

Motor Imagery Decoding in the Presence of Distraction Using Graph Sequence Neural Networks

Shengyuan Cai¹, Student Member, IEEE, Haoran Li, Qiang Wu², Member, IEEE, Ju Liu³, Senior Member, IEEE, and Yu Zhang, Senior Member, IEEE

Abstract—In this study, we propose a graph sequence neural network (GSNN) to accurately decode patterns of motor imagery from electroencephalograms (EEGs) in the presence of distractions. GSNN aims to build subgraphs by exploiting biological topologies among brain regions to capture local and global relationships across characteristic channels. Specifically, we model the similarity between pairwise EEG channels by the adjacency matrix of the graph sequence neural network. In addition, we propose a node domain attention selection network in which the connection and sparsity of the adjacency matrix can be adjusted dynamically according to the EEG signals acquired from different subjects. Extensive experiments on the public Berlin-distraction dataset show that in most experimental settings, our model performs considerably better than the state-of-the-art models. Moreover, comparative experiments indicate that our proposed node domain attention selection network plays a crucial role in improving the sensibility and adaptability of the GSNN model. The results show that the GSNN algorithm obtained superior classification accuracy (The average value of Recall, Precision, and F-score were 80.44%, 81.07% and 80.54%) compared to the state-of-the-art models. Finally, in the process of extracting the intermediate results, the relationships between important brain regions and channels were revealed to different influences in distraction themes.

Index Terms—Attention, electroencephalogram, graph sequence neural network, motor imagery, visual and auditory distractions.

I. INTRODUCTION

ELECTROENCEPHALOGRAPH (EEG) is an important way for the decoding of brain signals that reflect the dynamic activity of neurons originating in the central nervous system and respond quickly to different brain states [1]. Motor imagery (MI) is a commonly used noninvasive brain-computer interface (BCI) paradigm [2], [3]. When a person imagines or mentally rehearses body movements, the corresponding MI response is evoked in the brain, and a large number of neuronal activities occur in the motor cortex [4]. People with disabilities can control auxiliary robots [5], [6] or wheelchairs [7] and perform daily activities by implementing MI-based BCI system, which has been shown to contribute to stroke rehabilitation [3], [8], [9]. However, recent studies found that the performance of the MI-based BCI system depends heavily on the behaviors of subjects, such as the case of performing kinesthetic MI as visual MI. Precisely, the expected EEG response for kinesthetic MI requires participants to focus on imagining motion feedback, not just imagining the picture of the action [10]. According to Friedrichetal *et al.* [11], passive auditory distraction improves the subjects' performance in MI. By contrast, Alvarsson *et al.* proposed that both positive and negative auditory interferences can lead to the extension of imagery time, and found that there is a slight difference between positive distraction and negative distraction, thus indicating that different forms of auditory interferences (natural and noisy environments) have different effects on individual well-beings and arousal levels [12]. Studies have examined the differences in brain activation patterns when eyes were open and closed by considering the effects of distraction themes. Even if these studies were conducted under the same conditions, different brain activation patterns could be found in both cases [13]. In the open state of eyes, an increase in the level of awakening and increased eye movements are observed when the attention load increases [14]. Compared with the closed state, the “external sensory” network in the open state, including the attention, eye movement, and arousal systems, are more activated. In contrast, the “endosensory” network,

Manuscript received 22 December 2021; revised 22 April 2022; accepted 5 June 2022. Date of publication 14 June 2022; date of current version 1 July 2022. This work was supported in part by the National Key Research and Development Program under Grant 2020YFC0833204, in part by the Shandong Province Key Innovation Project under Grant 2020CXGC010903, in part by the Natural Science Foundation of Shandong Province under Grant ZR2021MF056, in part by the Clinical Research Center of Shandong University under Grant 2020SDUCRCB002, and in part by the National Training Programs of Innovation and Entrepreneurship under Grant 202110422135. The work of Yu Zhang was supported by Lehigh's Accelerator, FIG, and CORE Grants. (Corresponding author: Qiang Wu.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the Charite-Universitätsmedizin Berlin under Approval No. EA4/012/12, and performed in line with the Declaration of Helsinki.

Shengyuan Cai, Haoran Li, Qiang Wu, and Ju Liu are with the School of Information Science and Engineering, Shandong University, Qindao 266237, China, and also with the Institute of Brain and Brain-Inspired Science, Shandong University, Jinan 250012, China (e-mail: bert.cai.sdu@gmail.com; lihaoran16@mail.sdu.edu.cn; wuqiang@sdu.edu.cn; juliu@sdu.edu.cn).

Yu Zhang is with the Department of Bioengineering, Lehigh University, Bethlehem, PA 18015 USA (e-mail: yuzi20@lehigh.edu).

Digital Object Identifier 10.1109/TNSRE.2022.3183023

including the visual, auditory, and sensory systems, along with some default mode networks, is more active when eyes are closed than when eyes are open [15]. Therefore, subject to the influence of different distraction paradigms, accurate decoding of MI-associated EEG patterns becomes highly challenging, and thus require more sophisticated algorithms that are robust to noise disturbances caused by distraction [16].

So far, many traditional machine learning algorithms have been used for MI EEG classification, such as linear discriminant analysis (LDA), random forest (RF), and support vector machine (SVM). However, owing to the low signal-to-noise ratio (SNR) values of the EEG signals, the original EEG data are usually contaminated by noise. Therefore, the distribution of the original EEG samples is chaotic. In particular, this situation is more evident in the case of the MI task in the presence of distraction. To solve this problem, numerous research efforts have been devoted to the development of feature extraction methods to uncover MI-driven EEG features before classification. Common spatial pattern (CSP) [3] is a popular feature extraction method for MI BCI tasks. In the past decade, many CSP variants have been proposed to optimize EEG features for improved MI decoding. For example, Ang *et al.* [5] proposed the filter bank common space pattern (FBCSP), which uses bandpass filters and the CSP algorithm to extract optimal spatial features, while maintaining high-subject sensitivity. Inspired by the idea of multisubband input, Jing *et al.* [17] proposed the Correlation-based channel selection CSP (CCS-CSP) method, to select the channels that contained more correlated information in the MI tasks. However, this algorithm depends on frequency band selection and time window selection, which can not extract the ideal spatial features from the feature space lack diversity [18], [19]. Although these approaches showed strengths in improving MI decoding, all of them have been studied under a distraction-free experimental paradigm, wherein the SNR of the EEG was guaranteed by the full engagement of subjects. However, in natural and complex environments with various distractions, use of these traditional algorithms for EEG-based MI decoding may face the following challenges:

1. Given that subjects are affected differently by the same independent distraction theme, the EEG signals of different subjects vary greatly, which hinders the generality of classifiers in specific distraction themes.

2. The topological structure of EEG channels may not be effectively utilized to learn specific EEG representations.

3. Participants may not always produce the expected imagery under certain distraction themes [20]. Consequently, the distribution of the original EEG in the collected data may be noisy, thus making it impossible to extract valid features.

To overcome these problems, researchers have proposed many end-to-end MI EEG decoding methods based on deep learning techniques. For example, Li *et al.* [21] proposed an end-to-end framework called channel projection mixed scale convolution neural network (CPMixedNet) to improve the performance of MI EEG decoding. Long- and short-term memory (LSTM) model is suitable for time series feature learning, although it ignores spatial information that is important for neural pattern decoding [22]. Wang *et al.* [23] proposed long- and short-term memory based EEG classification

(Em-LSTM) to solve the problem of gradient disappearance or gradient explosion in the training process of convolution neural networks. Generally, the method based on deep learning can achieve better performance than traditional methods because of the advantages provided by learning and embedding of feature separation and classification in a single network. However, existing EEG coding methods based on deep learning do not consider the graph structure of EEG signals. Based on the research of graph theory, there are some related work to analyze brain signals, including topological structures information in cognitive tasks and rehabilitation engineering [24], [25]. The spectral method based on the redefinition of the Fourier convolution operation on the graph can effectively extract feature information. Kipf *et al.* [26] proposed a hierarchical propagation rule, which uses the Chebyshev expansion method to simplify the approximation of the graph Laplacian [27]. The graphical attention network [28] may be used to calculate the representation of nodes in an entire community based on the attention mechanism [29]. Wu *et al.* proposed a active framework named Simplifying Graph Convolutional Networks (SGC) [30]. These methods improve the performance of graph-related tasks, while ignoring the some characteristics of low frequency in graph structures [31]. Nevertheless, none of these studies can ensure that the found input features can be reconstructed reasonably in the MI research, that is, the provided important input features are guaranteed to ensure realization of MI task in the presence of distraction. Conclusively, the previous graph convolutional networks unable to adapt to the MI task well and can not have good generalization ability. In this work, we aim to reconstruct features and interpret distracted MI pattern from the channel standpoint. Simultaneously, the fact that the subjects behave differently from each other is rooted in individual differences in brain connectivity [32]. The domain-invariant feature learning method may be used to align the features of the source and target domains to the common feature space. This alignment can be achieved by minimizing bifurcation [33] or maximizing reconstruction [34], or by adversarial training [35]. The node-wise domain adaptation (NodeDAT) regularizer proposed by Miao *et al.* [36] extends domain adversarial training into graph neural networks. By minimizing the difference between the features in the source and target domains of each node, NodeDAT achieves finer-grained regularization. In this study, we propose a novel graph sequence neural network adapted a nodal domain attention selection network for the MI task subject to the influences of various distractions. With nodal domain confrontation training, our proposed model achieves a stronger generalization ability and can effectively solve the problems of BCI blindness and poor goal realization of MI subjects. The primary contributions of this study are as follows:

1. A preprocessing regularization map algorithm is proposed, which is suitable for preprocessing EEG signals in the presence of distraction and improves the robustness of the model against noisy data.

2. A gated graph convolutional network is proposed for task related EEG decoding, which can effectively capture the essential patterns of MI subjects in distraction and decode brain regions and interchannel structures in motor imagery tasks.

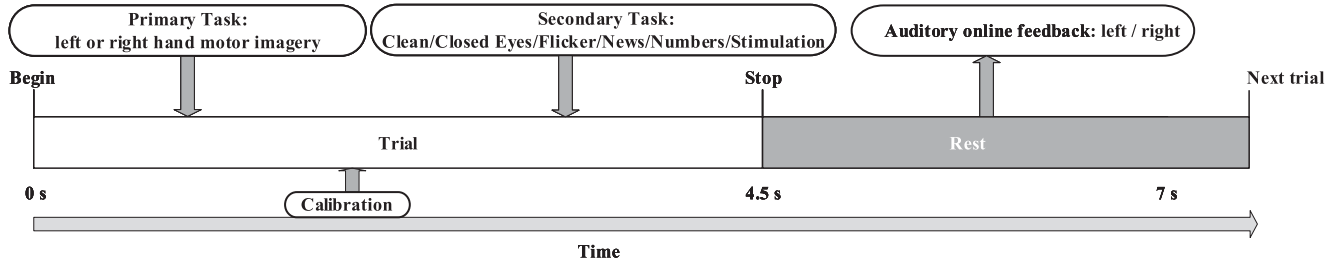


Fig. 1. Sequence of task paradigms. The participants performed left- and right-hand motor imagery tasks to the influences of different secondary distraction themes (none, clean, eye closure, flicker, news, and number).

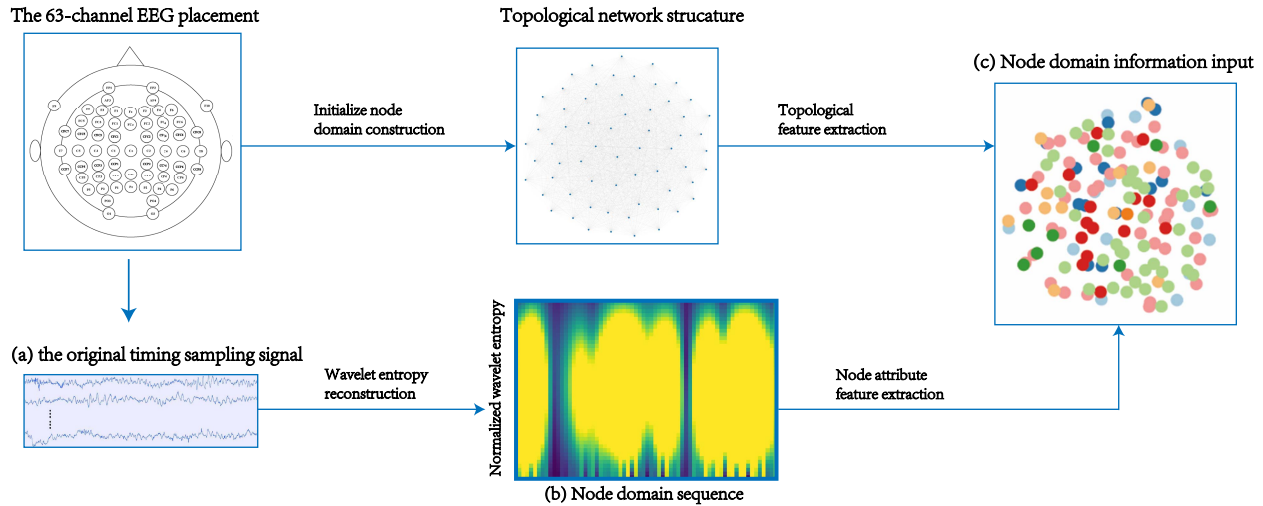


Fig. 2. Schematic showing the working principle of channel-to-graph processing conversion. (a). Baseline correction and bandpass filtering processing of the original signal. (b) Extraction of node features in wavelet domain. (c) Construction of a feature map framework.

3. A node domain adversarial training method for different independent distraction themes is proposed to improve the self adaptability and classification accuracy of graph sequence neural network (GSNN).

On Berlin Institute of Technology public datasets, our proposed model was validated with extensive experiments in comparison with benchmark models. The remainder of this paper is organized as follows. In section 2, we introduce the datasets and the baseline methods commonly used for graph representation learning and our designed GSNN network. Section 3 presents the experimental results of the GSNN. Section 4 discusses some detail analysis and potential applications.

II. MATERIALS AND METHODS

A. Dataset

A public dataset collected by the Technical University of Berlin from 16 healthy subjects between the ages of 24 and 28 was used for our study [37]. These participants were assigned to exercise imagery experiments in the presence of different independent distraction themes. For each subject, the main experiment was divided into seven runs for 10 min, and 72 secondary experiments were run each time; one experiment

trial lasted 4.5 s (as shown in Fig. 1). The participants were asked to choose a tactile hand movement subject to the influence of six secondary tasks (clean, eye-closed, news, numbers, flicker, stimulation). There are several MI strategies for participants to choose from. For example, squeezing a soft ball, turning on the faucet, playing the piano, and using a saltshaker. The experiment used the Fast EasyCap (Easy-CapGmbH) system, which consisted of 63 wet Ag/AgCl electrodes placed in a symmetrical position according to the International 10-20 system [38], with reference to the nose. Moreover, two 32-channel amplifiers were used to amplify the signal at a sampling rate of 100 Hz. The original EEG signal of the subjects was recorded in detail in the dataset and was applied to the main task tag of each subject to record whether the task was left-handed or right-handed. This study was conducted according to the declaration of Helsinki and was approved by the Ethics Committee of the Charite-Universitätsmedizin Berlin (approval number: EA4/012/12).

To cross-compare outcomes with other models, we adopted the same preprocessing method as the original dataset, sampled the EEG data at 100 Hz (as shown in Fig. 2 (a)), calculated the Laplace filter of C3 and C4, and used the fifth-order Butterworth filter to filter the data in the range of 9-13 and 18-26 Hz [37].

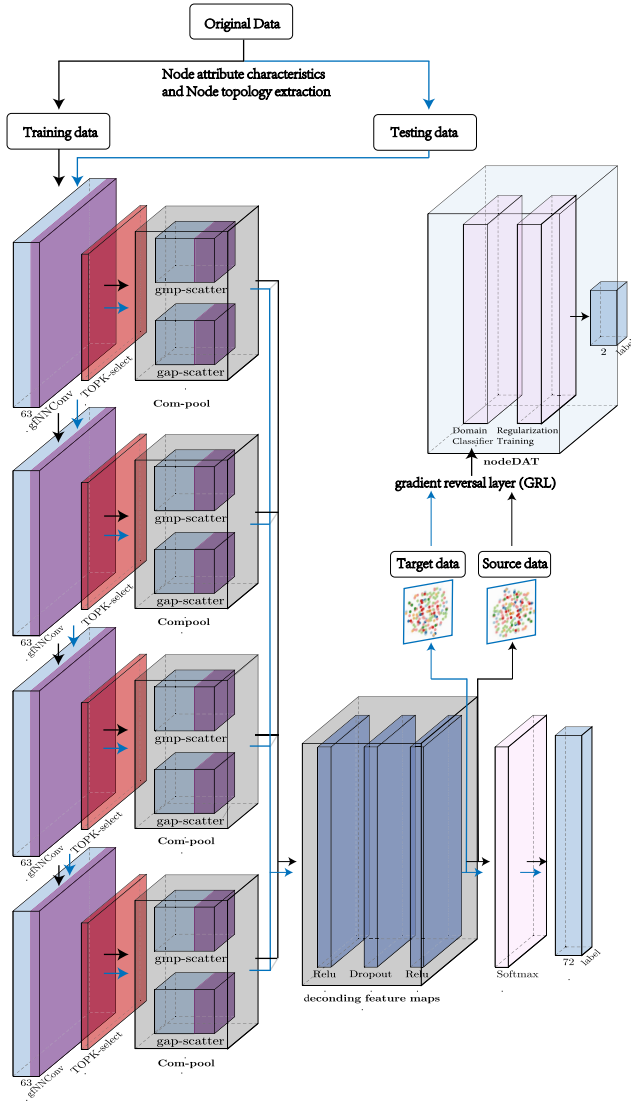


Fig. 3. Architectural design of the proposed graph sequence neural network (GSNN). The structure is a shrinking form, that is, the GSNN network inputs 63×128 node attribute features, and 63×63 node connection features. Finally, GSNN network outputs a class label through softmax to evaluate the model.

B. Overall Structure of Framework

In this study, we propose a novel framework, known as the graph sequence neural network (GSNN), which adapts the graph convolution network with self-attention mechanism and node domain adversarial, to predict the downstream task results of MI in the presence of distractions. The overall structure of our proposed framework is illustrated in Fig. 3. Specifically, the framework consists of three modules:

1. Graph reconstruction module for establishing a brain channel based graph (cf. Section II-C).
2. Gated graph convolutional network module for task-related EEG decoding which captures the essential patterns of MI subjects in distraction (cf. Section II-D).
3. Node domain adversarial training method for improving the classification accuracy in each task which provides the self adaptability of GSNN from a brain channel perspective (cf. Section II-E).

C. Graph Reconstruction

We use the notch filter to filter power frequency interference [38], and use wavelet entropy reconstruction as the node attribute feature. (as shown in Fig. 2) The dynamic distance ratio is reconstructed as the node connection parameter to construct the adjacency matrix of the edge connection.

1) Node Attribute Characteristics:

We reconstruct the signal based on the time-domain sequence signal after preprocessing. According to Rosso *et al.* [39], wavelet entropy can measure the degree of order and disorder of the signal and provide potential dynamic process information related to the signal. It selects the appropriate wavelet basis based on the wavelet transform with multiple scales and directions and expands the original signal at different scales [22]. Therefore, we used wavelet entropy reconstruction in each channel as the nodal attribute feature of the EEG channel. The relative wavelet energy of j th signal at each scale is defined as follows:

$$p_{hj} = \frac{E_h}{E_{total}} = \frac{\sum_{k=1}^{L_h} |C_h(k)|^2}{\sum_h \sum_{k=1}^{L_h} |C_h(k)|^2} \quad (1)$$

where $C_h(k)$ is the decomposition coefficient of each scale h , L_h denotes the number of coefficients at each decomposition scale, E_h is the wavelet energy, and E_{total} is the total energy. The wavelet entropy is calculated by the following formula:

$$X_j = -p_{hj} \log(p_{hj}) \quad (2)$$

where $X_j \in \mathbb{R}^{1 \times 128}$ is one of the element of $X \in \mathbb{R}^{N \times F}$, which is the input features of the graph with 63 nodes and 128-dimensional features of wavelet entropy (as shown in Fig. 2 (b)). Additional details are provided in [22].

2) *Node Topology Reconfiguration*: The topological structure of EEG channels is necessary for graphic representation learning. The adjacency matrix A represents the topological pattern, where f represents the number of channels in EEG signals. The weight between channels m and n is indicated by each learnable entry A_{mn} . According to Salvador *et al.* [20], the strength of connections between brain regions attenuates as an inverse square function of physical distance. Therefore, we initialized the relationship between local channels in the adjacency matrix as follows,

$$A_{mn} = \min\left(1, \frac{\delta}{q_{mn}^2}\right), \quad (3)$$

where $\delta > 0$ denotes a calibration constant, q_{mn} , $m, n = 1, 2, \dots, f$, denotes the physical distance between channels m and n , which is computed from their three-dimensional coordinates obtained from the data sheet of the recording device. Achard *et al.* [40] observed that sparse functional magnetic resonance imaging networks (~20% of all possible connections) can typically maximize the efficiency of network topology. Therefore, Setting δ to 5 implies that ~20% of the entries in A are non-negligible. Finally, to reduce overfitting, we set A_{mn} as a symmetric matrix.

D. Gated Graph Convolutional Network

1) *Graph Filter Neural Network*: We consider a graph $\mathcal{G} = (\mathcal{X}, \mathcal{A})$, where \mathcal{X} denotes a set of nodes and \mathcal{A} denotes a set of edges between nodes in \mathcal{X} . If the graph convolution network (GCN) formula [26] is used, the form of score $G \in \mathbb{R}^{N \times 1}$ is as follows.

$$G = \sigma(\tilde{D}^{-\frac{1}{2}} A \tilde{D}^{-\frac{1}{2}} X \Theta_{att}) \quad (4)$$

where σ is the activation function (e.g. *sigmoid*), $A \in \mathbb{R}^{N \times N}$ is the adjacency matrix with self-connections, which can represent the edge set \mathcal{A} . $\tilde{D} \in \mathbb{R}^{N \times N}$ is the degree matrix of A after renormalization trick, $X \in \mathbb{R}^{N \times F}$ is the input features of the graph with N nodes and F -dimensional features, which can represent data on \mathcal{X} . $\Theta_{att} \in \mathbb{R}^{F \times 1}$ is the only parameter of the convolution-layer. In general, GCNs learn a feature transformation function for X and produces the output $G \in \mathbb{R}^{N \times F'}$, where F' denotes the output feature dimension. However, when the noise is so large that GCN cannot be reduced by the first-order low-pass filter, it has a risk of overfitting to the noise. In particular, the inner weight Θ_{att1} and A may be trained using high noisy features.

According to Hoang *et al.* [31], graph filter neural network (gfNN) is a faster structure compared with GCN after the correction of the convolution layer, which is more suitable for the filtering of noisy signals. The form of multilayer (two-layer) perceptrons is as follows,

$$\begin{aligned} G_{gfNN} &= h_{MLP}(X | \Theta_{att1}, \Theta_{att2}) \\ &= \sigma_2(\sigma_1(H \Theta_{att1}) \Theta_{att2}) \end{aligned} \quad (5)$$

where σ_1 is the entry-wise rectifier linear unit function, and σ_2 is the softmax function. Note that both σ_1 and σ_2 are contraction maps. By using the graph convolution kernel, $H = \tilde{D}^{-\frac{1}{2}} A \tilde{D}^{-\frac{1}{2}} X$.

2) *Self-Attention Pooling*: Pooling layer enables the gfNN model to lessen the number of parameters by reducing the scale of representation, thus avoiding overfitting. According to Zhang *et al.* [41], SortPool sorts the embedding of nodes and feeds the sorted embedding to the next layer. Conversely, Ying *et al.* [42] proposed DiffPool, which is a hierarchical graph pooling method for end-to-end learning of allocation matrices.

Inspired by these studies, self-attention pooling is divided into two parts: global and hierarchical pooling methods [43]. The global pooling method considers the graphic function and uses neural networks to aggregate all representations of nodes in each layer. Diagrams with different structures can be processed because the global pool method collects all representations. The hierarchical pooling method is used to construct a model with node assignments based on features or topologies that can be learnt in each layer. $P^{(l)} \in \mathbb{R}^{N_l \times N_{l+1}}$ contains the probability values of nodes in layer l , which is assigned to clusters in the next layer $l+1$. A denotes an assignment matrix which updates in layer $l+1$ based on $P^{(l)}$. N_l denotes the number of nodes in layer l . Specifically, the form of assigned nodes is as follows:

$$P^{(l)} = G_{gfNN}^l(A^{(l)}, X^{(l)}) \quad (6)$$

$$A^{(l+1)} = P^{(l)\top} A^{(l)} P^{(l)} \quad (7)$$

Top $[kN]$ nodes are selected based on the value of G_{gfNN} . The pool ratio $k \in (0, 1]$ is a super parameter used to determine the number of nodes to be retained. The select score is calculated as follows

$$idx = top - rank(G, [kN]), \quad G_{mask} = G_{idx} \quad (8)$$

where idx is an indexing operation and G_{mask} is the feature attention mask. The listing top-rank is the function that returns the indices of the top $[kN]$ values. The select module form is as follows

$$\begin{aligned} X^{mask} &= X_{idx} \\ \text{Com-pool}(X_{out}) &= X^{mask} \odot G_{mask} \\ \text{Com-pool}(A_{out}) &= A_{idx, idx} \end{aligned} \quad (9)$$

where X_{idx} is the nodes matrix by node-wise indexing. For node pooling, indexed nodes are embedded into graph convolution mask in the sparse space. $A_{idx, idx}$ is adjacency matrix which is pooled by row-wise and col-wise indexing. Finally, where X denotes the node feature matrix. The output probability of class \mathbf{Y}_i can be computed based on passing each feature matrix X_i into model. The form is as follows

$$p(\mathbf{Y}_i | X_i, \theta) = \text{softmax}_{\mathbf{Y}_i}(\text{Com-pool}(\sigma(G_i)) \Theta_{att}^O) \quad (10)$$

where Com-pool denotes the combined pooling across all nodes on the graph, $\mathbf{Y}_i \in \{0, 1, \dots, L-1\}$ denotes the label class index, L denotes the number of classes, and $\Theta_{att}^O \in \mathbb{R}^{F' \times L}$ denotes the output weight matrix.

E. Node Domain Adversarial Training

When participants conduct the main task of prompting, owing to the influence of distraction, they may not always produce the expected MI, which may have a negative impact on the performance of the model [38]. Inspired by the node-wise domain distributed learning method (NodeDAT) [36], it is possible to learn the distribution of classes, which can solve the influence of different subjects to a certain extent. Specifically, the model converts each training label $\mathbf{Y}_i \in \{0, 1, \dots, L-1\}$ into a prior probability distribution of all classes $\hat{\mathbf{Y}}_i \in \mathbb{R}^L$, where $\hat{\mathbf{Y}}_{iL}$ denotes the probability of class L in $\hat{\mathbf{Y}}_i$. We have migrated this method to make it suitable for the task of motor imagery under distraction. Label has two classes: left and right class indices 0, 1, respectively. We convert \mathbf{Y} as follows

$$\hat{\mathbf{Y}}_i = \begin{cases} (1 - \frac{\epsilon}{2}, \frac{\epsilon}{2}), & \mathbf{Y}_i = 0 \\ (\frac{\epsilon}{2}, 1 - \frac{\epsilon}{2}), & \mathbf{Y}_i = 1 \end{cases} \quad (11)$$

Subsequently, the model establishes two domains, which are labeled as source/training data $X^O \in \mathbb{R}^{B \times N \times F}$ (in this subsection, we denote X by X^O for better clarity) and unlabelled target/testing data $X^E \in \mathbb{R}^{B \times N \times F}$, where in practice X^E can be either oversampled or downsampled to have the same number of samples as X^O [9], B denotes the number of training samples. The domain classifier aims to minimize the sum of the two binary cross-entropy losses,

$$\Phi_D = - \sum_{i=1}^B \sum_{j=1}^N (\log(p_j(0|X_i^O, \theta_D)) + \log(p_j(1|X_i^E, \theta_D))), \quad (12)$$

where 0 and 1 denote source and target domains respectively, θ_D denotes the parameters of the domain classifier. Intuitively, the domain classifier learns to classify source data as 0 and target data as 1. O is the labelled target, E is the unlabelled target. p_j is the domain probability of the j th node calculated by Eq.(5). To make the domain classifier less sensitive to noise input in the early stage of the training process. Zhong *et al.* implemented gradient reversal layer (GRL) in domain training. With the training, the scaling factor β gradually increases from 0 to 1, thus further scaling the inverse gradient. Additional details are provided in [36].

F. Dynamics of GSNN

The characteristics of the reconstructed EEG data from pre-processing regularization map algorithm are used to train the GSNN network (as shown in Fig. 2 (c)). Firstly, gfNN realizes the denoising of node features in a high-dimensional space. After 63-dimensional gfNN convolution, the self-attention parameter G of node selection is obtained. According to the node selection method [44], even when inputting graphs of different sizes and structures, part of the nodes of the input graph are retained. Therefore, we used self-attention pooling structures to preserve graphic features and topologies. Respectively, the global pooling and hierarchical merging were executed. The label output of the sparse node domain was used as the input of nodeDAT. That is, the source domain label and the target domain label are separated as much as possible in the node domain classifier. Graph convolutional feature extractor is combined with a domain classifier for adversarial training (using gradient reversal layer (GRL)). Finally, the model fuses the testing data features to achieve generalization of feature extraction. Through the regularization of the smooth node label, the embedding is conducted by using 63 nodes, and the subsequent round is executed after supervised evaluation (as shown in Fig. 3). The optimization of the model is determined by the following loss function, including graph network and nodeDAT.

$$\Phi' = \sum_{i=1}^B \text{KL}(p(\mathbf{Y}|X_i, \theta), \hat{\mathbf{Y}}_i) + \alpha \|A\|_1, \quad (13)$$

where θ denotes the model parameters that need to be optimized, and α denotes the weight of L1 sparse regularization for adjacency matrix A . $p(\mathbf{Y}|X_i, \theta)$ denotes the output probability distribution computed via Eq.(10). By combining both NodeDAT and graph sequence network, the overall loss function Φ'' of GSNN is as follows,

$$\Phi'' = \Phi' + \Phi_D. \quad (14)$$

The detailed algorithm for training GSNN is presented in Algorithm.1. The entire neural network is fine-tuned after GSNN training.

III. EXPERIMENTAL RESULTS

A. Experiment Protocols

According to Glorot *et al.* [45], even if the evaluation is conducted on the same database, different evaluation schemes

Algorithm 1 The Training Process of Graph Sequence Neural Network (GSNN)

Input: Number of epochs E , training samples X and $\hat{\mathbf{Y}}$, learning rate η , batch size B , unlabelled testing samples X^E , other super-parameters;

Output: The parameters in learned GSNN model;

- 1: Randomly initialize parameters in GSNN model;
 - 2: Initialize adjacency matrix A based on Eq.(3);
 - 3: **for** $i = 1: T$ **do**
 - 4: **repeat**
 - 5: Select training samples X_B and $\hat{\mathbf{Y}}_B$ in one batch from X and $\hat{\mathbf{Y}}$
 - 6: Select testing samples X_B^E in one batch from X^E ;
 - 7: Compute degree matrix D based on Eq.(4)
 - 8: Compute normalized adjacency matrix A based on Eq.(7);
 - 9: Compute output representation G based on Eq.(5);
 - 10: Select top-nodes to represent the graph sparsely G based on Eq.(9);
 - 11: Use X_B and $\hat{\mathbf{Y}}_B$ to compute KL loss Φ' based on Eq.(13);
 - 12: Use X_E and X_B^E to compute domain loss Φ_D based on Eq.(12);
 - 13: Compute gradient reversal layer scaling factor β ;
 - 14: Update $\Theta_{att}^D \leftarrow \Theta_{att}^D - \eta \frac{\partial \Phi_D}{\partial \Theta_{att}^D}$;
 - 15: Update $\Theta_{att}^O \leftarrow \Theta_{att}^O - \eta \frac{\partial \Phi'}{\partial \Theta_{att}^O}$;
 - 16: Update $\Theta_{att} \leftarrow \Theta_{att} - \eta (\frac{\partial \Phi'}{\partial \Theta_{att}} - \beta \frac{\partial \Phi_D}{\partial \Theta_{att}})$;
 - 17: Update $A \leftarrow A - \eta (\frac{\partial \Phi'}{\partial A} - \beta \frac{\partial \Phi_D}{\partial A})$;
 - 18: **until** all samples in X have been drawn;
-

can lead to significant differences in results. Generally, the influence of verification methods on test results proceeds as follows: (1) compared with theme-independent evaluation methods, subject-independent evaluation methods are more objective because subjects rely on evaluation methods that disregard individual differences. (2) Subject-level k-fold cross-validation (CV) can evaluate better the sensitivity of the model among different subjects. (3) Independent distraction theme-level k-fold CV can evaluate better the adaptive adjustment ability of the model at different secondary tasks.

Therefore, to avoid data leakage in the evaluation process, we conducted six-fold CV topic independent evaluation protocol under the independent distraction theme of the database. The training and test data used to evaluate the individual difference sensitivity of the model were produced from the same subjects, and there is no information overlap between the subjects. Specifically, at each independent distraction theme, we cross-validated 16 subjects at six-fold CV, and repeated the verification process until each secondary task of subjects was treated as a test dataset. In other words, for each independent distraction theme, we repeated it 16 times and calculated the final CV performance as the average of all the test results obtained. This verification method provides a fair evaluation of the performance of the cross-object model and can accurately

TABLE I

MOTOR IMAGERY PERFORMANCE (%) COMPARISON WITH DIFFERENT NETWORK CONFIGURATIONS USING ALL INDEPENDENT DISTRACTION THEMES ON THE BERLIN-MOTOR IMAGERY DATABASE. THE STATISTICAL SIGNIFICANCE OF PERFORMANCE DIFFERENCE BETWEEN OUR METHOD AND OTHER APPROACHES WERE EXAMINED USING PAIRED SAMPLE T-TEST

Precision (mean+std)								
Model	calibration	clean	eyes	flicker	news	numbers	stimulation	p-value
FBCSP [5]	70.16±13.97	79.02±14.22	69.05±19.60	70.11±19.21	71.03±15.45	58.33±19.74	60.39±20.32	0.0347
CCS-CSP [19]	71.11±14.05	79.19±13.37	69.52±18.71	70.27±20.54	73.81±20.91	59.78±20.23	61.75±20.11	0.0231
Em-LSTM [23]	53.35±18.31	55.12±19.71	52.68±18.59	55.36±19.77	56.95±23.02	48.26±22.64	50.99±23.09	0.0162
CPMixedNet [21]	71.80±12.93	77.37±15.76	71.74±15.62	70.87±15.31	72.4±11.13	70.13±15.73	69.72±14.71	0.0153
GSNN(Our model)	78.81±9.01	82.14±9.69	79.31±9.59	82.41±9.34	80.40±10.19	79.59±8.91	78.35±8.97	-
F-score (mean+std)								
Model	calibration	clean	eyes	flicker	news	numbers	stimulation	p-value
FBCSP [5]	69.08±14.17	79.13±11.16	69.81±20.19	71.13±20.17	71.06±15.89	59.14±19.83	58.3±20.94	0.0311
CCS-CSP [19]	69.21±14.37	79.15±11.29	70.62±19.02	71.94±20.62	73.81±19.93	60.32±19.78	59.01±20.01	0.0216
Em-LSTM [23]	53.2±19.24	54.85±21.11	52.51±19.13	54.92±19.75	55.66±22.52	46.26±22.39	49.85±21.05	0.0110
CPMixedNet [21]	71.36±13.76	77.5±15.84	72.23±15.37	70.31±15.43	72.41±11.04	69.48±15.70	69.8±14.69	0.0157
GSNN(Our model)	79.24±9.38	82.89±9.81	79.51±9.53	82.65±9.11	79.44±10.16	79.62±8.82	78.36±8.89	-
Recall (mean+std)								
Model	calibration	clean	eyes	flicker	news	numbers	stimulation	p-value
FBCSP [5]	68.14±14.44	79.07±14.22	68.73±19.78	71.29±20.04	71.74±15.92	59.07±20.43	58.85±20.32	0.0331
CCS-CSP [19]	69.42±14.19	79.29±13.37	70.16±20.76	71.99±20.27	73.63±21.16	61.79±20.84	61.74±20.08	0.0205
Em-LSTM [23]	53.34±19.9	55.07±21.02	52.63±18.57	55.21±19.81	56.42±21.62	50.21±21.76	50.92±21.11	0.0113
CPMixedNet [21]	70.31±13.82	77.37±15.89	72.54±15.52	70.55±15.22	72.91±11.16	69.66±15.64	69.86±14.74	0.0142
GSNN(Our model)	78.31±9.94	82.14±9.40	79.79±9.59	82.44±9.01	79.13±10.18	79.33±8.91	78.41±8.89	-

estimate the possible recognition accuracy of new data from new objects. To compare findings with the results of previous studies, we used four indicators, namely accuracy (P_{acc}), recall (P_{re}), precision (P_{pre}), and F1-score (P_f), to evaluate the performance of the model. The F1-score was introduced as a comprehensive index to balance the influence of accuracy and recall and to evaluate the classifier in a more comprehensive manner.

B. Network Training

In the process of training, according to Shchur *et al.* [46], different data segmentations affect the model performance of graph neural network (GNN). In our experiment, the weight parameters in the convolution layer were initialized with 20 random seeds and six-fold CV. In a round of cross-validation in the distraction database, samples from 16 subjects \times 12 secondary experiments were randomly used as test data (a total of 16 subjects \times 12 secondary tasks \times 63 nodes \times 1953 edges), and samples from other 16 subjects \times 60 secondary experiments were used as training data (16 subjects \times 60 secondary tasks \times 63 nodes \times 3969 edges). To reduce the time complexity of the computation, the momentum of the ADAM optimizer was 0.9. If the verification loss was not improved after 50 epoch intervals under the epoch termination condition (up to 100 k epochs), we stopped training. The model weight that generated the lowest loss of the validation set was saved as the final parameter. A small batch

random gradient descent method was adopted with a fixed learning generator rate of 0.001 and a discriminator of 0.0002. Herein, the small batch size was equal to 12. All models were trained on NVIDIA GeForce RTX 2080 graphics processing unit (GPU), and CUDA 11.0 used Pytorch API.

C. Evaluation on Database

Table I presents a comparison of the results of the proposed method with those of the previous studies. The corresponding verification methods are elucidated. The experimental results show that, compared with other supervised methods, the self-attention framework (based on GSNN) yields a better performance of MI recognition. For the supervised method proposed in [21], the initial evaluation scheme is based on the CV of the subjects under all independent distraction themes, and the performance of the cross-subjects is separated. There is a fair comparison between our proposed model and other methods. Based on [21], we used six-fold CV to evaluate their work conducted at each distraction theme (the same method we proposed) and report the results in Table I. The compared results show that the proposed method is better than that reported by Darvish *et al.* [19]. In the case of the theme “calibration”, the valence of Recall increases from 70.31% to 78.31%, Precision from 71.80% to 78.81%, F-score from 71.36% to 79.24%; In the case of the theme “clean”, the valence of Recall increases from 77.37% to 82.14%, Precision from 77.37% to 82.14%, F-score from 77.50% to 82.89%. In the cases of the

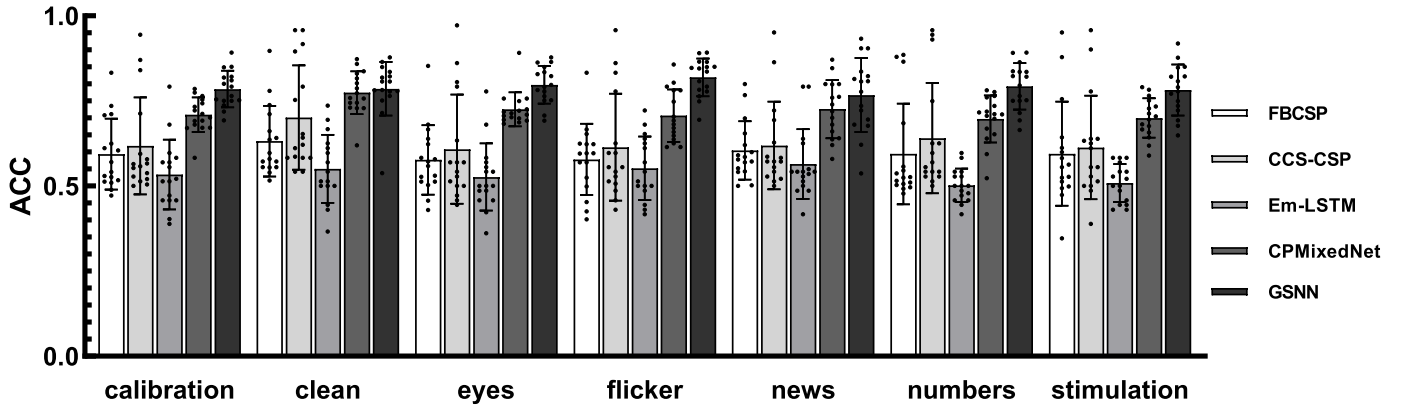


Fig. 4. Results show that our proposed method achieves a performance comparable to the other supervised methods, whereby the recognition accuracies (ACC) of calibration, clean, eye close, flicker, news number, and stimulation are 78.53 %, 82.41 %, 79.79%, 81.91%, 79.69%, 79.32%, and 78.35%, respectively. Each point represents the accuracy of a subject performing a motor imagery task at an independent distraction theme.

TABLE II
PERFORMANCES (%) OF GRAPH SEQUENCE NEURAL NETWORK AT INDEPENDENT CONDITIONS

Subjects	Accuracy-GSNN						
	Calibration	Clean	Eyes	Flicker	News	Numbers	Stimulation
P1	81.05	80.67	75.24	85.05	83.62	89.24	85.05
P2	69.33	53.81	69.24	74.86	62.10	66.48	77.81
P3	85.05	78.10	76.38	86.57	68.00	79.52	80.76
P4	71.90	83.52	83.43	77.81	81.43	86.38	91.90
P5	75.43	80.67	78.19	84.95	90.48	83.71	82.10
P6	81.90	71.43	77.81	89.05	80.76	75.24	77.14
P7	76.48	81.90	84.95	84.57	80.67	76.48	73.52
P8	79.24	84.95	82.86	69.43	69.52	82.29	64.86
P9	80.00	75.14	75.14	76.38	77.71	71.43	67.71
P10	73.71	77.81	79.71	89.24	73.52	89.14	84.76
P11	75.05	86.48	79.14	83.62	53.71	82.10	87.71
P12	82.10	76.48	87.81	79.24	72.19	83.52	69.33
P13	75.05	82.19	70.67	78.29	67.71	75.05	74.95
P14	84.86	87.81	86.48	80.76	90.48	75.05	79.05
P15	75.24	78.19	86.29	84.67	93.33	83.52	72.19
P16	89.24	77.90	82.29	86.48	82.29	69.33	82.19

remaining six distraction themes, the accuracy of the model was improved to a certain extent. The average value of Recall, Precision, and F-score were 80.44%, 81.07% and 80.54%. The standard deviation of each independent motion imagery task was also calculated. As shown in Table I, compared with other methods, the proposed method achieves a lower standard deviation. This indicates that the proposed method can alleviate overfitting to a great extent and help achieve improved and more stable performances than other state-of-the-art models (paired t-test, $p < 0.05$, with seven degrees of freedom). Fig. 4 shows the accuracy assessment of our model and other algorithms. The average growth rates of P_{acc} and P_f were 8.35% and 14.09%, respectively. The results also showed that, compared with the manual features used in [5], the deep topology and node features characterized by GSNN were less sensitive to individual differences.

D. Evaluation on Different Subjects

Table II shows the classification effect of the GSNN model for different subjects. On the basis of 16 subjects participating in each distraction experiment, we conducted independent

model training and evaluation for each subject. It can be seen that in subjects 1, 3, 6, 12, 14 and 16, the model has good classification accuracy under each distraction theme. In the rest of the subjects, it also showed good performance.

IV. DISCUSSION

In this section, we conduct ablation experiment and analyze the sensitivity of the GSNN model based on the distraction effect. The relationship between the important brain regions and MI channels is analyzed. Finally, we determine the potential prospects of GSNN.

A. Ablation Analysis

To evaluate the sensitivity of the proposed GSNN method proposed, we conducted experiments on datasets with different data enhancement methods. First, we conducted a controlled experiment with no data extension for each distraction theme. Subsequently, considering the same conditions, after the node domain adversarial enhancement method, the accuracy of each independent motion imagery task was calculated. As shown

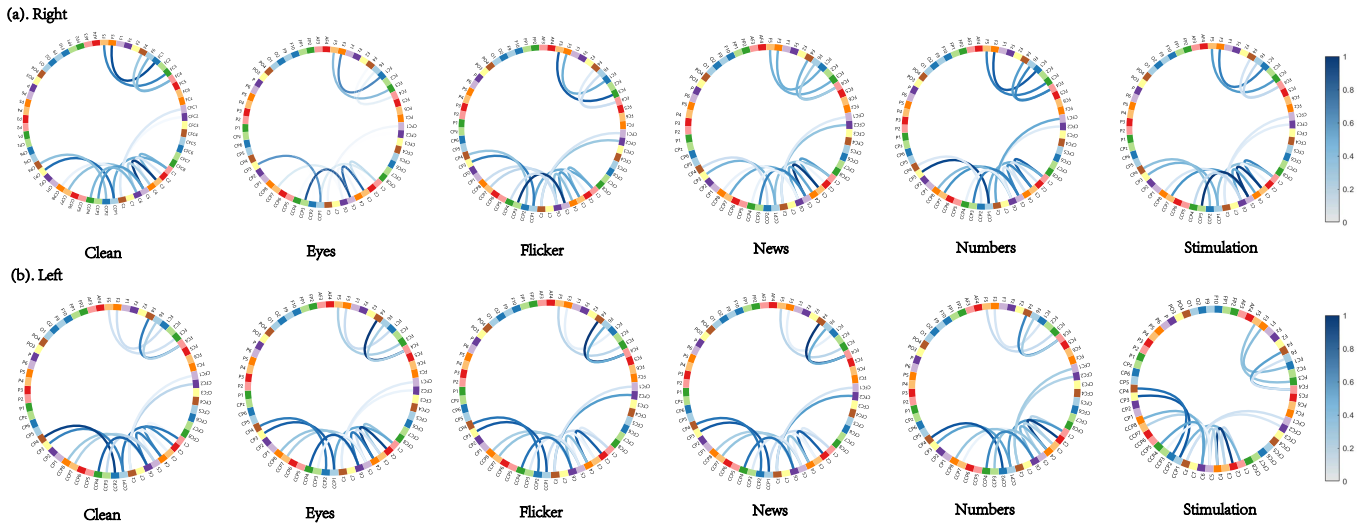


Fig. 5. (a) Top 17 edges between channels in the adjacency matrix A of right imagery, excluding global connections for better clarity. (b) Top 17 edges between channels in the adjacency matrix A of left imagery in subject-dependent classification on the Berlin-motor imagery at the distraction database.

TABLE III
ACCURACY (%) OF MOTOR IMAGERY RECOGNITION COMPARISON AT DIFFERENT NETWORK CONFIGURATIONS USING ALL INDEPENDENT DISTRACTION THEMES ON THE BERLIN-MOTOR IMAGERY DATABASE

Model	calibration	clean	eyes	flicker	news	numbers	stimulation
GNN	72.50	73.54	73.74	70.06	71.46	73.66	70.21
Regularization Map+GNN(GSNN)	75.81	76.87	75.56	75.03	75.65	75.48	73.46
GSNN+nodeDAT (proposed method)	79.05	82.14	79.39	82.59	76.77	78.57	78.41

TABLE IV
ACCURACY (%) COMPARISON OF THE GSNN WITH DIFFERENT COV-KERNEL TYPES

Cov-kernel type	Accuracy (%)		
	GCN	SGC	gfNN
calibration	74.04	77.84	78.39
clean	71.36	76.29	81.05
eyes	74.24	73.66	79.53
flicker	74.46	75.40	80.76
news	69.95	75.46	76.43
numbers	69.45	72.11	79.91
stimulation	74.81	75.22	78.44

TABLE V
ACCURACY (%) COMPARISON OF THE GSNN WITH DIFFERENT BLOCK NUMBERS

Block number	Accuracy (%)			
	2	3	4	5
calibration	70.78	75.24	78.47	79.71
clean	60.49	75.48	81.05	76.06
eyes	67.06	72.29	79.72	73.44
flicker	61.11	70.43	81.93	79.75
news	64.00	74.14	76.72	75.45
numbers	71.02	72.51	79.38	80.66
stimulation	71.05	75.43	78.19	78.16

in Table III, compared with each part of data enhancement method, the proposed method achieves a higher accuracy rate. This indicates that properly extracting regularization map is beneficial to model performance. Node domain adversarial training also proves to be useful in reducing overfitting and improving accuracy.

B. Experiments Using Different Structures of GSNN

By constructing convolution kernels with different types, it can be found that gfNN is more suitable for the classification of motor imagery tasks under independent distraction than other convolution kernels. The accuracy is shown in Table IV,

where it can be seen that higher performance can be achieved when gfNN blocks is applied.

By constructing different blocks, the optimal number of convolutional channels can be filtered out. The accuracy is shown in Table V, in which it can be seen that higher performance can be achieved when four gfNN blocks are applied. It can therefore be concluded that selecting appropriate parameters is conducive to improving classification performance.

C. Analysis of Important Brain Regions and Interchannel Structures

We identified important brain regions of MI based on independent distraction themes. Fig. 5 presents a heat map of the

diagonal elements in the adjacency matrix we learned, where each subgraph depends on the classification of independent distraction themes. These values were scaled to the $[0, 1]$ range for better visualization. Intuitively, diagonal values represent the contribution of each channel to the calculation of the final EEG representation. Fig. 5 depicts that there is strong activation in the prefrontal, parietal, and temporal lobes of all frequency bands, suggesting that these areas may be closely related to MI in the brain. Our findings are consistent with existing research. Fig. 5 demonstrates the first 17 connections between the channels with the largest edge weight in the adjacency matrix. After A is learnt and sparse selection is completed based on the selection parameter G value, global link-ages remain the strongest, confirming that the relationship between global channels is essential for MI. Additionally, Fig. 5(a) shows that the connection between the channel pairs (C3, C5) is the strongest when imagining the right side at the independent distraction theme, followed by (CCP3, C5) and (CCP1, CP3). The connection between the channel pairs (CCP4, C6) is the strongest when imagining the left side of the task [10], in the case of the independent distraction theme, followed by (C2, C4) and (CCP2, CP4). We also found that the topological connection strength of (F5, FC3) and (F4, FC5) on the prefrontal lobe are related to the distraction on the left and right side imagery [47], indicating that the relationship between channels in the local frontal region may be important for MI.

D. Other Prospects of the Model

In the process of representation learning with atlas, a better pooling method is needed for learning and selecting important topological features. Because pooling is not a decisive factor, it is difficult to determine whether a global pooled architecture or a hierarchical pooled architecture is beneficial to graph classification in MI. The global pool architecture minimizes information loss, making this architecture perform better than the hierarchical pool architecture on datasets with fewer nodes. However, layering is more effective on datasets with numerous nodes because it can effectively extract useful information from large-scale graphs. Therefore, for the data set of distracted MI, the pooling weight super-parameter setting of the two parts still requires a more precise model.

E. Conclusion

In this study, we trained the sample datasets specific to each subjects at each independent distraction theme. The model adaptively adjusted the subjects' influence at different distraction themes. The problem of noise samples was solved by using the feature standard of regularization and network method in the node domain to solve the influence difference of subjects. The sensitivity of the model to different subjects was reduced. Finally, the results showed that pertaining to the task of MI during distraction, our model performances were significantly better than those of the state-of-the-art models. Moreover, the interchannel structures of active brain regions were revealed at different distraction themes,

to a certain extent, which provides basic knowledge for additional research.

CONFLICTS OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

- [1] M. Xu, J. Han, Y. Wang, T.-P. Jung, and D. Ming, "Implementing over 100 command codes for a high-speed hybrid brain-computer interface using concurrent P300 and SSVEP features," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 11, pp. 3073–3082, Nov. 2020.
- [2] K. P. Thomas, C. Guan, C. T. Lau, A. P. Vinod, and K. K. Ang, "A new discriminative common spatial pattern method for motor imagery brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2730–2733, Nov. 2009.
- [3] G. Pfurtscheller, C. Guger, and H. Ramoser, "EEG-based brain-computer interface using subject-specific spatial filters," in *Engineering Applications of Bio-Inspired Artificial Neural Networks* (Lecture Notes in Computer Science), vol. 1607, J. Mira and J. V. Sánchez-Andrés, Eds. Berlin, Germany: Springer, 1999, doi: 10.1007/BFb0100491.
- [4] J. Decety and D. H. Ingvar, "Brain structures participating in mental simulation of motor behavior: A neuropsychological interpretation," *Acta Psychol.*, vol. 73, no. 1, pp. 13–34, Feb. 1990.
- [5] 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., IEEE World Congr. Comput. Intell.*, Jun. 2008, pp. 2390–2397.
- [6] Y. Liu *et al.*, "Motor-imagery-based teleoperation of a dual-arm robot performing manipulation tasks," *IEEE Trans. Cognit. Develop. Syst.*, vol. 11, no. 3, pp. 414–424, Oct. 2018.
- [7] R. Zhang *et al.*, "Control of a wheelchair in an indoor environment based on a brain-computer interface and automated navigation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 128–139, Jan. 2016.
- [8] L. M. Alonso-Valerdi, R. A. Salido-Ruiz, and R. A. Ramirez-Mendoza, "Motor imagery based brain-computer interfaces: An emerging technology to rehabilitate motor deficits," *Neuropsychologia*, vol. 79, pp. 354–363, Dec. 2015.
- [9] S. Bajaj, A. J. Butler, D. Drake, and M. Dhamala, "Brain effective connectivity during motor-imagery and execution following stroke and rehabilitation," *NeuroImage, Clin.*, vol. 8, pp. 572–582, Jan. 2015.
- [10] V. A. Maksimenko *et al.*, "Nonlinear analysis of brain activity, associated with motor action and motor imaginary in untrained subjects," *Nonlinear Dyn.*, vol. 91, no. 4, pp. 2803–2817, Mar. 2018.
- [11] E. V. C. Friedrich, R. Scherer, K. Sonnleitner, and C. Neuper, "Impact of auditory distraction on user performance in a brain-computer interface driven by different mental tasks," *Clin. Neurophysiol.*, vol. 122, no. 10, pp. 2003–2009, Apr. 2011.
- [12] J. J. Alvarsson, S. Wiens, and M. E. Nilsson, "Stress recovery during exposure to nature sound and environmental noise," *Int. J. Environ. Res. Public Health*, vol. 7, no. 3, pp. 1036–1046, Mar. 2010.
- [13] E. Marx *et al.*, "Eye closure in darkness animates sensory systems," *NeuroImage*, vol. 19, no. 3, pp. 924–934, Jul. 2003.
- [14] K. Hübner *et al.*, "Differences in saccade-evoked brain activation patterns with eyes open or eyes closed in complete darkness," *Exp. Brain Res.*, vol. 186, no. 3, pp. 419–430, Apr. 2008.
- [15] P. Xu *et al.*, "Different topological organization of human brain functional networks with eyes open versus eyes closed," *NeuroImage*, vol. 90, pp. 246–255, Apr. 2014.
- [16] T. Hohmann, M. Exner, and N. Schott, "The role of vision and auditory distraction on the temporal congruence between physical execution and motor imagery," *J. Imag. Res. Sport Phys. Activity*, vol. 11, no. 1, pp. 25–33, Jan. 2016.
- [17] 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, Oct. 2019.
- [18] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, 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. 2020.

- [19] K. D. Ghanbar, T. Y. Rezaii, A. Farzamnia, and I. Saad, "Correlation-based common spatial pattern (CCSP): A novel extension of CSP for classification of motor imagery signal," *PLoS ONE*, vol. 16, no. 3, Mar. 2021, Art. no. e0248511.
- [20] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. D. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, Jan. 2005.
- [21] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.
- [22] R. Li, Q. Wu, J. Liu, Q. Wu, C. Li, and Q. Zhao, "Monitoring depth of anesthesia based on hybrid features and recurrent neural network," *Frontiers Neurosci.*, vol. 14, p. 26, Feb. 2020.
- [23] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-based EEG classification in motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2086–2095, Nov. 2018.
- [24] S. Zheng, D. Punia, H. Wu, and Q. Liu, "Graph theoretic analysis reveals intranasal oxytocin induced network changes over frontal regions," *Neuroscience*, vol. 459, pp. 153–165, Apr. 2021.
- [25] C. Chen, K. Yuan, W. C.-W. Chu, and R. K.-Y. Tong, "The effects of 10 Hz and 20 Hz tACS in network integration and segregation in chronic stroke: A graph theoretical fMRI study," *Brain Sci.*, vol. 11, no. 3, p. 377, Mar. 2021.
- [26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
- [27] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2016. [Online]. Available: <https://proceedings.neurips.cc/paper/2016/file/04df4d434d481c5bb723be1b6df1ee65-Paper.pdf>
- [28] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *Stat*, vol. 1050, p. 20, 2017.
- [29] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
- [30] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6861–6871.
- [31] H. Ni and T. Maehara, "Revisiting graph neural networks: All we have is low-pass filters," 2019, *arXiv:1905.09550*.
- [32] S. Mueller *et al.*, "Individual variability in functional connectivity architecture of the human brain," *Neuron*, vol. 77, no. 3, pp. 586–595, 2013.
- [33] A. Gretton, K. M. Borgwardt, M. J. Rasch, A. Smola, B. Schölkopf, and A. S. Gretton, "A kernel two-sample test bernhard schölkopf," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, 2012.
- [34] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 597–613.
- [35] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, May 2015.
- [36] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, early access, May 11, 2020, doi: [10.1109/TAFFC.2020.2994159](https://doi.org/10.1109/TAFFC.2020.2994159).
- [37] S. Brandl, L. Frölich, J. Höhne, K.-R. Müller, W. Samek, and F. Heinrich, "Brain-computer interfacing under distraction: An evaluation study," *J. Neural Eng.*, vol. 13, no. 5, p. 56012, 2016.
- [38] S. Brandl and B. Blankertz, "Motor imagery under distraction—An open access BCI dataset," *Frontiers Neurosci.*, vol. 14, p. 967, Oct. 2020.
- [39] O. A. Rosso *et al.*, "Wavelet entropy: A new tool for analysis of short duration brain electrical signals," *J. Neurosci. Methods*, vol. 105, no. 1, pp. 65–75, Jan. 2001.
- [40] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS Comput. Biol.*, vol. 3, no. 2, p. e17, Feb. 2007.
- [41] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, "An end-to-end deep learning architecture for graph classification," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1.
- [42] Z. Su, Z. Hu, and Y. Li, "Hierarchical graph representation learning with local capsule pooling," in *Proc. ACM Multimedia Asia*, Dec. 2021, pp. 1–7.
- [43] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3734–3743.
- [44] C. Cangea, P. Veličković, N. Jovanović, T. Kipf, and P. Lio, "Towards sparse hierarchical graph classifiers," 2018, *arXiv:1811.01287*.
- [45] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
- [46] O. Shchur, M. Mumme, A. Bojchevski, and S. Günnemann, "Pitfalls of graph neural network evaluation," 2018, *arXiv:1811.05868*.
- [47] S. Pilgramm *et al.*, "Motor imagery of hand actions: Decoding the content of motor imagery from brain activity in frontal and parietal motor areas," *Hum. Brain Mapping*, vol. 37, no. 1, pp. 81–93, Jan. 2016.