Mathematical Biosciences and Engineering

MBE, 20(12): 20624–20647.
DOI: 10.3934/mbe.2023912
Received: 11 October 2023
Revised: 01 November 2023
Accepted: 07 November 2023
Published: 15 November 2023

http://www.aimspress.com/journal/MBE

Research article

A multi-band centroid contrastive reconstruction fusion network for motor imagery electroencephalogram signal decoding

Jiacan Xu[1], Donglin Li[2], Peng Zhou[1,*], Chunsheng Li[2,*], Zinan Wang[1] and Shenghao Tong[1]

[1] The College of Engineering Training and Innovation, Shenyang Jianzhu University, Shenyang 110000, China
[2] The College of Electrical Engineering, Shenyang University of Technology, Shenyang 110000, China

* **Correspondence:** Email: zhoupeng@sjzu.edu.cn, lichunsheng@sut.edu.cn.

Abstract: Motor imagery (MI) brain-computer interface (BCI) assist users in establishing direct communication between their brain and external devices by decoding the movement intention of human electroencephalogram (EEG) signals. However, cerebral cortical potentials are highly rhythmic and sub-band features, different experimental situations and subjects have different categories of semantic information in specific sample target spaces. Feature fusion can lead to more discriminative features, but simple fusion of features from different embedding spaces leading to the model global loss is not easily convergent and ignores the complementarity of features. Considering the similarity and category contribution of different sub-band features, we propose a multi-band centroid contrastive reconstruction fusion network (MB-CCRF). We obtain multi-band spatio-temporal features by frequency division, preserving the task-related rhythmic features of different EEG signals; use a multi-stream cross-layer connected convolutional network to perform a deep feature representation for each sub-band separately; propose a centroid contrastive reconstruction fusion module, which maps different sub-band and category features into the same shared embedding space by comparing with category prototypes, reconstructing the feature semantic structure to ensure that the global loss of the fused features converges more easily. Finally, we use a learning mechanism to model the similarity between channel features and use it as the weight of fused sub-band features, thus enhancing the more discriminative features, suppressing the useless features. The experimental accuracy is 79.96% in the BCI competition IV-IIa dataset. Moreover, the classification effect of sub-band features of different subjects is verified by comparison tests, the category propensity of different sub-band features is verified by confusion matrix tests and the distribution in different classes of each sub-band feature and fused feature are showed by visual analysis, revealing the importance of different sub-band features

for the EEG-based MI classification task.

Keywords: brain computer interface (BCI); motor imagery (MI); electroencephalogram (EEG); feature reconstruction; feature fusion

1. Introduction

The brain-computer interface (BCI) system is a new approach to human-computer interaction that works by decoding brain activity patterns that generate user-specific imagery tasks [1] and translating these patterns into computer commands that control external devices [2]. Motor imagery electroencephalogram (MI-EEG) is the electrical activity signal in motor-related neurons due to the triggering of intercellular communication between associated neurons of the brain when users imagine that a part of their body is moving. It has the advantages of being non-invasive and having high temporal resolution [3]. The decoding algorithm has become a popular research topic in the field of brain-computer interfaces to extract brain activity information to achieve control of external devices in MI-EEG. Therefore, decoding EEG signals effectively is crucial in the practical application of MI-BCI systems.

The decoding of MI-EEG signals remains challenging due to the non-smooth dynamic and complex nature [4]. Considering the changes in relevant the temporal and frequency domain, Pham et al. [5] combined the wavelet-time and wavelet-image scattering features of brain signals for learning feature characterization of EEG signals. Hsu et al. [6] proposed wavelet-based temporal-spectral-attention correlation coefficient (WTS-CC), to consider the features and their weighting in spatial, EEG-channel, temporal and spectral domains. Liu et al. [7] proposed a SincNet-based hybrid neural network (SHNN) that splits the data into different time windows and maps them to the common spatial pattern (CSP) feature space for feature representation. Yin et al. [8] propose an optimal channel-based sparse time-frequency block common spatial pattern (OCSB-CSP) method, which uses the one-dimensional EEG data of the optimal channel to calculate the discriminant ability of each time-frequency block. These algorithms that fuse features from various sub-bands not only consider the frequency characteristics of EEG but also retain its temporal features, which can solve the information loss caused by only using single modal features, all of which enhanced feature expression to different degrees and were effective ways to improve the final recognition performance.

However, the electrical signals collected on the cerebral cortex are not potential changes in a single neuron but generated by the activity of many neurons, so that the EEG signal consists of superimposed signals of different frequencies and is highly rhythmic [9–11]. Considering the discriminative information contained in different sub-bands, traditional EEG signal feature extraction methods usually use band-pass filters to extract the main frequency range of the EEG signal from 8–30 Hz or select the μ and β bands of sensorimotor rhythms [12,13]. Herman et al. [14] used power spectral density (PSD) to extract power features in the 8–12 Hz and 18–25 Hz bands for the motor imagery classification task. Orset et al. [15] proposed a method for decoding the termination of a motor imagery task by extracting the frequency features in both the μ (8–13 Hz) and β (20–26 Hz) rhythms. Zhang et al. [16] developed a temporally constrained sparse group spatial pattern (TSGSP) algorithm to derive spectrum-specific signals that have multiple overlapping sub-bands from raw EEG data through bandpass filter. Lee et al. [17] selected the weighted phase lag index (wPLI) and the directed transfer function (DTF) of EEG signals in the 8–12 Hz and 13–30 Hz frequency bands, which were related with

motor imagery as functional connectivity characteristics of the brain network. Miao et al. [18] presented a common time-frequency-spatial pattern (CTFSP) algorithm for extracting and filtering features from multi-band filtered EEG data with multiple time windows. These methods map the complete time series signal to the feature space, ignoring the dynamic time characteristics of the signal that may contain valuable information.

Convolutional neural networks can capture the feature representation of EEG signals from different modalities and facilitate the exchange of information between heterogeneous features. The discriminative ability of the model can be improved by fusing the multi-source information of EEG signals, which is a direct solution to break the performance bottleneck of a single model [19–22]. To retain the dynamic time characteristics of EEG signals while to fuse the multi-source information of EEG signals and ensure the complementarity of multi-band information, Zhang et al. [23] supplemented frequency and time dimensions for EEG feature maps based on the Morlet wavelet transform and proposed the R3DCNN algorithm to learn EEG features simultaneously from three dimensions: spatial, spectral and temporal. Sakhavi et al. [24] utilized FBCSP to obtain multi-band spatial features of EEG signals and extracted spatial-temporal data representations of EEG signals through multiple one-dimensional convolutions. Brenda et al. [25] spliced the spatial features of the refined and optimal sub-bands filtered by the CSP method and extracted their convolutional neural network (CNN) features. Ma et al. [26] proposed a time-distributed attention network (TD-Atten) to improve the representativeness of the final features to the target task by increasing the attention of the time series and fusing multi-band CSP features. Zhang et al. [27] proposed a CNN, combined with a frequency-time band common spatial pattern (FTBCSP), to fuse multi-band features and obtain channel weights and enhance spatial information. However, due to the differences in experimental conditions and subjects, the EEG frequency bands of the corresponding rhythm are not the same [28–32]. Even for the same subject, these rhythm bands fluctuate according to physical health and mental state. It is possible that the discriminatory information is provided by some specific sub-bands more than others or the overall data [33,34]. The fusion task can be completed by utilizing multiple priori features, but large distribution gaps between different features can affect the effectiveness of fusion [35]. Fusing sub-band features by splicing not only fails to consider the similarities and differences between sub-bands features, but also ignores the degree of their contribution to the recognition task, resulting in an exponential increase in feature dimensionality and affecting the computational efficiency of the network model [36,37].

The cerebral cortical potentials are highly rhythmic and sub-band features, and different subjects have different categories of semantic information in specific frequency bands. To improve the decoding performance of the MI-EEG-BCI-based motor imagery recognition task, the complementarity of feature multi-source information is considered. In this paper, we propose a multi-band centroid contrastive reconstruction fusion network (MB-CCRF) for EEG signal decoding. More specifically, our contribution can be summarized as follows:

1) Considering the task-related rhythmic features in different EEG signals, a multi-stream cross-layer connected convolutional network is used to represent the deep features of each sub-band separately, integrating the features of different sub-bands and making full use of the multi-source information of EEG signals.

2) We propose the centroid contrastive reconstruction module, which maps different sub-band and category features into the same shared embedding space by centroid contrast in a single sub-band

feature domain, helping to assign a similar semantic structure to each sub-band feature in the shared embedding space and thus enhancing the feature discrimination.

3) The model establishes correlation between different features, which is weighted and fused according to the classification contribution of these features. It helps to highlight more discriminative features and suppress useless features.

The rest of the paper is organized as follows: Section 2 introduces the data preparation phase and describes the model of the multi-band centroid contrastive reconstruction fusion network (MB-CCRF), including the parameters and implementation details of the model. The experimental results are described in Section 3, including ablation experiments and the comparison of the classification results of different sub-band features and other baseline algorithms. The visualization analysis is presented in Section 4. Finally, the paper is concluded in Section 5.

2. Methods

In this section, we first describe the multi-band EEG data representation process, which focuses on features in the frequency, temporal and spatial domains. Then, we proceed the multi-band centroid contrastive reconstruction fusion network MB-CCRF, including the pre-trained feature extractor modules and centroid contrastive reconstruction fusion modules. Finally, the parameters and implementation details of the network are presented. Figure 1 illustrates the structure of MB-CCRF.

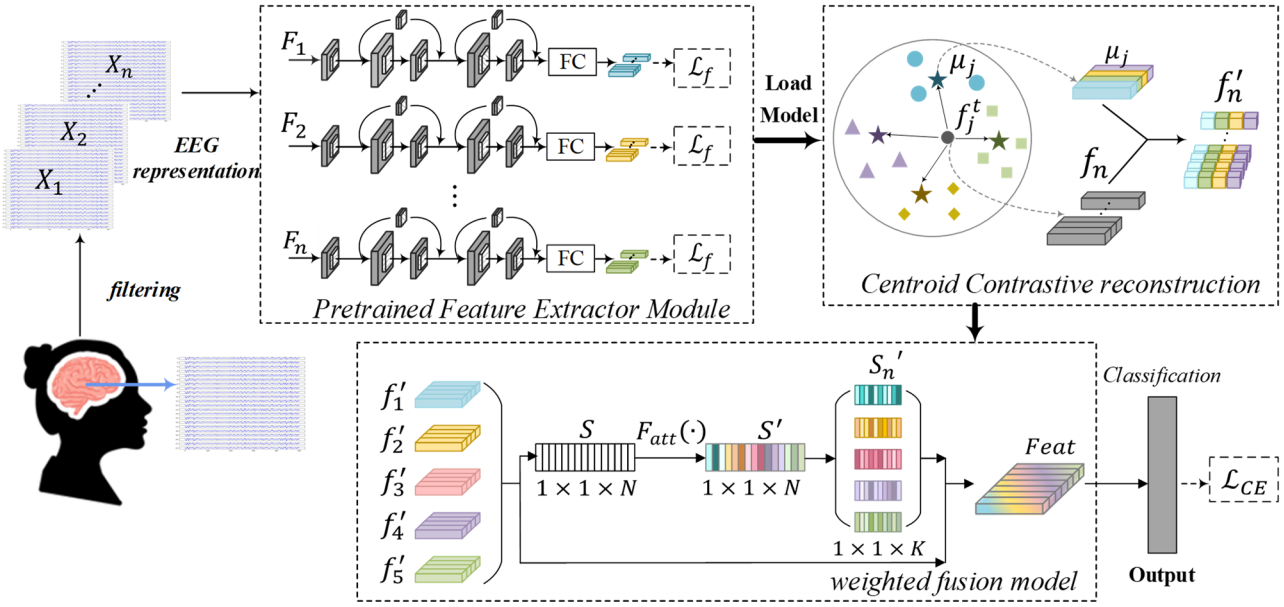

Figure 1. Graphical representation depicting the methods used for processing the imagined digits from multichannel EEG signals and classifying the multiclass EEG signals using the MB-CCRF network.

2.1. EEG representation

In EEG representation stage, we considered the information in each EEG signal frequency band and extracted data representations in the temporal and spatial domains from each frequency band signal. First, the EEG temporal domain signals are extracted for five frequency bands, 8–12 Hz, 12–16 Hz, 16–20 Hz, 20–24 Hz and 24–30 Hz. Then, the CSP spatial filtering algorithm [13,38,39] is used to spatially filter the EEG signals in each frequency band to obtain a multi-band data representation.

CSP is a data-driven filtering algorithm widely used in brain-computer interfaces, which tries to find spatial filters by linear transformation to maximize the differences of the two types of characteristics from multichannel EEG data. The one-vs-rest (OVR) strategy is used as an assistant of the CSP filtering methods to process multi-band temporal signals $X_n, n = (1,2,3,4,5)$. The optimal CSP filter can be achieved by maximizing the following objective function [24].

$$W^* = \arg\max_{w} \frac{W^T \Sigma_{c_1} W}{W^T (\Sigma_{c_1} + \Sigma_{c_2}) W}, \tag{1}$$

where Σ_{c_1}, Σ_{c_2} are the covariance matrices for classes c_1 and c_2 from X_n and W is the spatial filtering projection matrix learned from the two classes of sub-band features. This objective function came from generalized Rayleigh quotient and has an analytical solution, which is equivalent to solving a generalized eigenvalue decomposition problem. The result W is the eigenvectors corresponding to the maximum and minimum eigenvalues of the two covariance matrices. Using the filter W to extract spatial features can maximize the feature difference between the two types of data. So, we choose the first m rows and the last m rows $(m = 2)$ of the spatial filter matrix W as the spatial filters [33]. Finally, the spatial features of the EEG signal in each frequency band are obtained by using the CSP spatial filter W_n. The final data representation is expressed as:

$$F_n = W_n X_n. \tag{2}$$

2.2. Pretrained feature extractor module

Traditional neural network models generally use layer-by-layer connected convolutional layers to extract deep features. When the number of network layers is shallow, the features have a more detailed representation ability, but the receptive field is smaller, the semantics are weaker and there is more noise. As the number of network layers increases, the receptive field of the network gradually becomes more extensive, and the semantic expressiveness increases, but the perception of details worsens [40]. However, there are shallow and deep features in the MI-EEG, which both characterize the abstract representations of the input data at their own levels. Therefore, we choose the ResNet model [41] to fuse the shallow detail information and deep global structure of the input data by supplemented cross-layer connection paths to the layer-by-layer connected CNN network. It ensures maximum information flow and enhances the feature representation of small-scale targets.

As shown in Figure 1, the input of the pre-trained network is EEG features $F_n \in \mathbb{R}^{E \times T}$, where E denotes the number of feature channels after data representation and T denotes the sample points. We extract the feature in the temporal and spatial domains by using two kinds of one-dimensional convolution operations, respectively, as shown in Figure 2. For the spatial domain convolution $Conv1$, a convolution layer of $(E \times 1)$ kernel size is used to learn the spatial information of multi-electrodes;

for the temporal domain information, we construct multiple layer basic residual modules including $Conv2$ and $Conv3$. The output of the lth layer is represented as [41]:

$$Y_l = h(F_l) + \mathcal{F}(F_l, W_l), \tag{3}$$

where $h(\cdot)$ is the direct mapping, i.e., the part of the curve in Figure 2. Since the feature dimension of Y may be different from F after the convolution operation, $h(\cdot)$ uses 1×1 convolution mapping to raise or reduce the dimension of the feature F. $\mathcal{F}(\cdot)$ is the residual part consisting of the convolutions $Conv2$ and $Conv3$, i.e., the straight-line part in Figure 2. The temporal domain feature representation is extracted using a convolutional layer with kernel size (1×25).

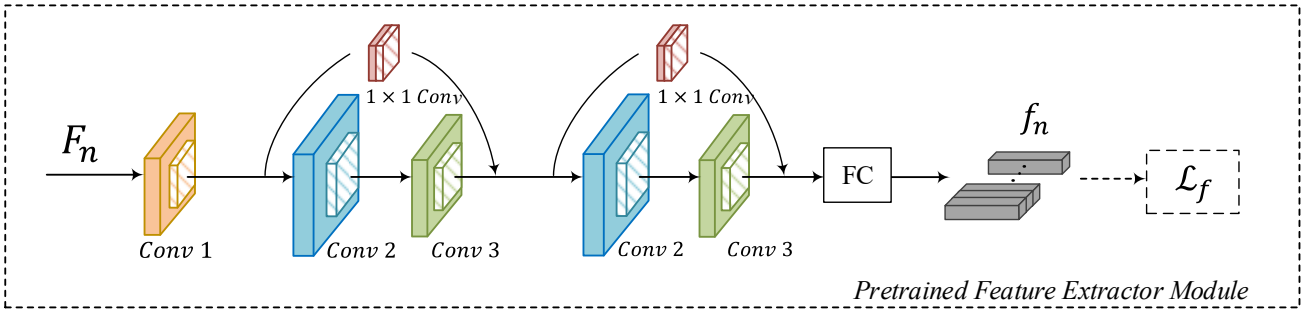

Figure 2. The structure of the specific pre-training feature extraction module.

Through forward propagation of the network, the final output of the residual module can be expressed as [41]:

$$\begin{aligned} Y_{l+1} &= F_{l+1} + \mathcal{F}(F_{l+1}, W_{l+1}) \\ &= f(Y_l) + \mathcal{F}(f(Y_l), W_{l+1}). \end{aligned} \tag{4}$$

Finally, we use two fully connected (FC) layers $(1024, cls)$ to generate depth features represented as f. Details of the network parameters are shown in Table 1.

To ensure that the features are discriminative and to prevent overfitting during training, we used cross-entropy classification loss $\mathcal{L}_f$ with labeled smoothing to minimize the difference between the predictions of the network model and the corresponding ground-truth. The equation of the loss function is as follows:

$$y^{LS} = y(1 - \alpha) + \frac{\alpha}{K}, \tag{5}$$

$$\mathcal{L}_f = -\frac{1}{N}\sum_i \sum_{c=1}^{K} \mathbb{1}_{(y_i^{LS} == c)} \log(p_{ic}), \tag{6}$$

where y^{LS} denotes the smoothed label, α denotes the hyperparameter of the smoothed quantity, K denotes the category label, $\mathbb{1}$ is the indicator function, which is set to be 1 if $y^{LS} == c$ is satisfied or 0 if not, and p_{ic} denotes the predicted probability that sample i belongs to category c.

Table 1. The MB-CCRF network parameters.

Layers	MB-CCRF network
Conv 1	$(16 \times 1), 128$, stride $(16, 1)$
Conv 2	$\begin{bmatrix} 1 \times 25, 128 \\ 1 \times 25, 128 \end{bmatrix} \times 2$
Max-pooling1	1×25， stride $(1, 25)$
Conv 3	$\begin{bmatrix} 1 \times 25, 256 \\ 1 \times 25, 256 \end{bmatrix} \times 2$
Max-pooling2	$(1 \times 20), stride\ (1, 20)$
FC	$[1024; 4]$， SoftMax

2.3. Centroid contrastive reconstruction fusion module

In the pre-trained network model, the output of each branch is the EEG feature of each frequency band, respectively. However, the sensorimotor rhythms of different subjects may be presented in different frequency bands. Fusing multi-band features by stitching will increase the dimensions and affect the computational efficiency of the network. Also, it can negatively affect the final recognition effect of features in the shared embedding space due to the influence of bands with weaker discriminative information. To make extensive use of the discriminative information in multi-band features, we consider the differences and complementarities between features, recode them in the shared embedding space using a self-supervised k-means model, reconstruct their category semantic structure and fuse multi-band features with a learning mechanism to improve the feature representation capability of the network model.

2.3.1. Centroid contrastive reconstruction module

First, we obtain the features f^t of the entire training set by invoking the pre-training model. The step is repeated over five frequency band branches.

$$f^t = [f_1^t, \cdots, f_{N_{train}}^t], \tag{7}$$

where f_n is the feature vector of X_n. N_{train} represents the number of features in the training set.

Then, in order to rebuild the semantic features by embedding multi-band features into the same feature space, k-means clustering [42] is performed for the training set features f^t, finding each sub-band feature cluster $\boldsymbol{C}^t = \{C_1, C_2, \ldots, C_K\}$ with normalized feature prototypes $\{\mu_j\}_{j=1}^k$. Specifically, $\mu_j = \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|}$, where $\mathbf{u}_j = \frac{1}{|C_j|}\sum_{f_i^t \in C_j} f_i^t$.

Finally, we compare the features from the training set with the category prototypes to construct the affinity matrix and reconstruct the semantic structure of the features. The final feature f' is expressed as:

$$Z_{i,j} = \frac{f_i^t \cdot \mu_j}{\|f_i^t\| \|\mu_j\|}, \tag{8}$$

$$f_i' = [Z_1; Z_2; \cdots; Z_k]. \tag{9}$$

2.3.2. Weighted fusion module

After reconstructing the features, each band's preloaded model output features are mapped into the same shared embedding space, providing unique feature discriminative information for each band. To better fuse them, we weight the reconstructed features to increase the impact of feature channels that are more useful for the current task and to suppress the less helpful features. Figure 1 shows a detailed process diagram of the weighted fusion module.

First, to learn the interrelationships from frequency band features, we fuse the multi-band convolutional features f_n' of the five branches using row dimensional splicing and denote them as S.

$$\begin{aligned} S &= [f_1' \quad f_2' \quad f_3' \quad f_4' \quad f_5'] \\ &= [S_1, S_2, S_3, \cdots, S_N], \end{aligned} \tag{10}$$

where the recoded features $S \in \mathbb{R}^N$, $N = 5 \times K$.

Then, to learn the non-linear relationship between channels, we use a feature weight learning strategy, which can automatically obtain the weight corresponding to the classification contribution of each feature channel by recoding the features S.

$$\begin{aligned} S' &= F_{att}(S, \Omega) = \sigma(\omega_2 \cdot ReLU(BN(\omega_1 S))) \\ &= [S_1' \quad S_2' \quad S_3' \quad S_4' \quad S_5'], \end{aligned} \tag{11}$$

where $\omega_1 \in \mathbb{R}^{\frac{K}{r} \times K}$, $\omega_2 \in \mathbb{R}^{K \times \frac{K}{r}}$. ω_1 is the weight of the dimensionality-reducing layer to decrease the dimensions of the feature data and r is the dimensionality-reducing hyperparameter. ω_2 is the weight of dimensionality-increasing layer to recover the original data dimensions by 1×1 convolution. Sigmoid function (denoted as F_{att}) is used to activate and learn the global dynamic attention mechanism. And the global dymanic weights S' are repartitioned into ${S_n}' \in \mathbb{R}^K$ corresponding to the attention of each branch feature f_n'.

Finally, we use the weights ${S_n}'$ to sequentially boost the semantic structure of branch features f_n' to obtain the weighted fusion feature $Feat$, as shown in Eq (12).

$$Feat = \sum_{n=1}^{5} {S_n}' \cdot {f_n}'. \tag{12}$$

By weighted fusing five branches frequency band features f_n', the final feature $Feat$ has stronger discriminative ability than each branch feature f_n'.

2.4. Optimization training

In the pre-training of the MB-CCRF network, we set the smoothing hyperparameter $\alpha = 0.2$ for the cross-entropy loss function $\mathcal{L}_f$ for labelled smoothing, and the training process is as follows:

$$\mathcal{L}_{pre} = \mathcal{L}_f(f), \tag{13}$$

$$\hat{\theta}_n = \underset{\theta_n}{arg\,min}\, \mathcal{L}_{pre}(f_n|\theta_n), \tag{14}$$

where θ_n denotes the parameters of the MB-CCRF network, and the network model of five bands are trained independently.

In the centroid contrastive reconstruction module, we obtain the centroids of each frequency band data by self-supervised k-means and minimize the cross-entropy loss function $\mathcal{L}_{CE}$ to optimize the fusion network parameters θ_ω.

$$\theta_\omega = \underset{\theta_\omega}{argmin}\mathcal{L}_{CE}\left(Feat\middle|\theta_{\omega_1}, \theta_{\omega_2}\right), \tag{15}$$

where θ_{ω_1} is the dimensionality-reducing layer parameter for constructing the global feature S and θ_{ω_2} is the dimensionality-increasing layer parameter for constructing the attention mechanism S'.

During the testing process, we acquire the category centroid from the training dataset and reconstruct the final semantic structure of the testing dates by comparing them with the category centroid.

The method in this paper is implemented using the PyTorch library in Python, an AMD 5900X CPU and a GTX 3080ti GPU. In the proposed MB-CCRF network, the pre-training batch size is 64, the ReLU function is chosen as the activation function and the dimensionality reducing hyperparameter $r = 4$ in the weighted fusing module is determined by training iteratively. In addition, batch normalization is added at each layer connection. The number of iterations (Epoch) is set to 600, the initial learning rate is set to 10^{-5} and the learning rate in the Optimizer is updated and adjusted by using the Scheduler.

3. Experiments and results

In this section, we verify the effectiveness of the proposed MB-CCRF method. First, we present the multi-class EEG dataset of the BCI Competition IV IIa used in the experiments. Second, we compare the MB-CCRF method proposed in this paper with other baseline methods. Finally, we analyze the performance of each module through ablation experiments.

3.1. Dataset and pre-processing

The Brain-Computer Interface Competition IV IIa dataset contains 22-channel EEG signals from nine subjects with a sampling frequency of 250 Hz (refer to A1–A9). The dataset was collected on four different motor imagery tasks, including left hand (class0), right hand (class1), feet (class2) and tongue (class3). The timing scheme consists of a fixed 2 s, a reminder time of 1.25 s, followed by a period of MI of 4 s. For each subject, two sessions of data were collected with 288 trials (72 trials per MI task) for each session. More details of the dataset can be seen in [43] and it is available to download at http://www.bbci.de/competition/iv/#dataset2a.

The motor imagery signals from the time segment of 0.5 to 2.5 s after the cue instructed in each trial are used for the experiments. We remove the data from erroneous experiments. For the null-value problem, we chose linear interpolation to fill in the missing values. In this study, the hold-out strategy is used to divide the dataset, which randomly divides the available trials into 8 sections. Six sections are used as the training set and the remaining two sections are used as the testing set. The best pre-trained model for each subject is selected to verify the testing set, and the performance of the MB-CCRF network model proposed in this paper is evaluated using the classification accuracy of the testing set.

3.2. Analysis of experimental results

To verify the effectiveness of the MB-CCRF algorithm proposed in this paper, we use the BCI Competition IV IIa dataset to analyze the classification effect of the algorithm and compare it with other baseline methods, as shown in Table 2, where the classification results are from the original manuscript [12,24,25,44,45].

Table 2. Classification accuracy of different frequency band features and fusion features of THE BCI Competition IV IIa dataset.

Subject	A1	A2	A3	A4	A5	A6	A7	A8	A9	Ave
8–12 Hz	77.37	55.47	94.03	61.16	46.62	57.94	77.04	83.46	79.03	70.24
12–16 Hz	82.48	61.31	58.96	70.25	50.38	45.79	80.00	74.44	80.65	67.14
16–20 Hz	56.20	55.47	70.15	65.29	56.39	51.40	80.74	80.45	82.26	66.48
20–24 Hz	76.64	58.39	77.61	57.85	61.65	53.27	77.78	71.43	77.42	68.00
24–30 Hz	79.56	51.09	87.31	58.68	48.12	55.14	70.37	87.97	87.10	69.48
8–30 Hz	81.02	62.77	89.55	66.12	56.39	64.49	84.44	86.47	83.06	74.92
MB-CCRF	88.32	72.26	89.55	76.86	65.41	61.68	87.41	90.22	87.90	79.96

1) FBCSP-SVM [24] and GLRCSP [12] are the improved algorithms of CSP, and they both use CSP to maximize the variance between features in different classes. Among them, the average classification accuracy of the GLRCSP algorithm differs from the MB-CCRF algorithm by only 1.76%. But it is less effective for some subjects due to individual differences in EEG signals that cause the algorithm to bias towards certain datasets. Compared with FBCSP-SVM, the MB-CCRF algorithm incorporates a cross-layer connected network training module and a centroid contrastive reconstruction fusion module based on the refined frequency bands, which improves the average accuracy by 8.78%. The MB-CCRF algorithm not only enhances the classification accuracy in general, but also improves the classification results of the subject's dataset with poor classification and has higher generalizability.

2) ConvNet [44] is a traditional CNN algorithm, and Multi-Branch 3D CNN is an improved algorithm of CNN. The MB-CCRF algorithm appends cross-layer connection paths to the traditional CNN network for learn deeping and shallow features in sub-bands. Comparing with the two algorithms, the average classification accuracy of the MB-CCRF algorithm is improved by 7.43% and 4.94%,

respectively. It shows that constructing data representation for the EEG signal and fusing the shallow and deep features of the network are beneficial to retain more discriminative information, thus improving the classification performance.

3) CW-CNN, C2CM [24] and monolithic [25] network are integrated algorithms of CSP and CNN. Comparing with these three networks, the MB-CCRF algorithm has a better average accuracy with 6.89%, 5.50% and 1.55% differences, respectively. It implies that mapping the features into the shared embedding space and weighting the fusion based on the difference in contribution of each band feature can ensure the maximum fusion gain.

Table 3. Classification accuracy comparison with other published results applied to the BCI Competition IV IIa dataset.

Methods	FBCSP -SVM [24]	GLR CSP [12]	ConvNet [44]	Multi-branch 3D CNN [45]	CW-CNN [24]	C2CM [24]	Monolithic network [25]	MB-CCRF
A1	82.29	86.11	76.39	77.40	86.11	87.5	83.13	88.32
A2	60.42	58.33	55.21	60.14	60.76	65.28	65.45	72.26
A3	82.99	93.75	89.24	82.93	86.81	90.28	80.29	89.55
A4	72.57	67.36	74.65	72.29	67.36	66.67	81.60	76.86
A5	60.07	55.56	56.94	75.84	62.50	62.5	76.70	65.41
A6	44.10	65.28	54.17	68.99	45.14	45.49	71.12	61.68
A7	86.11	81.25	92.71	76.04	90.63	89.58	84.00	87.41
A8	77.08	93.75	77.08	76.85	81.25	83.33	82.66	90.22
A9	75.00	88.19	76.39	84.67	77.08	79.51	80.74	87.90
Ave	71.18	78.20	72.53	75.02	73.07	74.46	78.41	79.96

To analyze the effect of multi-band features on classification performance, we compared the classification accuracy of five sub-bands features, 8–30 Hz band features and centroid contrast reconstruction fusion features for nine subjects, respectively, as shown in Table 3. For different band features in the same subjects, the classification accuracies varied widely. It indicates that the MI-EEG features have different abilities to express motor imagery task in different frequency bands. The optimal classification bands also differed for different subjects. It shows that the differences in the classification contribution of features in different frequency bands are caused by the different response frequencies of EEG signals when different subjects perform the motor imagery task.

In the same subject, the classification accuracy of some sub-band features is improved compared with the 8–30 Hz features, but the fusing features achieved more significant classification results. In A2, A4 and A5 datasets with poor classification, the fused classification effects improved by 9.49%, 10.74% and 9.02%, respectively, compared with 8–30 Hz features. Therefore, refining the frequency bands of MI-EEG, reconstructing the semantic structure of features by centroid contrast and fusing multi-band features weighted according to the classification contribution can obtain more significant sensory-motor related information and be more competitive in the classification recognition of MI-EEG.

3.3. Results of ablation experiment

To analyze the performance of new-added modules in the model, we constructed three ways through ablation experiments to compare and analyze the classification accuracy of nine subjects with added fusion model, weighted fusion model and reconstruction weighted fusion model, as shown in Figure 3. We fused the pre-trained multi-band features by directly adding the MB-CCRF model without the other module, named as added fusion model. In weighted fusion model, we weighted and fused the multi-band features by using the learning mechanism, that is the MB-CCRF without the reconstruction module. In MB-CCRF model, both modules are used to fuse features from the centroid contrastive reconstruction module and weighted fusion module.

The average decoding accuracies of the three models are 69.6%, 76.8% and 79.9%, respectively. The weighted fusion model and MB-CCRF model improved by 7.2–10.3% over the added fusion model, which is due to the discriminative information contained in the multi-band features being fully utilized. In particular, the discriminative characteristics of the sub-band features are enhanced by weighted reconstruction, which strengthens the decoding accuracy of the model. On the contrary, the added fusion approach is affected by the sub-band features with less discriminative information, which weaken the representation of the strong semantic sub-band features after embedding in the common space, and negatively affect the final recognition effect of the model.

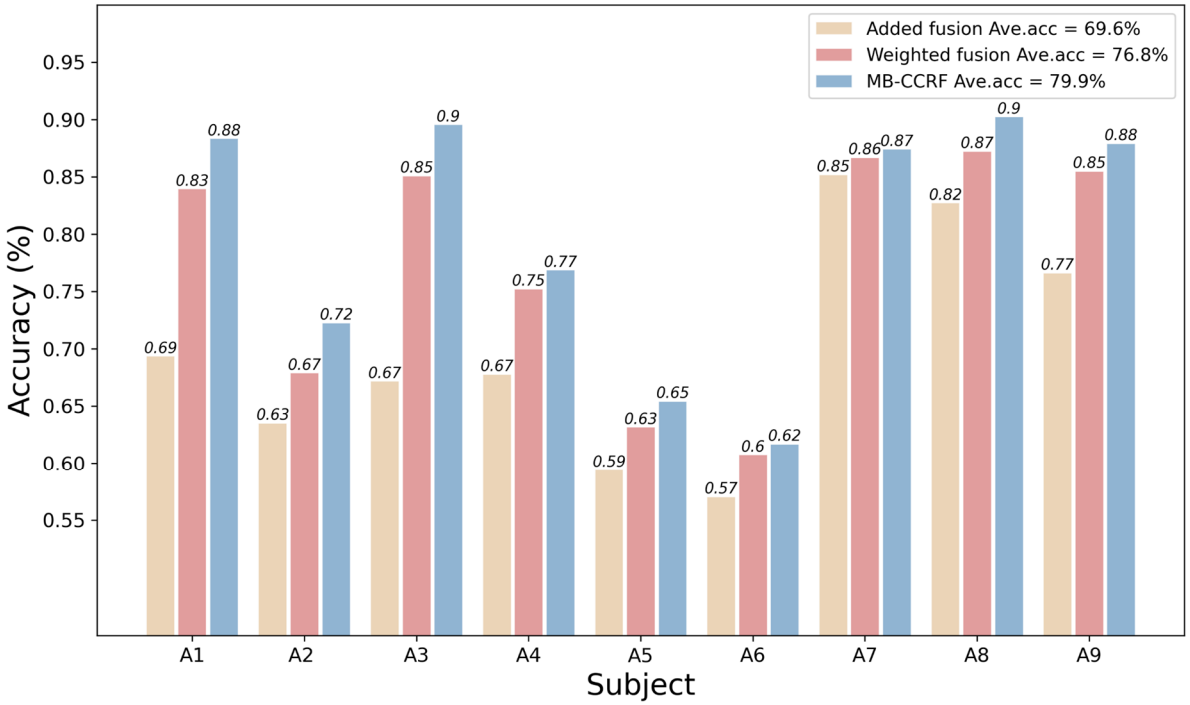

Figure 3. Compare the accuracy of different methods with added fusion, weighted fusion and MB-CCRF.

4. Analysis

In this section, we use three methods to discuss the effectiveness of the MB-CCRF network,

specifically modeling complexity and computation time, the feature clustering visualization, the confusion matrix and the sub-band feature activation interval.

4.1. Analysis of model complexity and computation time

In practice, large models can lead to overfitting of simple samples, while small models are not enough to cope with complex samples. For EEG signals, we need to ensure model performance and practical efficiency. Table 4 shows the number of parameters and computational complexity of the model. The MACs (multiply accumulate calculations) are the cumulative number of multiplications and additions performed per second, which can be used to measure the computational complexity of the model. It is independent for the environment in which the model is running (CPU, GPU or TPU). Where the MAC of the model is proportional to the inference time and the computational complexity, and the number of parameters for each part of the model's modules are small and suitable for real-world environments.

Table 4. The number of parameters and computational complexity of the MB-CCRF.

Model	MACs (G)	Parameters (M)
Pre-trained feature extracted	0.94	7.68
Weighted fusion	4e-08	4e-05

In addition, we analyze the computation time of the MB-CCRF model and the pre-trained model for different frequency bands, as shown in Table 5. The training time is the average training time of 100 iterations for one batch sample and the inference time is the average testing time for 100 random test samples. The training and testing times of the pre-trained models are roughly the same because the framework and the parameter number of pre-trained models are roughly the same. The computational time for the MB-CCRF model is the time to recall the pre-trained model and to build the reconstructed fusion model, so it does not increase exponentially.

Table 5. The training time and inference time of the MB-CCRF model.

Model	8–12 Hz	12–16 Hz	16–20 Hz	20–24 Hz	24–30 Hz	MB-CCRF
Training time (ms)	237.55	232.55	237.98	233.99	234.45	471.61
Inference time (ms)	3.59	3.64	3.59	3.59	3.54	41.18

4.2. Visualization of feature clustering effects

In this paper, t-distributed stochastic neighbor embedding (t-SNE) is a reduction and data visualization technique that transforms the similarity relationship between high-dimensional data into probability distributions to map them into a low-dimensional space for easier visual analysis of features. In the t-SNE visualization graph, we map the high-dimensional features of the EEG signal to two-dimensional spatial coordinates, as shown in Figure 4. Each colored spot in the graph represents the feature vector of the data sample that has been mapped into the low-dimensional space. The relative

distances of these points in the visualization graph reflect the similarity relationship between the original high-dimensional data samples.

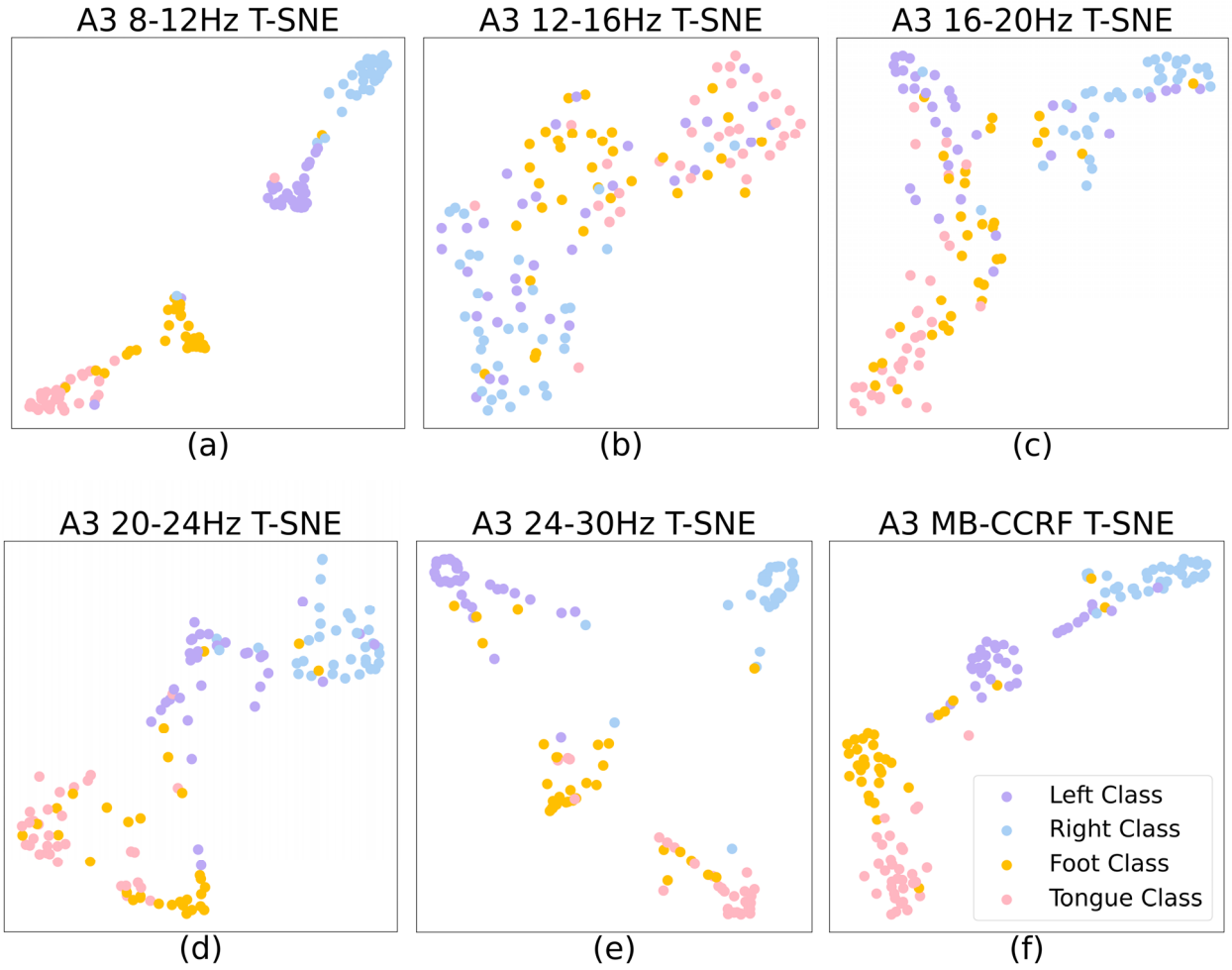

Figure 4. T-SNE visibility graph: sub-band features that are obtained by the centroid contrastive reconstruction module and fusion feature that is obtained by the weighted fusion model, mapped to 2-D plane. (a) 8–12 Hz; (b) 12–16 Hz; (c) 16–20 Hz; (d) 20–24 Hz; (e) 24–30 Hz; (f) MB-CCRF.

In Figure 4, we compare the classification effects of sub-band features and fusing features by visualizing the features of the A3 dataset that have significant differences in the classification effects of sub-band features. Among them, in the feature frequency bands of 8–12 Hz and 24–30 Hz, the distribution clusters of the four classes are clearly divided, but the intra-class distances are compact. In 16–20 Hz and 20–24 Hz bands, there is less distinction between left and right hand category clusters, and the features of the foot and tongue were heavily confounded. In the 12–16 Hz band, the category clusters of the left and right hand were severely confused, and some of the three category features, which are left hand, right hand and foot, entered the tongue class cluster. In contrast, as shown in Figure 4(f), the fusing features after the MB-CCRF network were not affected by the heavily confused

band features. They absorbed the advantages of the sub-bands with better classification effects, aggregated the same category features and separated the different category features, so that each class feature formed its own class cluster and improved its recognizability.

4.3. Confusion matrix of sub-band features

In this section, we use confusion matrices to discuss the classification results of different frequency band features and MB-CCRF features. In Figure 5, we choose the sub-band features and MB-CCRF fusion features of the A8 subject for discussion analysis. As shown in Figure 5(a)–(e), the categories with the best classification accuracy are different for different frequency band features, but they are often accompanied by spurious improvement of classification accuracy due to misclassification. This leads to a category propensity for single band features. In Figure 5(f), the classification accuracies are improved for four motor imagery tasks compared with single sub-band features. There is no such thing as an increase of the classification accuracy in a category due to category propensity.

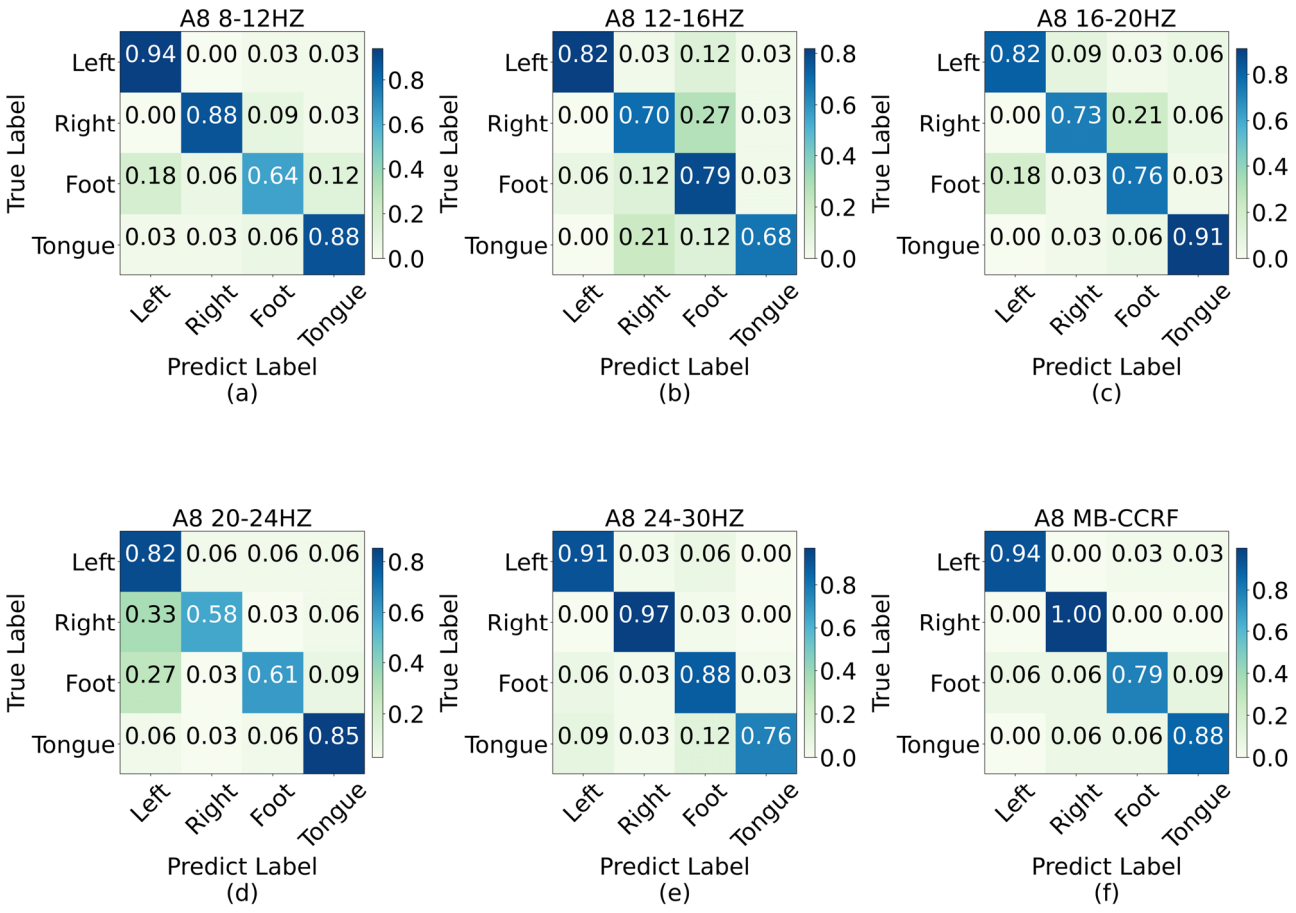

Figure 5. Confusion matrix analysis diagrams of different sub-band features: (a) 8–12 Hz; (b) 12–16 Hz; (c) 16–20 Hz; (d) 20–24 Hz; (e) 24–30 Hz; (f) MB-CCRF.

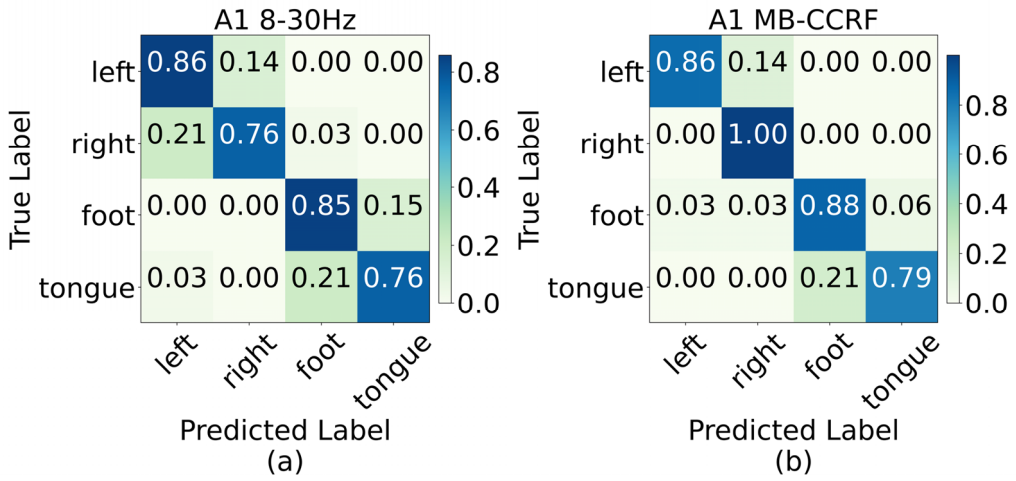

Figure 6. Confusion matrices of the subject A1: (a) the feature of 8–30 Hz; (b) the fusion feature.

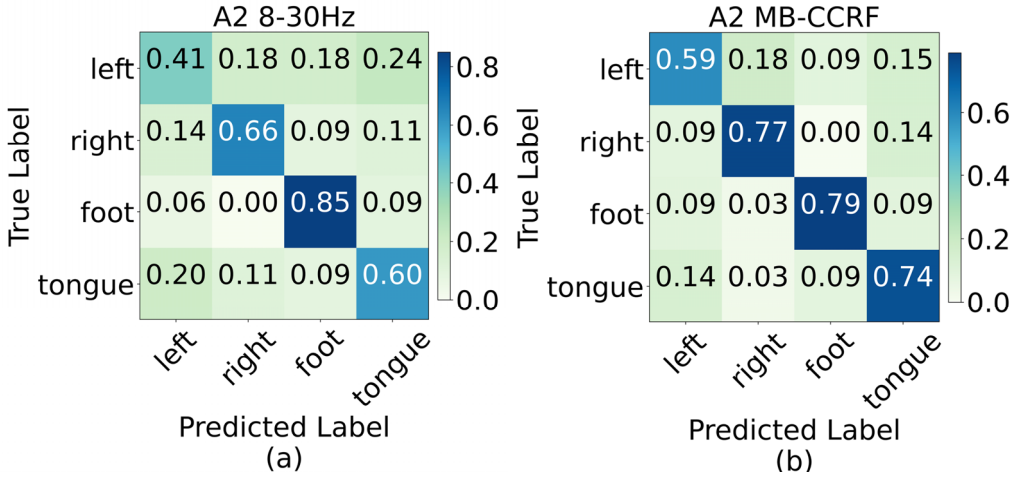

Figure 7. Confusion matrices of the subject A2: (a) the feature of 8–30 Hz; (b) the fusion feature.

As shown in Figures 6 and 7, we compare the 8–30 Hz band features and MB-CCRF fusion features of two subjects, which are A1 and A2. The recall rates of four classes have improved significantly. Moreover, for the difficult category samples, such as the left hand motor imagery task category samples of A2 subjects, the MB-CCRF algorithm not only improves the classification accuracy of this category but also reduces the misrecognition. It indicates that the MB-CCRF algorithm can neutralize the sensorimotor information in each sub-band feature, reduce the confusion between soft and hard samples and more precisely respond to the subject's brain activity.

4.4. Sub-band feature activation interval

Grad-CAM (gradient-weighted class activation mapping) is a technique for producing visual explanations for decisions made by convolutional neural network (CNN)-based models. It uses the gradients of target concept flowing into the final convolutional layer to produce an energy localization map, highlighting the regions of the image that are most relevant for a particular class prediction. It

assigns importance values to each neuron in the last convolutional layer based on the gradient information flowing into it, allowing for the explanation of output layer decisions [46].

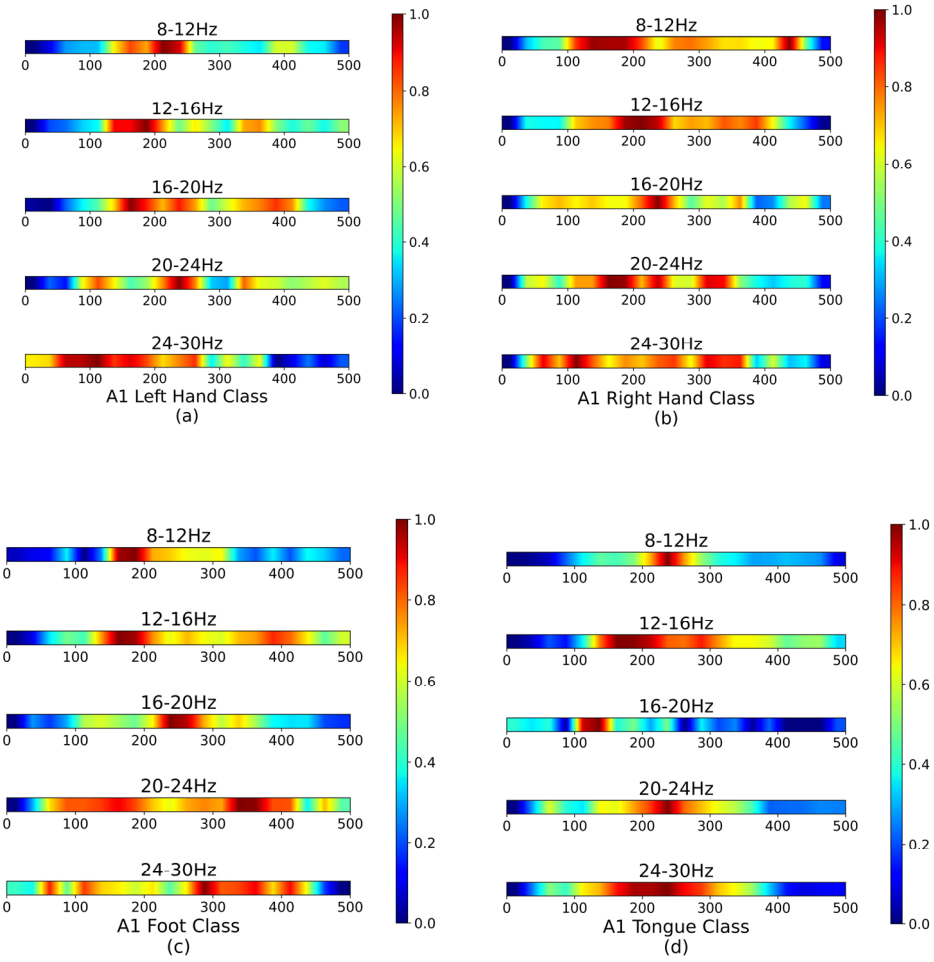

Figure 8. Grad-CAM heat map for sub-band feature activation interval: (a) left hand class; (b) right hand class; (c) foot class; (d) tongue class.

To observe the discriminative information of each sub-band in different motor imagery classes, we used Grad-CAM to visualize and to discuss the feature map based on the last layer of the pre-trained network, which is rich in highly abstracted class features. Deeper convolutional layers in the network capture more advanced signal features and retain some spatial information compared to fully connected layers. As a result, the final convolutional layer achieves an optimal balance between high-level semantics and detailed spatial information. To obtain class activation mapping (CAM), we replace the CNN classifier with global average pooling (GAP) and a fully connected layer equal to the class number, and then retrain the model features to account for the degree of category correlation between different regions. The portion of the original input data corresponding to highly correlated features can be determined by up-sampling. By the above method, we were able to explain the distribution of attention and the category correlation of the rhythm signal features extracted by the CNN model.

Figure 8(a)–(d) show the feature expression ability of the four classes of A1 subject in the original

signal interval. In the figure, the features extracted in the sub-band model are mapped into the intervals carrying key discriminative information in the original signal. It shows that the peak value of feature activation is mostly around the 200th time point. But the feature activation regions were significantly different for different motor imagery tasks, and it was significantly higher for left and right hand compared to tongue. Moreover, the activation regions of different sub-bands are varied under the same motor imagery task, but the regions with higher activation levels were partially overlapping and complementary. It indicates that with the MB-CCRF approach, sub-band features can complement each other to provide complete category information.

4.5. Validity analysis of the training module

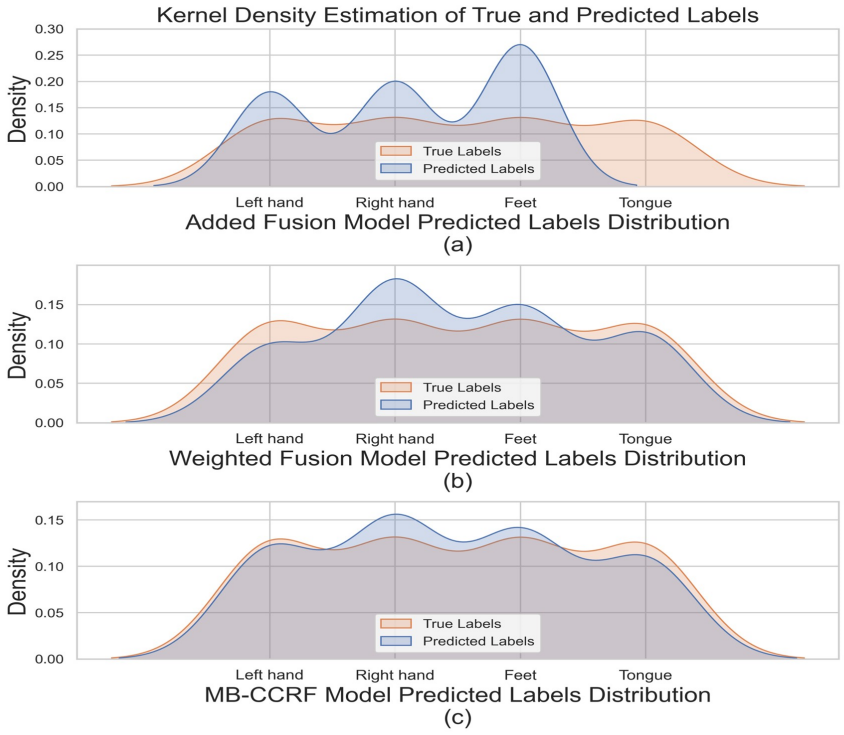

Figure 9. Kernel density estimation graph of different methods: (a) added fusion model; (b) weighted fusion model; (c) MB-CCRF model.

In this section, we use kernel density estimation to analyze the effectiveness of the weighted fusion module and centroid contrastive reconstruction module in the MB-CCRF model. Three different fusion models were used to analyze the EEG data of A1 subjects. The probability dense distributions of their predicted and actual labels are visualized, as shown in Figure 9. In Figure 9(a), the dense distribution of predicted labels is mainly clustered in the foot class region. There is also a slightly increased number of predicted labels in other classes, while the predicted labels in the tongue class are empty. It indicates that the added fusion features lead to a spurious increase in single class accuracy due to the categorization tendency. In Figure 9(b), it can focus more on the sub-bands of strong semantic categories by weighted fusion of the different sub-bands features, effectively avoiding the influence of anomalous features. In Figure 9(c), by utilizing the centroid comparative reconstruction

features, there is a large overlap in the distribution regions of the predicted and actual labels, and the distribution of multiple peaks is more even. It indicates that the centroid comparative reconstruction module can make full use of the pre-trained sub-band features to further improve the recognition effect of the weighted fusion model.

5. Conclusions

This paper proposes a multi-band centroid contrastive reconstruction fusion network, MB-CCRF, for the intention recognition task of MI-EEG signals. First, considering the individual band variability of EEG signals, we compensate for the incompleteness of the single-band data information by extracting the spatiotemporal feature matrices of the five sub-bands of MI-EEG. Second, we pre-train the multi-band features using a multi-branch cross-layer connected network model. The cross-layer connected feature extraction approach not only enhances the receptive field of the model and the detail sensibility of the network, but also ensures the maximum information flow of the network is transmitted. Then, we extract category centroid features for each class feature in a single sub-band and reconstruct the categorical semantic structure of the features by comparing with the centroid to ensure the similarity of features in a shared embedding space. Finally, considering the redundancy and variability among different features, we build a weighted fusion model according to the correlations between multiband features and class labels, thus adaptively adjusting the weight of each band feature so that the generated fused features are more discriminative.

In this paper, we demonstrated three issues through ablation experiments and visualization analysis. First, the single-band features do not provide complete sensorimotor information because the best response frequency of MI-EEG varies between different subjects or experimental settings. The MB-CCRF fusion algorithm can compensate for the information loss and overcome the problem of different subjects with different best response bands. Second, recognizing features within a single frequency band tends to a certain category. These tendentious optimal accuracies cannot be used as an evaluation criterion for the effectiveness of recognition in that category. The MB-CCRF fusion algorithm improves the category tendency problem, reduces the confusion between soft and hard samples and reflects the subject's current motor intention accurately. Third, MI-EEG features carry different discriminative information in each frequency band. The MB-CCRF network adopts a learning mechanism to adaptively align the sub-band features and learns more information from sub-band features with more important discriminative information, which effectively improves the recognizability of the fused feature.

In conclusion, the MB-CCRF model shows great potential in decoding MI-EEG signals, but also faces some important challenges. Its main disadvantage is the complexity of the model, which requires careful tuning of multiple steps and many hyperparameters, and increases the cost of development and maintenance. Then, the method requires a large amount of high-quality, multi-band MI-EEG data. This limits its availability for practical applications, especially when data collection is difficult or expensive. In addition, it can lead to poor generalization ability of deep learning models that lack data samples with diversity information and feature representations with high generalization capabilities [22,47]. Therefore, it is unclear whether the MB-CCRF model has sufficient cross-subject generalization performance.

In the future, we need to explore some methods to reduce the complexity of the model, such as transforming regularization parameters into trainable parameters to decrease the requirements for manual hyperparameter tuning [20]. Moreover, we can improve the data processing techniques to

increase data availability or enhance the cross-subject generalization performance of the MB-CCRF method. The MB-CCRF method is applicable not only for EEG but also for non-invasive optical techniques for imaging and monitoring of brain and retinal neurology, such as optical coherence tomography (OCT) and other biomedical imaging applications [48–51]. Therefore, future research and experiments will be necessary to comprehensively evaluate the performance of the MB-CCRF method and determine its applicability in different fields and applications.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 62173238, 62273095, 52205117 and 52207134), Liaoning Provincial Natural Science Foundation of China (Nos. 2022-KF-21-06 and 2021-KF-12-11). The authors would like to acknowledge the BCI Competition IV dataset 2a which were used to test the algorithms proposed in this study.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, et al., Brain-computer interface technology: A review of the first international meeting, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **8** (2000), 164–173. https://doi.org/10.1109/TRE.2000.847807
2. J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, Brain–computer interfaces for communication and control, *Clin. Neurophysiol.*, **113** (2002), 767–791. https://doi.org/10.1016/S1388-2457(02)00057-3
3. V. Mihajlović, B. Grundlehner, R. Vullers, J. Penders, Wearable, wireless EEG solutions in daily life applications: What are we missing, *IEEE J. Biomed. Health Inf.*, **19** (2015), 6–21. https://doi.org/10.1109/JBHI.2014.2328317
4. Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, et al., Sparse group representation model for motor imagery EEG classification, *IEEE J. Biomed. Health Inf.*, **23** (2018), 631–641. https://doi.org/10.1109/JBHI.2018.2832538
5. T. D. Pham, Classification of motor-imagery tasks using a large EEG dataset by fusing classifiers learning on wavelet-scattering features, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **31** (2023), 1097–1107. https://doi.org/10.1109/TNSRE.2023.3241241
6. W. Y. Hsu, Y. W. Cheng, EEG-Channel-Temporal-Spectral-Attention correlation for motor imagery EEG classification, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **31** (2023), 1659–1669. https://doi.org/10.1109/TNSRE.2023.3255233

7. C. Liu, J. Jin, I Daly, S Li, H. Sun, Y. Huang, et al., SincNet-based hybrid neural network for motor imagery EEG decoding, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **30** (2022), 540–549. https://doi.org/10.1109/TNSRE.2022.3156076
8. X. Yin, M. Meng, Q. She, Y. Gao, Z. Luo, Optimal channel-based sparse time-frequency blocks common spatial pattern feature extraction method for motor imagery classification, *Math. Biosci. Eng.*, **18** (2021), 4247–4263. https://doi.org/10.3934/mbe.2021213
9. S. Vaid, P. Singh, C. Kaur, EEG signal analysis for BCI interface: A review, in *Fifth International Conference on Advanced Computing & Communication Technologies*, (2015), 143–147. https://doi.org/10.1109/ACCT.2015.72
10. Y. Li, X. D. Wang, M. L. Luo, K. Li, X. F. Yang, Q. Guo, Epileptic seizure classification of EEGs using time–frequency analysis based multiscale radial basis functions, *IEEE J. Biomed. Health Inf.*, **22** (2017), 386–397. https://doi.org/10.1109/JBHI.2017.2654479
11. J. W. Li, S. Barma, P. U. Mak, F. Chen, C. Li, M. Li, et al., Single-channel selection for EEG-based emotion recognition using brain rhythm sequencing, *IEEE J. Biomed. Health Inf.*, **26** (2022), 2493–2503. https://doi.org/10.1109/JBHI.2022.3148109
12. F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms, *IEEE Trans. Biomed. Eng.*, **58** (2010), 355–362. https://doi.org/10.1109/TBME.2010.2082539
13. H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **8** (2000), 441–446. https://doi.org/10.1109/86.895946
14. P. Herman, G. Prasad, T. M. McGinnity, D. Coyle, Comparative analysis of spectral approaches to feature extraction for EEG-Based motor imagery classification, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **16** (2008), 317–326. https://doi.org/10.1109/TNSRE.2008.926694
15. B. Orset, K. Lee, R. Chavarriaga, J. Millán, User adaptation to closed-loop decoding of motor imagery termination, *IEEE Trans. Biomed. Eng.*, **68** (2020), 3–10. https://doi.org/10.1109/TBME.2020.3001981
16. Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, A. Cichocki, Temporally constrained sparse group spatial patterns for motor imagery BCI, *IEEE Trans. Cyber.*, **49** (2018), 3322–3332. https://doi.org/10.1109/TCYB.2018.2841847
17. M. Lee, Y. H. Kim, S. W. Lee, Motor impairment in stroke patients is associated with network properties during consecutive motor imagery, *IEEE Trans. Biomed. Eng.*, **69** (2022), 2604–2615. https://doi.org/10.1109/TBME.2022.3151742
18. Y. Y. Miao, J. Jin, L. Daly, C. Zuo, X. Wang, A. Cichocki, et al., Learning common time-frequency-spatial patterns for motor imagery classification, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **29** (2021), 699–707. https://doi.org/10.1109/TNSRE.2021.3071140
19. D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, J. Chanussot, Graph convolutional networks for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.*, **59** (2021), 5966–5978. https://doi.org/10.1109/TGRS.2020.3015157
20. C. Li, B. Zhang, D. Hong, J. Yao, J. Chanussot, LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection, *IEEE Trans. Geosci. Remote Sens.*, **61** (2023), 1–12. https://doi.org/10.1109/TGRS.2023.3279834

21. J. Yao, B. Zhang, C. Li, D. Hong, J. Chanussot, Extended Vision Transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework, *IEEE Trans. Geosci. Remote Sens.*, **61** (2023), 1–15. https://doi.org/10.1109/TGRS.2023.3284671
22. D. Hong, B. Zhang, H. Li, Y. Li, J. Yao, C. Li, et al., Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks, *Remote Sens. Environ.*, **299** (2023). https://doi.org/10.1016/j.rse.2023.113856
23. P. Zhang, X. Wang, W. Zhang, J. Chen, Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **27** (2019), 31–42. https://doi.org/10.1109/TNSRE.2018.2884641
24. S. Sakhavi, C. Guan, S. Yan, Learning temporal information for brain-computer interface using convolutional neural networks, *IEEE Trans. Neural Networks Learn. Syst.*, **29** (2018), 5619–5629. https://doi.org/10.1109/TNNLS.2018.2789927
25. B. E. Olivas-Padilla, M. I. Chacon-Murguia, Classification of multiple motor imagery using deep convolutional neural networks and spatial filters, *Appl. Soft Comput.*, **75** (2019), 461–472. https://doi.org/10.1016/j.asoc.2018.11.031
26. X. Ma, S. Qiu, H. He, Time-distributed attention network for EEG-based motor imagery decoding from the same limb, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **30** (2022), 496–508. https://doi.org/10.1109/TNSRE.2022.3154369
27. R. Zhang, N. L. Zhang, C. Chen, D. Y. Lv, G. Liu, F. Peng, et al., Motor imagery EEG classification with self-attention-based convolutional neural network, in *7th International Conference on Intelligent Informatics and Biomedical Science (ICIIBMS)*, (2022), 195–199. https://doi.org/10.1109/ICIIBMS55689.2022.9971698
28. J. Zheng, M. Liang, S. Sinha, L. Ge, W. Yu, A. Ekstrom, et al., Time-frequency analysis of scalp EEG with Hilbert-Huang transform and deep learning, *IEEE J. Biomed. Health. Inf.*, **26** (2022), 1549–1559. https://doi.org/10.1109/JBHI.2021.3110267
29. H. Fang, J. Jin, I. Daly, X. Wang, Feature extraction method based on filter banks and Riemannian tangent space in motor-imagery BCI, *IEEE J. Biomed. Health. Inf.*, **26** (2022), 2504–2514. https://doi.org/10.1109/JBHI.2022.3146274
30. F. Lotte, L. Bougrain, M. Clerc, Electroencephalography (EEG)-based brain-computer interfaces, in *Wiley Encyclopedia of Electrical and Electronics Engineering*, Wiley, (2015). https://doi.org/10.1002/047134608X.W8278
31. G. Pfurtscheller, C. Neuper, D. Flotzinger, M. Pregenzer, EEG-based discrimination between imagination of right and left hand movement, *Electroencephalogr. Clin. Neurophysiol*., **103** (1997), 642–651. https://doi.org/10.1016/S0013-4694(97)00080-1
32. R. Chai, S. H. Ling, G. P. Hunter, Y. Tran, H. T. Nguyen, Brain–computer interface classifier for wheelchair commands using neural network with fuzzy particle swarm optimization, *IEEE J. Biomed. Health. Inf.*, **18** (2014), 1614–1624. https://doi.org/10.1109/JBHI.2013.2295006
33. K. K. Ang, Z. Y. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, in *2008 IEEE International Joint Conference on Neural Networks*, (2008), 2390–2397. https://doi.org/10.1109/IJCNN.2008.4634130

34. K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, K. K. Ang, A new discriminative common spatial pattern method for motor imagery brain–computer interfaces, *IEEE Trans. Biomed. Eng*., **56** (2009), 2730–2733. https://doi.org/10.1109/TBME.2009.2026181
35. D. Hong, J. Yao, C. Li, D. Meng, N. Yokoya, J. Chanussot, Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion, *IEEE Trans. Geosci. Remote Sens.*, **61** (2023), 1–12. https://doi.org/10.1109/TGRS.2023.3324497
36. Y. Yuan, G. Xun, K. Jia, A. Zhang, A multi-view deep learning framework for EEG seizure detection, *IEEE J. Biomed. Health Inf.*, **23** (2019), 83–94. https://doi.org/10.1109/JBHI.2018.2871678
37. D. Zhang, K. Chen, D. Jian, L. Yao, Motor imagery classification via temporal attention cues of graph embedded EEG signals, *IEEE J. Biomed. Health Inf.*, **24** (2020), 2570–2579. https://doi.org/10.1109/JBHI.2020.2967128
38. W. Wu, X. Gao, B. Hong, S. Gao, Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL), *IEEE Trans. Biomed. Eng.*, **55** (2008), 1733–1743. https://doi.org/10.1109/TBME.2008.919125
39. F. Qi, Y. Li, W. Wu, RSTFC: A novel algorithm for spatio-temporal filtering and classification of single-trial EEG, *IEEE Trans. Neural Networks Learn. Syst.*, **26** (2015), 3070–3082. https://doi.org/10.1109/TNNLS.2015.2402694
40. D. Li, J. Xu, J. Wang, X. Fang, Y. Ji, A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **28** (2020), 2615–2626. https://doi.org/10.1109/TNSRE.2020.3037326
41. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *arXiv preprint*, (2015), arXiv:1512.03385. https://doi.org/10.48550/arXiv.1512.03385
42. D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, (2007), 1027–1035.
43. K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV Datasets 2a and 2b, *Front. Neurosci.*, **6** (2012), 39. https://doi.org/10.3389/fnins.2012.00039
44. R. T. Schirrmeister, J. T. Sprongenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, et al., Deep learning with convolutional neural networks for EEG decoding and visualization, *Hum. Brain Mapp.*, **38** (2017), 5391–542. https://doi.org/10.1002/hbm.23730
45. X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, L. Sun, A multi-branch 3D convolutional neural network for EEG-based motor imagery classification, *IEEE Trans. Neural Syst. Rehabil. Eng*., **27** (2019), 2164–2177. https://doi.org/10.1109/TNSRE.2019.2938295
46. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), 618–626. https://doi.org/ 10.1109/ICCV.2017.74
47. D. Hong, N. Yokoya, J. Chanussot, X. Zhu, An augmented linear mixing model to address spectral variability for hyperspectral unmixing, *IEEE Trans. Image Process.*, **28** (2019), 1923–1938. https://doi.org/10.1109/TIP.2018.2878958
48. R. K. Meleppat, C. R. Fortenbach, Y. Jian, E. S. Martinez, K. Wagner, B. S. Modjtahedi, et al., In vivo imaging of retinal and choroidal morphology and vascular plexuses of vertebrates using swept-source optical coherence tomography, *Transl. Vision Sci. Technol.*, **11** (2022), 11. https://doi.org/10.1167/tvst.11.8.11

49. K. M. Ratheesh, L. K. Seah, V. M. Murukeshan, Spectral phase-based automatic calibration scheme for swept source-based optical coherence tomography systems, *Phys. Med. Biol.*, **61** (2016), 7652–7663. https://doi.org/10.1088/0031-9155/61/21/7652
50. R. K. Meleppat, E. B. Miller, S. K. Manna, P. Zhang, E. N. Pugh, R. J. Zawadzki, Multiscale hessian filtering for enhancement of OCT angiography images, in *Ophthalmic Technologies XXIX*, (2019), 64–70. https://doi.org/10.1117/12.2511044
51. R. K. Meleppat, P. Prabhathan, S. L. Keey, M. V. Matham, Plasmon resonant silica-coated silver nanoplates as contrast agents for optical coherence tomography, *J. Biomed. Nanotechnol.*, **12** (2016), 1929–1937. https://doi.org/10.1166/jbn.2016.2297

AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)