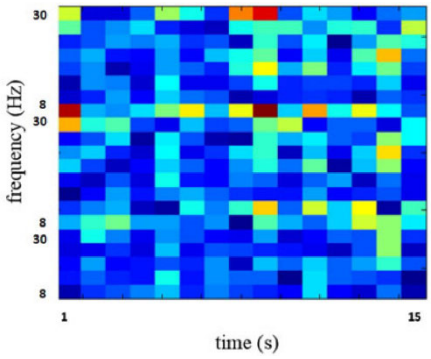


FIGURE 2. Time-frequency image of C3.

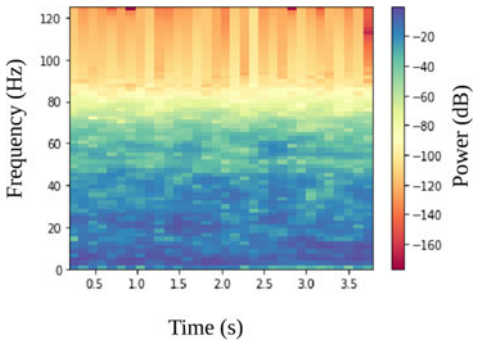


FIGURE 3. Time-frequency image of three electrodes.

The time-frequency images were also produced by STFT in [17], but the frequency bands were selected as 6∼13 Hz and 17∼30 Hz. In each window of STFT, the spectral power was calculated as the feature value. Images of three electrodes were also merged in one figure. Fig. 4 shows an example of the image in [17].

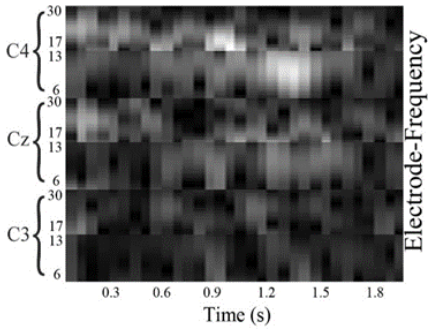


FIGURE 4. Example of an STFT-based image.

C. WT-BASED IMAGING METHOD

Each channel of the MI-EEG signal was multi-scale-analyzed by WT, and a time-frequency image was obtained [19]. The frequency band was in the range from 0∼60 Hz. The images generated from three channels (C3, Cz and C4) were finally combined as the final input images of a CNN. The wavelet time-frequency image of C3 is shown in Fig. 5, and the final image is demonstrated in Fig. 6.

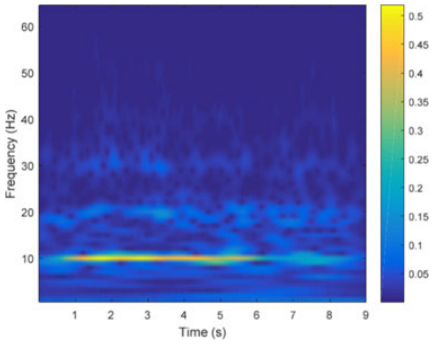


FIGURE 5. Wavelet Time-frequency image of C3.

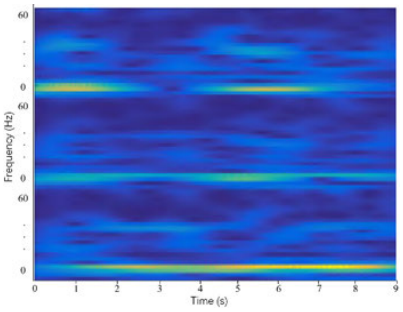


FIGURE 6. Combined wavelet time-frequency image of three electrodes.

III. METHODOLOGY

The imaging method and the framework design of DCNN will be developed. In Section III-A, Clough-Tocher interpolation algorithm is reviewed and how it is employed in MI-EEG imaging. In Section III-B, the feature calculation and interpolation operation are presented in steps. The framework of Visual Geometry Group Network (VGG) and its modification are presented in Section III-C.

A. CLOUGH-TOCHER INTERPOLATION ALGORITHM

Clough-Tocher (CT) interpolation algorithm was first introduced in 1960's [24], and it has been the most widely used multi-dimensional scattered data interpolant. Consider the scattered points $(\boldsymbol{x_n}, \boldsymbol{y_n})$ located in the x-y plane and their values $\boldsymbol{z_n}$ over the plane, the triangulation of the scattered points in the x-y plane induces a piecewise triangular surface over the plane, whose nodes are the points $(\boldsymbol{x_n}, \boldsymbol{y_n}, \boldsymbol{z_n})$. In CT algorithm, a piecewise cubic function is employed as the interpolant for each triangle. Specifically, each triangle is further divided into three equivalent subtriangles, where a cubic function in the form of B'ezier surface is estimated. The finite element method based CT algorithm can preserve the local spatial relation and reconstruct the relation of the original input. Therefore, it is selected to construct the MI-EEG image with time-frequency-space features in this paper.

B. TIME DOMAIN POWER AND CLOUGH-TOCHER INTERPOLATION-BASED IMAGING (TPCT)

To sufficiently utilize both time-frequency feature of MI-EEG and the location of electrodes, the novel imaging method is proposed based on time domain power on interest frequency bands and CT interpolation algorithm, noted as TPCT. The frequency band 8∼30 Hz is taken as the region of interest for motor imagery [25]. It is divided into three sections, on which the time-frequency features are independently extracted by using FFT. CT algorithm is further applied to guarantee the time-frequency features locating at corresponding electrode position.

1) TIME-FREQUENCE FEATURE CALCULATION

Suppose that $x_m \in R^{1\times N_s}$ is the MI-EEG signal over the mth electrode, where $m \in \{1, 2, 3, \ldots, N_c\}$, N_c is the number of

electrodes, N_s represents the number of sampling points. So, X_m can be described by the following Eq.:

$$x_m = \left[x_m(1), x_m(2), \ldots, x_m(N_s) \right] \in R^{1\times N_s} \quad (2)$$

Then, the x_m is divided into N_D windows, $N_D \in N^+$. The data in each window are

$$x_{m,j}^D \in R^{1\times \frac{N_s}{N_D}}, \quad (3)$$

where j is the serial number of windows, $j \in \{1, 2, 3, \ldots, N_D\}$. So, $x_{m,j}^D$ concludes the sampling points:

$$x_{m,j}^D = \left[x_m^{((j-1)\times \frac{N_s}{N_D})} \ldots x_m^{(j\times \frac{N_s}{N_D})} \right] \in R^{1\times \frac{N_s}{N_D}} \quad (4)$$

FFT was used to transfer the data from each time window to the frequency domain. In order to improve the frequency resolution, the discrete time series $x_{m,j}^D$ is zero-padded to length N_{FFT}, the transferred sequence was denoted as $X_{m,j}^F$.

$$X_{m,j}^F \in C^{1\times N_{FFT}}, \quad (5)$$

Furthermore, the frequency band 8∼30 Hz was divided into three sub-bands, i.e. 8∼13 Hz, 13∼21 Hz and 21∼30 Hz, and the three corresponding sub-sequences of $X_{m,j}^F$ were obtained as shown in Eq. (6):

$$X_{m,j,f}^F \in C^{1\times N_{F,f}}, \quad (6)$$

where $f \in \{1, 2, 3\}$ is the serial number of frequency bands. The length of each sub-sequences is $N_{F,f}$, which is calculated by Eq. (7).

$$N_{F,f} = (F_{H,f} - F_{L,f}) \times \frac{N_{FFT}}{\frac{f_s}{2}}, \quad (7)$$

where $F_{H,f}$ and $F_{L,f}$ are the upper limit and lower limit of one frequency band, and f_s is the sampling frequency.

Then, $X_{m,j,f}^F$ is transferred into time domain independently based on Inverse FFT (IFFT). Thus, $x_{m,j,f}^I$ is obtained as follows.

$$x_{m,j,f}^I = IFFT\left(X_{m,j,f}^F\right) \in R^{1\times N_{F,f}}, \quad (8)$$

It is further described as shown in Eq. (9).

$$x_{m,j,f}^I = (x_{m,j,f}^I(1), x_{m,j,f}^I(2), \ldots, x_{m,j,f}^I(N_{F,f})), \quad (9)$$

The average power is calculated independently for each frequency band sequence according to Eq. (10).

$$x_{m,j,f}^P = \frac{1}{N_{F,f}} \sum_{T_f=1}^{N_{F,f}} x_{m,j,f}^{I^2}(T_f), \quad (10)$$

The feature $x_{m,f}^F$ is obtained by averaging three power values over N_D time windows as shown in Eq. (11).

$$x_{m,f}^F = \frac{1}{N_D} \sum_{j=1}^{N_D} x_{m,j,f}^P \in R^1, \quad (11)$$

So the feature matrix of one trail is constructed by Eq. (12):

$$x^F = \begin{bmatrix} x_{1,1}^F & x_{1,2}^F & x_{1,3}^F \\ \vdots & \vdots & \vdots \\ x_{N_c,f}^F & x_{N_c,2}^F & x_{N_c,3}^F \end{bmatrix} \in R^{N_c\times 3}, \quad (12)$$

2) INTERPOLATION IMAGING

The N_c coordinates were acquired from the electrode's distribution figure of a BCI system, defined as $M \in R^{2\times N_C}$. Then, in view of the construction of time-frequency feature x^F, slight movements were adopted twice for these coordinates by adding and subtracting a small value in both the x and y axes. Therefore, the new coordinate system of electrodes was obtained, defined as $M' \in R^{2\times(N_C\times 3)}$.

A 64*64 pixel resolution grid system, denoted as $G \in R^{64*64}$, was established based on four points (x_{max}, y_{max}), (x_{max}, y_{min}), (x_{min}, y_{max}), and (x_{min}, y_{min}), which were composed of the maximum and minimum values of $M' \in R^{2\times(N_C\times 3)}$ on the x and y axes. Then, X_f^F was interpolated into $G \in R^{64*64}$ by $M' \in R^{2\times(N_C\times 3)}$. Thus, an image $G_f \in R^{64*64}$, including tri-channel time-frequency features and electrode coordinate information, was finally obtained.

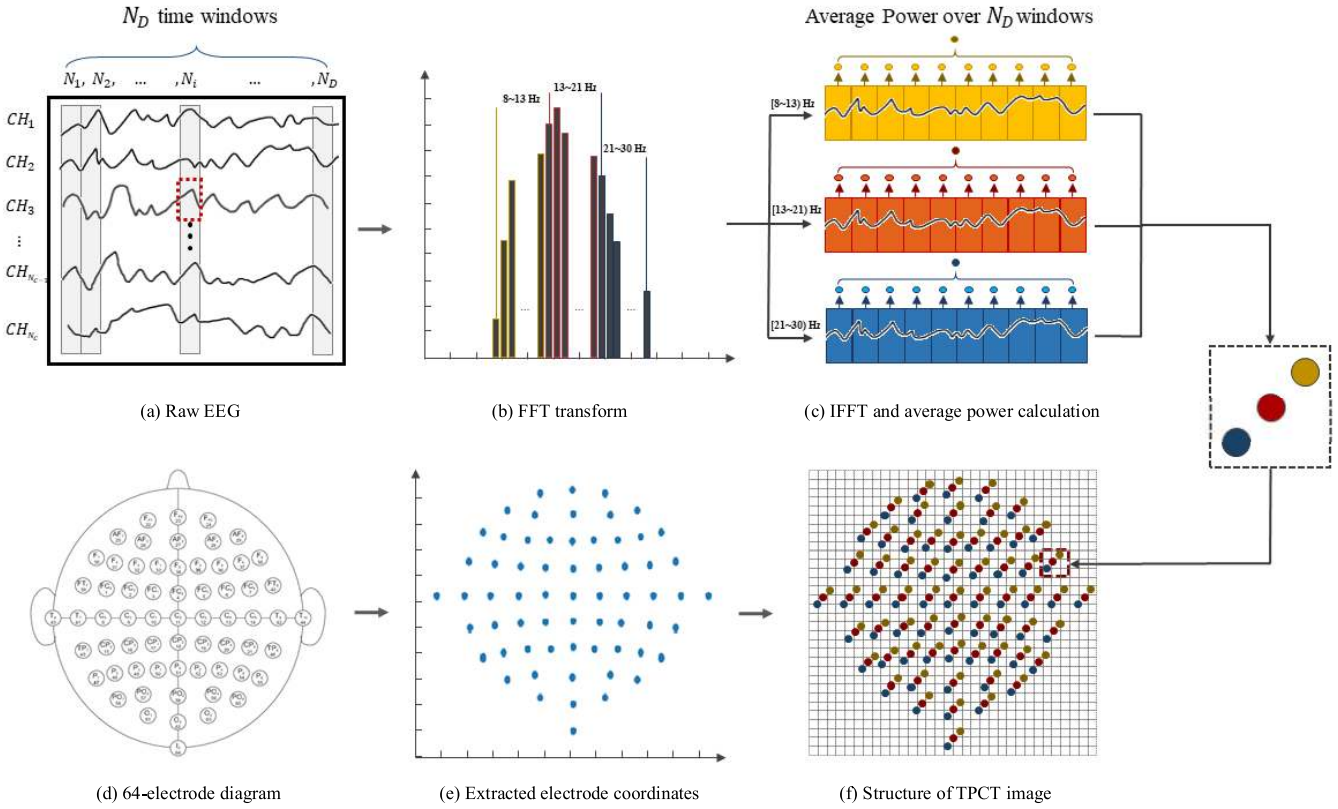
Fig. 7 shows the generation process for the TPCT image. Fig. 7 (a) shows a schematic diagram of the raw EEG signal, in which CH_i represents the ith channel, and $i = \{1, 2, \cdots, N_C\}$. Fig. 7 (b) shows the process of FFT transform of each time window for each channel. The average power of the time domain for the three frequency bands was calculated independently, as shown in Fig. 7 (c). Fig. 7 (d) shows the 64-electrode distribution diagram. Fig. 7 (e) presents the extracted coordinates of the 64-electrode diagram, and Fig. 7 (f) shows the internal structure of the TPCT image. In the feature extraction, the three frequency-bands of interest are considered independently. The μ rhythm, including 8-13Hz, is preserved intact, and the β rhythm is divided into 13-21Hz and 21-30Hz to obtain the ample time-frequency information. Thus, according to TPCT method, three average powers of time domain from three frequency bands were calculated, and their positions were located by using CT interpolation. The detailed process of TPCT is as follows: A 64 × 64 grid system was established in advance, and the boundaries of features' coordinate and the grid system were coincided, then the values of all grid points were computed by CT interpolation algorithm to expand the time-frequency features into a 2D space, generating a complicated MI-EEG image.

C. THE MODIFIED VISUAL GEOMETRY GROUP NETWORK

1) VISUAL GEOMETRY GROUP NETWORK

The Visual Geometry Group Network (VGG) was published in ICLR 2015 and received proxime accessit in the ImageNet Challenge 2014 [26]. Due to its good structural adaptability, VGG is widely used in the fields of image feature extraction and transfer learning [27]. The structure of VGG-16 layer is shown in Fig. 8 (a). The study of VGG verified that the deep stacking of small 3 × 3-size convolution kernels could increase the nonlinear mapping combination and reduce the computational complexity, which was also verified in the large-scale image classification database [26].

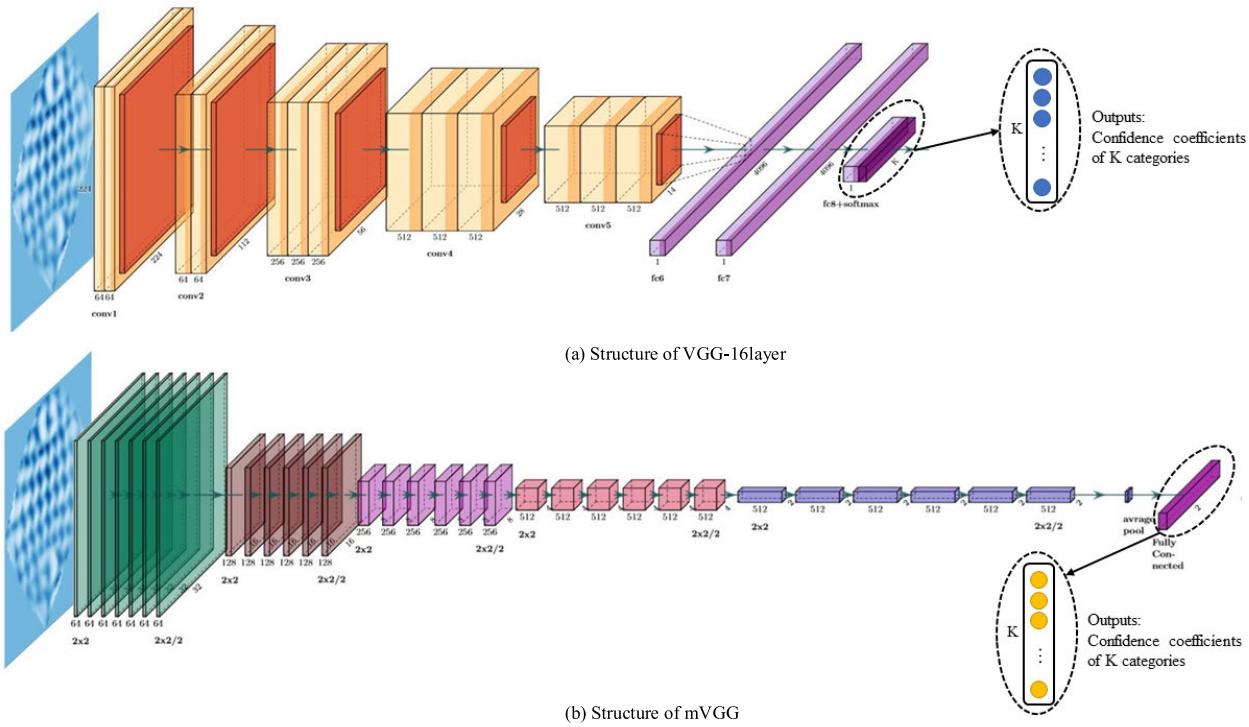
The input and output dimensions of VGG were adjusted to fit the MI-EEG data, and the Fully-Connected (FC) layer



(a) Raw EEG (b) FFT transform (c) IFFT and average power calculation

(d) 64-electrode diagram (e) Extracted electrode coordinates (f) Structure of TPCT image

FIGURE 7. Generation process of the TPCT.



(a) Structure of VGG-16layer

(b) Structure of mVGG

FIGURE 8. Structures of two CNN frameworks. The input is TPCT image of MI-EEG.

length was changed from 4096 to 256. VGG was used as a baseline to validate the viewpoints in this paper. The TPCT images with the 64 × 64 pixels were input into the network, and the 3 × 3 convolutional kernel was adopted to extract the relationship of the raw pixels. The overall convolutional structure contained five sections, and the spatial resolution of

the feature map in each section was preserved by using the padding of 1-pixel. The configuration is presented in Table 1. At the end of each section, Max-pooling was performed over a 2 × 2-pixel window to abstract features by preserving the maximum pixel value.

TABLE 1. Configurations of VGG and mVGG.

CNN Configuration		
Model:	VGG	mVGG
Input:	64x64 image	
1th:	(conv3-64)x2	(conv3-64)x6
DS:	Max-pool	conv2-64(stride=2)
2nd:	(conv3-128)x2	(conv2-128)x5
DS:	Max-pool	conv2-128(stride=2)
3rd:	(conv3-256)x4	(conv2-256)x5
DS:	Max-pool	conv2-256(stride=2)
4th:	(conv3-512)x4	(conv2-512)x5
DS:	Max-pool	conv2-512(stride=2)
5th:	(conv3-512)x4	(conv2-512)x5
DS:	Max-pool	conv2-512(stride=2)
FC	FC:256	Average-pool
FC	FC:256	FC:2
FC	FC:2	
Soft-max		

The Max-pooling operation demonstrated invariance properties in translation, rotation and scale, i.e., when the image is shifted within a few pixels, the deviation will not affect the final feature map, which can benefit the generation ability. However, simply discarding the data also led to instability of the training and a reduction of model accuracy. The feature map in the last section with a resolution of 2 × 2 was stacked followed by three Fully-Connected (FC) layers: the first two had 256 kernels each, and the third had 2 kernels for each class. The final layer was the soft-max layer. Regularization [28] and Dropout [29] techniques were used in the FC layers to reduce overfitting problems and model complexity.

2) THE MODIFIED VGG NETWORK

The VGG Network was modified to match dense information images, like TPCT images, and the modified network was referred to as mVGG. The structure of mVGG is shown in Fig. 8 (b). The mVGG was mainly adjusted in four aspects. The configurations of the VGG and mVGG are shown in Table 1. The overall structure of mVGG consisted of five sections of convolution labeled as 1th, 2nd… 5th. Moreover, 'DS' denotes the Down Sampling layer, and 'FC' represents the Fully Connected layer.

In the description of configuration, for example, '(conv3-64) ×2' means the size of the kernel is 3 × 3, the amount of the kernel is 64 and there are 2 convolution layers in this section. Concerning the remaining tags, the 'M-pool' represents the Max-pooling layer, the 'A-pool' represents the average pooling layer and 'FC: 256' indicates the length of the FC layer is 256. The Soft-max function, shown in Eq. (13), was employed for the final output to obtain the corresponding probability.

$$P(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{13}$$

where j ∈ {1, 2, 3, . . . , K}. The e^{z_j} is the output of each kernel of the last layer, and K is equal to the amount of classes.

a: REPLACING MAX POOLING WITH THE CONVOLUTION LAYER

The max-pooling layer in the VGG structure was replaced by the convolution operation. In studies of EEG recognition and CNN, many researchers have described the instability of Max-pooling. In our study, the convolution operation via the 2 × 2 kernel with stride 2 was used for sub-sampling; thus, each obtained pixel represented a nonlinear combination of the input data. Fig. 9 shows the Max-pooling operation in the third convolutional segment. It can be seen that a Max-pooling kernel with a size of 2 × 2 covers the area of 9 electrodes, and 75% of the information would be discarded. Hence, this process would cause instability in the training process.

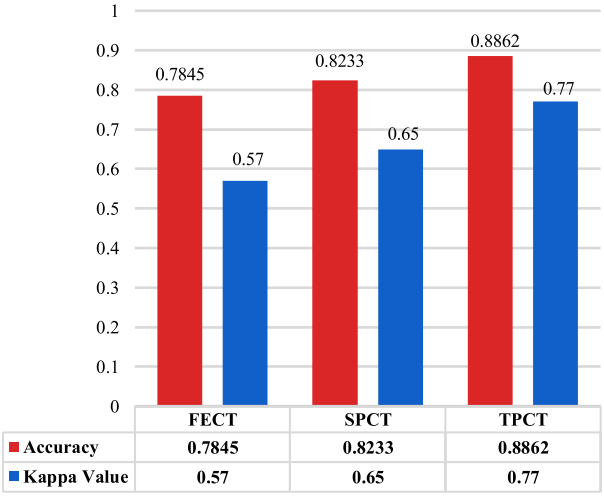


FIGURE 9. Impact of different representative features and the division of frequency band on the recognition accuracies and Kappa values.

b: MODIFYING THE SIZE OF THE CONVOLUTION KERNEL

In contrast to the redundant information for the natural scene image, the interpolation image-based electrode position showed a greater relationship to the size of the convolution kernel. We reserved the 3 × 3 kernel in the first convolution layer, which helped to achieve a larger sensing field in the low-order feature maps and changed all the subsequent convolution layers to the 2 × 2 kernel, which could provide a more non-linear combination of electrodes. Thus, the mVGG could make full use of the space information for MI-EEG, which is more suitable for recognition compared with the original VGG.

c: DEEPENING NETWORK LAYERS

The convolutional layers close to the input contained a small receptive field when projected back to the original image, which represented specific features, while the convolutional layer close to the output had a large receptive field in the original image, which represented abstract features. Due to the large number of electrodes in the BCI2000 dataset, a single electrode only occupied a small number of pixels. Therefore, 7 convolution layers were arranged in the first convolution segment of mVGG to facilitate specific feature extraction and non-linear combination on such a small pixel scale. There were 5 layers in the remaining segments.

d: USING AVERAGE POOLING INSTEAD OF THE FULLY CONNECTED LAYER

To avoid the overfitting problem caused by the small amount of neural biological signal datasets, the average pooling operation was adopted to replace the FC layers, which has obvious redundancy. Two convolutional kernels were adopted to calculate 512 feature maps, which were output from the last convolutional segment. Through the average-pooling layer, the values, which represented information for the categories, were obtained. This adjustment was able to alleviate overfitting and computation complexity.

IV. RESULTS

A. RESULTS BASED ON THE BCI2000 DATASET

1) DESCRIPTION OF THE BCI2000 DATASET

This dataset includes 109 subjects, who performed different motor imagery tasks while recording 64-channel EEG. This multi-electrode system (BCI2000 system) can better show the overall information of electrode linkage. The EEGs were recorded from 64 electrodes as per the international 10-10 system (excluding electrodes Nz, F9, F10, FT9, FT10, A1, A2, TP9, TP10, P9, and P10), as shown in Fig. 3 (d)). The numbers below each electrode name indicate the order in which they appear in the records. Note that signals in the records are numbered from 0 to 63, while the numbers in the figure range from 1 to 64.

The dataset includes 4702 trails, which contains half of the imaging for opening and closing both fists and half of the imaging for both feet. Each trail includes 800 samples in 5 seconds, sampled at 160 samples per second. Moreover, the first second corresponded to the rest that was applied in the baseline elimination [30].

It is worth noting that all the data from 109 subjects were arranged to merge into a uniform dataset to meet the requirement of DCNN. In this way, the DCNN, as a universal model, can be applied to a variety of scenarios without manually selecting features or optimizing parameters for each subject. Meanwhile, it will request significantly higher performance than the personalized recognition methods.

2) EXPERIMENTAL ENVIRONMENT AND PERFORMANCE INDEXES

a: EXPERIMENTAL ENVIRONMENT

The experiment in this study was performed in the TensorFlow environment on NVDIA GTX1080Ti and Intel 2.1Ghz Xeon Silver 4110 CPU with 64G RAM.

The network was designed in TensorFlow 1.6.

b: PERFORMANCE INDEXES

Two evaluation criterions were used to demonstrate the model's performance, which are the Accuracy and Kappa value calculated by equations (14) and (15), respectively.

$$\text{Accuracy} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{Kappa} = \frac{Accuracy - R}{1 - R} \tag{15}$$

where "TP" was the True-Positives field in the confusion matrix, "FN" was the False-Negatives field. The R in Eq. (15) denotes the random classification rate for the problem. The R is 0.5 in two-category classification problem.

Moreover, the receiver operating characteristic curve (ROC) graph is applied to visualize the model's performance. The horizontal axis of the ROC is the false positive rate, and the vertical axis is the true positive rate. According to different thresholds, the ROC curve acquires different Ture Positive rates on the vertical axis, reflecting the response of the model to stimuli at various points. We can judge the model by the Area Under the Curve (AUC) and determine the stability and robustness of the discriminant probability.

3) COMPARISON OF DIFFERENT REPRESENTATIVE FEATURES

In previous studies, many representative features were selected and had a significant impact on the experimental results. To represent the effectiveness of time domain power as time-frequency features in proposed TPCT, the other commonly used two types of feature values were adopted to form two comparison imaging methods, called the FFT energy and Clough-Tocher interpolation-based imaging (FECT) and the Spectral power and Clough-Tocher interpolation-based imaging (SPCT). In FECT and SPCT, the length of FFT and the spectral power were utilized as the representative features of MI-EEG, respectively, and the frequency band from 8∼30 Hz was also divided into three sub-frequency bands, i.e., 8∼13 Hz, 13∼21 Hz and 21∼30 Hz, and three feature values were obtained for one electrode of MI-EEG. The interpolation processing in the two imaging methods were followed the Section 3.1, which was same as the proposed TPCT. Three technical indicators, namely accuracy, the Kappa value and AUC, were adopted. All experiments in this section were based on mVGG.

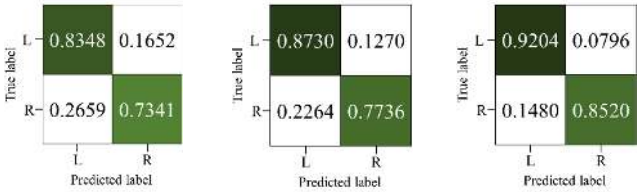
a: ACCURACY AND KAPPA VALUE

The experiment results in this study were based on 10-fold cross-validation, and the average accuracies and kappa values of FECT, SPCT and TPCT were shown in Fig. 9.

From Fig. 9 we can see that the accuracy of FECT is 78.45%, and the Kappa value was 0.57. When the spectral power is used as the feature value, the accuracy the Kappa value of SPCT increased to 82.33% and 0. 65, respectively. And in the proposed TPCT method, the accuracy reaches 88.62%, demonstrating the best performance among the three methods. At the same time, the Kappa value achieves 0.77, showing almost perfect consistency. The overall experiment indicates that the time-domain power of interest frequency bands is a superiority representative feature of MI-EEG, and the proposed TPCT achieves the obvious improvements of accuracy 6.29% and 10.17%, respectively, compared to the SPCT and FECT imaging methods.

b: CONFUSION MATRIX

Fig. 10 shows the confusion matrix of the three methods.



(a) FECT (b) SPCT (c) TPCT

FIGURE 10. Confusion matrices of the different imaging methods.

It's is clear that the difference between the two motor imagery tasks in FECT exceeded 10%. The difference is still large when using spectral power as the feature value, which also shows a 10% difference. However, the resulted difference by TPCT decreases to 6.84%. In addition, the accuracy of every category increases among FECT, SPCT and TPCT imaging methods. Therefore, TPCT achieves the best result and shows the most stable performance in consistency, demonstrating the great advantage of time domain power as feature value in TPCT.

c: ROC

Furthermore, Fig. 11 (a) shows the ROC curves of three imaging methods.

It can be seen that TPCT, SPCT and FECT from the top down in terms of ROCs indicates that time domain power of the interest frequency band is a superior characteristic to the MI-EEG signal. Although the spectral feature-like FFT energy and spectral power can reflect the variation trend of the spectral component combination of a signal, it neglects the time dynamics. Ultimately, the basic response of MI-EEG signal is power in the time domain. In TPCT, three-channel time domain power as time-frequency features can reflect the dynamic characteristics of a signal in time and frequency domains, and combined with the interpolation method,

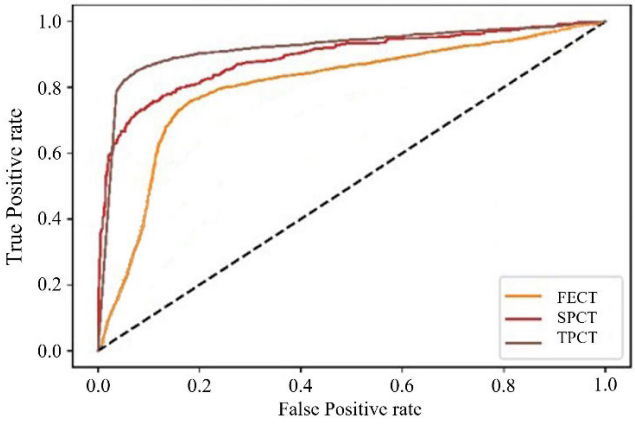


FIGURE 11. ROCs of the different imaging methods.

they can reveal the activated space information of motor imagery as well, endowing mVGG with more information for identification. In the following experiments, TPCT is used as a feature extraction and imaging method.

4) COMPARISON OF DIFFERENT DIVISION OF FREQUENCY BAND

The division of the frequency band of interest had several schemes in the previous studies, the frequency band of 8-30 Hz was mainly preserved as a single channel or was considered as several independent sub-bands. In this section, there were four segmentation approaches about the frequency band 8-30 Hz. One case was that the frequency band of interest was preserved, this means only single time domain power was calculated for each electrode of MI-EEG signal, and the imaging method was named as TPCTS; The second case was that the frequency band of interest were divided to μ rhythm 8-13Hz and β rhythm 13-30Hz, namely, two time domain powers would be computed as the features of an electrode, and the imaging method was called as TPCTT; The third case was that the frequency band 8-30Hz was divided to μ rhythm 8-13Hz, lower β rhythm 13-21Hz and higher β rhythm 21-30Hz, the imaging method was just TPCT; The last case was that the μ rhythm 8-13Hz remained intact, and the β rhythm was divided to lower β rhythm 13-19Hz, middle β rhythm 19-25Hz and higher β rhythm 25-30Hz, four time domain powers were used as the reprehensive features of an electrode, thus the imaging method was denoted as TPCTF. For the four cases, the average recognition accuracies and Kappa values of 10-fold CV. were shown in Fig. 12.

Fig.12 shows that in TPCTS, it is insufficient to express the information in 8-30Hz by using only one feature, 67.82% of accuracy and 0.36 of Kappa value are obtained. So, the performance is relatively poor, this perhaps is because of the information super-compression in feature computation. In the TPCTT, the 8-30Hz is divided into two sub-bands that caused the augment of information, the accuracy and Kappa value are increased to 86.54% and 0.73, respectively. Due to the wide range of β rhythm, when it is further divided into two sub-bands, called TPCT method, which can best represent the

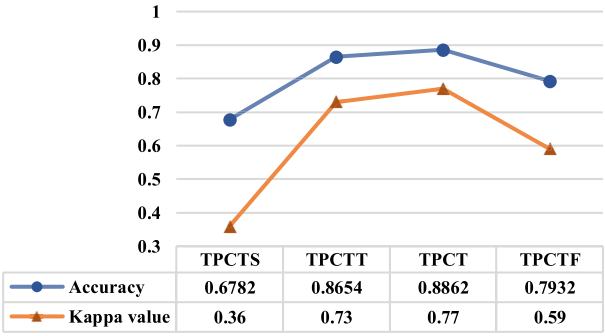


FIGURE 12. Impact of the division of the frequency band on the recognition accuracies and Kappa values.

characteristics of MI-EEG signal and its performance indexes reaches the highest level. The accuracy is 88.62% and the Kappa value is 0.77. When the β rhythm is further divided into three sub-bands, namely, TPCTF, there may be redundancy of information, performance indexes are declining, 79.32% of accuracy and 0.59 of Kappa value are obtained. Besides, the MI-EEG image's performance is very sensitive to the density of electrodes' coordinates distribution, it can be speculated that there is an optimal matching relationship between the density of coordinate points and the spatial characteristic distribution of MI-EEG signal. When the frequency band 8-30Hz is divided to three sub-bands in TPCT, both spatial characteristics of the signal and the information contained in the frequency band may be properly described.

5) INFLUENCE OF ELECTRODE POSITION INFORMATION ON THE CLASSIFICATION PERFORMANCE

a: ACCURACY AND KAPPA VALUE

The TPCT with proper electrodes position was compared with the TPCT with out-of-order electrodes (named as TPCTO) to demonstrate the impact of electrodes position information on the recognition performance. The TPCTO undergone the same imaging process as TPCT but with a scrambled electrode number corresponding to the coordinate. Meanwhile, VGG and mVGG were also added into experiments. Thus, four situations were considered, and their results were shown in Fig. 13.

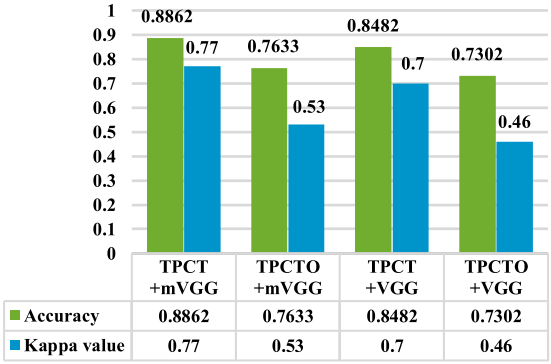


FIGURE 13. Impact of the electrode position on the recognition accuracies and Kappa values.

It is clear that for the proposed "TPCT+mVGG" method, the classification accuracy is 88.62%, and the Kappa value is 0.77, indicating that substantial consistency is achieved in "TPCT+mVGG". And when the order of the electrodes is scrambled, the accuracy of "TPCTO+mVGG" has a decline of 12.29%, and the Kappa value drops to 0.53, reflecting that the correct electrode coordinate information significantly benefits the classification accuracy and could also improve the data consistency. Moreover, the results of VGG show the same trend. The accuracy of "TPCT+VGG" is 84.82%, which is higher than the accuracy of "TPCTO+VGG", 73.02%. This declination is also shown in the Kappa value, moving down from 0.7 to 0.46. Therefore, TPCT has superiority compared to TPCTO by using VGG and mVGG, this indicates the effectiveness of the electrodes' position in the imaging.

In addition, the mVGG based results has a significant improvement compared to that of VGG, advance of 3.8% is yielded by TPCT and 3.31% by TPCTO. Perhaps mVGG can better recognize the spatial characteristics of MI-EEG, and the position of electrodes has much contribution to the performance of imaging, this would be why the experiment of "mVGG+TPCT" yields the best result.

b: CONFUSION MATRIX

Figures 14 (a) - (d) show the confusion matrix for the four combinations, which are successively "TPCT+mVGG", "TPCTO+mVGG", "TPCT+VGG" and "TPCTO+VGG". The "BF" represents the MI-task of "imaging both fists", and the "BFE" represents the MI-task of "imaging both feet".

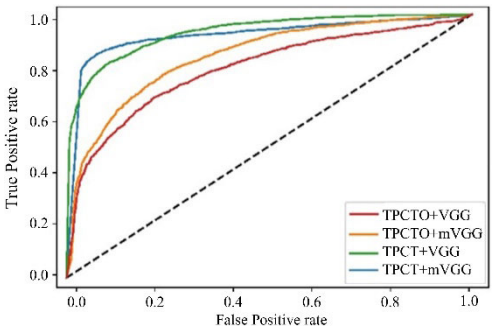


FIGURE 14. ROC of four combinations: TPCT + mVGG, TPCTO + mVGG, TPCT + VGG and TPCTO + VGG.

Fig. 14 indicates that TPCT yields significantly higher recognition accuracies for BF and BFE than TPCTO by using both VGG and mVGG, and the result of TPCTO+mVGG has better consistency compared to TPCTO+VGG. This means that the introduction of electrode position not only can increase the overall performance of classification, but also can simultaneously improve the accuracy of each MI-task. Furthermore, from figures 14 (a) and (c), we can see that when TPCT is applied, mVGG only has a slight improvement in contrast with VGG for feet motor imagery task. However,

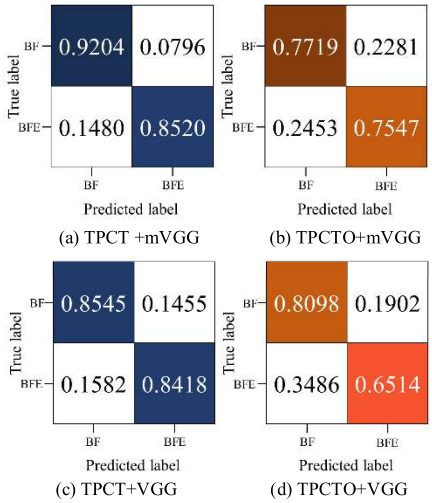


FIGURE 15. Confusion matrix of the four approaches.

for fists motor imagery task, the accuracy is increased from 85.45% to 92.04%. This may be because that BF can generate the more intensive activation than BFE, and mVGG is helpful to explore and utilize the spatial information of MI-EEG images produced by TPCT. So, the combination of TPCT and mVGG shows the potential and advantage, particularly in recognizing BF task.

c: ROC

To fully evaluate the real performance of the trained CNN model, we also provide the ROC comparison of the four combinations, as shown in Fig. 15.

Whether mVGG or VGG, the overall AUC of TPCT is better than that of TPCTO, which indicates that regardless of the classification threshold selection, the MI-EEG imaging method with correct electrode position always has superiority in contrast to that with scrambled electrode location. It is mainly because the electrode position is very important in MI-EEG feature representation. Moreover, the AUC of mVGG is greater than VGG for TPCTO, and the ROC of mVGG is closer to the top-left corner than that of VGG for TPCT. This illustrates that mVGG could achieve better results when the threshold is approximately 0.5, which is practically used in the classification. The trained model of TPCT+mVGG gets the best balance between sensitivity and specificity, yielding the advantage of its classification performance.

6) COMPARISON OF DIFFERENT MODIFIED STRUCTURES

To further discuss the modifications of VGG, the comparison experiments were conducted in four aspects, including the modification of max pooling layer, kernel size, fully connected layer and model's depth. Single variable analysis was the main qualitative analysis method used in the following experiments. Imaging data was based on the TPCT of BCI2000 data set.

a: MARGINAL DATA

Table 2 shows modified items, accuracy and Kappa value. The "VGG" represents the VGG-16 layers model, which proposed by [26] and only changed the length of the softmax layer from 1024 to 2 to accommodate the dichotomy task. The "M" model represents replacing the Max-pooling layer with the convolutional layer to down sampling. The "C" model refers to adopt the 2×2 convolutional kernel in each layer.

The "F" model refers to remove the Fully Connected layer. The "D" model refers to different depth of model, for example, "D-7" refers to the 7-layer convolution. The "mVGG" refers to our proposed network which is integrated all the adjustments. The configurations of each modified network were shown as Table 2.

TABLE 2. Configurations of different modified structures.

Section	S	P	C	F	D7	D11
1th:	(conv3-64)x2	(conv3-64)x2	(conv2-64)x2	(conv3-64)x2	conv3-16	conv3-64
DS:	Max-pool	conv2-128 (stride=2)	Max-pool	Max-pool	Max-pool	Max-pool
2nd:	(conv3-128)x2	(conv3-128)x2	(conv2-128)x2	(conv3-128)x2	conv3-32	conv3-128
DS:	Max-pool	conv2-128 (stride=2)	Max-pool	Max-pool	Max-pool	Max-pool
3rd:	(conv3-256)x4	(conv3-256)x4	(conv2-256)x4	(conv3-256)x4	conv3-64	(conv3-256)x2
DS:	Max-pool	conv2-128 (stride=2)	Max-pool	Max-pool	Max-pool	Max-pool
4th:	(conv3-512)x4	(conv3-512)x4	(conv2-512)x4	(conv3-512)x4	conv3-128	(conv3-512)x2
DS:	Max-pool	conv2-128 (stride=2)	Max-pool	Max-pool	Max-pool	Max-pool
5th:	(conv3-512)x4	(conv3-512)x4	(conv2-512)x4	(conv3-512)x4		(conv3-512)x2
DS:	Max-pool	conv2-128 (stride=2)	Max-pool	Max-pool		Max-pool
FC	FC:256	FC:256	FC:256	conv2-2	FC:256	FC:256
FC	FC:256	FC:256	FC:256	Average-pool	FC:256	FC:256
FC	FC:2	FC:2	FC:2	FC:2	FC:2	FC:2
			Softmax layer			

b: BASELINE

As can be seen from Table 3, the ''VGG'' yields the accuracy of 84.71% and Kappa value of 0.72 that is used as the baseline in this experiment.

TABLE 3. The accuracy and Kappa value of different modified VGG.

Modified models	Accuracy (%)±std	Kappa value
VGG	84.71±2.98	0.70
M	86.92±1.53	0.74
C	87.55±1.41	0.75
F	85.38±1.62	0.71
D-11	74.84±2.78	0.50
D-7	72.02±1.58	0.44
mVGG	88.62±1.65	0.77

c: RESULTS OF ''M'' MODEL

The ''M'' model gets 2.21% improvement of accuracy and 0.04 of Kappa value compare to the VGG. This result verifies that the information discarding by the Max-pooling operation does cause the loss of accuracy. After using convolution operation as substitute, all the information is taken into account, so that the advantage of multi-electrode interpolation imaging in the expression of position information can be maximized.

d: RESULTS OF ''C'' MODEL

The ''C'' model has changed the convolution kernel's size from 3×3 to the 2×2 that has improved accuracy from 84.71% to 87.55%, and the Kappa value is also improved from 0.70 to 0.75. Reducing the size of the convolution kernel would greatly increase the nonlinear combination of features and reduce the computational complexity. In the VGG, the number of parameters in the first layer is $(3\times3\times3+1)\times64 = 1792$, and after 2×2 convolution kernel is used, the number of parameters is reduced to $(2 \times 2 \times 3 + 1) \times 64 = 832$.

The number of parameters was reduced to 46% and increasing the nonlinear combination of position information between electrodes. The increasing number of combinations of different electrodes can be considered as basis vectors that may have great help to improve the model's performance.

e: RESULTS OF ''F'' MODEL

The adopting of convolutional layer instead of the fully connected layer has become an important branch of the development of CNN in recent years [31] that could greatly reduce parameters of model so as overfitting problem.

The Fully Connected layer contains most parameters of CNN. Take the VGG model as an example. The final pooling layer has $(2 \times 2 + 1) \times 512 \times 256 = 655360$ parameters. Where the 512 is the number of neurons in the pooling layer, and 256 is the number of neurons in the fully connected layer. In the ''F'' model, two 2×2 convolution kernels are used to replace the Fully Connected layer, and $(2 \times 2 + 1) \times 512 \times (2 \times 2 + 1) \times 2 = 25600$ parameters would be calculated. Thus, the number of parameters is reduced by 96% only at the connection between the FC layer and Max-pooling layer.

The advantage of using convolution to replace the entire FC layer lies in the absence of flatten operation, which would not affect the spatial structure of the feature map. Although in the training process, the lack of FC layer would cause slight instability of gradient descent, the reduction of model parameters still makes the training more efficient and improves the final classification accuracy from 84.71% to the 85.38%.

f: RESULTS OF ''D'' MODEL

The enhancement of convolutional layers can improve the model's ability and finally show up in the model's performance. This model aims to verify the depth of our network is necessary, when the shallow CNNs are adopted, the performance shows significant declines. The modified ''D'' models consist of ''D-7'', ''D-11'', and the original 3×3 convolution kernel, the Max-pooling layer and Fully Connected layer are retained. The ''D-7'' is designed because its scale is similar to the mainstream method for classifying MI-EEG signals, and the accuracy of 72.02% and Kappa value of 0.44 are obtained in our experiment. The ''D-11'' model is also a typical VGG [26], it yields accuracy of 74.84% and the Kappa values of 0.70. As a baseline, 84.71% is acquired on VGG.

When the deeper CNN is applied, the better descriptive power and superior accuracy are obtained. The result illustrates the multi-electrode interpolation-imaging image contains a large number of comprehensive information of spatial domain, and the shallow CNN lacks in the abstraction and combination ability of both specific and abstract features. Therefore, when the depth is deepened, the CNN shows higher classification accuracy.

B. RESULTS BASED ON BCI COMPETITION IV 2a AND 2b DATASETS

In this section, our method is verified in two public datasets, i.e., BCI Competition IV 2a and 2b datasets, to compare with state-of-the-art methods.

1) RESULTS ON BCI COMPETITION IV 2a DATASET

a: BCI COMPETITION 2a DATASET

The BCI Competition IV dataset 2a includes one session for the training set and one session for test set. Each session has 288 trails for one subject, the overall dataset contained 5,184 trials. In Experimental paradigm, the cue-based BCI paradigm consisted of four different motor imagery tasks, namely the imagination of left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4) movements. And the dataset includes 9 subjects. The Motor-Imagery length is set to three seconds, and the sampling frequency is 250Hz. A trial consists of 750 sampling points. The feature extraction process and the imaging method are followed Section 3.1, and the coordinate information of the interpolation imaging was extracted according to official 10-20 electrode system.

b: RESULTS

To further verify the performance of TPCT+mVGG, it was compared with other related methods based on BCI competition IV 2a dataset. To guarantee the comparability, the original training set and test set in the dataset need to be re-divided in the following experiments, and two schemes were considered. The first scheme, noted as S1, is that the training set and test set are merged and evaluated by 10-fold CV. The second scheme, noted as S2, is that the original training set and test set are applied to train and test the model, respectively. The experiment results were shown in Table 4.

TABLE 4. Comparison with relative study on 2a dataset.

Imaging Method	Data partition	Accuracy (%)	Kappa value
FBCSP[32]	BCI Competition	67.75	0.57
TPCT^{S2}	BCI Competition	88.87	0.78
ACSP[12]	5×5-fold CV.	69.27	0.59
WT[19]	5-fold CV.	85.59	0.71
FBCSP[13]	10-fold CV.	74.46	0.66
FFTEM[15]	10-fold CV.	68.00	0.57
TPCT^{S1}	10-fold CV.	92.13	0.90

In Table 4, the winner method of the BCI Competition IV [32] utilized Filter-Band CSP(FBCSP) matrix as the feature, and its data partition was followed Competition's request, namely S2, the accuracy and Kappa value are 67.75% and 0.57, respectively. In the same data partition, the proposed TPCT yields 88.87% of accuracy and 0.78 of Kappa value, which is greatly increased comparing to FBCSP [32]. The work in [12] adopted the Augmented-CSP(ACSP) method and combined with a three-layer CNN to do the extraction and classification, the accuracy achieves 69.27% based on 5 × 5-fold CV. The state-of-the-art method [19], in which WT was employed to generate a time-frequency image and the data of each electrode was taken as an independent input channel, yields 85.59% of accuracy and 0.71 of Kappa value by 5-fold CV, this shows the superiority of time-frequency features. In [13], a new representation of FBCSP was combined with CNN, this results in the relatively higher accuracy 74.46% and Kappa value 0.66 by 10-fold CV in contrast to the original FBCSP [32] and ACSP [12]. In [15], 2D-CNN is used to classify the FFTEM image, and the accuracy of 68% is obtained based on 10-fold CV, which verifies the availability of CNN combining spatio-frequency imaging. However, the recognition is poor. Finally, the experiment results are significantly improved when the proposed TPCT is combined with mVGG, and 92.13% of accuracy and 0.90 of Kappa value are obtained. This is mainly due to the introduction of real location of electrodes in TPCT and the modification of VGG as well.

It is worth to note that the data of all subjects are merged in recognition of TPCT based images, the overall results are obtained by training one mVGG model. In the other methods of Table 4, each subject needs a CNN model trained by using his or her MI-EEG data and has conducted a CV, and the results of all subjects are averaged as the final result. Besides, our method can avoid personalized hyperparameters, thus enhancing the adaptability of classification model among the subjects as well as the stringency of our results.

Our method also has a good adaptability to the classification of each category. The true positive rate, false positive rate, true negative rate and false negative rate of the specific category are shown in Fig. 16.

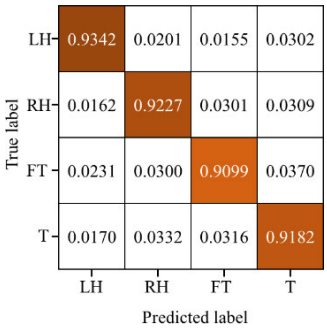


FIGURE 16. Confusion matrix of 2a data set.

Fig. 16 shows the confusion matrix of 2a dataset in the scheme S1. The accuracies of four categories, which include "LH", "RH", "FT" and "T", are 93.42%, 92.27%, 90.99% and 91.82%. Here, "LH", "RH", "FT", "T" in Fig. 16 represent four MI tasks: imaging left hand, imaging right hand, imaging both feet and imaging tongue. The average accuracy and Kappa value of four categories can be calculated, and they are 92.13% and 0.90, respectively. It means our method achieves almost perfect in consistency. Besides, the accuracies of "LH" and "RH" are higher than other tasks which are 93.42% and 92.27%. It may demonstrate that the ERS/ERD's contribution in separability of spatial is still effective. The accuracy of task "T" is the third high in the results, and the task of "Feet" is hard to distinguish relatively. It can be seen that the accuracies of four MI tasks are all higher than 90.00%, and they are very close, which shows the excellent consistency of the proposed method, and the competitive result is obtained by this benefit. This reveals that TPCT is superior in feature fusion expression of time-frequency-spatial-domain. In addition, our modified mVGG is more conducive to matching the multi-domain fusion imaging method with qualified feature extraction ability, and finally the whole framework achieves excellent performance. We suggest that TPCT is an excellent way to show the spatial information sufficiently as well as the guaranteed time-frequency features.

2) RESULTS ON BCI COMPETITION IV 2b DATASET

a: BCI COMPETITION 2b DATASET DESCRIPTION

The BCI Competition IV 2b dataset consisted of two classes, namely the motor imagery (MI) of left hand (class 1) and right hand (class 2). There were three sessions in the training set, and two sessions in the test set. 9 subjects were included, and the overall dataset contained 6300 trials. Each trial started

with a fixation cross and an additional short acoustic warning tone (1kHz, 70ms). Some seconds later a visual cue (an arrow pointing either to the left or right, according to the requested class) was presented for 1.25 seconds. Afterwards the subjects had to imagine the corresponding hand movement over a period of 4 seconds. The data of 4s to 7s was used as the MI data.

b: RESULTS

In order to compare with the related studies based on same dataset, we conduct our experiments in two situations: the first scheme S3 is that the training set and test set mentioned above are merged and evaluated with 10-fold CV; the second scheme S4 is that only the training set is utilized for the 10-fold CV. The experiment results are shown in the Table 5.

TABLE 5. Comparison with relative study on 2b dataset.

Imaging Method	Data partition	Accuracy (%)	Kappa value
CSP[14]	9-subject CV.	73.00	0.46
STFT[21]	5-fold CV.	92.28	0.85
TPCT^{S3}	10-fold CV.	96.82	0.94
STFT[17]	10×10-fold CV. on training set	75.10	0.50
STFT[18]	10×10-fold CV. on training set	78.20	0.56
TPCT^{S4}	10×10-fold CV. on training set	96.48	0.93

It can be seen, in [14], by encoding the CSP matrix of multi-channel data independently via a separated channel convolution, this method acquires accuracy of 73.00% based on 9-subject CV. In [21], the STFT-based time-frequency imaging is combined with a dCNN model, ample time-frequency features are exploited, yielding the-state-of-the-art result of 92.28% based on 5-fold CV. The proposed method (TPCT) yields accuracy of 96.82% and Kappa value of 0.94 in scheme S3, which has an improvement of 4.54% compared to [21]. And due to the successful application of 1D-convolution layer, [17] obtains the accuracy of 75.10% based on the STFT imaging and CNN combining SAE method, in which only the training set of 2b dataset was used and 10 × 10-fold CV was adopted. The structural innovation in [18] obtains 3.1% improvement comparing to [17] by replacing the SAE with VAE based on the same data partition. In scheme S4, namely, only training set is considered, the TPCT yields accuracy of 96.48% by 10 × 10-fold CV and has an increment of 18.28% in contrast to [18]. The excellent robustness of TPCT is self-evident, so no obvious regression is generated when the training set is reduced. The results may reflect the great coordination between TPCT and mVGG, and imply the effectiveness in application of the tri-electrode system like BCI Competition IV 2b dataset.

Fig. 17 shows the confusion matrix in scheme S3. The "L" represents the MI task of imaging left hand and the "R" represents that of imaging right hand. The accuracies of "L" and "R" are 95.31% and 98.32%, respectively.

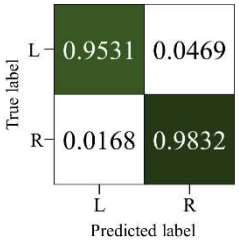


FIGURE 17. Confusion matrix of 2b data set.

The average accuracy of the two categories is 96.82%, and the Kappa value is 0.94, which achieves almost perfect results.

From the above experimental results and analysis, we know that whether in the BCI Competition IV 2a or 2b data set, "TPCT + mVGG" demonstrates the excellent performance for two and four categories MI-tasks, and it is insensitive to the number of electrodes, which shows the generality and applicability of our presented method.

V. DISCUSSION

This study aimed to improve the classification accuracy of MI-EEG signals by introducing the relative position of electrodes to imaging process and promote the optimization of DCNN in MI-EEG research. The public data set of imaging both fists and feet recorded by 64-channel system was utilized to discuss the proposed imaging method (TPCT) and the modified VGG (mVGG). The TPCT method can sufficiently utilize the MI-EEG's characteristics by promoting the fusion of temporal, frequency and spatial features. The representation of feature value and the division of frequency bands are essential in the imaging process of MI-EEG signal. Fig. 9-11 show the comparison of different feature values, the accuracy of FECT, SPCT and TPCT are 78.45%, 82.33% and 88.62%, respectively. This means the time domain power on interest frequency band (8-30Hz) is prominent than the frequency domain length or power. Further, the interest frequency band (8-30Hz) is preserved as an intact channel or divided into two to four channels (namely, sub-bands), the trend of performance is exhibited in Fig. 12. The best result (88.62%) is obtained in the three-channel scheme and begins to decline in the four-channel scheme. Presumably, there is a proper matching degree among the spatial characteristic of MI-EEG, the density of interpolation-based MI-EEG image and the DCNN. When the number of channels is too small, ample information is compressed and cannot do enough contribute to the result; when the number of channels is too large, the density of interpolation-based image may break the tolerance, which increases the interference between two adjacent electrodes, causing the recognition accuracy going down. The experimental results demonstrate that it is the most proper scheme when the interest frequency band 8∼30Hz is divided to three channels, and it is applied in the following experiments.

Duo to the obvious spatial activation for MI tasks, the MI-EEG signal has the characteristic of spatial distribution.

Therefore, the location of electrodes has much contribution to the decoding of MI tasks. By shuffling the relation of the electrodes and their serial numbers, the location information of electrodes is lost, and the resulted accuracy declines about 12% in Figures 14 and 15. It indicates the imaging with the location of electrodes is matching with the physical basis of MI-EEG signal and is helpful to display the spatial and dynamic information.

To matching the TPCT imaging, VGG, a typical DCNN, is considered to be modified in four aspects, Table 3 shows the performance of different modified models. It implies that when replacing Max pooling with the convolution layer, modifying the size of the convolution kernel, and using average pooling instead of the fully connected layer, the accuracies are 86.92%, 87.55% and 85.38%, respectively, which are higher than the original VGG's result (84.71%); Besides, when the network layers decreases to "D-7" and "D-11", the accuracies significantly drop to 72.02% and 74.84%, respectively. The modified VGG (called mVGG) achieves the highest accuracy of 88.6% and the best Kappa value of 0.77 as well. This indicates the necessaries of the modification for VGG. In addition, from the ROC in Fig. 14, we know that mVGG also ensures the best performance when the threshold is 0.5.

To fully verify our method, extensive experiments are conducted on BCI Competition IV 2a and 2b data set, and the accuracies of 10-fold CV are 92.13% and 96.82%, respectively. This means TPCT imaging integrated with mVGG has superiority and competitiveness in the generalization and adaptive ability to different data sets, no matter how many electrodes are recorded. (see Tables 4 and 5). To compare with the related methods objectively, two data partition schemes are adopted in 2a or 2b data set. Comparative results are obtained, showing the robustness of our method in both two-category and four-category classification.

Because the interpolation-based imaging method with position of electrodes is rarely applied to decode MI-EEG in current research, we should not ignore the outstanding advantages of the utilization of location information. We will further explore the improvement of the combination of the MI-EEG and DCNN and enhance the utilization of time information based current imaging method. In addition, the modification of VGG is also a notable aspect in this paper. It fully considers the physical basis of the MI-EEG signal and provides a foundation and guarantee for the classification of more multi-electrode MI-tasks. After the adjustment of VGG's convolution kernel, down-sampling mode and fully connection layer, mVGG achieves great improvement in rich experiments and achieves the best results on three public data sets. With widespread application of deep learning in MI-EEG, our TPCT imaging method establishes a solid basis for combining time, frequency and position features of MI-EEG signal into images and will promote the integration of MI-EEG signals with more convolutional models.

VI. CONCLUSION

A novel imaging method, called as TPCT, is proposed to generate MI-EEG images for decoding MI tasks. Through the interpolation algorithm, TPCT can fix the relevant position of electrodes in MI-EEG image as well as assign the time-frequency feature values to pixels, which may promote the natural feature fusion in time-frequency-space domains. To fully explore the features with deep dimensionality produced by TPCT, the original VGG is modified, which mainly focuses on the unify of VGG's structure and MI-EEG's physical characteristics. The extensive experiment results on three public data sets suggest that TPCT combined with mVGG achieves excellent performance in MI-EEG recognition, and it has a prominent advantage in accuracy, Kappa value, confusion matrix and ROC in contrast to the existing related methods. In the future work, we will devote developing more interpolation-based imaging methods, and hope they can be applied in areas such as disease diagnosis, rehabilitation engineering and brain function research etc.

ACKNOWLEDGMENT

The authors would like to thank the provider of dataset and all of the people who have given us helpful suggestions and advice. The authors are obliged to the anonymous referee for carefully looking over the details and for useful comments that improved this paper.

REFERENCES

[1] X. Zhang, L. Yao, Q. Z. Sheng, S. S. Kanhere, T. Gu, and D. Zhang, "Converting your thoughts to texts: Enabling brain typing via deep feature learning of EEG signals," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Athens, Greece, Mar. 2018, pp. 1–10, doi: 10.1109/percom.2018.8444575.

[2] H.-C. Chou, N. Prataksita, Y.-T. Lin, and C.-H. Kuo, "P300 and motor imagery based brain–computer interface for controlling wheelchairs," *J. Med. Devices*, vol. 8, no. 3, Sep. 2014, Art. no. 030906, doi: 10.1115/1.4027100.

[3] S. Bhattacharyya, A. Konar, and D. N. Tibarewala, "Motor imagery, P300 and error-related EEG-based robot arm movement control for rehabilitation purpose," *Med. Biol. Eng. Comput.*, vol. 52, no. 12, pp. 1007–1017, Dec. 2014, doi: 10.1007/s11517-014-1204-4.

[4] A. S. Royer, A. J. Doud, M. L. Rose, and B. He, "EEG control of a virtual helicopter in 3-dimensional space using intelligent control strategies," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 6, pp. 581–589, Dec. 2010.

[5] R. Van Der Lubbe and C. Utzerath, "Lateralized power spectra of the EEG as an index of visuospatial attention," *Adv. Cogn. Psychol.*, vol. 9, no. 4, pp. 184–201, Dec. 2013.

[6] B. R. Cahn and J. Polich, "Meditation states and traits: EEG, ERP, and neuroimaging studies.," *Psychol. Bull.*, vol. 132, no. 2, pp. 180–211, Mar. 2006.

[7] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Proc. NIPS*, 2016.

[8] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012.

[9] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI–EEG signal classification for BCIs," *Expert Syst. Appl.*, vol. 114, pp. 532–542, Dec. 2018.

[10] 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, doi: 10.1109/access.2019.2895688.

[11] S. Sakhavi, C. Guan, and S. Yan, "Parallel convolutional-linear neural network for motor imagery classification," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 2736–2740.
[12] 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)*, Milan, Italy, Aug. 2015, pp. 2620–2623.
[13] 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.
[14] 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.
[15] T. Uktveris and V. Jusas, "Application of convolutional neural networks to four-class motor imagery classification problem," *Inf. Technol. Control*, vol. 46, no. 2, pp. 260–273, 2017, doi: 10.5755/j01.itc.46.2.17528.
[16] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition," *Concurrency Comput., Pract. Exper.*, vol. 30, no. 23, Dec. 2018, Art. no. e4413.
[17] 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, Feb. 2017, Art. no. 016003.
[18] 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, Jan. 2019.
[19] B. Xu, L. Zhang, A. Song, C. Wu, W. Li, D. Zhang, G. Xu, H. Li, and H. Zeng, "Wavelet transform time–frequency image and convolutional network–based motor imagery EEG classification," *IEEE Access*, vol. 7, pp. 6084–6093, 2019.
[20] Y. Zhang, X. Zhang, H. Sun, Z. Fan, and X. Zhong, "Portable brain-computer interface based on novel convolutional neural network," *Comput. Biol. Med.*, vol. 107, pp. 248–256, Apr. 2019.
[21] Z. Tayeb, "Validating deep neural networks for online decoding of motor imagery movements from EEG signals," *Sensors*, vol. 19, no. 1, p. 210, 2009.
[22] S. Chaudhary, S. Taran, V. Bajaj, and A. Sengur, "Convolutional neural network based approach towards motor imagery tasks EEG signals classification," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4494–4500, Jun. 2019.
[23] J. Li, Z. Struzik, L. Zhang, and A. Cichocki, "Feature learning from incomplete EEG with denoising autoencoder," *Neurocomputing*, vol. 165, pp. 23–31, Oct. 2015.
[24] P. Alfeld, "A trivariate clough—Tocher scheme for tetrahedral data," *Comput. Aided Geometric Des.*, vol. 1, no. 2, pp. 169–181, Nov. 1984.
[25] B. R. Cahn and J. Polich, "Meditation states and traits: EEG, ERP, and neuroimaging studies," *Psychol. Bull.*, vol. 132, no. 2, pp. 180–211, Mar. 2006.
[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
[27] P. Smith and C. Chen, "Transfer learning with deep CNNs for gender recognition and age estimation," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 2564–2571, doi: 10.1109/BigData.2018.8621891.
[28] Q. Zheng, M. Yang, and J. Yang, "Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process," *IEEE Access*, vol. 6, pp. 15844–15869, 2018.
[29] 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. 1, pp. 1929–1958, Jun. 2014.
[30] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
[31] X. Wang, "Architecture optimization, training convergence and network estimation robustness of a fully connected recurrent neural network," Ph.D. dissertation, Gradworks, Coleraine, U.K., 2010.
[32] 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," *Front. Neurosci.*, vol. 6, p. 39, Mar. 2012.

MING-AI LI received the B.Sc. and M.Sc. degrees from the Daqing Petroleum Institute, Heilongjiang, China, in 1987 and 1990, respectively, and the Ph.D. degree from the Beijing University of Technology, Beijing, in 2006. She is currently a Professor with the Faculty of Information Technology, Beijing University of Technology. Her current research interests mainly include brain-computer interface, intelligent control, pattern recognition, and rehabilitation robot.

JIAN-FU HAN received the B.Sc. degree from the Beijing University of Technology, Beijing, in 2017, and the M.Sc. degree in control science and engineering from the Beijing University of Technology, in 2019. His research interests include brain-computer interface, EEG imaging technology, and deep learning.

LI-JUAN DUAN received the B.Sc. and M.Sc. degrees in computer science from the Zhengzhou University of Technology, Zhengzhou, China, in 1995 and 1998, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2003. She is currently a Professor with the Faculty of Information Technology, Beijing University of Technology, China. Her research interests include artificial intelligence, image processing, machine vision, and information security. She has published more than 70 research articles in refereed journals and proceedings on artificial intelligence, image processing, and computer vision.