

# A Novel Classification Framework Using the Graph Representations of Electroencephalogram for Motor Imagery Based Brain-Computer Interface

Jing Jin<sup>ID</sup>, Senior Member, IEEE, Hao Sun, Ian Daly<sup>ID</sup>, Shurui Li<sup>ID</sup>, Chang Liu<sup>ID</sup>, Xingyu Wang, and Andrzej Cichocki, Fellow, IEEE

**Abstract**—The motor imagery (MI) based brain-computer interfaces (BCIs) have been proposed as a potential physical rehabilitation technology. However, the low classification accuracy achievable with MI tasks is still a challenge when building effective BCI systems. We propose a novel MI classification model based on measurement of functional connectivity between brain regions and graph theory. Specifically, motifs describing local network structures in the brain are extracted from functional connectivity graphs. A graph embedding model called Ego-CNNs is then used to build a classifier, which can convert the graph from a structural representation to a fixed-dimensional vector for detecting critical structure in the graph. We validate our proposed method on four datasets, and the results show that our proposed method produces high classification accuracies in two-class classification tasks (92.8% for dataset 1, 93.4% for dataset 2, 96.5% for dataset 3, and 80.2% for dataset 4) and multiclass classification tasks (90.33% for dataset 1). Our proposed method achieves a mean Kappa value of 0.88 across nine participants, which is superior to other methods we compared it to. These results indicate that there

is a local structural difference in functional connectivity graphs extracted under different motor imagery tasks. Our proposed method has great potential for motor imagery classification in future studies.

**Index Terms**—Motor imagery (MI), electroencephalogram (EEG), functional connectivity, graph representation.

## I. INTRODUCTION

**M**OTOR imagery (MI) is the process of imagining the movement of some parts of the body without sensory stimulation [1]. Motor imagery (MI) classification can be used in the control of brain-computer interfaces (BCIs) [2]. Specifically, the BCI attempts to decode the motor intention of the user without the user making any physical movement [3]. MI-based BCI system can help patients with motor dysfunction to control external rehabilitation equipment [4], such as, but not limited to, wheelchair start and stop functions and exoskeleton rehabilitation training systems.

However, identifying neural markers of motor imagery in the brain is not an easy task and usually requires the use of specific feature extraction methods and classification algorithms [5], which must be sufficiently robust to deal with the complexity of the EEG signals [6]. Common Spatial Patterns (CSP) [7] is one of the most effective methods for extracting features for MI classification and, consequently, many extended versions of the original CSP algorithms have been developed [8]–[10].

For classifying MI tasks, traditional machine learning methods like support vector machines (SVMs), canonical correlation analysis (CCA) [11], k-means clustering [12], and Fisher's discriminant analysis (FDA) [13] may be applied to the EEG. Although these methods can be effectively used in traditional BCI systems to a certain extent, they may neglect some critical information in the raw EEG signals which can limit their performance. In one attempt to address this, a growing number of researchers are turning to measures of brain network connectivity for classification of MI activity from the EEG [14]–[16].

The human brain is a large complex network containing many billions of neurons [17], and its function depends on the real-time dynamic interaction between many spatially

Manuscript received July 6, 2021; revised November 10, 2021; accepted December 13, 2021. Date of publication December 28, 2021; date of current version January 28, 2022. This work was supported in part by the National Key Research and Development Program under Grant 2017YFB13003002, in part by the Grant National Natural Science Foundation of China under Grant 61573142 and Grant 61773164, in part by the Program of Introducing Talents of Discipline to Universities (the 111 Project) under Grant B17017, and in part by the “ShuGuang” Project by Shanghai Municipal Education Commission and Shanghai Education Development Foundation under Grant 19SG25. (Corresponding author: Jing Jin.)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Jing Jin, Hao Sun, Shurui Li, Chang Liu, and Xingyu Wang are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200231, China (e-mail: jinjingat@gmail.com; y20180078@mail.ecust.edu.cn; shuruili1008@163.com; liuchang@mail.ecust.edu.cn; xywang@ecust.edu.cn).

Ian Daly is with the Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, U.K. (e-mail: i.daly@essex.ac.uk).

Andrzej Cichocki is with the Skolkovo Institute of Science and Technology (SKOLTECH), 143026 Moscow, Russia, also with the RIKEN Brain Science Institute, Wako 351-0198, Japan, and also with Nicolaus Copernicus University (UMK), 87-100 Torun, Poland (e-mail: a.cichocki@riken.jp).

Digital Object Identifier 10.1109/TNSRE.2021.3139095

distributed regions [18]. Network science, and in particular graph theory, has the potential to describe these interactions and, consequently, has become increasingly used in the fields of neuroscience and neurology. Network neuroscience [19] is a relatively new area of research, which provides researchers with a unique opportunity to evaluate, quantify, and ultimately understand the characteristic information embodied by complex brain networks when individuals perform cognitive tasks. It may also be used to construct BCI systems based on characteristics of the brain networks while users perform specific cognitive tasks. The resulting brain network characteristics can be sent to a classifier as features, potentially resulting in improved classification performance. Connectivity in the brain is measured in three ways, functional connectivity, anatomical connectivity, and effective connectivity [20].

Functional connectivity evaluates the statistical relationships between different brain regions, it may provide a new perspective that helps people to understand the neural mechanism underlying motor execution (ME) and motor imagery (MI). A large number of methods have been proposed to measure functional connectivity in the brain via functional magnetic imaging (fMRI) and EEG. For example, an algorithm based on K-means clustering of the functional connectivity graphs which were obtained from the phase-locking value (PLV) metric has been applied to high-resolution EEG to study the brain networks properties modify during visual tasks [21]. Functional connectivity can be seen as a neurophysiological biomarker to assess Alzheimer's disease in patients [22]. Ge *et al.* found that functional connectivity between the medial temporal lobe, the lateral parietal, and lateral temporal regions increased after the process of motor imagery training [23]. Gonuguntla *et al.* analyzed the network mechanisms related to motor imagery tasks based on PLV in the alpha frequency band of the EEG [15]. In reference [14], Daly *et al.* pointed out that prior to motor execution and motor imagery there is an increase in the level of PLV-measured functional connectivity at the Mu rhythm and that this may be used successfully as the control signal for building a highly accurate BCI.

Graph theory has great advantages in studying the working characteristics of brain networks. In [16], the author combined spectral graph theory and a quantum genetic algorithm to find an effective set of channels for motor imagery classification. Stefano Filho *et al.* modeled interactions among EEG electrodes during a motor imagery task and classified the signals using LDA and SVM classifiers [6]. Gao *et al.* explored the network connectivity differences between core brain regions during ME and MI through conditional granger causality and In-Out degrees [24]. Xu *et al.* explored and compared the functional connectivity between ME and MI by calculating betweenness centrality (BC), a measure of graph theory [25]. In [26], the author used the spectral decomposition of a graph defined by a geometrical distribution of electrodes to achieve EEG signal dimensionality reduction. In [27], a graph-based method for EEG biometric identification was proposed, which consisted of a network estimation module and a graph analysis module.

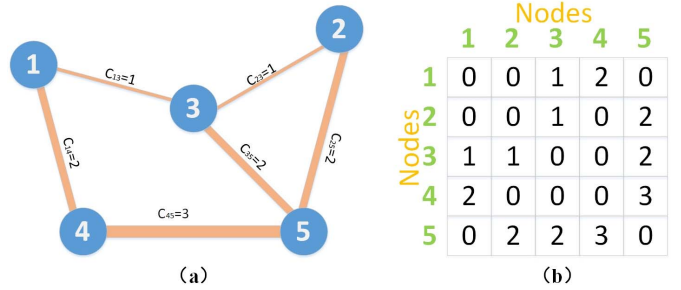


Fig. 1. An example of a weighted undirected graph and the corresponding adjacency matrix, where Fig.1(a) shows the weighted connections of five nodes and Fig.1(b) shows the adjacency matrix. In this example, the strength of connectivity is divided into three levels.

To solve the low performance of EEG-based MI classification, we propose a novel classification framework for identifying MI tasks using the graph representations transferred from EEG signals. In this paper, we use functional connectivity analysis to convert the time series of EEG signals to graph data. We then use the resulting graph data for classifying different motor imagery states. In our method, the graph built by measuring functional connectivity is an undirected weighted graph and we extract the distinguished local structure through a local structure extraction matrix  $E_{LS}$ . To identify the structure in the graph during different motor imagery tasks, we choose a graph embedding model, called Ego-CNNs [28] which is used for learning discriminative features from graphs.

## II. METHOD

### A. Graph Representation

A weighted undirected graph can be defined as  $\mathcal{G} = (V, E)$ , in which  $V$  represents the set of nodes with the number of  $|V| = N$  and  $E$  denotes the set of edges connecting these nodes. The term  $e_{ij} \in E$  denotes the edge connection between node  $v_i$  and node  $v_j$ . If the connection between node  $v_i$  and node  $v_j$  exists  $e_{ij} = 1$  otherwise  $e_{ij} = 0$ . The structure of a graph also can be represented by an adjacency matrix, in which each cell represents the connection between a pair of nodes. The adjacency matrix of the undirected graph is symmetric. Different node pairs may have different connectivity strengths and, in this case,  $c_{ij}$  denotes the strength of connectivity between node  $v_i$  and node  $v_j$ . Figure.1 illustrates an example of a weighted undirected graph with five nodes and the strength of connectivity between node pairs, as well as the adjacency matrix associated with the graph. In Fig.1(a), different sizes of edges represent the different strengths of connectivity between pairs of nodes, whereas Fig.1(b) illustrates the corresponding adjacency matrix of the graph.

In the case of EEG, the recording electrodes can be seen as the nodes, but the methods to measure the connection strength between different nodes are diverse. In this paper, we focus on the characteristic of functional connectivity of EEG signals in the time domain. Therefore, we use mutual information to measure connection strength between nodes.

## B. Normalized Mutual Information

The normalized mutual information quantifies the amount of information that two signals share with each other and is based on information theory [29].

Let  $\rho(x) = P_r\{X=x\}$  and  $\rho(y) = P_r\{Y=y\}$  be the probability density functions of random variables  $X$  and  $Y$ . The joint probability density is defined as  $\rho(x, y) = P_r\{X=x, Y=y\}$ . The Shannon entropy  $H(X)$  and  $H(Y)$  measure the average information obtained from the observations of random variables  $X$  and  $Y$  [30]. They are defined as:

$$H(X) = -\sum_x \rho(x) \log(\rho(x)) \quad (1)$$

$$H(Y) = -\sum_y \rho(y) \log(\rho(y)) \quad (2)$$

The joint entropy  $H(X, Y)$  is:

$$H(X, Y) = -\sum_x \sum_y \rho(x, y) \log(\rho(x, y)) \quad (3)$$

The conditional entropy of  $X$  given  $Y$  is defined as:

$$H(X|Y) = -\sum_x \sum_y \rho(x, y) \log(\rho(x|y)) \quad (4)$$

where  $\rho(x|y) = P_r\{X=x|Y=y\}$  is the conditional probability. The joint entropy is similar in form to the Shannon entropy. The conditional entropy can be represented by  $H(X, Y) = H(X|Y) + H(Y)$ . The mutual information  $I(X; Y)$  measures the amount of shared information between  $X$  and  $Y$ :

$$I(X; Y) = \sum_x \sum_y \rho(x, y) \log\left(\frac{\rho(x, y)}{\rho(x)\rho(y)}\right) \quad (5)$$

It can also be expressed as:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y) \\ &= H(X, Y) - H(Y|X) - H(X|Y) \end{aligned} \quad (6)$$

The mutual information needs to be normalized to measure connectivity between variables [31]:

$$\text{NorMI}(X; Y) = \frac{I(X; Y)}{\sqrt{H(X)H(Y)}} \quad (7)$$

The normalized mutual information between two EEG electrodes may be used as a measure of the strength of functional connectivity between those electrodes. Consequently, the structure of the graph describing brain connectivity is defined as follows. The nodes are electrodes and the weighted edges are the normalized mutual information between the nodes.

## C. Selection of Nodes and Edges

In general, when the number of electrodes is large, if we use all electrodes to build an adjacency graph, the graph is large and, as a result, contains many redundant edges and nodes. We aim to select a subset of edges and nodes which is most informative for discriminating between different motor imagery tasks. When participants perform different motor imagery tasks, the connectivity pattern of the brain looks

different [6]. We choose a subset of weighted edges (calculated by measuring functional connectivity) with a large difference between different imagery tasks.

For one trial recorded during one MI task, the adjacency matrices obtained from all electrodes are defined as:  $A_p^q \in \mathbb{R}^{N_{all} \times N_{all}}$ , which represents the  $q^{th}$  trial of the  $p$  type motor imagery task. Here,  $N_{all}$  denotes the number of all electrodes,  $V_{all}$  represents the set of nodes, and  $|V_{all}| = N_{all}$ .

$$A_p^q = \begin{bmatrix} 1 & c_{12} & c_{13} & c_{14} & \dots \\ c_{21} & 1 & c_{23} & c_{24} & \dots \\ c_{31} & c_{32} & 1 & c_{34} & \dots \\ c_{41} & c_{42} & c_{43} & \ddots & \vdots \\ \vdots & \dots & \dots & \dots & 1 \end{bmatrix} \quad (8)$$

$$c_{ij} = \text{NorMI}(v_i; v_j), \quad v_i, v_j \in V_{all} \text{ and } i \neq j$$

$A_1$  represents the adjacency matrix of one motor imagery task, we assume the number of trials for one type of MI task(class1) is  $T_1$  and the number of trials for the other type ( $A_2$ ) of MI task (class2) is  $T_2$ . We assume  $\{c_{ij}^l \in A_1, l = 1, 2, \dots, T_1\}$  follows the distribution  $P_1$ , and  $\{c_{ij}^r \in A_2, r = 1, 2, \dots, T_2\}$  follows the distribution  $P_2$ . The differential functional connectivity network graph is defined as:

$$A_{DFCN} = \begin{bmatrix} 0 & d_{12} & d_{13} & d_{14} & \dots \\ d_{21} & 0 & d_{23} & d_{24} & \dots \\ d_{31} & d_{32} & 0 & d_{34} & \dots \\ d_{41} & d_{42} & d_{43} & \ddots & \vdots \\ \vdots & \dots & \dots & \dots & 0 \end{bmatrix} \in \mathbb{R}^{N_{all} \times N_{all}} \quad (9)$$

where  $d_{ij} \geq 0$  ( $i \neq j$ )

$$d_{ij} = \sum_{k=1}^T w_{ij}^k \quad (10)$$

where  $w_{ij}$  is the Wasserstein distance of  $c_{ij}$  in different MI tasks. This is defined as:

$$w_{ij} = \frac{\inf_{\gamma \in \prod[P_1, P_2]} \sum_{c_{ij}^l} \sum_{c_{ij}^r} \gamma(c_{ij}^l, c_{ij}^r) d(c_{ij}^l, c_{ij}^r)}{\quad} \quad (11)$$

where  $\gamma \in \prod[P_1, P_2]$  is joint probability distribution of  $P_1$  and  $P_2$ ,  $d(f_i^l, f_i^r)$  is the distance of every sample pair of the joint distribution, in this paper, we used  $d(c_{ij}^l, c_{ij}^r) = \|c_{ij}^l - c_{ij}^r\|$ .  $T = C(p, 2)$  denotes the number of different unordered pairs of MI tasks. For example, in four-class classification,  $p = 4, T = C(4, 2) = 6$ . The term  $d_{ij}$  measures the difference in connectivity between two electrodes during different motor imagery tasks. Based on  $A_{DFCN}$ , we created a local structure extraction matrix  $E_{LS} \in \mathbb{R}^{N_{all} \times N_{all}}$ .  $E_{LS}$  is a symmetric square matrix only consisting of elements 1 and 0. All the elements  $d_{ij}$  in matrix  $A_{DFCN}$  are then sorted in descending order and we select the  $2N_{edge}$  elements that are ranked highest. We record the positions of these elements in matrix  $A_{DFCN}$  and set the  $E_{LS}$  elements at those positions

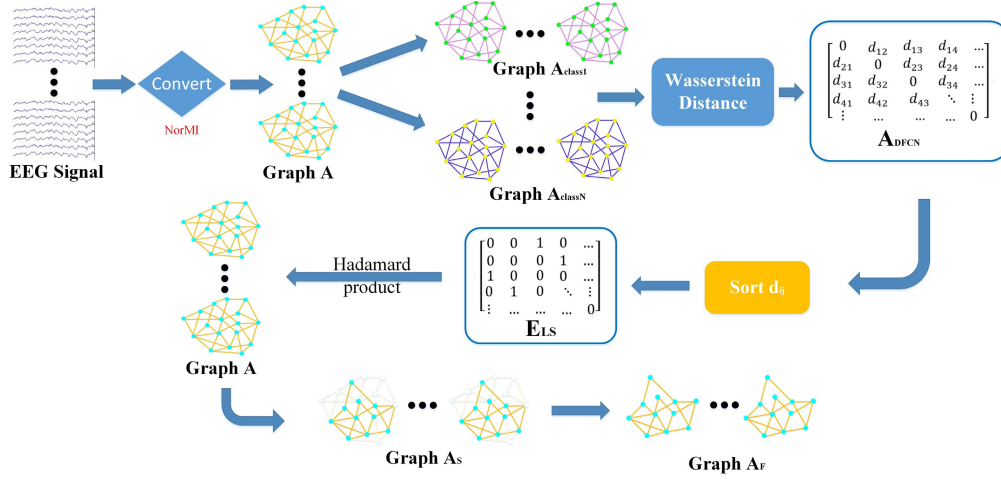


Fig. 2. The process of extracting locally important structures by  $E_{LS}$ .

to 1.  $N_{edge}$  represents the number of edges selected from the set of all edges.

$$E_{LS} = \begin{bmatrix} 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \dots & \dots & \dots & 0 \end{bmatrix} \in \mathbb{R}^{N_{all} \times N_{all}} \quad (12)$$

$$A_S = A \circ E_{LS}, A_S \xrightarrow{\text{transform}} A_F \in \mathbb{R}^{N_{node} \times N_{node}} \quad (13)$$

Under the function of matrix  $E_{LS}$ , important structures of the graph, containing all nodes and all edges, can be preserved, while the redundant nodes and edges are eliminated.  $A \in \mathbb{R}^{N_{all} \times N_{all}}$  denotes a generic graph containing all nodes and edges,  $A_S \in \mathbb{R}^{N_{all} \times N_{all}}$  represents the motif extracted from the graph with an important structure. In formula (13), the calculating symbol  $\circ$  denotes the Hadamard product [32].  $A_S$  is a symmetric sparse matrix, and it can then be transformed to  $A_F \in \mathbb{R}^{N_{node} \times N_{node}}$  by eliminating rows and columns with all zeros. The dimension of  $A_F$  is the number of selected nodes  $N_{node}$ , which depends on the number of selected edges and structure formed by the selected edges. Even if the number of selected edges ( $N_{edge}$ ) is the same, the number of selected nodes ( $N_{node}$ ) may be different, because different participants exhibit different patterns of connectivity. However, the number of edges is the same.

$A_F$  denotes the adjacency matrix, which can represent the motif with the most informative structures. The motif is extracted from the graph with all nodes and all edges by  $E_{LS}$ . The next step is to build a model that can classify the graph data described by  $A_F$ . Fig. 2 describes the processing of selection of edges and nodes.

#### D. Ego-CNNs Classification Model

The graph embedding algorithm converts graphs from structural representations to a fixed-dimensional vector for detecting critical structures in the graph. The graph described by  $A_F$  typically has a distinct structure, so we decided to

use a graph embedding model to build the classifier. Ego-CNN is a novel graph embedding model [28], that employs ego-convolutions at each layer and stacks up layers using an ego-centric approach to detect precise critical structures efficiently. The Ego-CNNs are a generalization of a node embedding model called a 1-head-attention graph attention network (1-head GATs) [33]. One challenge for identifying critical structure is that critical structure is task-specific and participant-specific. In other words, the shape and location of the critical structure may vary from person to person. For one participant, every  $A_F$  describes the graph during different tasks, and has the same location but different structure.

Given a graph  $\mathcal{G} = (V, E)$ ,  $V$  and  $E$  represent the set of nodes and edges, respectively. The dimensionality of feature embedding is  $D$  and the number of convolution layers is  $L$ . If a node  $n$  has  $K$  neighbors, the features of node  $n$  can be represented by the features of these neighbors. The Ego-CNN we use is developed from the Patchy-San model [34], which can detect precise critical graph structures at a local scale. In the Patchy-San model, the neighborhood of node  $n$  at the input layer is defined as the  $K \times K$  adjacency matrix of the  $K$  nearest neighbors of the node.

Let  $Nbr(n, k)$  be the  $k$ -th nearest neighbor of node  $n$  in the graph  $\mathcal{G}$ . The graph embedding output at the  $l$ -th layer by  $D$  filters  $W^{(l,1)}, \dots, W^{(l,D)}$  is defined as  $H^{(l)} \in \mathbb{R}^{N \times D}$ . For  $l = 1, \dots, L$ , filter  $W^{(l,d)}$  scans through the adjacency matrix of a node  $n$  to generate a graph embedding as:

$$H_{n,d}^{(l)} = \sigma \left( E^{(n,l)} * W^{(l,d)} + b_d^{(l)} \right) \quad (14)$$

$$E^{(n,l)} = \left[ H_{n,:}^{(l-1)}, H_{Nbr(n,1),:}^{(l-1)}, \dots, H_{Nbr(n,K),:}^{(l-1)} \right]^T \quad (15)$$

The term  $E^{(n,l)} \in \mathbb{R}^{(K+1) \times D}$  is a matrix denoting the neighborhood of node  $n$  at the  $l$ -th layer,  $b^d$  is the bias term,  $*$  is the Frobenius inner product defined as  $X * Y = \sum_{i,j} X_{i,j} Y_{i,j}$ , and  $\sigma$  is the activation function. In this study, we use the edge weights to determine the  $K$  nearest neighbors of a node  $n$ . The term  $H_{n,:}^{(0)} \in \mathbb{R}^K$  defines the adjacency vector between node  $n$  and its  $K$  nearest neighbors.



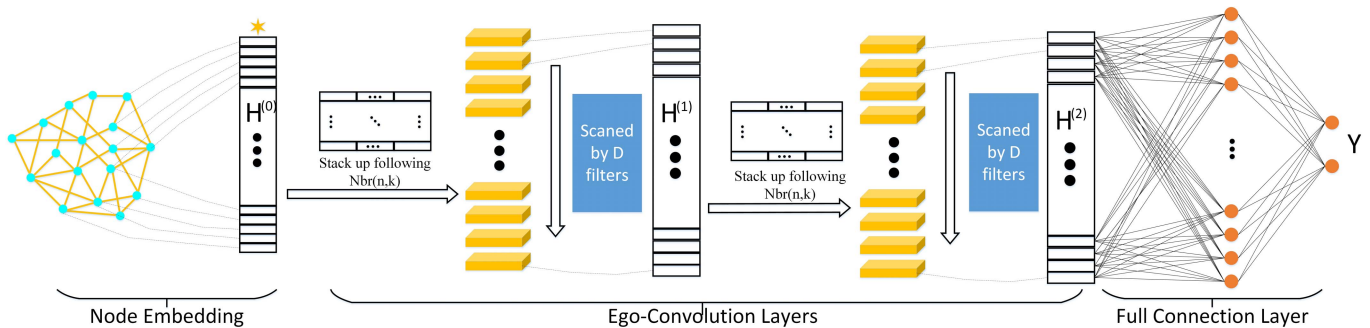


Fig. 3. The structure of the ego-convolution network model.

### III. EXPERIMENTS

In this section, we describe our experiments on four publicly available MI EEG datasets that are commonly used in EEG MI classification to evaluate the effectiveness of our proposed graph representation method. We used dataset 1 to test the effectiveness and performance of our proposed method under low-density channel conditions. We then used dataset 2, dataset 3 and dataset 4 to evaluate the performance of our method under high-density channel conditions.

#### A. Dataset 1

Dataset 1 is taken from the BCI competition IV Dataset IIa [35] and contains EEG signals recorded from nine healthy participants, recorded via 22 electrodes. During the recording process, the participants were instructed, with visual cues, to perform one of four motor imagery tasks: the imagination of movement of their left hand (class 1), right hand (class 2), both feet (class 3), and tongue (class 4). Every MI task contained 72 trials, each participant performed 288 trials in total. The EEG signals were sampled at 250Hz and band-pass filtered between 0.5Hz and 100 Hz. The data set can be download from the website: <http://www.bbc.de/competition/iv/>.

#### B. Dataset 2

Dataset 2 comes from the BCI competition IV [36] provided by the Berlin BCI group. All data were recorded from seven healthy participants via an Ag/AgCl electrode cap with 59 channels. The experimental paradigm used was a standard MI paradigm without feedback. For every participant, the signal was selected from two kinds of MI tasks (left-hand MI and right-hand MI). Two runs of the experiment were performed. Each run contains 100 trials. The data was down-sampled to 100 Hz. The data set can be download from the website: <http://www.bbc.de/competition/iv/>.

#### C. Dataset 3

Dataset 3 is taken from the BCI Competition III Dataset IVa [37], which was recorded from the 5 healthy participants via 118 electrodes. During the recording process, the participants were instructed to perform one of two motor imagery tasks: right hand and foot MI. In total, 280 MI trials were requested from each participant, and all EEG data was down-sampled to 100Hz. The data set can be download from the website: <http://www.bbc.de/competition/iii/>.

#### D. Dataset 4

Dataset 4 is a popular open-access motor movement/imagery dataset, available in Physionet Resource [38, 39]. It consists of 64-channel EEG data recorded at a 160 Hz sampling rate from 109 volunteers. We removed the data from participants S88, S89, S92, and S100 because of the damaged recording with multiple consecutive “rest” sections. As a result, only 105 participants’ data were used in this experiment, and each participant had 42 or 44 trials (the mean value is 43.6) with a balanced ratio in the right and left fist motor imagery conditions. The data set can be download from the website: <https://physionet.org/content/eegmmidb/1.0.0/>.

#### E. Data Processing

All EEG data used in the four datasets were band-pass filtered using a fifth-order Butterworth filter from 8Hz to 30Hz. Because of the differences among the paradigms in the three datasets, different time windows were used in this study: 0.5-3.5s for Dataset 1, 0.5-4s for Dataset 2, 0.5-3.5s for Dataset 3 and 0-2s for Dataset 4.

In the process of selecting nodes and edges from the graph with all nodes and all edges, the key factor is  $N_{edge}$  (the number of edges selected), which determines the extracted local structure and the number of selected nodes  $N_{node}$ . Taking into account the differences in the number of electrodes used in different data sets, the different  $N_{edge}$  values were as follows for each of the datasets  $N_{edge} = 20$  for Dataset 1,  $N_{edge} = 30$  for Dataset 2 and Dataset 4, and  $N_{edge} = 40$  for Dataset 3. The weight  $c_{ij}$  is rounded to an integer and used as the label of the edge between node  $v_i$  and node  $v_j$ . The labels of the nodes are set to the order of electrodes for recording the EEG data.

#### F. Model Setting

In this study, we want to demonstrate the ability of the graph represented by  $A_F$ , to distinguish between different types of MI. The network architecture of our Ego-CNN implementation used in this study follows the recommendations by Tzeng and Wu [28] and remains the same for all datasets. The Ego-CNNs contains 1 node embedding layer (Patchy-San with 128 filters and  $K = 10$ ), 2 Ego-Convolution layers (both with 128 filters and  $K = 10$ ) and 2 fully connected layers. Fig. 3 describes the structure of the Ego-CNNs model. The Dropout (drop rate 0.2) and Batch Normalization were used in the input

TABLE I  
AVERAGE TEST ACCURACY (%) ON FOUR DATASETS UNDER THE DEFAULT CONDITION

	$N_{part}$	$N_{channel}$	$N_{trial}$	$N_{fold}$	$N_{edge}$	$\overline{N}_{node}$	$\overline{Test\ ACC}$
Dataset 1 (two-class)	9	22	144	6	20	13.2	92.8%
Dataset 1 (four-class)	9	22	288	6	20	14.8	90.33%
Dataset 2	7	59	200	10	30	22.4	93.4%
Dataset 3	5	118	280	10	40	22	96.5%
Dataset 4	105	64	43.6	6	30	25.4	80.20%

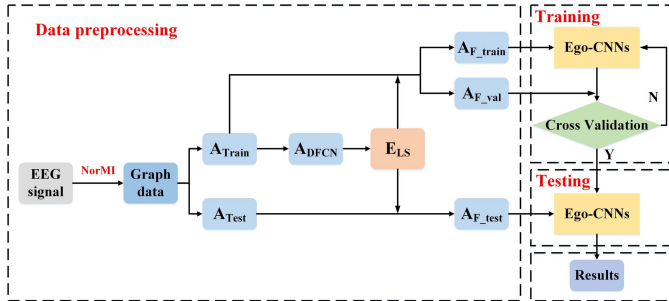


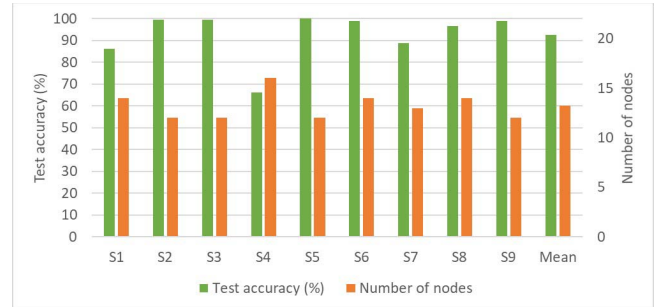
Fig. 4. General scheme of the proposed method.

and Ego-Convolution layers. We apply the Adam algorithm to train the Ego-CNNs with a learning rate of 0.0001. If the node has less than  $K$  neighbors, zero vectors will be used to denote non-existing neighbors. Since the amount of data from one participant contained in each dataset is different we set different cross-validation parameters for each dataset. For datasets 1 and 4, we use a 6-fold cross-fold training and testing strategy, for datasets 2 and 3, we use the test accuracy from a 10-fold cross-fold training and testing strategy. Figure 4 describes the general method proposed in this paper. The graph dataset derived from the EEG dataset was divided into training, validation, and testing set.

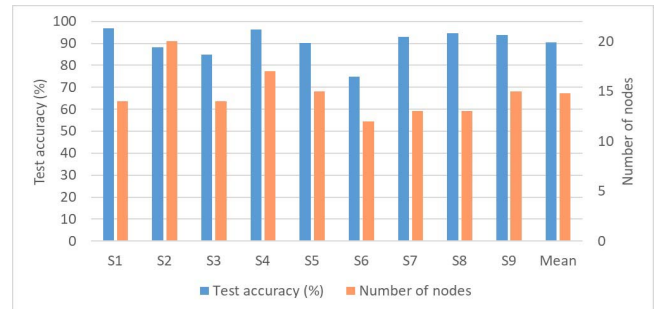
## IV. RESULTS

### A. Generic Performance of Classification

Table I shows the details of the datasets and the test accuracies achieved with the default parameters. The details listed include the number of participants ( $N_{part}$ ), the number of channels used in recording the EEG signals ( $N_{channel}$ ), the number of trials for one participant ( $N_{trial}$ ), the number of folds for cross-validation ( $N_{fold}$ ), the number of edges selected ( $N_{edge}$ ), the average number of nodes selected ( $\overline{N}_{node}$ ), and the average test accuracy ( $\overline{Test\ ACC}$ ) for each dataset. Note, for dataset 1, we performed both two-class and four-class classification experiments. The two-class classification result shown in the table is calculated from only left-hand motor imagery (72 trials) and right-hand motor imagery (72 trials). As can be seen from Table I, for two-class classification experiments, the proposed method can achieve more than 90% average test accuracies on first three datasets. But the dataset 4 only obtained 80.2% test accuracy, this is caused by the too-small number of trials. For the multi-class classification task, the proposed framework also can achieve 90.33% average accuracy. The results shown in table I illustrate the effectiveness of the method proposed in this paper on different datasets (with different numbers of electrodes and different numbers of trials).



(a)



(b)

Fig. 5. Six-Fold CV test accuracy (%) and number of nodes determined for dataset 1. (a) Two-class classification (Left hand/Right hand). (b) Four-class classification (Left hand, Right hand, Foot and Tongue).

Fig. 5 illustrates the 6-Fold CV test accuracies and the numbers of nodes determined for every participant for dataset 1. Fig. 5(a) shows the test accuracy and the number of nodes determined in binary classification. Only participant ‘S4’ has “bad” performance, with a classification accuracy lower than the 70% that is often described as the threshold necessary to control practical BCIs [40]. Fig. 5(b) shows the test accuracy and nodes setting in multiclass classification. The number of nodes is similar across all participants.

The relationship between the training loss and the number of iterations is shown in Fig. 6. After each training step, we also use validation set to evaluate the Ego-CNNs model, and the classification accuracy curve of validation set is overlaid in Fig. 6. As we can see, the training loss, the training loss is less than 0.1 after 20 iterations, and the accuracy curve of validation set converges to around 0.9.

### B. Comparison Results

Dataset 1 and dataset 3 were tested with existing classification models. We compared our proposed method with pre-existing motor imagery EEG classification methods. We used the same preprocessing steps outlined above. We first made a comparison with the classic MI classification algorithm

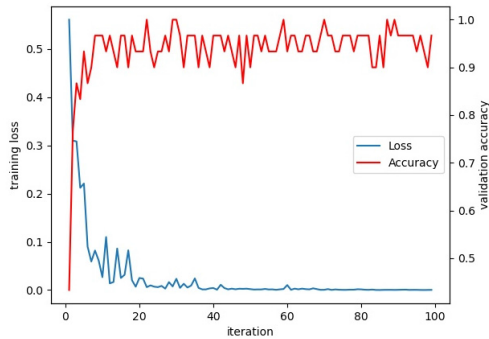


Fig. 6. Loss of the training set and classification accuracy of the validation set. (With 100 iterations).

TABLE II

TWO-CLASS CLASSIFICATION PERFORMANCE COMPARISON OF DIFFERENT METHODS APPLIED ON DATASET 1 AND DATASET 3 (ACCURACY)

Dataset 1 (BCI Competition IV, dataset 2a)						
	CSP	FBCSP	DNN	EEGNet	W-CNN	Proposed
	[8]	[42]	[43]	[44]	[45]	
S1	81.25	90.27	86.81	71.43	76.67	86.10
S2	65.28	54.86	66.70	78.51	72.00	<b>98.90</b>
S3	95.14	96.52	95.83	<b>100.00</b>	90.00	98.90
S4	72.22	65.27	<b>76.39</b>	64.28	73.33	66.10
S5	56.25	88.89	57.64	71.43	83.33	<b>100.00</b>
S6	62.50	71.52	68.06	78.57	80.00	<b>98.90</b>
S7	72.92	<b>97.91</b>	75.00	71.43	82.67	85.00
S8	97.22	97.22	93.75	92.86	80.00	<b>98.30</b>
S9	79.86	79.16	77.08	<b>100.00</b>	80.00	<b>100.00</b>
Mean	75.85	82.40	77.47	80.95	79.78	<b>92.47</b>
Std	$\pm 14.0$	$\pm 15.6$	$\pm 12.7$	$\pm 13.4$	<b><math>\pm 5.4</math></b>	$\pm 11.5$
Dataset 3 (BCI Competition III, dataset IVa)						
aa	80.00	93.97	82.50	67.85	89.29	<b>100.00</b>
al	97.86	<b>99.03</b>	96.07	92.86	98.57	92.70
av	49.29	69.00	53.93	57.14	65.00	<b>97.30</b>
aw	83.21	<b>95.10</b>	89.64	71.43	82.50	90.00
ay	86.43	93.82	87.50	67.85	92.86	<b>99.70</b>
Mean	79.36	90.18	81.93	71.43	85.64	<b>95.94</b>
Std	$\pm 18.1$	$\pm 12.0$	$\pm 16.4$	$\pm 13.1$	$\pm 12.9$	<b><math>\pm 4.4</math></b>
$\rho$ -value	0.0011	0.0395	0.0025	0.0001	0.0004	

(common spatial patterns) CSP [7]. CSP uses diagonalization of the covariance matrix to find a set of optimal spatial filters for classification. Then we compare our proposed method with the state-of-the-art methods (filter bank common spatial patterns) FBCSP [41], which is extended from CSP. The FBCSP algorithm addresses the problem of selecting the appropriate operational frequency band to extract optimal CSP features. We also compare several deep learning approaches with various model structures and feature embedding strategies. Specifically, a deep neural network (DNN) model, as reported in [42], is used. This model uses an adaptive method to determine the classification threshold. We perform a further comparison with the EEGNet [43] approach. The EEGNet encapsulates well-known EEG feature extraction concepts for BCI to construct a uniform approach for different paradigms. Lastly, we compare our proposed model with the wavelet-CNN (W-CNN) model [44]. The wavelet transform is introduced to generate the input images for the CNN model.

Table II summarizes the overall comparison results. The last row of the table presents the  $\rho$ -values obtained from the paired

$t$ -test between the results of our proposed method and other methods we compare against our method. It can be seen from Table II that our method outperforms other existing methods on the evaluation datasets in terms of mean classification accuracy. According to the results, our proposed method yielded an average improvement of 13.21% (on dataset 1) and 14.64% (on dataset 3) in terms of mean classification accuracy. The improvement in the classification accuracy for some participants such as ‘S2’ and ‘av’ is substantial (around 32.6% and 38.7%). This showed that our proposed method is capable of extracting distinctive local structure features for classification and that these features improve classification accuracy substantially.

To evaluate the effectiveness of our proposed framework in the multi-class classification task, we compared our framework with machine learning methods [45]–[54] and deep learning methods [55]–[57]. Table III displays the mean Kappa value using the proposed framework and other existing methods at each subject from Dataset 1. The last columns present the average kappa value, standard deviation, and the  $t$ -test  $p$ -value between our proposed method and other approaches. Although the performance is different between the nine participants, our proposed framework is superior to existing methods in general.

The primary reason that our proposed method surpasses traditional algorithms such as CSP, and FBCSP is the multiple nonlinear transformation processes, which is an advantage of deep learning methods. When compared with other deep learning methods, our proposed model has three main advantages that help to achieve superior performance. The first one is that we convert the EEG time series to graph data by first measuring functional connectivity between channel pairs. Compared with the classical deep learning methods which directly apply the convolutional operation on the raw EEG signal, like EEGNet, our proposed graph representation makes use of the spatial relationships and strengths of connectivity of EEG nodes, which facilitates the neural network in differentiating the different mental states. The second advantage of our proposed local structure extraction method is that it can reduce the redundant nodes and edges from the graph obtained from the raw EEG signals. The local structure extracted from the whole graph can hold distinctive structural features, which aids the classification process. Finally, the third advantage of our proposed method is the graph embedding method (Ego-CNNs), which employs the ego-convolutions at each layer and stacks up layers in an ego-centric way to detect important critical structures efficiently.

## V. DISCUSSION

### A. Rationality of the Local Structure Extraction Matrix

The graphs identified from the raw EEG signals have a lot of edges and nodes, whose structures cannot be differentiated easily under different motor imagery tasks. To solve this problem, we designed the local structure extraction matrix  $E_{LS}$ . Under the action of  $E_{LS}$ , the redundant edges and nodes are removed and the most informative structures are preserved. Fig. 7 represents the connectivity characteristics of the whole graph and the local structure extracted by the  $E_{LS}$ . Different colors denote different strengths of connectivity between pairs

TABLE III  
FOUR-CLASS CLASSIFICATION PERFORMANCE COMPARISON OF DIFFERENT METHODS APPLIED ON DATASET 2 (KAPPA VALUE)

Method	S1	S2	S3	S4	S5	S6	S7	S8	S9	Mean±std	<i>p</i> -value
SUSS-SRKDA [46]	0.83	0.51	0.88	0.68	0.56	0.35	0.90	0.84	0.75	0.70±0.19	0.0124
SRLDA [47]	0.84	0.55	<b>0.90</b>	0.71	0.66	0.44	<b>0.94</b>	0.85	0.76	0.74±0.17	0.0264
CBN+SVM [48]	0.69	0.51	0.87	0.85	0.78	0.42	0.54	<b>0.97</b>	0.45	0.68±0.20	0.0148
KPCA+CILK [49]	0.71	0.46	0.76	0.44	0.26	0.37	0.79	0.75	0.61	0.57±0.19	0.0017
sDPLM [50]	0.75	0.49	0.76	0.49	0.34	0.36	0.68	0.76	0.76	0.60±0.18	0.0011
NRSC [51]	0.87	0.62	<b>0.90</b>	0.77	0.62	0.53	0.80	0.82	0.76	0.74±0.13	0.0099
SS-MEMDBF [52]	0.86	0.24	0.70	0.68	0.36	0.34	0.66	0.75	0.82	0.60±0.23	0.0027
DST [53]	0.78	0.59	0.85	0.72	0.67	0.57	0.81	0.86	0.88	0.75±0.12	0.0117
MKSSP [54]	0.90	0.66	0.89	0.72	0.83	0.68	0.90	0.89	0.87	0.82±0.10	0.1745
CCSP [55]	0.72	0.40	0.70	0.55	0.20	0.35	0.66	0.78	0.77	0.57±0.21	0.0012
C2CM [56]	0.83	0.54	0.87	0.56	0.50	0.24	0.86	0.78	0.73	0.66±0.21	0.0070
CNN-LSTM [57]	0.85	0.54	0.87	0.78	0.77	0.66	0.95	0.83	0.90	0.79±0.13	0.1163
SCNN-Bilstm [58]	0.84	0.65	0.73	0.87	0.74	<b>0.69</b>	0.87	0.76	0.81	0.77±0.08	0.0071
Proposed	<b>0.97</b>	<b>0.89</b>	0.77	<b>0.99</b>	<b>0.86</b>	0.65	0.88	0.94	<b>0.94</b>	<b>0.88±0.11</b>	

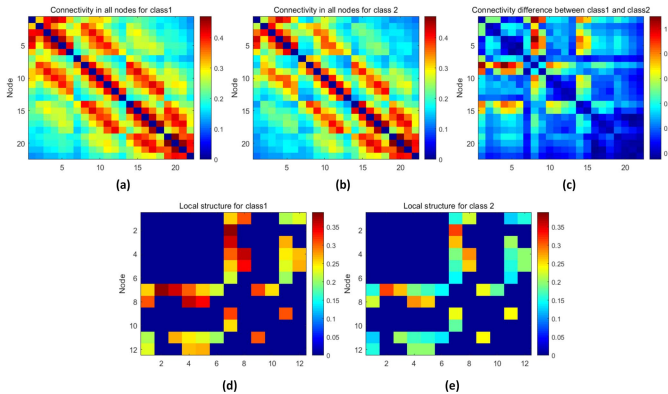


Fig. 7. Connectivity characteristics. (a)(b) represent the functional connectivity in all nodes. (c) represents the connectivity difference for two classes motor imagery tasks. (d) (e) represent local structures extracted by  $E_{LS}$ .

of nodes (electrodes). The value indicated by the color represents the functional connectivity between the corresponding nodes. Red indicates high levels of functional connectivity, blue indicates low levels of connectivity. We randomly selected one participant to plot the connectivity characteristics under different motor imagery tasks.

Fig. 7(a) and (b) represent the connectivity relationship (the adjacency graphs produced without extracting the local structure) between all electrodes under the two kinds of motor imagery tasks from the three datasets. Fig. 7(d) and (e) represent the local structures extracted by  $E_{LS}$  from the Fig. 7(a) and (b). Fig. 7(c) represents the connectivity difference between different tasks. It can be seen that Fig. 7(a) and (b) represent different tasks that are too similar to distinguish, but Fig. 7(d) and (e) represent local structures extracted by  $E_{LS}$  that can be easily distinguished. From Fig. 7, we can also see the effectiveness of using  $E_{LS}$  to extract a local structure for classification under different motor imagery tasks. Simultaneously, the dimensionality of the graph is reduced, which is beneficial for the construction of a classification network. Both low-density and high-density electrode montages can make use of the local structure of the graph to represent features for classification.

## B. Feature Distribution

Discriminative feature of graph features can be extracted from raw EEG data by the Ego-CNN model. In order to

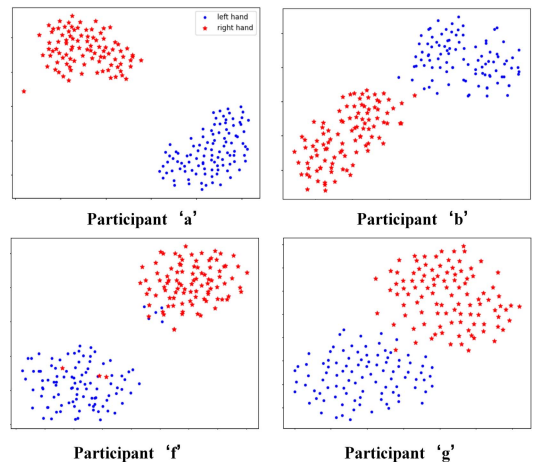


Fig. 8. Visualizations of high-dimensional data of two-class MI EEG features (Dataset 2).

further illustrate the validity of the model, we use data visualization technology to illustrate the extracted features. T-distributed Stochastic Neighbor Embedding (t-SNE) [58] is used to illustrate connectivity graph for each participant, as shown in Fig. 8 and Fig. 9. In the t-SNE scatter plot, the classification categories are represented by different colors and different shapes.

Fig. 8 shows the feature distributions of two-class MI EEG features from Dataset 2. It can be seen that the points of the same category are close and the points of different categories are clearly separated. Fig. 9 shows the feature distributions of four-class MI EEG features from Dataset 1. In the four-class classification task, some data from different categories is hard to identify. For example, the MI EEG feature distributions of left-hand and foot imagery tasks from participant ‘S2’ have some overlapping parts. However, in general, the distributions of four-class MI EEG features are clearly separated.

## C. Comparison With Other Graph Neural Network

We compared our model with classical graph neural network models Graph Convolutional Networks (GCN) [59] and Graph Isomorphism Networks (GIN) [60]. The experiment results are presented in the Table IV. It can be seen from Table IV that the Ego-CNNs model we used in this study has better performance than GCN and GIN. The traditional GCN and GIN



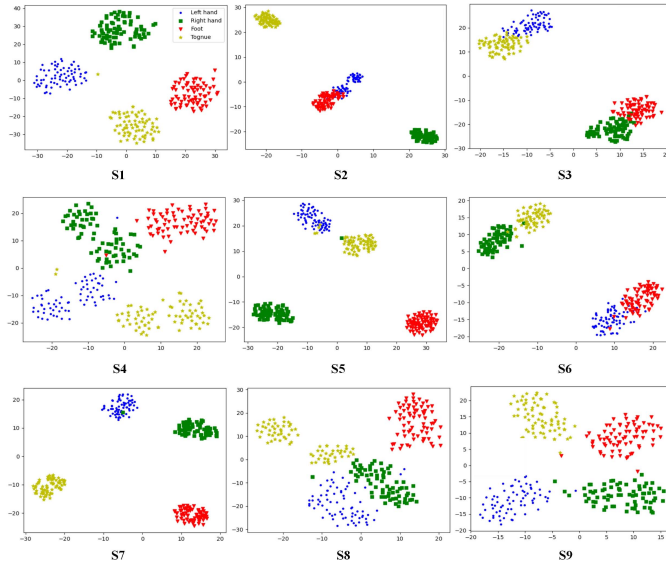


Fig. 9. Visualizations of high-dimensional data of four-class MI EEG features (Dataset 1).

TABLE IV  
CLASSIFICATION ACCURACY FOR DIFFERENT  
GRAPH NEURAL NETWORKS

	GCN	GIN	Ego-CNNs
Dataset 2	75.71%	79.29%	93.40%
Dataset 3	77.14%	77.86%	96.50%

models extract the features of the graphs by aggregating the characteristics of adjacent nodes. They pay more attention to the attributes of nodes and ignore the structural characteristics of connected edges. The Ego-CNNs model employs novel ego-convolutions to learn the latent representations at each network layer, and can efficiently extract task-dependent, important structural features of connected edges. In this study, we convert EEG signal to graph data and extract distinguishable local structures from the graph for classification tasks. In the process of extracting local structures, we preserve the important structural features of the connected edges. Therefore, in terms of graph data used in this study, the Ego-CNNs model is more suitable for the classification task of local structures than GCN and GIN.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel method for motor imagery tasks classification. The method can transform the EEG data to graph data by calculating functional connectivity between pairs of electrodes. A measure of the local graph structure is then extracted by  $E_{LS}$  and used by an Ego-CNNs model for classification. We tested our method on four datasets and the results showed that our proposed method can achieve more than 90% classification accuracy in the first three datasets and more than 80% in dataset 4. This high classification performance is most likely due to the following reasons:

1) The graph representation derived from the EEG signals can describe the synchronous collaboration between different regions of the brain. Functional connectivity is only related to the degree of coupling without direction or causality.

2) The local structure extraction matrix  $E_{LS}$  can effectively extract important local structures under different motor

imagery tasks, which can be used as features and then sent to the Ego-CNNs model for classification.

3) The Ego-CNNs model can detect critical meaningful structures in the graph and this results in good performance in task classification.

In the future, we will optimize our method on a small training set and continue to study functional connectivity during motor imagery tasks. In this paper, the local structural difference between multi kinds of motor imagery tasks was only constructed by measuring connectivity strength (edges in the graph), but the information from the electrodes (the nodes of the graph) wasn't considered. Therefore, in future work, we will consider the information contained in the node itself and continue to seek improvements to our proposed method.

## REFERENCES

- [1] H. E. Savaki and V. Raos, "Action perception and motor imagery: Mental practice of action," *Prog. Neurobiol.*, vol. 175, pp. 107–125, Apr. 2019.
- [2] J. Jin, Z. Chen, R. Xu, Y. Miao, X. Wang, and T.-P. Jung, "Developing a novel tactile P300 brain-computer interface with a cheeks-stim paradigm," (in English), *IEEE Trans. Biomed. Eng.*, vol. 67, no. 9, pp. 2585–2593, Sep. 2020.
- [3] G. Pfurtscheller and C. Neuperb, "Motor imagery activates primary sensorimotor area in humans," *Neurosci. Lett.*, vol. 239, nos. 2–3, pp. 65–68, 1997.
- [4] J. Pan *et al.*, "Prognosis for patients with cognitive motor dissociation identified by brain-computer interface," (in English), *Brain*, vol. 143, no. 4, pp. 1177–1189, Apr. 2020.
- [5] X. Xiao, M. Xu, J. Jin, Y. Wang, T.-P. Jung, and D. Ming, "Discriminative canonical pattern matching for single-trial classification of ERP components," (in English), *IEEE Trans. Biomed. Eng.*, vol. 67, no. 8, pp. 2266–2275, Aug. 2020.
- [6] C. A. Stefano Filho, R. Attux, and G. Castellano, "Can graph metrics be used for EEG-BCIs based on hand motor imagery?" *Biomed. Signal Process. Control*, vol. 40, pp. 359–365, Feb. 2018.
- [7] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," (in English), *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
- [8] 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," (in English), *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.
- [9] 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," (in English), *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4814–4825, Nov. 2021.
- [10] J. Feng *et al.*, "Towards correlation-based time window selection method for motor imagery BCIs," (in English), *Neural Netw.*, vol. 102, pp. 87–95, Jun. 2018.
- [11] M. Xu, X. Xiao, Y. Wang, H. Qi, T.-P. Jung, and D. Ming, "A brain-computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli," (in English), *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1166–1175, May 2018.
- [12] J. Jin *et al.*, "The study of generic model set for reducing calibration time in P300-based brain-computer interface," (in English), *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 3–12, Jan. 2020.
- [13] K. Wang, M. Xu, Y. Wang, S. Zhang, L. Chen, and D. Ming, "Enhance decoding of pre-movement EEG patterns for brain-computer interfaces," (in English), *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016033.
- [14] I. Daly, S. J. Nasuto, and K. Warwick, "Brain computer interface control via functional connectivity dynamics," *Pattern Recognit.*, vol. 45, no. 6, pp. 2123–2136, 2012.
- [15] V. Gonuguntla, Y. Wang, and K. C. Veluvolu, "Phase synchrony in subject-specific reactive band of EEG for classification of motor imagery tasks," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 2784–2787.
- [16] J. S. Kirar and R. K. Agrawal, "A combination of spectral graph theory and quantum genetic algorithm to find relevant set of electrodes for motor imagery classification," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105519.

- [17] 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," (in English), *Cerebral Cortex*, vol. 28, no. 10, pp. 3623–3637, Oct. 2018.
- [18] M. Hassan and F. Wendling, "Electroencephalography source connectivity: Aiming for high resolution of brain networks in time and space," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 81–96, May 2017.
- [19] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature Neurosci.*, vol. 20, no. 3, p. 353, 2017.
- [20] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," (in English), *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010.
- [21] A. Mheich, M. Hassan, M. Khalil, C. Berrou, and F. Wendling, "A new algorithm for spatiotemporal analysis of brain functional connectivity," *J. Neurosci. Methods*, vol. 242, pp. 77–81, Mar. 2015.
- [22] M. Hata *et al.*, "Functional connectivity assessed by resting state EEG correlates with cognitive decline of Alzheimer's disease—An eLORETA study," *Clin. Neurophysiol.*, vol. 127, no. 2, pp. 1269–1278, Feb. 2016.
- [23] R. Ge, H. Zhang, L. Yao, and Z. Long, "Motor imagery learning induced changes in functional connectivity of the default mode network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 1, pp. 138–148, Jan. 2015.
- [24] Q. Gao, X. Duan, and H. Chen, "Evaluation of effective connectivity of motor areas during motor imagery and execution using conditional Granger causality," *NeuroImage*, vol. 54, no. 2, pp. 1280–1288, Jan. 2011.
- [25] L. Xu *et al.*, "Motor execution and motor imagery: A comparison of functional connectivity patterns based on graph theory," *Neuroscience*, vol. 261, pp. 184–194, Mar. 2014.
- [26] T. Tanaka, T. Uehara, and Y. Tanaka, "Dimensionality reduction of sample covariance matrices by graph Fourier transform for motor imagery brain-machine interface," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jun. 2016, pp. 1–5.
- [27] M. Wang, J. Hu, and H. A. Abbass, "BrainPrint: EEG biometric identification based on analyzing brain connectivity graphs," *Pattern Recognit.*, vol. 105, Sep. 2020, Art. no. 107381.
- [28] R.-C. Tzeng and S.-H. Wu, "Ego-CNN: Distributed, egocentric representations of graphs for detecting critical structures," Tech. Rep., 2019.
- [29] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461–464, Jul. 2000.
- [30] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, Dec. 2010.
- [31] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, no. 3, pp. 583–617, 2003.
- [32] M. Pavlovic, "Hadamard product in  $Q(p)$  spaces," *J. Math. Anal. Appl.*, vol. 305, no. 2, pp. 589–598, May 2005.
- [33] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," Tech. Rep., 2017.
- [34] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 2014–2023.
- [35] M. Naem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," *J. Neural Eng.*, vol. 3, no. 3, pp. 208–216, 2006.
- [36] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
- [37] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
- [38] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, Jun. 2000.
- [39] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
- [40] A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. P. Birbaumer, "Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication," *Arch. Phys. Med. Rehabil.*, vol. 82, pp. 1533–1539, Nov. 2001.
- [41] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.
- [42] S. Kumar, A. Sharma, K. Mamun, and T. Tsunoda, "A deep learning approach for motor imagery EEG signal classification," in *Proc. 3rd Asia-Pacific World Congr. Comput. Sci. Eng. (APWC CSE)*, Dec. 2016.
- [43] 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, Oct. 2018, Art. no. 056013.
- [44] B. Xu *et al.*, "Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification," (in English), *IEEE Access*, vol. 7, pp. 6084–6093, 2018.
- [45] L. F. Nicolas-Alonso, R. Corralajo, J. Gomez-Pilar, D. Álvarez, and R. Hornero, "Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain-computer interfaces," (in English), *Neurocomputing*, vol. 159, pp. 186–196, Jul. 2015.
- [46] L. F. Nicolas-Alonso, R. Corralajo, J. Gomez-Pilar, D. Álvarez, and R. Hornero, "Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces," (in English), *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 702–712, Jul. 2015.
- [47] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," (in English), *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.
- [48] K. Sadatnejad and S. S. Ghidary, "Kernel learning over the manifold of symmetric positive definite matrices for dimensionality reduction in a BCI application," (in English), *Neurocomputing*, vol. 179, pp. 152–160, Feb. 2016.
- [49] A. Davoudi, S. S. Ghidary, and K. Sadatnejad, "Dimensionality reduction based on distance preservation to local mean for symmetric positive definite matrices and its application in brain-computer interfaces," (in English), *J. Neural Eng.*, vol. 14, no. 3, Jun. 2017, Art. no. 036019.
- [50] S. U. Kumar and H. H. Inbarani, "PSO-based feature selection and neighborhood rough set-based classification for BCI multiclass motor imagery task," (in English), *Neural Comput. Appl.*, vol. 28, no. 11, pp. 3239–3258, 2017.
- [51] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," (in English), *Expert Syst. Appl.*, vol. 95, pp. 201–211, Nov. 2018.
- [52] S. Razi, M. R. K. Mollaei, and J. Ghasemi, "A novel method for classification of BCI multi-class motor imagery task based on Dempster-Shafer theory," (in English), *Inf. Sci.*, vol. 484, pp. 14–26, May 2019.
- [53] S. Galindo-Noreña, D. Cárdenas-Peña, and Á. Orozco-Gutierrez, "Multiple kernel stein spatial patterns for the multiclass discrimination of motor imagery tasks," (in English), *Appl. Sci.*, vol. 10, no. 23, p. 8628, Dec. 2020.
- [54] 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," (in English), *PLoS ONE*, vol. 16, no. 3, Mar. 2021, Art. no. e0248511.
- [55] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.
- [56] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *J. Neural Eng.*, vol. 16, no. 6, Oct. 2019, Art. no. 066004.
- [57] S. Lian, J. Xu, G. Zuo, X. Wei, and H. Zhou, "A novel time-incremental end-to-end shared neural network with attention-based feature fusion for multiclass motor imagery recognition," (in English), *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–16, Feb. 2021.
- [58] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Spidlen, and J. E. Snyder-Cappione, "Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets," (in English), *Nature Commun.*, vol. 10, no. 1, pp. 1–12, Nov. 2019.
- [59] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
- [60] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–17.