Graph Convolution Neural Network Based End-to-End Channel Selection and Classification for Motor Imagery Brain–Computer Interfaces

Biao Sun , *Senior Member, IEEE*, Zhengkun Liu , Zexu Wu , Chaoxu Mu , *Senior Member, IEEE*, and Ting Li

Abstract—Classification of electroencephalogram-based motor imagery (MI-EEG) tasks is crucial in brain–computer interface (BCI). EEG signals require a large number of channels in the acquisition process, which hinders its application in practice. How to select the optimal channel subset without a serious impact on the classification performance is an urgent problem to be solved in the field of BCIs. This article proposes an end-to-end deep learning framework, called EEG channel active inference neural network (EEG-ARNN), which is based on graph convolutional neural networks (GCN) to fully exploit the correlation of signals in the temporal and spatial domains. Two channel selection methods, i.e., edge-selection (ES) and aggregation-selection (AS), are proposed to select a specified number of optimal channels automatically. Two publicly available BCI Competition IV 2a (BCICIV 2a) dataset and PhysioNet dataset and a self-collected dataset (TJU dataset) are used to evaluate the performance of the proposed method. Experimental results reveal that the proposed method outperforms state-of-the-art methods in terms of both classification accuracy and robustness. Using only a small number of channels, we obtain a classification performance similar to that of using all channels. Finally, the association between selected channels and activated brain areas is analyzed,

which is important to reveal the working state of brain during MI.

Index Terms—Brain computer interface (BCI), channel selection, graph convolutional network (GCN), motor imagery (MI).

I. INTRODUCTION

BRAIN–COMPUTER interface (BCI) systems that capture sensory-motor rhythms and event-related potentials from the central nervous system and convert them to artificial outputs have shown great value in medical rehabilitation, entertainment, learning, and military applications [1], [2], [3], [4]. Motor imagery (MI) can evoke SMR, which shares common neurophysiological dynamics and sensorimotor areas with the corresponding explicit motor execution (ME), but does not produce real motor actions [5], [6]. As a functionally equivalent counterpart to ME, MI is more convenient for BCI users with some degree of motor impairment who cannot perform overt ME tasks, making it important to study BCI. However, MI still faces two major challenges. First, improving the performance of MI-based classification poses a huge challenge for BCI design and development. Second, existing algorithms usually require a large number of channels to achieve good classification performance, which limits the practicality of BCI systems and their ability to be translated into the clinic.

Because of the nonstationary, time-varying, and multichannels of EEG signals, traditional machine learning methods such as Bayesian classifier [7] and support vector machine (SVM) have limitations in achieving high classification performance. Recently, deep artificial neural networks, loosely inspired by biological neural networks, have shown a remarkable performance in EEG signal classification. An et al. [8] proposed to use multiple deep belief nets as weak classifiers and then combined them into a stronger classifier based on the Ada-boost algorithm, achieving a 4–6% performance improvement compared to the SVM algorithm. A framework combining conventional neural network (CNN) and autoencoder was proposed by Tabar et al. [9] to classify feature which was transformed by short time distance Fourier transform (STFT) with more significant results. The lately proposed EEGNet [10] employed a novel scheme that

combined classification and feature extraction in one network, and achieved relatively good results in several BCI paradigms. Sun et al. [11], [12] added an attention mechanism to a CNN designed to give different attention to different channels of EEG data, achieving state-of-the-art results in current BCI applications. Although CNN models have achieved good results for MI classification, it is worth noting that traditional CNN are better at processing local features of signals such as speech, video, and images, where the signals are constantly changing [13]. CNN approaches may be less suitable for EEG signals, as EEG signals are discrete and noncontinuous in the spatial domain.

Recent work has shown that graph neural network (GNN) can serve as valuable models for EEG signal classification. GNN is a novel network that use the graph theory to process data in the graph domain, and has shown great potential for non-Euclidean spatial domains such as image classification [14], channel classification [15], and traffic prediction [16]. Cheb-Net [14] was proposed to speed up the graph convolution operation while ensuring the performance by parameterizing the graph convolution using the Chebyshev polynomials. Based on ChebNet, Kipf et al. [17] proposed the graph convolutional network (GCN) by combining CNN with spectral theory. GCN is not only better than ChebNet in terms of performance, but also highly scalable [15]. Compared with CNN models, GCN has the advantage in handling discriminative feature extraction of signals [18], and more importantly, GCN offers a way to explore the intrinsic relationships between different channels of EEG signals. GCN has been widely used in brain signal processing and its effectiveness has been proved. Some current methods based on GCN made some innovations in the adjacency matrix. Zhang et al. [19] used prior knowledge to transform the 2-D or 3-D spatial positions of electrodes into adjacency matrix. Li et al. [20] used mutual information to construct the adjacency matrix. Du et al. [21] used spatial distance matrix and relational communication matrix to initialize the adjacency matrix. However, most of the existing work has focused on the design of adjacency matrices to improve the decoding accuracy, which often requires manual design or requires a priori knowledge.

The use of dense electrodes for EEG recordings increases the burden on the subjects, it is becoming increasingly evident that novel channel selection approaches need to be explored [22]. The purpose of channel selection is to select the channels that are most critical to classification, thereby reducing the computational complexity of the BCI system, speeding up data processing, and reducing the adverse effects of irrelevant EEG channels on classification performance. The activity of brain areas still varies from subject to subject in the same MI task despite the maturity of brain region delineation. Therefore, the selection of EEG channels that are appropriate for a particular subject on an individual basis is essential for the practical application of MI-BCI. There have been some studies on channel selection, including filters, wrappers, and embedded methods [23], [24], [25]. Among these methods, the common spatial pattern (CSP) algorithm and its variants [26], [27], [28] have received much attention for their simplicity and efficiency. Meng et al. [29] measured channel weight coefficients to select channels via CSP, whose computational efficiency and accuracy cannot be satisfied at the same time. In order to solve the channel selection problem,

Yong et al. [30] used ℓ_1 parametric regularization to enable sparse space filters. It transforms the optimization problem into a quadratically constrained quadratic programming problem. This method is more accurate, but the calculation cost is high. Based on the hypothesis that the channels related to MI should contain common information, a correlation-based channel selection is proposed by Jing et al. [31]. Aiming to improving classification performance of MI-based BCI, they also used regularized CSP to extract effective features. As a result, the highly correlated channels were selected and achieve promising improvement. Zhang et al. [11] proposed to use deep neural networks for channel selection, which automatically selects channels with higher weights by optimizing squeeze and excitation blocks with sparse regularization. However, it does not sufficiently take into account the spatial information between channels.

To address the above issues, this article proposes a EEG channel active inference neural network (EEG-ARNN), which not only outperforms the state-of-the-art (SOTA) methods in terms of accuracy and robustness, but also enables channel selection for specific subjects. The main contributions are as follows:

1) An end-to-end EEG-ARNN method for MI classification, which consists of temporal feature extraction module (TFEM) and channel active reasoning module (CARM), is proposed. The TFEM is used to extract temporal features of EEG signals. The CARM, which is based on GCN, eliminates the need to construct an artificial adjacency matrix and can continuously modify the connectivity between different channels in the subject-specific situation.

2) Two channel selection methods, termed as edge-selection (ES) and aggregation-selection (AS), are proposed to choose optimal subset of channels for particular subjects. In addition, when using selected channels to train EEG-ARNN, classification performance close to that of full channel data can be obtained by using only 1/6 to 1/2 of the original data volume. This will help to simplify the BCI setup and facilitate practical applications.

3) We explore the connection between the EEG channels selected by ES and AS during MI and the brain regions in which they are located, offering the possibility to further explore the activity levels in different brain regions during MI and paving the way for the development of practical brain–computer interface systems.

The rest of this article is organized as follows: Section II introduces the EEG-ARNN model, ES and AS methods. In Section III, experimental results are presented and the relationship between the brain regions is explored. Finally, Section IV concludes this article.

II. METHODS

By simulation of human brain activation with GCN and extracting the EEG features of temporal domain with CNN, a novel MI-EEG classification framework is built in this work. As shown in Fig. 1, EEG-ARNN mainly consists of two parts: the CARM based on CNN and the TFEM based on GCN. In this section, CARM, TFEM, and the whole framework detail are described. After that, the CARM-based ES and AS methods are described in detail.
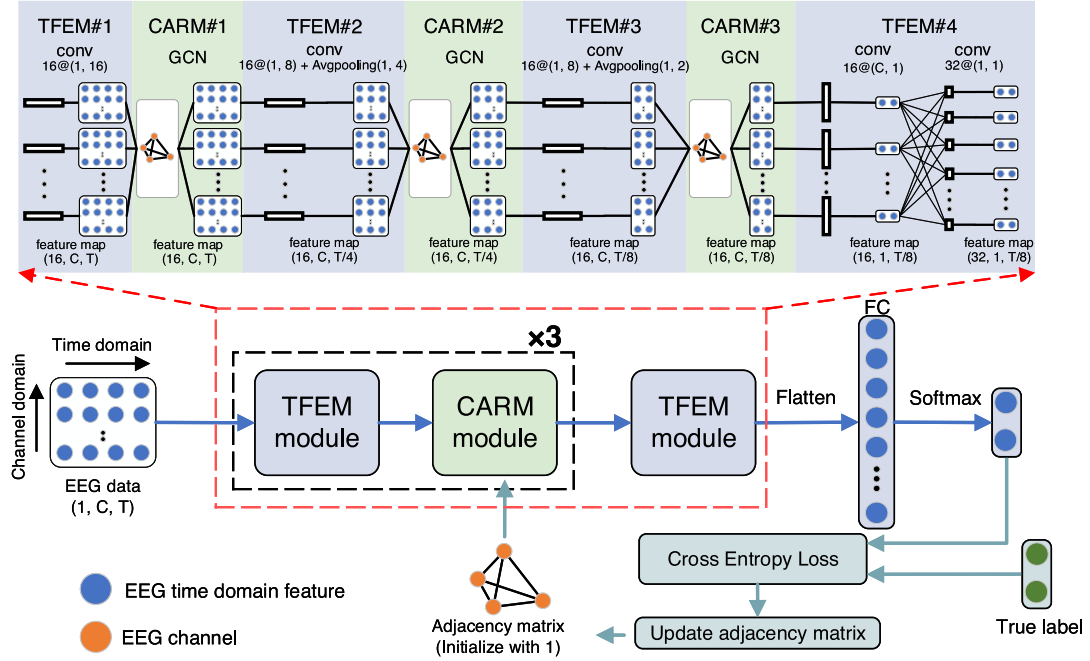
Fig. 1.   Proposed EEG-ARNN framework.

A. Channel Active Reasoning Module

GCN performs convolution operations on graph data in non-Euclidean space. The graph is defined as $\mathcal{G} = (V, E)$, where V, E represent the nodes and edges of the graph, respectively. The connection relationship between different nodes is described by the adjacency matrix $\mathbf{W} \in R^{N \times N}$. A complete EEG signal is composed of the channel and time-domain features, and the information in the EEG channel dimension is discrete and irregular in spatial distribution, so the use of graph convolution to extract features in the EEG channel dimension is important for improving model performance. Constructing an adjacency matrix between EEG channels requires access to the connectivity relationships between channels, but the complexity of the human brain's activation states during MI makes it difficult to construct an artificial adjacency matrix using existing knowledge. To address this issue, the CARM that extracts the connectivity of different channels automatically is proposed.

The Laplacian matrix of the graph $\mathcal{G}$ is defined as $\mathbf{L}$, which can be written as

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \in R^{N \times N} \tag{1}$$

where adjacency matrix $\mathbf{W} \in R^{N \times N}$ is used to represent the connection relationship between EEG channels. $\mathbf{D} \in R^{N \times N}$ is the degree matrix of graph $\mathcal{G}$. The graph Fourier transform (GFT) of a given spatial signal $\mathbf{x} \in R^N$ is expressed as

$$\widehat{\mathbf{x}} = \mathbf{U}^T \mathbf{x} \tag{2}$$

where $\widehat{\mathbf{x}}$ represents the transformed frequency domain signal. The real symmetric matrix $\mathbf{L}$ can be obtained by orthogonalizing and diagonalizing the following formula:

$$\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \tag{3}$$

where the orthonormal matrix $\mathbf{U}$ is the eigenvector matrix of $\mathbf{L}$, $\mathbf{U}\mathbf{U}^T = \mathbf{I}_N$, and $\mathbf{\Lambda} = \mathrm{diag}([\lambda, \ldots, \lambda_{N-1}])$ is a diagonal matrix whose elements on the diagonal are the eigenvalues of $\mathbf{L}$. From (3), the inverse of GFT for the spatial signal $\mathbf{x}$ is

$$\mathbf{x} = \mathbf{U}\widehat{\mathbf{x}} = \mathbf{U}\mathbf{U}^T \mathbf{x}. \tag{4}$$

Then, the graph convolution operation for the signals $\mathbf{x_1}$ and $\mathbf{x_2}$ can be written as

$$\begin{aligned}
\mathbf{x_1} *_\mathcal{G} \mathbf{x_2} &= \mathbf{U}\left(\left(\mathbf{U}^T \mathbf{x_1}\right) \odot \left(\mathbf{U}^T \mathbf{x_2}\right)\right) \\
&= \mathbf{U}\left(\widehat{\mathbf{x}}_1 \odot \left(\mathbf{U}^T \mathbf{x_2}\right)\right) \\
&= \mathbf{U}(\mathrm{diag}\left(\widehat{\mathbf{x}}_1\right)\left(\mathbf{U}^T \mathbf{x_2}\right)) \\
&= \mathbf{U}\mathrm{diag}(\widehat{\mathbf{x}}_1)\mathbf{U}^T \mathbf{x_2}
\end{aligned} \tag{5}$$

where $\odot$ denotes the Hadamard product.

Let filter function $g_\theta = \mathrm{diag}(\theta)$, the convolution operation can be written as

$$g_\theta *_\mathcal{G} \mathbf{x} = \mathbf{U}\mathrm{diag}(\theta)\mathbf{U}^T \mathbf{x}. \tag{6}$$

Let g_θ be the function $g_\theta(\mathbf{\Lambda})$ of the eigenvalue matrix of Laplace $\mathbf{L}$. Since computing the expression of $g_\theta(\mathbf{\Lambda})$ directly is difficult, the polynomial expansion of $g(\mathbf{\Lambda})$ will be replaced by a Chebyshev polynomial of order K, which can speed up the computing speed. Specifically, the largest element in the diagonal term of $\mathbf{\Lambda}$ is denoted by $\lambda_{\max}$ and the normalized $\mathbf{\Lambda}$ is denoted by $\bar{\mathbf{\Lambda}}$, i.e., $\bar{\mathbf{\Lambda}} = 2\mathbf{\Lambda}/\lambda_{\max} - \mathbf{I}_N$, by the above operation, the diagonal elements of $\bar{\mathbf{\Lambda}}$ are in the interval [-1, 1], where $\mathbf{I}_N$ is the identity matrix of dimension $N \times N$.

$g(\Lambda)$ can be approximated in the framework of K order Chebyshev polynomial as

$$g(\mathbf{\Lambda}) = \sum_{k=0}^{K-1} \theta_k T_k(\bar{\mathbf{\Lambda}}) \tag{7}$$

where θ_k is the coefficient of Chebyshev polynomials, and the Chebyshev polynomial $T_k(\mathbf{\Lambda})$ can be defined in a recursive manner as

$$\begin{cases} T_0(\bar{\mathbf{\Lambda}}) = 1 \\ T_1(\bar{\mathbf{\Lambda}}) = \bar{\mathbf{\Lambda}} \\ T_k(\bar{\mathbf{\Lambda}}) = 2\bar{\mathbf{\Lambda}} T_{k-1}(\bar{\mathbf{\Lambda}}) - T_{k-2}(\bar{\mathbf{\Lambda}}). k \geq 2. \end{cases} \tag{8}$$

According to (6) and (7), we have

$$g_\theta *_{\mathcal{G}} \mathbf{x} = \sum_{k=0}^{K} \theta_k T_k(\bar{\mathbf{\Lambda}}) \mathbf{x} \tag{9}$$

where θ_k is the coefficient of Chebyshev polynomials. With the order K of the Chebyshev polynomial set to 1 and $\lambda_{\max}$ approximated to 2, the convolution operation can be written as

$$g_\theta *_{\mathcal{G}} \mathbf{x} = \theta_0 \mathbf{x} + \theta_1 (\mathbf{\Lambda} - \mathbf{I}_N) \mathbf{x}$$
$$= \theta_0 \mathbf{x} + \theta_1 \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}} \mathbf{x}. \tag{10}$$

The above (10) has two trainable parameters, using $\theta = \theta_0 = \theta_1$ to further simplify (9), the following formulas can be obtained

$$g_\theta *_{\mathcal{G}} \mathbf{x} = \theta \left(\mathbf{I}_N + \bar{\mathbf{\Lambda}}^{-\frac{1}{2}} \mathbf{W} \bar{\mathbf{\Lambda}}^{-\frac{1}{2}} \right) \mathbf{x}. \tag{11}$$

Using the normalized $\mathbf{I}_N + \bar{\mathbf{\Lambda}}^{-\frac{1}{2}} \mathbf{W} \bar{\mathbf{\Lambda}}^{-\frac{1}{2}}$ to avoid the gradient disappearing or exploding, set $\widetilde{\mathbf{W}} = \mathbf{W} + \mathbf{I}_N$, and $\widetilde{\mathbf{D}}_{ii} = \sum_j \widetilde{\mathbf{W}}_{ij}$, so the operation of graph convolution is represented as

$$g_\theta *_{\mathcal{G}} \mathbf{x} = \theta \left(\widetilde{\mathbf{D}}^{-\frac{1}{2}} \widetilde{\mathbf{W}} \widetilde{\mathbf{D}}^{-\frac{1}{2}} \right) \mathbf{x}. \tag{12}$$

Input from the spatial domain will be extended to the spatiotemporal domain to obtain the signal $\mathbf{X} \in R^{N \times T}$, and the signal at the time point t is denoted as $\mathbf{X}_t \in R^N$. The graph convolution operation is

$$\mathbf{H}_t = \widetilde{\mathbf{D}}^{-\frac{1}{2}} \widetilde{\mathbf{W}} \widetilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}_t \Theta_t \tag{13}$$

where $\mathbf{H}_t$ is the output of graph convolution, $\Theta_t \in R^{T \times T^\ell}$ is a trainable parameter for linear transformation of the signals in the time domain. Let $\hat{\mathbf{W}} = \widetilde{\mathbf{D}}^{-\frac{1}{2}} \widetilde{\mathbf{W}} \widetilde{\mathbf{D}}^{-\frac{1}{2}}$, the graph convolution operation can be written as

$$\mathbf{H}_t = \hat{\mathbf{W}} \mathbf{X}_t \Theta_t. \tag{14}$$

It has been shown that the brain does not activate only one area during MI, but the several areas work together. In some previous studies, Sun et al. [11] proposed to construct the adjacency matrix of graph by connecting on channel to the surrounding neighboring channels in the standard 10/20 system arrangement, Zhang et al. [19] proposed to construct the adjacency matrix using the 3-D spatial information of the natural EEG channel connections. Although the abovementioned methods provide some rough descriptions of the connectivity of the brain regions,

where the EEG channels are located, they require the input of artificial prior knowledge. These static adjacency matrices do not reflect the connectivity of brain regions during MI in real-world situations on a subject-specific basis, for which the CARM initially connects one channel to all remaining channels as

$$\mathbf{W}_{ij}^* = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases} \tag{15}$$

where $\mathbf{W}_{ij}^*$ denotes the adjacency matrix of CRAM, ith and jth represent the rows and columns of $\mathbf{W}_{ij}^*$. Furthermore, the normalized adjacency matrix $\hat{\mathbf{W}}^*$ is derived using the graph convolution formula from the above. The purpose of setting up the adjacency matrix in this way is to assume that each channel plays the same role in the initial state, which is subsequently updated for $\hat{\mathbf{W}}^*$ during the training process. It is well known that back-propagation (BP) will be used to iteratively update the parameter gradients in deep neural network, and the CRAM also makes use of the BP as well. The calculation of the partial derivative of the $\hat{\mathbf{W}}^*$ is key to enabling the network to make active inference about channel connectivity relationships, and the partial derivative of $\hat{\mathbf{W}}^*$ can be expressed as

$$\frac{\partial \text{Loss}}{\hat{\mathbf{W}}^*} = \begin{pmatrix} \frac{\partial \text{Loss}}{\partial \hat{\mathbf{W}}^*_{11}} & \cdots & \frac{\partial \text{Loss}}{\partial \hat{\mathbf{W}}^*_{1N}} \\ \vdots & \vdots & \vdots \\ \frac{\partial \text{Loss}}{\partial \hat{\mathbf{W}}^*_{N1}} & \cdots & \frac{\partial \text{Loss}}{\partial \hat{\mathbf{W}}^*_{NN}}, \end{pmatrix} \tag{16}$$

where $\hat{\mathbf{W}}^*_{ij}$ denotes ith row and jth column element of $\hat{\mathbf{W}}^*$. After obtaining the partial derivative of $\frac{\partial \text{Loss}}{\partial \hat{\mathbf{W}}^*}$, $\hat{\mathbf{W}}^*$ can be updated using the following rules:

$$\hat{\mathbf{W}}^* = (1 - \rho)\hat{\mathbf{W}}^* - \rho \frac{\partial \text{Loss}}{\partial \hat{\mathbf{W}}^*} \tag{17}$$

where ρ is a scalar with a value of 0.001. Therefore, CARM gives the final formulas as

$$\mathbf{H}_t = \hat{\mathbf{W}}^* \mathbf{X}_t \Theta_t. \tag{18}$$

CARM does not require the prior knowledge of the adjacency matrix, and can also correct the connection relations between different EEG channels in the subject-specifical situation, improving the ability of graph convolution to extract EEG channel relationships.

B. Temporal Feature Extraction Module

In previous work, the amplitude–frequency features due to their high discriminability are widely used for EEG signal classification. However, the extraction of amplitude–frequency features increases the computation time of the model and may lose the information of important frequency bands. So, we design the CNN-based TFEM, which directly performs feature extraction in the time domain. There are four TFEM in our framework. The first TFEM consists of convolution, batch normalization (BN), exponential linear unit (ELU), and a dropout. The kernel size and the stride of the first TFEM are (1, 16) and (1, 1), respectively. The input data dimension is specified as $(N, C, 1, T)$, where N is the number of trials, C denotes the

---

Algorithm 1: Training Procedure of EEG-ARNN.

Input: EEG trial E; data label L; initial adjacency matrix $\hat{\mathbf{W}}^*$; parameter ρ; training epoch n

Output: Model prediction L^p; trained adjacency matrix $\hat{\mathbf{W}}^*$

1:    Initialization of model parameters
2:    $epoch = 1$
3:    **repeat**
4:      $k = 1$
5:      **repeat**
6:        Calculating the results of the k-th TFEM
7:        Calculating the results of the k-th CARM
8:        $k = k + 1$
9:      **until** k reaches to 3
10:    Calculating the results of the final TFEM
11:    Flattening the feature obtain in step 10 and calculating the predictions of the full connect layer
12:    Calculating $\frac{\partial \text{Loss}}{\partial \hat{\mathbf{W}}^*}$ using (16)
13:    Updating the model parameters include the learnable matrix

$$\hat{\mathbf{W}}^* = (1 - \rho)\hat{\mathbf{W}}^* - \rho\frac{\partial \text{Loss}}{\partial \hat{\mathbf{W}}^*}$$

14:    $epoch = epoch + 1$
15:   **until** $epoch$ reaches to n

---

number of channels, and T denotes the number of time samples. The dimension of output obtained by the first TFEM remains unchanged. Moreover, TFEM does not convolve the channel dimension, which preserves the physiological significance of the channel dimension for CARM simulations of human brain activity. The second TFEM and the third TFEM are based on the first TFEM with average pooling to preserve its global features in time domain. Note that the fourth TFEM contains two convolutions, the first with a kernel of (60, 1) and a stride of (1, 1), which is intended to fuse the EEG channel features in order to facilitate the output of the fourth TFEM into the fully connected layer.

C. Network Architecture Details

The EEG-ARNN consists of three main modules: CARM, TFEM, and a full connected layer. Except the forth TFEM, each TFEM, which extracts the EEG temporal features is connected to a CARM called TFEM-CARM block. The forth TFEM is used to compress the channel features and feed them into the full connected layer. Since Softmax activation function is applied to the output of the EEG-ARNN, the cross-entropy loss $\text{CE}(L, L^p)$ is used to measure the similarity between the actual labels L and the predictions L^p. ELU is used as activation function in both CARM and TFEM. To avoid overfitting, the Dropout is also applied in CARM and TFEM.

D. EEG Channels Selection

How to select the EEG channels which are beneficial for the MI-EEG tasks is important to BCI systems. CARM solves the
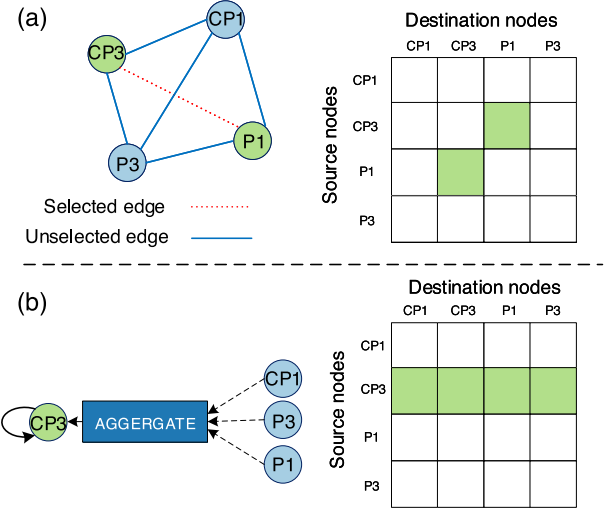

Fig. 2. Schematic representation of the results of selecting 4 channels from 64-channel EEG data using (a) ES and (b) AS methods. The corresponding adjacency matrices are illustrated as well.

problem of the lack of a priori knowledge of the graph structure constituted at the EEG channels. In addition, the dynamic adjustable adjacency matrix $\hat{\mathbf{W}}^*$ provides a description of the connection relationships between different channels. Inspired by this, we propose two graph-based channel selection methods, i.e., ES and AS. An example of ES and AS is shown in Fig. 2.

1) Edge-Selection: In the dynamic adjustable adjacency matrix $\hat{\mathbf{W}}^*$, the edge from node i to node j is designated as $e_{i,j}$, the value of edge from node i to node j is defined as $f_{i,j}$. A large edge weight indicates a strong action relationship between the EEG channels on either side of the edge, and this action relationship has a beneficial effect on the MI-EEG classification task through the adjustment of CARM. Considering the action between two nodes is mutual, we define the weight of a certain edge as

$$\delta_{i,j} = |f_{i,j}| + |f_{j,i}|, i \neq j. \tag{19}$$

where $i, j = 1, 2, \ldots, n$ and n is the number of channels. The edges with the largest value of group k in δ would be selected, and the EEG channels on both sides of the edge would be chosen, where k is the number of channels to be selected and should be set in advance.

2) Aggregation-Selection: The above ES roughly describes the strength of the connection relationship between two nodes but does not take into account the aggregating cooperation between the node and the all neighboring nodes. To circumvent this issue, AS method is brought up. For node i, the CARM aggregates the information from nodes $1, 2, \ldots, i-1, i+1, \ldots, 60$ via edges $e_{i,1}, e_{i,2}, \ldots, e_{i,i-1}, e_{i,i+1}, \ldots, e_{i,60}$, respectively, and the node's degree is taken into account as well. The ith node's information can be calculated as

$$\tau_i = \sum_{j=1}^{j=N} |f_{i,j}| + |d_i| \tag{20}$$

where d_i is the ith entry in the leading diagonal of the degree matrix. Therefore, the nodes with large τ values, representing the channels carrying more information, will be selected in AS method.

III. EXPERIMENTS AND RESULTS

A. Experimental Protocol and Data Preprocessing

TJU dataset: Experiments were conducted with 25 right-handed students (12 men and 13 women) at Tianjin University, their average age is 25.3 years (range, 19–32). None of them have personal or family history of neurological illness. Besides, participants were asked not to take psychotropic drugs two days before the experiment and to get at least 7 h of sleep the night before the experiment to avoid interference with the experiment. All procedures for recording experiments were approved by the China Rehabilitation Research Center Ethics Committee (No. CRRC-IEC-RF-SC-005-01). The EEG signals were acquired using the Neuroscan system, which consists of 64 Ag/AgCl scalp electrodes arranged according to the 10/20 system. The sampling frequency is set at 1000 Hz and can be downsampled during the preprocessing phase. Before the experiment, the electrode impedance would be tuned to below 5 kΩ through injecting conductive gel. Two of the 64 electrodes are used to detect all eye movements, and two are defined as reference electrodes. Subjects are asked to remain as still as possible throughout the experiment to avoid affecting with other movements or brain activity during experiment. During the preprocessing, the EEGLAB toolbox [32] was used to perform artifact correction, baseline correction, artifact removal, and common average referencing of the EEG data. The sampling frequency was reduced to 128 Hz and the EEG signal was bandpass-filtered at 0.5–50 Hz to eliminate powerline interference at 50 Hz and physiological noise at high frequencies. Then, the components closely related to EOG would be identified and removed by independent component analysis (ICA). Preprocessed EEG data containing 60 channels would be divided into nonoverlapping 4-s samples. Each subject participated in 320 trials, which included 160 trials involving right-hand imagery movements and 160 trials of foot imagery movements.

BCICIV 2a dataset [33]: The BCICIV 2a dataset collects EEG signals of 22 nodes recorded from nine healthy subjects. For each subject, two session of data are collected on two different days. Each session is comprised of 288 MI trials per subject. The signals were sampled with 250 Hz and bandpass-filtered between 0.5 and 100 Hz by the dataset provider before release. In our experiment, considering the fairness of comparison, left-hand movement, and right-hand movement are included in the dataset to validate the performance of the model, which results in 288 trials (144 trials × 2 sessions) per subject. The sampling rate was reduced to 128 Hz with 4 s resulting in 512 time points.

PhysioNet dataset [34]: The PhysioNet dataset contains EEG data collected from 109 healthy subjects who are asked to imagine the open and close of the left/right fist with 64 channels and a sampling rate of 160 Hz. However, due to the damaged recordings with multiple consecutive "rest" sections, the data of subject #88, #89, #92, #100 are removed. Thus, in this experiment, we have EEG data from 105 subjects, each providing approximately 43 trials, with a roughly balanced ratio of binary task. Each trial consist of 3.2 s, resulting in 512 time points. We do not perform any additional preprocessing on the EEG data.

B. Baselines and Comparison Criteria

The computer hardware resources used in this article include NVIDIA Titan Xp GPU and Intel Core I7 CPU. The proposed model is built and evaluated in PyTorch [35] and python 3.5 environments. For TJU and BCICIV 2a datasets, the data of each subject are used to train and evaluate the performance of the model separately. 10-fold cross-validation is applied to the tests of each model, and the trials are randomly divided into 10 equal-size parts. A total of nine parts are used as the training set and the remaining one part is used as the test set. The average of the classification accuracy of the 10 model test set is used as the final accuracy. For PhysioNet dataset, the data partitioning is consistent with [19], ten of the 105 subjects are randomly chosen as the test set and the rest as the training set. We run the experiments 10 times and report the averaged results.

A total of five baselines are chosen to evaluate the performance metrics of classification accuracy with the proposed EEG-ARNN, including FBCSP [36], CNN-SAE [9], EEG-Net [10], ACS-SE-CNN [11], and graph-based G-CRAM [19]. To ensure the reliability of our experiments, we set the batch size to 20 for 500 epochs in the following methods with deep learning. We use Adam optimizer with a learning rate of 0.001. The drop out rate is set to 0.25.

C. Classification Performance Comparisons

To evaluate the proposed EEG-ARNN, we first perform FBCSP, CNN-SAE, EEGNet, ACS-SE-CNN, G-CRAM, EEG-ARNN on TJU datasets of 25 subjects in sequence. The experimental results are shown in Table I. The average results of the six methods above are 67.5%, 74.7%, 84.9%, 87.2%, 71.5%, 92.3%. It is observed that the EEG-ARNN provides a 24.8% improvement concerning FBCSP, a 17.4% improvement to CNN-ASE in terms of average accuracy. Compared with these two methods, the improvement effect is significant. As for EEGNet and ACE-SE-CNN, the average accuracy improvement in EEG-ARNN is 7.4%, 5.1%. Compared with the graph-based G-CRAM method, our average accuracy improves by 17.2%. G-CRAM is designed to handle the cross-subject datasets, so the dataset size of a single subject limits the performance of G-CRAM. It is also proved that our method can deal with small datasets. Moreover, the average standard deviation (std) of 10-fold cross-validation accuracies for EEG-ARNN is 3.0%, which is less than that of FBCSP (std = 7.9%), EEGNet (std = 5.0%), CNN-SAE (std = 5.7%), ACE-SE-CNN (std = 5.0%), G-CRAM (std = 3.9%), thus proves that EEG-ARNN is quite robust in EEG recordings. Table I also illustrates the F1-score result, which indicates that the proposed model outperforms other methods. In addition, EEG-ARNN outperforms FBCSP, EEGNet, and CNN-SAE in all 25 subjects. It also performs better in 24 out of 25 subjects compared with ACS-SE-CNN

TABLE I
CLASSIFICATION ACCURACY (%), STANDARD DEVIATION (STD), AND F1-SCORE (%) RESULTS ON TJU DATASET

| Subject	Accuracy % (mean ± std) / F1-score %					
	FBCSP	CNN-SAE	EEGNet	ACS-SE-CNN	G-CRAM	EEG-ARNN
No.1	66.3±8.1 / 66.5	78.8±4.8 / 78.9	85.6±4.0 / 85.1	87.2±5.3 / 86.4	86.6±4.1 / 86.0	**95.6±4.2 / 95.5**
No.2	87.8±7.9 / 86.4	79.7±5.3 / 79.2	98.8±1.8 / 98.8	96.3±3.5 / 96.5	76.9±3.6 / 76.5	**98.8±1.9 / 98.9**
No.3	79.1±8.4 / 79.0	73.8±7.7 / 72.9	95.3±2.1 / 95.5	92.5±5.3 / 92.3	74.3±4.4 / 74.5	**95.9±4.2 / 96.0**
No.4	80.8±3.9 / 80.5	90.6±8.2 / 90.5	96.5±5.0 / 96.4	81.1±11.3 / 81.0	83.1±3.1 / 83.2	**92.2±5.6 / 92.3**
No.5	81.5±7.5 / 81.1	73.0±5.9 / 72.7	93.7±4.2 / 93.6	94.3±3.9 / 94.1	75.0±3.1 / 74.6	**98.4±1.6 / 98.5**
No.6	50.7±7.6 / 50.5	71.3±4.6 / 71.4	73.2±6.4 / 73.3	74.7±4.1 / 74.6	**87.6±3.8 / 87.7**	79.5±4.2 / 79.2
No.7	53.3±9.5 / 53.3	71.5±5.3 / 71.4	71.8±5.2 / 71.9	73.1±7.5 / 72.9	78.1±2.5 / 78.2	**79.3±3.1 / 79.4**
No.8	75.7±7.8 / 75.5	74.2±5.7 / 74.0	90.5±4.2 / 90.5	93.4±3.5 / 93.3	81.3±3.5 / 81.2	**97.8±2.4 / 97.7**
No.9	53.3±9.5 / 53.2	73.1±4.7 / 73.0	76.6±4.3 / 76.5	68.6±6.5 / 68.6	75.2±5.9 / 75.6	**78.8±3.9 / 78.7**
No.10	73.5±9.4 / 73.5	73.2±3.2 / 73.3	92.2±3.8 / 92.1	100.0±0.0 / 100.0	75.1±3.5 / 75.0	**100.0±0.0 / 100.0**
No.11	74.2±8.5 / 74.3	73.4±6.6 / 73.5	87.9±5.8 / 87.8	96.1±2.5 / 96.5	76.2±5.5 / 76.5	**99.6±1.2 / 99.4**
No.12	71.3±6.6 / 71.1	75.8±5.4 / 75.4	90.5±7.9 / 89.8	96.5±5.9 / 96.5	80.8±4.0 / 80.7	**99.5±1.4 / 99.6**
No.13	64.2±9.7 / 64.2	76.2±4.1 / 76.1	81.7±5.2 / 81.8	90.0±6.8 / 90.0	73.8±2.2 / 73.6	**99.2±1.7 / 99.0**
No.14	60.6±8.2 / 60.5	66.6±6.1 / 66.5	69.1±5.8 / 69.0	73.8±6.4 / 73.6	65.0±4.1 / 65.2	**79.4±6.1 / 79.4**
No.15	58.7±7.9 / 58.6	72.3±5.5 / 72.4	74.0±6.2 / 73.9	73.3±4.7 / 73.4	60.6±5.0 / 60.5	**80.3±2.8 / 80.4**
No.16	73.4±7.7 / 73.5	93.8±11.4 / 93.7	95.3±7.8 / 95.2	95.6±8.9 / 95.5	68.8±3.5 / 68.9	**96.6±6.3 / 96.7**
No.17	91.9±4.1 / 91.9	95.9±5.4 / 96.2	98.1±5.6 / 98.2	94.3±7.5 / 94.2	93.1±4.6 / 93.2	**98.4±3.9 / 98.5**
No.18	55.3±8.0 / 55.0	65.7±6.0 / 65.5	67.0±5.2 / 67.3	74.2±3.1 / 74.2	51.2±4.7 / 51.1	**88.7±4.3 / 88.8**
No.19	60.9±10.8 / 60.8	81.9±5.7 / 81.8	88.8±4.9 / 88.7	**92.8±3.7 / 92.9**	70.6±5.7 / 70.5	92.8±3.4 / 92.7
No.20	68.1±10.5 / 68.0	73.4±2.5 / 73.5	91.6±6.4 / 91.5	96.3±4.1 / 96.4	74.4±3.3 / 74.2	**98.4±1.6 / 98.3**
No.21	59.1±8.7 / 58.8	64.1±4.7 / 64.2	96.6±2.2 / 96.5	99.7±0.9 / 99.5	73.1±3.2 / 73.2	**99.7±1.6 / 99.6**
No.22	51.6±4.9 / 51.4	68.1±5.7 / 68.1	75.0±7.4 / 74.9	90.6±5.0 / 90.7	71.2±2.8 / 71.0	**92.8±0.9 / 92.8**
No.23	56.6±7.7 / 56.5	68.8±5.8 / 68.5	65.9±6.0 / 65.8	68.1±6.5 / 67.8	76.1±4.3 / 76.0	**76.3±2.5 / 76.5**
No.24	77.1±7.0 / 77.2	67.8±4.4 / 67.6	97.7±2.1 / 97.6	98.7±2.1 / 98.5	76.9±3.8 / 76.9	**99.7±1.0 / 99.6**
No.25	59.6±5.8 / 59.5	69.7±7.0 / 69.5	68.4±5.7 / 68.5	77.9±6.9 / 77.8	70.6±3.0 / 70.5	**88.6±3.5 / 88.5**
Average	67.5±7.9 / 67.2	74.9±5.7 / 74.8	84.9±5.0 / 84.8	87.2±5.0 / 87.1	75.1±3.9 / 75.0	**92.3±3.0 / 92.2**

The best value in each row is denoted in boldface.
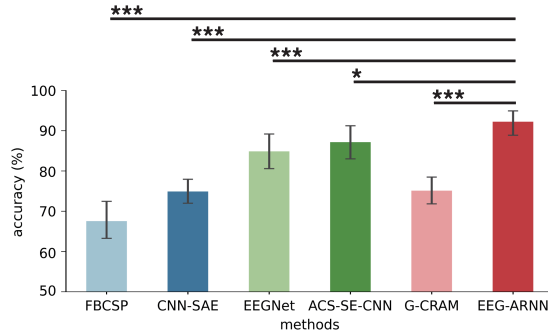
Fig. 3. Mean classification performance (%) of each algorithm averaged across all 25 subjects from the TJU dataset. *** and * above certain lines denote that the performance of EEG-ARNN was significantly better than that of the corresponding algorithm at the 0.005 and 0.1 level.

and G-CRAM. Moreover, statistical significance is assessed by Wilcoxon signed-rank test for each algorithm with EEG-ARNN as shown in Fig. 3. The results show that EEG-ARNN dominates among all algorithms in terms of average accuracy. The differences are significant except for EEG-ARNN versus ACE-SE-CNN, the EEG-ARNN performs slightly better than ACS-SE-CNN.

We also validate the performance of our proposed method on two widely used public datasets. Tables II and III illustrate the classification accuracy, standard deviation, and F1-score results of proposed and baseline methods on BCICIV 2a and PhysioNet dataset, respectively. It can be observed that the overall performance of our EEG-ARNN is also competitive on public datasets. For BCICIV 2a dataset, the average classification accuracy outperforms all other baseline methods, including traditional method, CNN-based methods, and graph-based method as well as the classification accuracy and F1-score of

more than two-thirds of subjects on EEG-ARNN are higher than other baselines. For PhysioNet dataset, as shown in Table III, the proposed method achieves the highest average accuracy and F1-score among all baseline methods. Furthermore, the average standard deviation of EEG-ARNN is lower than 4 of 5 baseline models in nine replicates. These indicate that our proposed method is also competitive on cross-subjects datasets.

D. Ablation Experiments

In this section, ablation experiments were conducted to identify the contribution of key components of the proposed method (the part inside the black dashed line in Fig. 1), the training method and parameter settings for the ablation experiments remained the same as those in Section III-B.

We considered three cases on TJU dataset, i.e., retaining TFEM or CARM only, using different number of TFEM-CARM blocks, switching the sequence of TFEM and CARM. The average classification accuracies, standard deviation, and F1-score in three cases for all subjects are illustrated in Table IV. The accuracies of the EEG-ARNN without CARM or TFEM decrease a lot compared to the proposed method. When the CARM is removed, the model loses the update mechanism on $\hat{\mathbf{W}}^*$ and the ability to make active reasoning about channel connectivity relations. The average accuracy of EEG-ARNN is 92.3%±3.0%, indicating 7.0% improvements compared to the model with TFEM only. On the other hand, when the TFEM is removed, the ability to extract temporal feature is excluded from the proposed method. It has an accuracy of 75.4%±5.1%, a decrease of 16.9% compared to the EEG-ARNN. To explore the optimal structure of the network, we evaluate the differences in results obtained using different number of TFEM-CARM blocks. Note that the model with i blocks is named as TFEM-CARM $\times i$, where $i = 1, 2, 3$. It can

Subject	Accuracy % (mean ± std) / F1-score %					
	FBCSP	CNN-SAE	EEGNet	ACS-SE-CNN	G-CRAM	EEG-ARNN
No.1	78.8±4.9 / 79.7	72.4±8.2 / 73.0	92.8±2.2 / **93.0**	81.0±5.6 / 82.2	75.8±3.2 / 76 .1	**93.2±5.4** / 92.6
No.2	59.3±9.3 / 47.8	58.6±6.8 / 57.4	91.7±3.4 / 91.7	71.3±5.6 / 71.6	76.2±2.8 / 76.8	**98.4±8.3 / 98.2**
No.3	88.4±7.2 / 88.6	85.2±5.5 / 87.3	**98.9±1.9 / 98.8**	94.1±3.5 / 94.1	87.5±4.5 / 86.2	96.2±3.3 / 96.5
No.4	70.8±6.4 / 69.5	65.5±10.0 / 63.2	96.5±1.8 / 96.3	73.9±6.3 / 74.6	69.2±3.6 / 69.3	**97.5±3.4 / 97.8**
No.5	54.4±9.0 / 47.6	66.8±8.4 / 65.5	95.9±2.0 / 95.7	69.6±5.9 / 69.5	71.3±5.9 / 72.0	**96.2±1.1 / 96.2**
No.6	65.3±5.4 / 65.6	62.1±8.0 / 63.2	94.4±2.0 / 94.4	76.3±7.1 / 77.0	73.9±4.8 / 73.4	**98.8±4.0 / 98.7**
No.7	72.6±6.9 / 71.9	63.5±7.6 / 66.3	96.4±3.5 / 96.1	75.8±6.5 / 76.7	84.2±3.8 / 83.8	**96.5±2.2 / 96.7**
No.8	93.4±5.7 / 93.2	93.5±5.1 / 92.2	**100.0±0.0 / 100.0**	93.0±4.7 / 93.2	91.4±4.7 / 91.9	99.0±2.2 / 98.9
No.9	83.7±5.4 / 83.0	72.4±7.5 / 71.3	96.3±1.3 / 96.3	95.1±3.1 / 95.2	86.4±3.9 / 86.1	**96.9±1.9 / 96.7**
Average	74.1±6.7 / 71.9	71.1±7.5 / 71.0	95.9±2.0 / 95.8	81.1±5.4 / 81.5	79.5±4.1 / 79.5	**96.9±3.5 / 96.9**

The best value in each row is denoted in boldface.

Methods	Accuracy	std	F1-score
FBCSP	59.0	**3.0**	58.7
CNN-SAE	62.8	6.4	62.5
EEGNet	69.9	3.8	70.2
ACS-SE-CNN	71.2	4.2	71.3
G-CRAM	74.2	4.5	73.8
EEG-ARNN	**82.0**	3.6	**82.1**

The best value in each column is denoted in boldface.

Methods	Accuracy	std	F1-score
TFEM × 3	85.3	3.8	84.8
CARM × 3	75.4	4.3	74.7
TFEM-CARM × 1	89.0	4.0	87.2
TFEM-CARM × 2	89.0	3.6	88.5
TFEM-CARM × 3	**92.3**	**3.0**	**92.2**
CARM-TFEM × 3	75.6	4.3	75.2

The best value in each column is denoted in boldface.

Method	Accuracy % (mean ± std)				
	top10	top20	top30	top40	all
AS-ARNN	87.9±4.3	89.3±3.6	89.8±3.6	89.7±3.7	92.3±3.0
AS-CNN	85.4±4.8	85.2±4.5	85.5±4.8	85.4±4.8	85.3±4.3
ES-ARNN	88.0±3.9	89.9±3.9	90.2±3.5	90.2±3.7	92.3±3.0
ES-CNN	75.1±3.9	85.0±4.1	85.2±4.1	85.2±4.3	85.3±4.3

be observed that even if one block is applied, the accuracy is 3.7% and 13.6% higher than the model only with TFEM and CARM, respectively. In addition, if we switch the order of TFEM and CARM (term as CARM-TFEM × 3), the accuracy drops to 75.6%, which is even lower than the model with one TFEM-CARM block.

Therefore, singular temporal or spatial feature is insufficient to describe complex physiological activities, and fewer TFEM-CARM blocks are not enough to extract effective spatiotemporal feature. Furthermore, the advantage of using TFEM and CARM alternately is to guarantee that corresponding spatiotemporal features can be extracted from the feature map at various scales, due to the fact that the neural activities of different subjects often exhibit diversified spatiotemporal interactions. The result of ablation experiments demonstrates that our EEG-ARNN is a preferable model to comprehensively leverage spatiotemporal feature for MI classification task.

E. Results of ES and AS

In order to further generalize the model, we use the trained $\hat{\mathbf{W}}^*$ to select the most important channels for BCI classification. The data obtained by channel reduction using ES and AS mentioned in Section II-D are retrained in EEG-ARNN. For this

experiment, we set four different stages (top$k$) using ES and AS, where $k = 10, 20, 30, 40$. Specifically, the EEG channels with the highest weight of k edges are selected by ES, and the k highest weighted EEG nodes are selected by the edge information aggregation capability of AS. All parameters are kept constant except for the channel of the input data to maintain the consistency of the experiment. To verify the effectiveness of the method, we also test the ES and AS using the network only with TFEM (term as CNN).

The results of AS method are shown in Table V. We observe that when the number of channels is reduced to 10, the average accuracy of the results is 87.9% ± 4.3%, which is a decrease of 4.4% compared to 60 channels data. Considering that only 10 channels are retained, the decrease is still within an acceptable range. As the number of channels increases to 20, 30, and 40, the accuracy increases to 89.3% ± 3.6%, 89.8% ± 3.6%, 89.7% ± 3.7%, a decrease of less than 3% compared to the 60 channels data, and it can be observed that the change in accuracies is not significant when the number of channels exceeds 20.

For the ES method, the average accuracy is 88.0% ± 3.9% when the 10 highest weighted edges are selected. The accuracy increases to 90.0% ± 3.9% when 20 edges are selected. When the number of selected edges reaches 30, the accuracy does not change significantly (90.2% ± 3.5%). When 40 edges are selected, the accuracy also remains at 90.2% ± 3.4%. The degradation of classification performance was not significant when comparing the four sets of experiments with channel selection to the full channel data experiments. Using EEG-ARNN, the average accuracy obtained with the 10 channels selected using the AS method is only 4.4% lower than that obtained with the full channels, still higher than the five baselines in Table I. Moreover, the amount of data is only 1/6 of the original data, implying that the channel selection process is important to save subjects' acquisition time and reduce the complexity of BCI experiments.
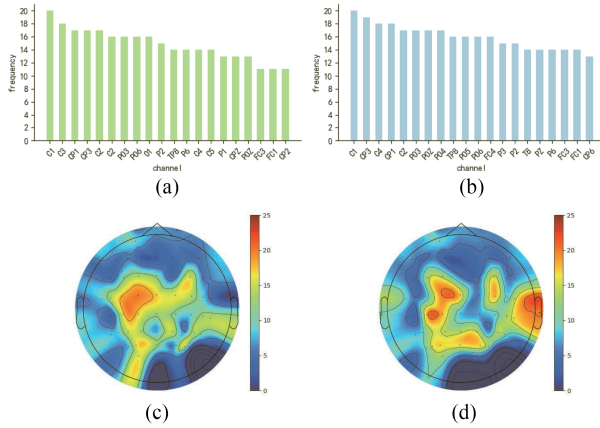
Fig. 4. Frequency and distribution of channels selected by ES/AS method among 25 subjects in TJU dataset. (a) Most frequency channel selected by ES. (b) Most frequency channel selected by. (c) Distribution of channels selected by ES. (d) Distribution of channels selected by AS.

AS method is a node-based selection method, which is a direct channel selection method. AS selects the number of channels equal to the specified value k. In contrast, ES is an indirect edge-based channel selection method. It selects the node corresponding to the largest edge of the group k at both ends, so the maximum number of nodes that may be selected is $2k$. However, since the activation regions of the brain are always similar under the fixed paradigm, leading to the case where a node is contained by several edges. In this case, we find that fewer than k nodes are selected by ES. With similar impact on classification accuracy, ES has less computational burden than AS, so ES is considered as a more efficient method.

F. Relation Between ES/AS and Neurology

To reveal which channel plays a major role in the EEG acquisition process and to explore the relationship between the brain region where the channel is located and the MI experiment, two ways to select the channels is designed in Section II-D, we further investigate what the EEG channels selected using ES and AS can indicate and whether the structures shown in Figs. 4 and 5 can match neurology concept. We first extracted the channels obtained from the top20 experiments of 25 subjects in the TJU dataset, and listed the channels selected more frequently by ES and AS methods in Fig. 4(a) and (c). Fig. 4(b) and (d) exhibit the distribution of these channels in scalp electrodes. It can be seen that "C1," "C3," "CZ," "CP1," "CP3" and other electrodes related to motor imagination are selected several times by the two methods, and some electrodes are chosen in more than two-thirds of the subjects, which indicates that the channel selection methods proposed have neurophysiological significance. Then, the edges/nodes structures of two subjects are selected and plotted using brainnetviewer [37]. According to Table I, it can be obtained that the data of the No.17 subject achieves excellent results on the six different classifiers. However, the No.23 subject has poor data quality. Based on this premise, we selected the 20 edges and 20 nodes with the highest weights following the method of Section II-D.

Methods	Accuracy % (mean ± std)	
	No.17	No.23
AS (EEG-ARNN)	98.7±3.0	73.8±4.0
AS (CNN)	98.7±3.0	73.1±6.0
ES (EEG-ARNN)	98.4±3.0	73.8±4.0
ES (CNN)	98.4±3.9	76.3±4.0

As shown in Fig. 5(a), the selected edges in No.17 subject are mainly in the left hemisphere, and the most frequent channels are "CP3," followed by "CP1," "CZ," and other channels. Human high-level senses (e.g., somatosensory, spatial sensation) are mainly performed by the parietal lobe, and electrode "CP3" is located in the parietal lobe. In the MI experiment, subjects did not produce actual movements, but only imagined movements based on cues on the screen, which required the sensation of movement. The electrode "CP3" is located in the parietal lobe, which is responsible for this sensation. For the AS selected channels shown in Fig. 5(c), the channel locations are similar to that of the channels selected using ES, with the channels mainly distributed in the left hemisphere. It is worth noting that ES selects the edge with the largest weight and then selects the EEG channels located on both sides of the edge. Therefore, the number of EEG channels selected by ES is usually less than the number of EEG channels selected by AS. For the No.17 subject, 20 nodes were selected using AS, while 11 nodes were selected using ES, but the corresponding accuracy decreased by only 0.3%, as shown in Table VI.

The channel connections of No.23 subject are shown in Fig. 5(b), with more channels located in the right hemisphere, except for the "CP3" channel, which still plays a important role. In contrast, No.17 subject selects 11 channels, which means that the distribution of channels of No.23 is more dispersed. The EEG channels selected using AS shows the same properties in Fig. 5(d), with channels mostly distributed in the right hemisphere, while a few related to the sensation of movement channels such as "FC5" and "PO3" are also selected. "C"-series channels (CZ, C1, C2,...) are mainly located in the precentral gyrus, and the neurons in this part are primarily responsible for human movements. It is obvious that most of the channels with high weights are "C"-series for No.17 subject. However, the distribution of the channels with a higher weight of No.23 subject is disorderly. The mean accuracies of No.17 and No.23 subjects are shown in Table VI. This further reveals the relationship between the selected channels of the ES/AS obtained through EEG-ARNN and the subjects performing the MI experiment. During the MI experiment, No.17 subject was energetically focused during the experiment, while No.23 subject had problems such as lack of concentration during the imagery. It can be confirmed that the vital feature of the MI is captured through the EEG-ARNN. It also demonstrates the importance of the EEG-ARNN proposed in revealing the working state of different brain regions of the subjects.
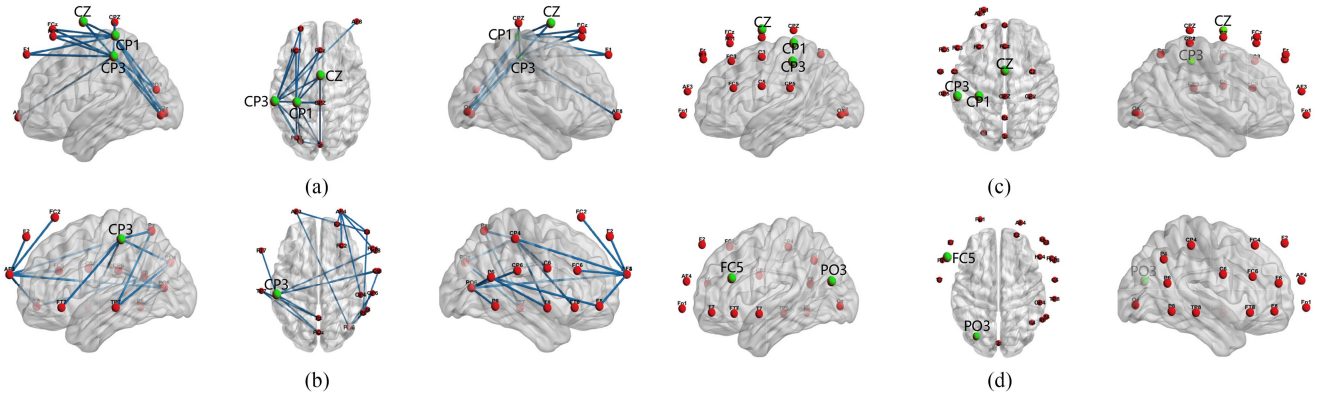
Fig. 5. Top20 edges/nodes drawn by the ES and AS method for subjects No.17 and No.23, respectively. (a) Edge-selection (Num17). (b) Edge-selection (Num23). (c) Aggregation-selection (Num17). (d) Aggregation-selection (Num23).

IV. CONCLUSION

This article proposed a novel hybrid deep framework called EEG-ARNN based on CNN and GCN for MI-EEG classification, which integrates the channel information dynamically and extracts the EEG signals in the time domain. Experimental results on three datasets showed that the proposed EEG-ARNN outperformed SOTA methods in terms of accuracy and robustness. In addition, two channel selection methods ES and AS were proposed to select the best channels. Finally, we compared the ES/AS-selected channels with active brain regions, which will help us further understand why subjects differed significantly in their performance in MI tasks.

The proposed model can be further improved by integrating convolution and graph convolution to reduce the computational complexity rather than simply stacking these two operations. In addition, the proposed method was only validated on the MI task. The future direction was to extend the EEG-ARNN to other paradigms, such as P300 and SSVEP, and continue to explore the connection relationship of channels in EEG data. Finally, it would be a meaningful work to incorporate our proposed model into a real-world BCI and evaluate its performance online.

REFERENCES

[1] B. J. Edelman et al., "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, 2019, Art. no. eaaw6844.

[2] J. Faller, J. Cummings, S. Saproo, and P. Sajda, "Regulation of arousal via online neurofeedback improves human performance in a demanding sensory-motor task," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 13, pp. 6482–6490, 2019.

[3] B. Sun, C. Mu, Z. Wu, and X. Zhu, "Training-free deep generative networks for compressed sensing of neural action potentials," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5190–5199, Oct. 2022.

[4] Y. Li, F. Wang, Y. Chen, A. Cichocki, and T. Sejnowski, "The effects of audiovisual inputs on solving the cocktail party problem in the human brain: An fMRI study," *Cereb. Cortex*, vol. 28, pp. 3623–3637, 2018.

[5] S.-Y. Dong, B.-K. Kim, and S.-Y. Lee, "EEG-based classification of implicit intention during self-relevant sentence reading," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2535–2542, Nov. 2016.

[6] S. Vyas, N. Even-Chen, S. D. Stavisky, S. I. Ryu, P. Nuyujukian, and K. V. Shenoy, "Neural population dynamics underlying motor learning transfer," *Neuron*, vol. 97, no. 5, pp. 1177–1186, 2018.

[7] M. Miao, H. Zeng, A. Wang, C. Zhao, and F. Liu, "Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: An sparse regression and weighted Naïve Bayesian classifier-based approach," *J. Neurosci. Methods*, vol. 278, pp. 13–24, 2017.

[8] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A deep learning method for classification of EEG data based on motor imagery," in *Proc. Int. Conf. Intell. Comput.*, 2014, pp. 203–210.

[9] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, 2016, Art. no. 016003.

[10] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.

[11] H. Zhang, X. Zhao, Z. Wu, B. Sun, and T. Li, "Motor imagery recognition with automatic EEG channel selection and deep learning," *J. Neural Eng.*, vol. 18, no. 1, 2021, Art. no. 016004.

[12] B. Sun, X. Zhao, H. Zhang, R. Bai, and T. Li, "EEG motor imagery classification with sparse spectrotemporal decomposition and deep learning," *IEEE Trans. Automat. Sci. Eng.*, vol. 18, no. 2, pp. 541–551, Apr. 2021.

[13] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul.–Sep. 2020.

[14] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.

[15] Z.-M. Chen, X.-S. Wei, P. Wang, and Y. Guo, "Multi-label image recognition with graph convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5177–5186.

[16] H. Zhang, Y. Song, and Y. Zhang, "Graph convolutional LSTM model for skeleton-based action recognition," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2019, pp. 412–417.

[17] Z. Diao, X. Wang, D. Zhang, Y. Liu, K. Xie, and S. He, "Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 890–897.

[18] A. Jeribi, "Spectral graph theory," in *Spectral Theory and Applications of Linear Operators and Block Operator Matrices*. Berlin, Germany: Springer, 2015, pp. 413–439.

[19] D. Zhang, K. Chen, D. Jian, and L. Yao, "Motor imagery classification via temporal attention cues of graph embedded EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2570–2579, Sep. 2020.

[20] Y. Li, N. Zhong, D. Taniar, and H. Zhang, "MutualGraphNet: A novel model for motor imagery classification," 2021, *arXiv:2109.04361*.

[21] G. Du et al., "A multi-dimensional graph convolution network for EEG emotion recognition," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, Sep. 2022, Art. no. 2518311.

[22] F. P. Such et al., "Robust spatial filtering with graph convolutional neural networks," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 884–896, Sep. 2017.

[23] F. Qi et al., "Spatiotemporal-filtering-based channel selection for single-trial EEG classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 558–567, Feb. 2021.

[24] T. N. Lal et al., "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, Jun. 2004.

[25] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, Jun. 2011.

[26] S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[27] E. A. Mousavi, J. J. Maller, P. B. Fitzgerald, and B. J. Lithgow, "Wavelet common spatial pattern in asynchronous offline brain computer interfaces," *Biomed. Signal Process. Control*, vol. 6, no. 2, pp. 121–128, 2011.

[28] J. Jin, R. Xiao, I. Daly, Y. Miao, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster-Shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4814–4825, Nov. 2021.

[29] J. Meng, G. Liu, G. Huang, and X. Zhu, "Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2009, pp. 2290–2294.

[30] X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 417–420.

[31] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, 2019.

[32] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[33] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schlogl, and G. Pfurtscheller, "BCI competition 2008–Graz data set A," *Inst. Knowl. Discov.*, vol. 16, pp. 1–16, 2008.

[34] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[35] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.

[36] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2008, pp. 2390–2397.

[37] M. Xia, J. Wang, and Y. He, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS One*, vol. 8, no. 7, 2013, Art. no. e68910.

Zexu Wu received the B.Eng. degree in control science and engineering in 2021 from Tianjin University, Tianjin, China, where he is currently working toward the Ph.D. degree in control science and engineering.

His research interests include video processing, deep learning, and brain–computer interface.

Chaoxu Mu (Senior Member, IEEE) received the Ph.D. degree in control science and engineering from the School of Automation, Southeast University, Nanjing, China, in 2012.

She was a visiting Ph.D. student with the Royal Melbourne Institute of Technology University, Melbourne, VIC, Australia, from 2010 to 2011. She was a Post-Doctoral Fellow with the Department of Electrical, Computer and Biomedical Engineering, The University of Rhode Island, Kingston, RI, USA, from 2014 to 2016. She is currently a Professor with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. Her current research interests include nonlinear system control and optimization, adaptive, and learning systems.

Biao Sun (Senior Member, IEEE) received the Diploma degree in electrical information science and technology from Central South University, Changsha, China, in 2004, and the Ph.D. degree in electrical science and technology from Huazhong University of Science and Technology, Wuhan, China, in 2013.

From 2015 to 2016, he was a Visiting Research Fellow with the Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. He is currently an Associate Professor with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His research interests include compressed sensing, machine learning, and brain–computer interface.

Zhengkun Liu received the B.Eng. degree in automation in 2021 from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, where he is currently working toward the master's degree in control science and engineering.

His research interests include brain computer interface, spiking neural network, and neural architecture search.

Ting Li was born in Wuhan, Hubei province, China, in 1982. She received the B.S., M.S., and Ph.D. degrees in biomedical engineering from Britton Chance Center for Biomedical Photonics, Huazhong University of Science and Technology, Hubei, China in 2004, 2006, 2010, respectively.

From 2010 to 2012, she got post-doc training in biomedical optics at University of Kentucky and Oregon Health & Science University, USA. Since 2012 to 2017, she has been an Associate Professor with college of microelectronics and solid electronics, University of Electronic Science and Technology of China continuing her research in electronics and medical optics. From year 2017 until now, she works as Professor with institute of biomedical engineering, Chinese Academy of Medical Sciences & Peking Union Medical College and is the Director of intelligent diagnosis and treatment laboratory, where focus is located on medical optoelectronics and clinical application, brain function research, rehabilitation robots and intelligent computation algorithms.

Dr. Li was a recipient of the Science and Technology Leading Talents in Tianjin in 2018 and Melvin H. Knisely Awardee in 2019 (1/world/year, the first awardee from China since 1983 the prize founded).