A Novel Artifact Removal Strategy and Spatial Attention-based Multiscale CNN for MI Recognition

Duan Li, Peisen Liu, Yongquan Xia
School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan, China

***Abstract*—The brain-computer interface (BCI) based on motor imagery (MI) is a promising technology aimed at assisting individuals with motor impairments in regaining their motor abilities by capturing brain signals during specific tasks. However, non-invasive electroencephalogram (EEG) signals collected using EEG caps often contain large numbers of artifacts. Automatically and effectively removing these artifacts while preserving task-related brain components is a key issue for MI de-coding. Additionally, multi-channel EEG signals encompass temporal, frequency and spatial domain features. Although deep learning has achieved better results in extracting features and de-coding motor imagery EEG (MI-EEG) signals, obtaining a high-performance network on MI that achieves optimal matching of feature extraction, thus classification algorithms is still a challenging issue. In this study, we propose a scheme that combines a novel automatic artifact removal strategy with a spatial attention-based multiscale CNN (SA-MSCNN). This work obtained independent component analysis (ICA) weights from the first subject in the dataset and used K-means clustering to determine the best feature combination, which was then applied to other subjects for artifact removal. Additionally, this work designed an SA-MSCNN which includes multiscale convolution modules capable of extracting information from multiple frequency bands, spatial attention modules weighting spatial information, and separable convolution modules reducing feature information. This work validated the performance of the proposed model using a real-world public dataset, the BCI competition IV dataset 2a. The average accuracy of the method was 79.83%. This work conducted ablation experiments to demonstrate the effectiveness of the proposed artifact removal method and SA-MSCNN network and compared the results with outstanding models and state-of-the-art (SOTA) studies. The results confirm the effectiveness of the proposed method and provide a theoretical and experimental foundation for the development of new MI-BCI systems, which is very useful in helping people with disabilities regain their independence and improve their quality of life.**

Keywords*—*Motor Imagery (MI); Brain Computer Interface (BCI); EEG signal; artifact removal; spatial attention; Convolutional Neural Network (CNN)

I. INTRODUCTION

Brain-computer interfaces (BCIs) have the capability of translating brain activity signals into commands to control external devices or communicate with the external environment [1]. One of the most common paradigms of BCIs is motor imagery (MI), which involves mentally simulating motor commands of specific body parts. The generation of MI signals is possible even in the presence of disabilities as the corresponding brain region is functioning properly. MI-BCIs have shown promising results in areas such as communication, control and rehabilitation [2-7]. Electroencephalography (EEG) is widely used for MI-BCI systems due to its convenience and low physical and psychological stress on the subject [8]. However, EEG signals can be affected by environmental factors, which can generate artifacts. The background noise will distort the signal of interest and consequently reduce the MI recognition accuracy. In recent years, deep learning has gradually received attention for MI recognition. However, extracting appropriate spatial and temporal information from EEG signals has always been a significant challenge, whether in deep learning or traditional studies.

Due to the fact that EEG signals are collected by electrodes in contact with the scalp, the signal-to-noise ratio of the EEG is relatively low. During the acquisition process, the EEG signals are easily contaminated by various factors, resulting in artifacts in the signal, such as eye movements, muscle movements and electric line noise [9,10]. Therefore, some studies exploring different types of artifact features and removal methods have achieved certain results. Independent component analysis (ICA) [11] is a widely used method for artifact removal in MI recognition. ICA can separate a specified number of components from the input signal. Typically, experienced researchers identify and exclude the components that are considered artifacts and then the remaining components are used to reconstruct the signal for further analysis. This approach has achieved some success in various studies [12-14]. However, classifying components extracted by ICA will require much time and effort from professionals, which is not feasible for large datasets. Therefore, the automatic classification of ICA components has been proposed in some literature. For example, Hesam et al. [15] used three features-based K-means clustering methods to automatically cluster and differentiate artifacts from brain states in ICA components, achieving a 3.95% accuracy improvement on the Physionet dataset. However, there is still a lack of effective exploration of different feature methods and the impact of different features on clustering is not yet clear. Therefore, further research in this area is necessary.

Traditional convolutional neural networks extract and learn features through convolutions. In contrast to the two-dimensional convolutions commonly used in image applications, one-dimensional convolutions are often employed in MI-related research to capture temporal information and electrode information from EEG signals. For example, well-known models such as EEGNet [16] and DeepConvNet [17], utilize one-dimensional convolutions to extract information from the signals. Recently, the TS-SEFFNet [18] also incorporates one-dimensional convolutions with multiple scales, which has been proven to be effective. The multi-scale module can effectively concentrate on multiple dimensional information. How to improve the useful multi-dimensional features and suppress useless information is a key issue for MI recognition. Attention mechanism [19], after its proposal in 2014, which has been widely used in both the computer vision (CV) [20] and natural language processing (NLP) [21] fields, provides an efficient way for multiple channel EEG recognition. The Attention mechanism mainly focuses limited attention on important information, thereby saving resources and quickly obtaining the most relevant information. There are three main types of attention models for multi-channel EEG: spatial attention, channel attention and a combination of spatial and channel attention [22]. Spatial attention pays attention to specific regions within the spatial domain, allowing the model to prioritize important spatial information. Channel attention assigns different weights to different channels based on their attention values, facilitating the model to focus on important channel-wise features. However, in MI-BCI research, constructing a robust CNN is still a challenging problem. The utilization of various modules in the network structure still requires exploration.

In this paper, a novel features selection ICA and K-means-based automatic artifact removal method and an end-to-end CNN that includes multi-scale convolutions and spatial attention mechanism were proposed to overcome the problem of low classification accuracy and enhance the network's robustness. It is well known that the artifacts can lead to the acquisition of task-irrelevant features by subsequent classifiers, resulting in a decrease in classification accuracy. Therefore, this work proposes an automatic artifact removal method that combines ICA and clustering algorithms. For the subsequent network architecture, this work introduces a modularized network called SA-MSCNN. It utilizes multi-scale block and spatial attention modules to extract different types of information. The method takes into account the deep network's ability to learn different features from the data and offers a novel approach for classifying motor imagery. This work evaluates the performance of the proposed method using 5184 trials with 22 EEG channels from nine subjects in the BCI-IV-2a dataset.

The remainder of this paper is organized as follows. Section II introduces the method for this study which includes the proposed artifact removal method and SA-MSCNN. In Section III, this work introduces the dataset and the evaluation metrics used in this work. And also describes preprocessing steps, experimental parameter settings and the experimental results. Section IV is the discussion and Section V concludes the paper.

II. Method

This section describes the method this work proposed in this paper. First, this work gives an overall framework of the proposed method and a brief introduction to the workflow. Then, this work describes in detail the proposed ICA+K-means artifact removal method and the SA-MSCNN. Finally, this work shows the training strategy for the proposed network.

A. Overall Framework

The proposed method is shown in Fig. 1. This work first pre-processed the data for all subjects and then performed artifact removal using the proposed ICA+K-means method. For each subject, this work pre-trained the proposed Spatial Attention-based Multi Scale Convolution Neural Network (SA-MSCNN) using data from all the subjects except the specific subject, then saved the SA-MSCNN's weights. Finally, the SA-MSCNN model was trained and adjusted by the specific subject data (on the right side of Fig. 1) and the following experiments were performed to obtain the classification results for comparison.

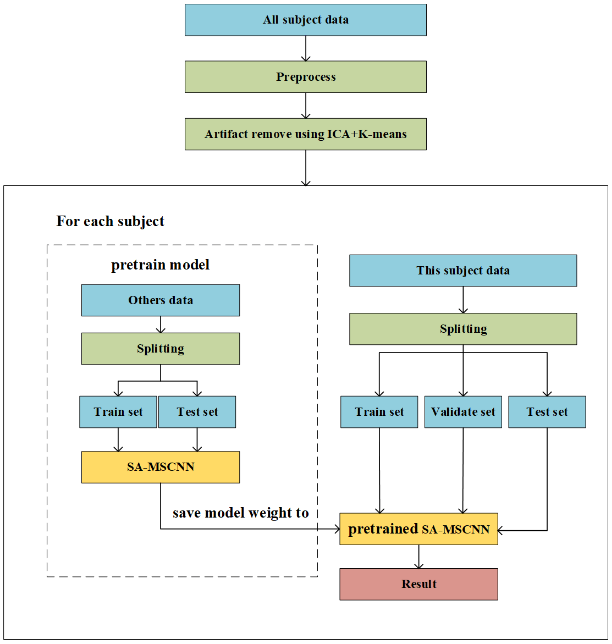

Fig. 1. Block diagram of the proposed framework.

B. Artifact Removal using ICA+K-means

Independent Component Analysis (ICA) is a statistical method used for signal processing and data analysis. It is widely used in MI to remove artifacts from EEG signals [23]. Unlike other traditional signal analysis methods such as Principal Component Analysis (PCA), ICA emphasizes the independence rather than the correlation of the signals. ICA decomposes signals according to Eq. (1) and (2).

$$\mathbf{X}_{m\times n}=\mathbf{A}_{m\times m}\mathbf{Y}_{m\times n} \tag{1}$$

$$\mathbf{Y}_{m\times n}=\mathbf{W}_{m\times m}\mathbf{X}_{m\times n} \tag{2}$$

Where X is the input EEG signal, A represents the mixing matrix, Y represents the original signal sources and W represents the inverse matrix of A. Here, m and n represent the number of EEG channels and the number of samples, respectively. The objective of the ICA algorithm is to find the weight matrix W that will decompose the EEG signals into ICs assuming temporal and spatial independence. Most artifact ICs in MI EEG signals are caused by eye movements, muscle movements and electric line noise [9, 10]. Traditionally, manual observation of ICs are used to differentiate artifact components from brain-related components and then remove the artifact components. This approach is feasible for datasets with a small number of subjects or a small amount of data for each subject. However, with the recent development of EEG measurement devices and experimental paradigms, large datasets like OpenBMI have emerged [24]. As the amount of data in the dataset increases, the manual selection of ICs becomes cumbersome and time-consuming. Therefore, the development of automated methods for ICs selection is necessary.

To address the above-mentioned issues, Hesam et al. [15] proposed using the K-means clustering method for component selection on extracted ICs from datasets including Physionet. They extracted three specific features from ICs and performed clustering, designating the class with the highest variance as the artifact class, and removing the components in that class. However, they only considered three specific feature selections and did not demonstrate the clustering results. Moreover, they did not explore the impact of different feature selections and multi-class datasets on K-means clustering from a broader perspective. Therefore, this study focuses on studying the impact of different feature selections on K-means clustering using representative datasets such as BCI-IV-2a. It provides recommendations for feature combinations and presents brain maps and ablation experiments after clustering. The seven selected statistical measures are variance, covariance, inverse covariance, correlation coefficient, kurtosis, skewness, and quartile range. Since the extracted ICs are vectors of certain lengths, these seven statistical measures can extract corresponding features of the ICs, as shown in Eq. (3) to (9) for their extraction method.

$$\sigma^2 = \frac{\sum (Y_i - \mu)^2}{n} \tag{3}$$

$$Cov(Y_i, Y_j) = E(Y_i, Y_j) - E(Y_i)E(Y_j) \tag{4}$$

$$invCov(Y_i, Y_j) = Cov(Y_i, Y_j)^{-1} \tag{5}$$

$$p = \frac{Cov(Y_i, Y_j)}{\sigma_{Y_i}\sigma_{Y_j}} \tag{6}$$

$$Kurt = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)^4}{\sigma^4} \tag{7}$$

$$Skew = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)^3}{\sigma^3} \tag{8}$$

$$IQR = Y_{75\%} - Y_{25\%} \tag{9}$$

Where σ^2 represents variance, *Cov* represents covariance, *E(A, B)* represents the expected value of a function involving two random variables, *A* and *B*, with their respective probability distributions, *invCov* represents inverse covariance, *p* represents correlation coefficient, *Kurt* represents kurtosis, *Skew* represents skewness, and *IQR* represents interquartile range. Y_i_represents the i-th sample, μ represents the sample mean, *n* represents the total number of samples, and $Y_{75\%}/Y_{25\%}$ represents the values at the 75th/25th percentile when the data is sorted in ascending order.

K-means is an unsupervised clustering algorithm where the parameter "K" in its name is set by the user to determine the number of clusters in the final result. The K-means algorithm iteratively maximizes the similarity among data points within each cluster while minimizing the similarity between clusters until the cluster centers no longer change or the predetermined number of iterations is reached. The most commonly used similarity measure is distance, such as Euclidean distance or Manhattan distance. The choice of K significantly impacts the clustering result and therefore, K-means clustering may converge to a locally optimal solution.

The different feature combinations for K-means algorithm will result in different clustering results. How to select optimal features for the automation of artifacts ICs identification is a key issue. Therefore, this study explores the different combinations of the aforementioned seven features and then adaptively achieves the best clustering results for denoising. The process is illustrated in Fig. 2. First, we extract the original ICA weights from the raw EEG signals of the first subject in the dataset.

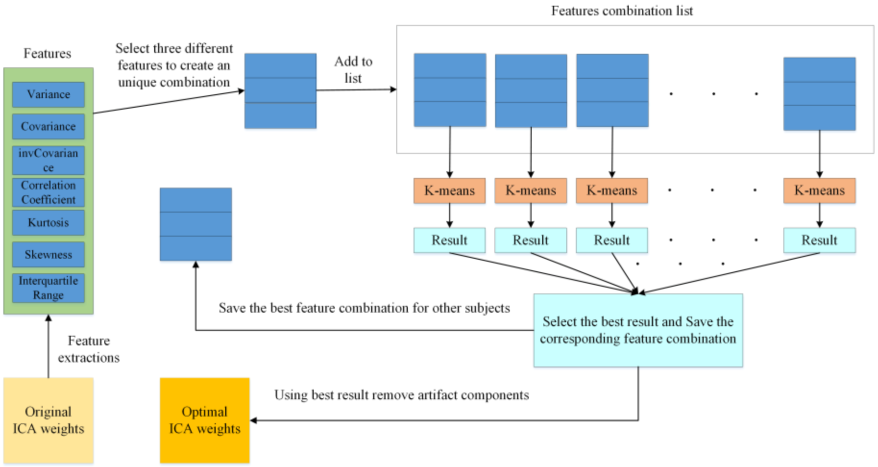
Features combination list
Features
Select three different features to create an unique combination
Add to list
Variance
Covariance
invCovariance
Correlation Coefficient
Kurtosis
Skewness
Interquartile Range
K-means
K-means
K-means
K-means
Result
Result
Result
Result
Save the best feature combination for other subjects
Select the best result and Save the corresponding feature combination
Feature extractions
Original ICA weights
Optimal ICA weights
Using best result remove artifact components

Fig. 2. Proposed ICA+K-means artifacts removal framework.

These weights are obtained by labeling the brain components and artifact components by experienced experts. Next, seven features were extracted from ICA weights. For these features, this work selects three different features to form a unique combination and then adds them to a feature's combination list. Each combination in the list is distinct from the others, resulting in a total of 35 unique feature combinations. For each combination, the K-means algorithm is to cluster using the three combined features. Then evaluate the clustering results based on the previous labels. Finally, select the best feature combination that achieves the optimal results. This work uses the best feature combination to obtain the optimal ICA weights in the following experiments for removing artifacts from the rest subjects. The proposed method is capable of automatically removing artifacts, reducing the expenditure of time and effort.

C. *Spatial Attention-based Multi Scale Convolution Neural Network (SA-MSCNN)*

The proposed SA-MSCNN is an end-to-end CNN, composed primarily of a multi scale temporal convolution block, a spatial convolution block, a spatial attention block, a separable convolution block and a classification block. The network structure is illustrated in Fig. 3.

As the 22 leads EEG data sampled at 250Hz and each sample has a length of 2.5s, the size of the sample is 22×625. The multi scale temporal convolution blocks and spatial convolution blocks of SA-MSCNN primarily capture the temporal and spatial features of the input EEG signals. The filters for the multi-scale temporal convolution block and spatial convolution block are set as [1, k] and [c, 1] respectively, where k and c are the lengths of the filters. Then, 2D convolution is performed accordingly. As mentioned in FBCSP [25], there may be some fluctuations in the MI frequency bands for different subjects. Therefore, to improve the classification accuracy, this work incorporates multi-scale convolutions with filters of different lengths in the temporal dimension to obtain information at different time scales. The obtained multi-scale features are then concatenated, normalized and passed through the spatial convolution blocks to extract spatial-scale information.

The subsequent Spatial Attention Block primarily utilizes the spatial attention module to obtain a refined feature map, further enhancing the model's attention and perception abilities in the spatial domain. The input feature, after undergoing max pooling and average pooling, has a shape of [batch_size, input_channels, height, width]. Then, a convolution operation is performed, resulting in a feature map with a shape of [batch_size, 1, height, width]. The values of the feature map are then mapped to a range between 0 and 1 using the Sigmoid function, achieving attention weights. Finally, these weights are multiplied element-wise with the original feature map, resulting in a weighted feature map, which is referred to as the refined feature map. In traditional feature extraction networks, the features at each position in the feature map are treated equally, without considering the importance of different positions. However, in real-world scenarios, the contribution of information from different positions varies. The spatial attention module introduces different weights to each position in the feature map, allowing the network to focus more on important positions and regions. This enables the network to capture richer and more accurate feature information thus improving the overall performance. Moreover, since the spatial attention module is introduced in the middle module, it does not directly process the raw EEG signals. It only requires the intermediate weights for feature extraction, thereby almost not increasing the complexity and computational cost of the network.

The refined feature map is performed batch normalization and ELU activation. Average pooling is applied to reduce the size of the feature map, while dropout is employed to prevent overfitting of the model. Then the feature map is fed into the depthwise separable convolution block. The depthwise separable convolution consists of depthwise convolution and pointwise convolution, with filters set as [1, k] and [1, 1], respectively. These convolutions aim to further shrink the feature map. Subsequently, batch normalization, ELU activation, average pooling and dropout are performed and the results are passed to the classification block. In the classification block of SA-MSCNN, the input features are flattened and then fed into a fully connected layer to compute the final output classification label.

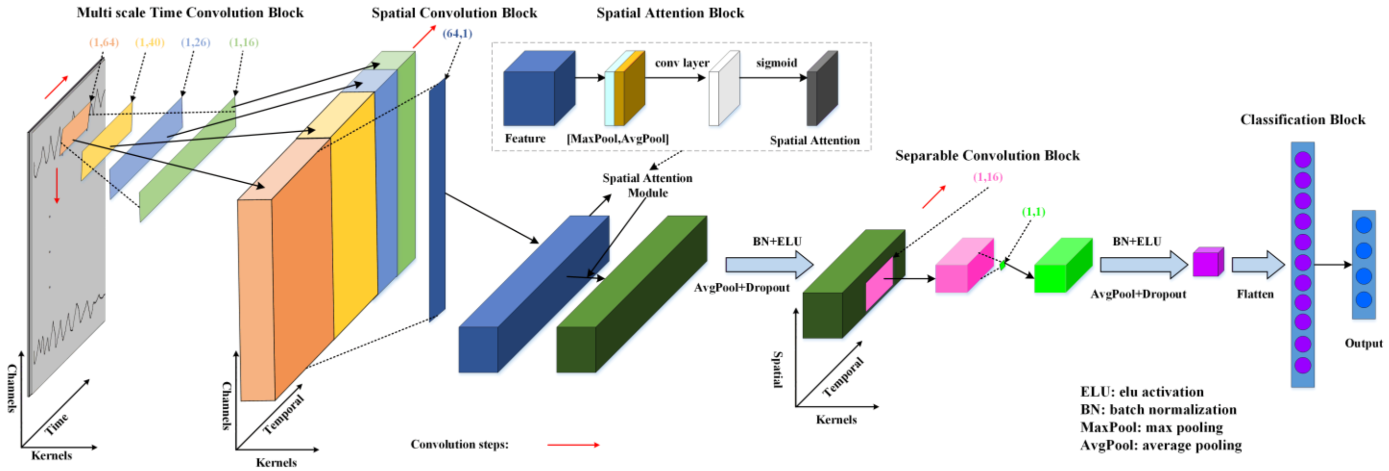

Fig. 3. Proposed SA-MSCNN architecture.

D. Training Strategy

For each subject, SA-MSCNN is pre-trained using data from all other subjects except the specific one and then trained, validated, and tested on the pre-trained model using the data of this specific subject. The reason for using the pre-training step is inspired by the idea of transfer learning. This training strategy aims to enable the model to learn features from the entire dataset as much as possible and increase the amount of data used for training. This allows the model parameters to converge faster when the data of the specific subject is fed into the pre-trained SA-MSCNN.

III. Experiments and Results

A. Data Sources and Evaluation Metrics

In this experiment, dataset 2a of the BCI competition IV, which contains 22 EEG channels and three monopolar EOG channels from nine subjects was used [26]. The 22 EEG electrodes were made by Ag/AgCl(with inter-electrode distances of 3.5 cm). The sampling frequency is 250Hz and bandpass filtering between 0.5Hz and 100Hz was applied when the dataset was recorded. The subjects were asked to perform four types of motor imagery tasks: left hand, right hand, tongue and both feet. Each category of the task was performed 72 times, resulting in 288 trials per session and each subject had two sessions. So, there are a total of 5184 trials in the dataset.

In this study, the performance of the proposed method was evaluated using its ac-curacy (represented by the symbol Acc) and Kappa value (represented by the symbol Kappa), defined by Eq. (10) and (11).

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

$$Kappa = 1 - \frac{1-P_o}{1-P_e} = \frac{P_o - P_e}{1-P_e} \tag{11}$$

Among them, TP, FP, TN and FN are true positive, false positive, true negative and false negative respectively. Po represents the total classification accuracy, and Pe is the sum of the product of the ground truth and predicted numbers for each category divided by the square of the total number of samples [27].

B. Pre-processing

In the data preprocessing stage, this work removed three EOG channels and retained 22 EEG channels to reduce their influence. MI tasks produce event-related synchronization (ERD) and event-related desynchronization (ERS) in EEG signals, corresponding to the sensorimotor rhythms of mu (8-13Hz) and beta (18-30Hz). The ERD/ERS patterns exhibit variability across subjects. Therefore, a wide-range band-pass filter of 4-40Hz was used to minimize band-pass filtering's influence on data while retaining MI-related features. For each task, the EEG data is segmented between 0 s to 2.5s after its start. Thus, each sample size is 22×625. No other operations are performed because this work wants to retain as much useful information as possible.

C. Experimental Setup

In the artifact removal stage, for the first subject, after ICA, the first two ICs were removed, and the remaining 20 components were manually classified into artifact and non-artifact categories. Then, the proposed method was used for clustering with three clusters, as the dataset consisted of four classes. The clustering results were recorded, excluding the features where artifacts and non-artifacts were clustered into the same category. The percentage of artifact ICs in the category with the most artifact ICs and the percentage of non-artifact ICs in the category with the most non-artifact ICs were recorded separately. The two percentages were added and divided by two to obtain the final clustering distinction percentage. A higher clustering distinction percentage suggests that the corresponding feature extraction method is more effective in artifact removal. For the remaining subjects, the most effective feature extraction method was applied to remove artifacts. To compare, two experiments were also conducted on all subjects: using ICA+manual artifact removal, and without using ICA for artifact removal.

ICA can be performed using the MNE package in Python [28] or the EEGLAB toolbox in MATLAB [29]. In the ICA artifact removal stage, for the first subject's ICs, the first two ICs were removed and the remaining 20 components were manually labeled as artifact and non-artifact categories. Hesam et al. [15] only used two classes of data from the Physionet dataset for their experiment and set the clustering to two classes. In this experiment, this work used a four-class dataset. Since each session for clustering involved four classes, in order to acquire more accurate clustering results, a three-class clustering approach was opted for. Then this work employed the proposed method for all the subjects using K-means clustering with three clusters. The clustering results were recorded, excluding the feature combinations where most artifacts and most non-artifacts ICs were clustered into the same category. The percentage of artifact ICs in the category with the maximum number of artifact ICs and the percentage of non-artifact ICs in the category with the maximum number of non-artifact ICs were calculated and recorded for each feature combination. The two percentages were added and divided by two to obtain the final clustering distinction. A higher clustering distinction suggests that the corresponding feature extraction method is more effective in artifact removal. For the remaining subjects, the three most effective feature extraction methods were applied to remove artifacts. To compare, the other two experiments were also conducted on all subjects: using ICA+manual artifact removal, and without artifact removal.

In SA-MSCNN, the kernel sizes for the multi-scale time convolutional blocks are set as [1,64], [1,40], [1,26], [1,16] and the kernel size for the spatial convolutional block is set as [64,1]. The depthwise separable convolution block consists of a depthwise convolution with a kernel size of [1,16] and a pointwise convolution with a kernel size of [1,1]. Each scale in the multi-scale temporal convolutional block has eight convolutional kernels, the spatial convolutional block has 16 convolutional kernels and the depthwise separable convolution block has 16 convolutional kernels. The dropout rate is set to

0.5 and the cross-entropy loss function is used. The Adam optimizer with a learning rate of 0.001 is utilized.

During the pretraining phase, for each subject, this work combined the data of the other eight subjects, totaling 4608 trials. This combined dataset was then divided into 75% for training and 25% for testing to pre-train the model parameters of SA-MSCNN. Then, the 576 trials from the current subject were divided into 50% for training, 25% for validation and 25% for testing, which were fed into the pre-trained SA-MSCNN to obtain results. Both the pre-training and training epochs are set to 200.

All the experiments were implemented on Windows 11 with an Nvidia RTX 3060 12GB GPU and the neural network was performed on the PyTorch platform.

D. Compared Methods

For comparison, this work also used three recent years' outstanding open-sourced models for EEG recognition including EEGNet, DeepConvNet and TS-SEFFNet.

- EEGNet [16]: EEGNet is a compact convolutional neural network specifically designed for processing EEG data. It extracts spatial and temporal features of EEG signals through one-dimensional convolutional layers and depthwise separable convolutional layers.
- DeepConvNet [17]: DeepConvNet is a model based on a deep convolutional neural network architecture used for classifying EEG signals. It employs multiple convolutional layers, pooling layers and fully connected layers to extract high-level features.
- TS-SEFFNet [18]: TS-SEFFNet is a time-frequency-based compressed and excitatory feature fusion network used for decoding motor imagery EEG. The network utilizes a novel time-frequency compression and excitatory feature fusion method for motor imagery EEG decoding

E. Result of Clustering

In Table I, Feature Combination represents the selected combinations of features. NBS MaxCate corresponds to the category where most of the artifact components are clustered. NBS MaxCatePercent represents the percentage of artifact components in this category out of the total artifact components. BS MaxCate corresponds to the category where most of the brain state components are clustered. BS MaxCatePercent represents the percentage of brain state components in this category out of the total brain state components. Clustering Distinction corresponds to the clustering distinctiveness, which is equal to half the sum of NBS MaxCatePercent and BS MaxCatePercent. In Table I， kurt represents kurtosis, skew represents skewness, cov represents covariance, iqr represents interquartile range, var represents variance, inv_cov represents inverse covariance and corr represents correlation coefficient. Three different feature extraction methods were selected from seven available methods to form a unique combination, resulting in a total of 35 combinations. Firstly, features that cluster artifact components and brain state components into the same category were excluded. Then, features with a clustering distinctiveness of less than 60% were excluded and the results are summarized in Table I. From Table I, it can be seen that the combination of kurt, skew and cov has the highest clustering distinction. Therefore, these three feature extraction methods were used in subsequent ICA+K-means automatic clustering to remove artifact components in other subjects.

TABLE I. CLUSTER RESULT OF SUBJECT 1

Feature Combination	**NBS MaxCate**	**NBS MaxCatePercent**	**BS MaxCate**	**BS MaxCatePercent**	**Clustering Distinction**
kurt, skew, cov	2	80%	0	86.67%	83.34%
kurt, skew, iqr	1	60%	0	86.67%	73.34%
kurt, skew,var	1	60%	0	86.67%	73.34%
inv_cov, iqr, var	0	100%	1	46.67%	73.34%
corr, kurt, cov	0	60%	1	73.33%	66.67%
corr,kurt, var	1	60%	0	73.33%	66.67%
kurt, cov, iqr	2	40%	0	86.67%	63.34%
corr, skew, inv_cov	2	80%	0	40%	60%

TABLE II. COMPARISON OF DIFFERENT EXPERIMENTS

Method	**Subjects**									
	A01	***A02***	***A03***	***A04***	***A05***	***A06***	***A07***	***A08***	***A09***	***Avg***
base model without artifacts remove	85.71%	64.86%	93.22%	69.86%	73.71%	61.33%	88.11%	81.31%	80.47%	77.62%
base model+ICA+ Manual select	89.71%	67.25%	95.9%	75.11%	77.21%	62.79%	89.25%	82.51%	83.22%	80.33%
base model+ ICA+K-means	89.71%	64.88%	94.97%	73.19%	78.68%	60.85%	90.93%	85.31%	79.97%	79.83%

F. Ablation Experiment of Artifact Removal Method

In this section, ablation experiments without artifacts, manual select and ICA combined K-means are performed respectively. The experimental results are summarized in Table II, the second column is the experimental results without artifact removal. The “ICA+Manual select” in the table represents the experimental results where manual observation was used to remove artifacts for all subjects. In the "ICA+K-means" experiments, the artifacts removal and optimal feature combinations for the remaining subjects were according to the first subject. The results are shown in Table II and Fig. 4.

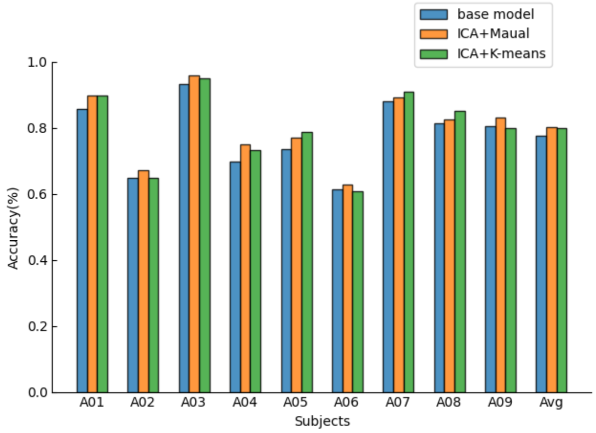

Fig. 4. Comparison of different experiments.

In the ICA+K-means method, the first subject still required manually classifying and removing artifact components. Therefore, the experimental results for the first subject, both in terms of manual artifact removal and cluster-based artifact removal, are the same. The results for the remaining subjects are different. From Fig. 4, it can be observed that both experiments using artifact removal methods outperformed the base model, which did not include artifact removal. This result demonstrates the effectiveness of artifact removal methods. Furthermore, Table II shows that the average classification accuracy for manual identification and exclusion of artifacts is 80.33%, while the average classification accuracy for using K-means clustering to identify and exclude artifacts is 79.83%, a difference of only about 1%.

Fig. 5 illustrate the training process for the third subject using the proposed method. Train_acc and Val_acc represent training and validating accuracy respectively, while Train_loss and Val_loss represent training and validation loss. Using the pre-training strategy, the model converges speed at a fast rate.

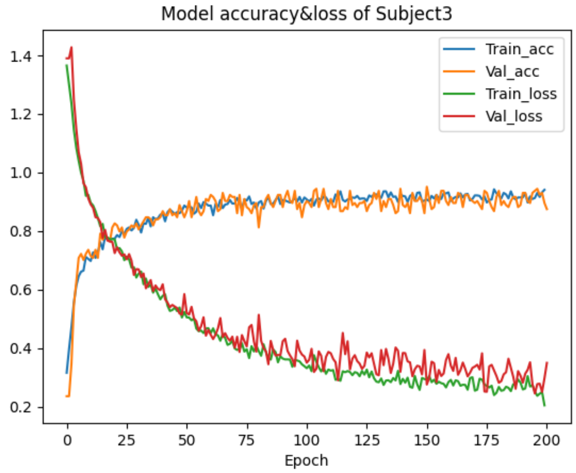

Fig. 5. Training process of subject 3.

G. Comparative Experiment with other Networks

EEGNet, DeepConvNet, and TS-SEFFNet are popular and excellent networks in the field of BCI decoding. From Table III, DeepConvNet achieves an average accuracy of 71.99%, EEGNet achieves 72.44% and TS-SEFFNet achieves 74.71%. By using the proposed artifact removal method and SA-MSCNN, the accuracy improves to 79.83%. Furthermore, two ablation experiments were also performed: 1) Removal of the multiscale block and the spatial attention block and 2) Removal of the artifact removal module. The results demonstrate the effectiveness of the proposed method. Fig. 6 is the confusion matrix of the four methods. The values of the matrix have been normalized by rows.

As can be seen in Table III, the proposed method outperforms these comparative methods in terms of both average classification accuracy and Kappa value. This work also conducted ablation experiments to compare and prove the effectiveness of the method, both SSA-MSCNN with multi-scale convolutional block and spatial attention block and ICA+K-means artifact removal method can improve the performance of the model. In Fig. 6, the classification accuracy of the method is improved on all four classes, especially on the left and right hands.

TABLE III. COMPARISON OF OTHER METHODS

Method	***Subjects' Acc***										***Kappa***
	A01	***A02***	***A03***	***A04***	***A05***	***A06***	***A07***	***A08***	***A09***	***Avg***	
DeepConvNet	80.90%	52.08%	84.72%	71.18%	70.49%	55.56%	69.10%	81.94%	81.94%	71.99%	0.627
EEGNet	81.25%	50.69%	91.67%	63.89%	70.14%	59.03%	79.17%	77.98%	78.18%	72.44%	0.632
TS-SEFFNet	82.29%	49.79%	87.57%	71.74%	70.83%	63.75%	82.92%	81.53%	81.94%	74.71%	0.663
SA-MSCNN without Multiscale Block and Spatial Attention Block	83.23%	61.07%	82.53%	65.23%	70.28%	57.47%	84.90%	78.16%	75.73%	73.18%	0.642
SA-MSCNN without ICA+K-means	85.71%	64.86%	93.22%	69.86%	73.71%	61.33%	88.11%	81.31%	80.47%	77.62%	0.702
SA-MSCNN with ICA+K-means	89.71%	64.88%	94.97%	73.19%	78.68%	60.85%	90.93%	85.31%	79.97%	79.83%	0.731

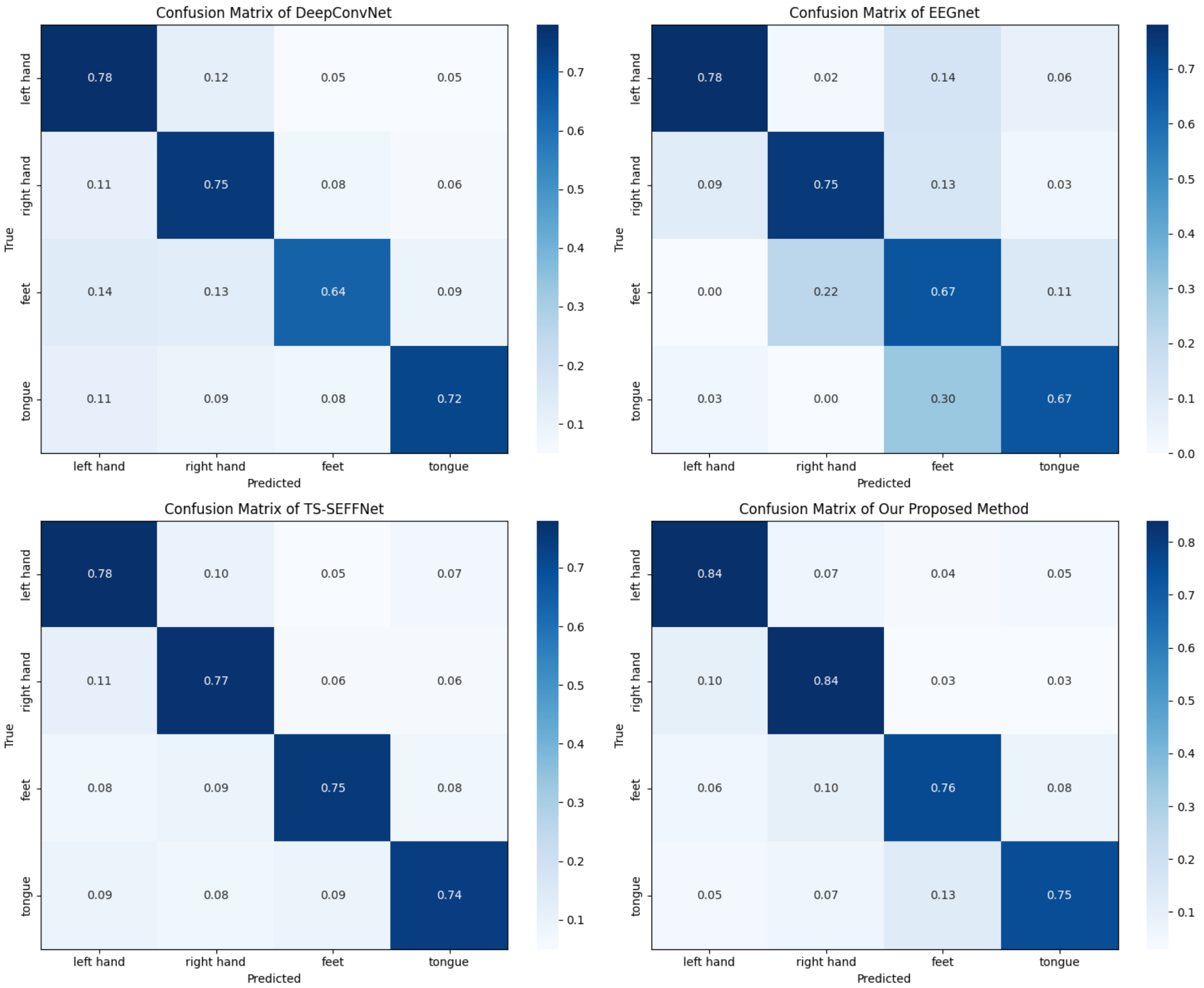

Fig. 6. Confusion matrix of the methods.

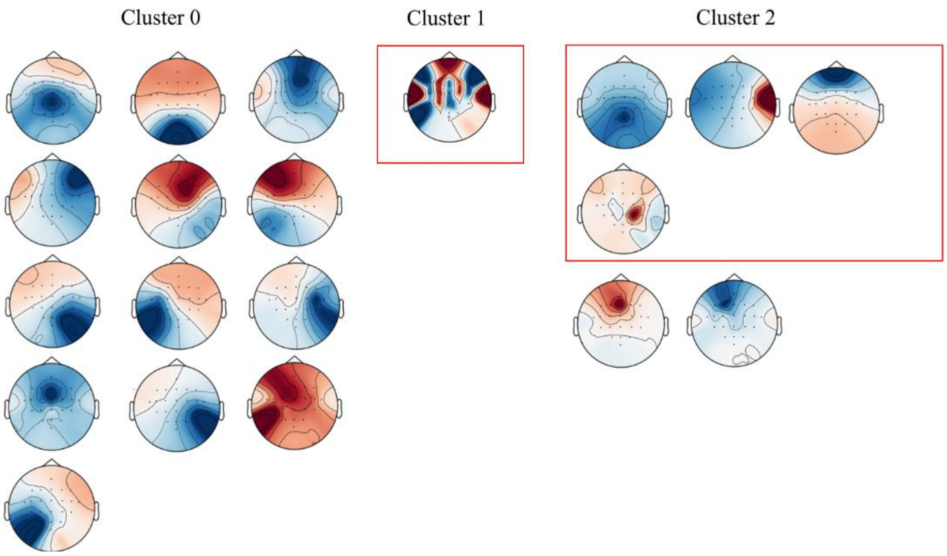

Fig. 7. Best clustering result for subject 1.

IV. Discussion

From the clustering results in Fig. 7, it can be seen that the combination of features (kurtosis, skewness, and covariance) can effectively cluster brain state components and non-brain state components into different clusters. The brain states within the red box are labeled as artifacts. However, in Cluster 2, two brain state components are still clustered together with artifacts. From the characteristics of these brain states, except for the central region of activation, the other regions of these two components are relatively inactive, that is similar to other artifact components in this category. This suggests that these two components may have some similarity in their features, resulting in clustering with artifacts. In the ablation experiment in Table I, the method of clustering and artifact removal using kurtosis, skewness and covariance features still achieves good average accuracy compared to not using this method. This proves that the selected features are still applicable to the remaining subjects. Compared to manually removing artifacts from nine subjects' ICs, the ICA+K-means method eliminates the time-consuming process of manually screening and removing artifact components for all subjects while achieving a performance loss of less than 1% in average accuracy. This is crucial for large datasets with many subjects and sessions. Manual identification and removal of ICs for each subject would be time-consuming and inefficient in large datasets. On the other hand, the automated process of the ICA+K-means method can quickly and accurately remove artifacts, saving significant labor and time costs.

As shown in Fig. 8, this work also performed a feature visualization of the comparison experiment. This was done by performing parameter extraction prior to the final classification of each network and then visualizing it via the t-SNE method. Different colors represent different parts of the motor imagery being performed: green for the left hand, purple for the right hand, blue for the tongue, and red for the feet. Different colors are used to distinguish the visualization results in order to represent them more intuitively. The visualization results of the proposed method have a smaller intra-class spacing than the other methods. This is consistent with the previous experimental results in Table III.

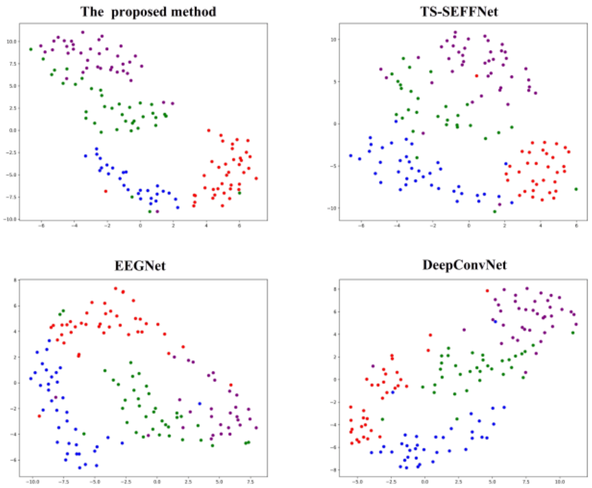

Fig. 8. Visualization results of comparison experiments.

EEG signals primarily require the extraction of temporal and spatial information. Previous studies have demonstrated that different people show specific electrical signal patterns in a certain range of frequency bands when imagining the same motor imagery task. However, the specific frequency bands may vary across individuals. Therefore, similar to the FBCSP approach, this study employs filters at multiple scales to capture temporal information, aiming to mitigate the impact of individual differences on temporal information extraction. For spatial information extraction, instead of electrode selection employed in some studies [30-34], the proposed method utilizes a spatial attention strategy after performing spatial convolution to automatically allocate weights to feature maps. The reason for this choice is that the dataset itself only has 22 EEG channels and performing electrode selection would result in the removal of some channels, which could have a negative impact on the result, especially in datasets with limited channels. To make full use of the feature maps, the proposed method utilizes spatial attention to automatically allocate weights to them. The results of the ablation experiment in Table III confirm the effectiveness of the multi-scale block and spatial block.

In addition to the above high-performance models, this work has also investigated some methods from recent MI-BCI studies that use the same dataset. Wang et al. [35] proposed an unsupervised domain adaptation framework called Iterative Self-training Multisubject Domain Adaptation (ISMDA) for the offline MI task, achieving an average classification accuracy of 69.51%. Liu et al. [36] proposed a SincNet-based hybrid neural network (SHNN) for MI-based BCIs to improve information utilization, achieving an average classification accuracy of 74.26%. She et al. [37] proposed an improved domain adaptation network based on Wasserstein distance, which utilizes existing labeled data from multiple subjects (source domain) to improve the performance of MI classification on a single subject (target domain), achieving an average classification accuracy of 77.6%. Fang et al. [38] proposed a fusion method combining Filter Banks and Riemannian Tangent Space (FBRTS) in multiple time windows to obtain more robust features, achieving an average classification accuracy of 77.7%. The comparative results are shown in Table IV.

TABLE IV. Comparison of Recent Studies

Method	***Acc***
ISMDA[35]	69.51%
SHNN[36]	74.26%
domain adaption network based on Wasserstein distance[37]	77.6%
FBRTS[38]	77.7%
Ours	79.83%

The optimal best feature combination used in this study achieved high performance on the BCI-IV-2a dataset. However, different datasets may have different parameters such as the number of electrodes, the number of subjects and the task types [39-41]. Therefore, when applying the proposed method to other publicly available datasets, parameters used in the experiment, such as the optimal feature combination, the number of training epochs for the network and the number of clusters for clustering should be adaptively adjusted further. Since deep neural networks require a large amount of data for training, some data augmentation or other methods may be required by the experimenter to avoid model overfitting if the method is to be reproduced on a smaller dataset. Moreover, the performance of the model can be further improved if more useful features are provided and optimal parameters are searched. However, as the number of features increases, combining and selecting them self-adaptively for a specific subject will be discussed in future work.

V. Conclusion

This study proposes a multi-scale CNN with a novel artifact removal strategy and spatial attention module for motor imagery recognition. By appropriately combining the selected features, it automatically removes artifacts using clustering algorithms on the components extracted by ICA, while ensuring high classification accuracy. The multi-scale convolutional blocks in SA-MSCNN, composed of different kernel sizes, extract multi-scale semantic features from the raw EEG data for classification purposes. The feature maps are then refined using a spatial attention module. The dense layer obtains the final classification results. To validate the effectiveness of this framework, the model has been applied to the BCI Competition IV-2a dataset. Compared to other existing excellent algorithms, this algorithm shows a significant improvement in classification accuracy. Experimental results demonstrate that this algorithm achieves high classification accuracy with an average accuracy of 79.83%. The current framework exhibits good classification performance and generalization. Compared to widely used EEGNet and DeepConvNet, the average classification accuracy improves by 7.39% and 7.84%, respectively. Compared to the newer state-of-the-art TS-SEFFNet, it achieves average classification accuracy improvements of 5.12%. This work also compares it with other recently published methods and the result shows the competitiveness of the proposed method. The proposed model can extract more effective features from EEG signals. This work contributes a novel method for automatic EEG artifact removal and an effective deep-learning model. It can be used to design efficient and accurate MI-based brain-computer interface frameworks to assist individuals with disabilities.

Acknowledgment

The work was supported by the Science and technology key project of Henan Province (232102210017).

Data Availability Statement

The data presented in this study are openly available at the following URL/DOI: https://bbci.de/competition/iv/.

References

[1] S. Kotchetkov, B. Y. Hwang, G. Appelboom, C. P. Kellner and E. S. Connolly Jr. “Brain-computer interfaces: military, neurosurgical, and ethical perspective.” Neurosurgical Focus, vol. 28 5, 2010, pp. E25.

[2] Xiaoqian Mao, Mengfan Li, Wei Li, Linwei Niu, Bin Xian, Ming Zeng and Genshe Chen, “Progress in EEG-Based Brain Robot Interaction Systems.” Computational Intelligence and Neuroscience, vol 2017, 2017.

[3] D. T. Bundy, L. Souders, K. Baranyai, L. Leonard, G. Schalk, R. Coker, Daniel W Moran, T. Huskey and E. C. Leuthardt, “Contralesional brain–computer interface control of a powered exoskeleton for motor recovery in chronic stroke survivors,” Stroke, vol. 48, no. 7, pp. 1908–1915, 2017.

[4] A. D. Moldoveanu, O. Ferche, F. Moldoveanu, R. G. Lupu, D. Cinteză, D. C. Irimia and C. Toader, “The TRAVEE system for a multimodal neuromotor rehabilitation,” IEEE Access, vol. 7, pp. 8151–8171, 2019.

[5] M. Staffa, M. Giordano, and F. Ficuciello, “A wisard network approach for a bci-based robotic prosthetic control,” International Journal of Social Robotics, vol. 12, pp. 749–764, 2020.

[6] R. Mane, T. Chouhan, and C. Guan, “BCI for stroke rehabilitation: motor and beyond,” Journal of Neural Engineering, vol. 17, no. 4, p. 041001, aug 2020.

[7] M. Sebastián-Romagosa, W. Cho, R. Ortner, N. Murovec, T. V. Oertzen, K. Kamada, B. Z. Allison and C. Guger, “Brain computer interface treatment for motor rehabilitation of upper extremity of stroke patients—A feasibility study,” Frontiers in Neuroscience, vol. 14, pp. 1–12, Oct. 2020.

[8] A. Biasiucci, B. Franceschiello, M. M. Murray, “Electroencephalography.” Current Biology, vol 29, 2019, pp. R80-R85.

[9] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, “Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features,” Psychophysiology, vol. 48, no. 2, pp. 229–240, 2011.

[10] M. Chaumon, D. V. Bishop, and N. A. Busch, “A practical guide to the selection of independent components of the electroencephalogram for artifact correction,” Journal of Neuroscience Methods, vol. 250, pp. 47–63, 2015.

[11] A. Hyvärinen and E. Oja, “Independent component analysis: Algorithms and applications.” Neural Networks vol.13, pp. 411-430, 2000.

[12] V. D. Calhoun, Jingyu Liu, T. Adali. “A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data.” Neuroimage vol. 45, 2009, pp. S163-S172.

[13] I. Winkler, S. Haufe and M. Tangermann. “Automatic classification of artifactual ICA-components for artifact removal in EEG signals.” Behavioral and Brain Functions vol. 7, 2011, pp. 1-15.

[14] A. M. Judith, S. B. Priya and R. K. Mahendran. “Artifact Removal from EEG signals using Regenerative Multi-Dimensional Singular Value Decomposition and Independent Component Analysis.” Biomedical Signal Processing and Control vol. 74, 2022, p. 103452.

[15] H. Varsehi, S. M. P. Firoozabadi. “An EEG channel selection method for motor imagery based brain-computer interface and neurofeedback using Granger causality.” Neural Networks, vol. 133, 2021, pp. 193-206.

[16] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordan, C. P. Hung, B. J. Lance. “EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces.” Journal of Neural Engineering vol. 15(5), 2018, p. 056013.

[17] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger and M. Tangermann, et al. “Deep learning with convolutional neural networks for EEG decoding and visualization.” Human Brain Mapping vol. 38,11 (2017): 5391-5420.

[18] Yang Li, Lianghui Guo, Yu Liu, Jingyu Liu and Fangang Meng. “A Temporal-Spectral-Based Squeeze-and- Excitation Feature Fusion Network for Motor Imagery EEG Decoding.” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, 2021, pp. 1534-1545.

[19] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, “Attention is all you need.” Advances in Neural Information Processing Systems. 2017.

[20] S. Woo, J. Park, JY. Lee, I.S. Kweon, “CBAM: Convolutional Block Attention Module.” In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. Springer, Cham.

[21] A. Galassi, M. Lippi and P. Torroni, “Attention in Natural Language Processing,” in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 10, pp. 4291-4308, Oct. 2021.

[22] G. A. Altuwaijri and G. Muhammad, “Electroencephalogram-Based Motor Imagery Signals Classification Using a Multi-Branch Convolutional Neural Network Model with Attention Blocks.” Bioengineering (Basel, Switzerland), vol. 9(7), p. 323, 2022.

[23] L. Liu, C. Shi and X. Wu, “Low Quality Samples Detection in Motor Imagery EEG Data by Combining Independent Component Analysis and Confident Learning,” 2022 21st International Symposium on Communications and Information Technologies (ISCIT), Xi'an, China, 2022, pp. 269-274.

[24] MH. Lee, OY. Kwon, YJ Kim, HK. Kim, YE. Lee and J. Williamson, et al. “EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy.” GigaScience vol. 8,5 (2019): giz002

[25] Kai Keng Ang, Zhang Yang Chin, Haihong Zhang and Cuntai Guan. “Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface,” IEEE International Joint Conference on Neural Networks. IEEE, 2008.

[26] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, “BCI competition 2008—Graz data set A,” Inst. Knowl. Discovery, Lab. Brain-Comput. Interfaces, Graz Univ. Technol., Graz, Austria, Tech. Rep., 2008, pp. 136–142.

[27] G. Dornhege, J.D.R. Mill´an, T. Hinterberger, D. McFarland, K.R. Müller, Toward brain-computer interfacing, MIT press, Cambridge MA, 2007.

[28] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier and C. Brodbeck, et al. “MEG and EEG data analysis with MNE-Python.” Frontiers in Neuroscience vol. 7 267. 26 Dec. 2013.

[29] A. Delorme and S. Makeig, “EEGLAB: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis,” Journal of Neuroscience Methods, vol. 134, no. 1, pp. 9–21, 2004.

[30] Jiazhen Hong, F. Shamsi and L. Najafizadeh. “A Deep Learning Framework Based on Dynamic Channel Selection for Early Classification of Left and Right Hand Motor Imagery Tasks.” Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference vol. 2022 (2022): 3550-3553.

[31] Z. A. A. Alyasseri, O. A. Alomari, S. N. Makhadmeh, S. Mirjalili, M. A. Al-Betar, S. Abdullah, et al., “EEG Channel Selection for Person Identification Using Binary Grey Wolf Optimizer,” in IEEE Access, vol. 10, pp. 10500-10513, 2022.

[32] Wei Mu, Tao Fang, Pengchao Wang, Junkongshuai Wang, Aiping Wang, Lan Niu, et al, “EEG Channel Selection Methods for Motor Imagery in Brain Computer Interface,” 2022 10th International Winter Conference on Brain-Computer Interface (BCI), Gangwon-do, Korea, Republic of, 2022, pp. 1-6.

[33] J. Wang, L. Shi, W. Wang and Z. -G. Hou, “Efficient Brain Decoding Based on Adaptive EEG Channel Selection and Transformation,” in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 6, pp. 1314-1323, Dec. 2022.

[34] Abdullah, I. Faye, and M. R. Islam, “EEG Channel Selection Techniques in Motor Imagery Applications: A Review and New Perspectives.” Bioengineering (Basel, Switzerland), vol. 9(12), p. 726, 2022.

[35] He Wang, Peiyin Chen, Meng Zhang, Jianbo Zhang, Xinlin Sun and Mengyu Li et al. “EEG-Based Motor Imagery Recognition Framework via Multisubject Dynamic Transfer and Iterative Self-Training.” IEEE Transactions on Neural Networks and Learning Systems, vol. PP 10.1109/TNNLS.2023.3243339. 20 Feb. 2023.

[36] Chang Liu, Jing Jin, Ian Daly, Shurui Li, Hao Sun and Yitao Huang et al. “SincNet-Based Hybrid Neural Network for Motor Imagery EEG Decoding.” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30 (2022): 540-549.

[37] Qingshan She, Tie Chen, Feng Fang, Jianhai Zhang, Yunyuan Gao and Yingchun Zhang. “Improved Domain Adaptation Network Based on Wasserstein Distance for Motor Imagery EEG Classification.” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. PP 10.1109/TNSRE.2023.3241846. 1 Feb. 2023.

[38] Hua Fang, Jing Jin, Ian Daly and Xingyu Wang. “Feature Extraction Method Based on Filter Banks and Riemannian Tangent Space in Motor-Imagery BCI.” IEEE Journal of Biomedical and Health Informatics, vol. 26,6 (2022): 2504-2514.

[39] Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon and Sung Chan Jun. (2017). “EEG datasets for motor imagery brain-computer interface.” GigaScience, vol. 6(7), pp. 1–8.

[40] Jun Ma, Banghua Yang, Wenzheng Qiu, Yunzhe Li, Shouwei Gao and Xinxing Xia. (2022). “A large EEG dataset for studying cross-session variability in motor imagery brain-computer interface.” Scientific Data, vol. 9(1), p. 531.

[41] J. Shin, A. Luhmann, B. Blankertz, DW. Kim, J. Jeong, HJ. Hwang, et al. “Open Access Dataset for EEG+NIRS Single-Trial Classification.” IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017; vol. 25(10), pp. 1735-1745.