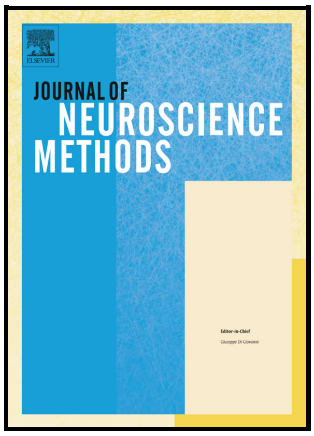
Feature learning framework based on EEG graph self-attention networks for motor imagery BCI systems

Hao Sun, Jing Jin, Ian Daly, Yitao Huang, Xueqing Zhao, Xingyu Wang, Andrzej Cichocki



PII: S0165-0270(23)00188-7

DOI: https://doi.org/10.1016/j.jneumeth.2023.109969

Reference: NSM109969

To appear in: *Journal of Neuroscience Methods*

Received date: 26 May 2023
Revised date: 18 August 2023
Accepted date: 3 September 2023

Please cite this article as: Hao Sun, Jing Jin, Ian Daly, Yitao Huang, Xueqing Zhao, Xingyu Wang and Andrzej Cichocki, Feature learning framework based on EEG graph self-attention networks for motor imagery BCI systems, *Journal of Neuroscience Methods,* (2023) doi:https://doi.org/10.1016/j.jneumeth.2023.109969

Available online at www.sciencedirect.com

ELSEVIER

00 (2012) 000–000

Feature learning framework based on EEG graph self-attention networks for motor imagery BCI systems

Hao Sun[a], Jing Jin[a,b,*], Ian Daly[c], Yitao Huang[a], Xueqing Zhao[a], Xingyu Wang[a], Andrzej Cichocki[d,e]

[a]*Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai, China.*
[b]*Shenzhen Research Institute of East China University of Technology, Shen Zhen, 518063, China.*
[c]*Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, U.K.*
[d]*RIKEN Brain Science Institute, Wako 351-0198, Japan.*
[e]*Nicolaus Copernicus University (UMK), 87-100 Torun, Poland.*

Abstract

Learning distinguishable features from raw EEG signals is crucial for accurate classification of motor imagery (MI) tasks. To incorporate spatial relationships between EEG sources, we developed a feature set based on an EEG graph. In this graph, EEG channels represent the nodes, with power spectral density (PSD) features defining their properties, and the edges preserving the spatial information. We designed an EEG based graph self-attention network (EGSAN) to learn low-dimensional embedding vector for EEG graph, which can be used as distinguishable features for motor imagery task classification. We evaluated our EGSAN model on two publicly available MI EEG datasets, each containing different types of motor imagery tasks. Our experiments demonstrate that our proposed model effectively extracts distinguishable features from EEG graphs, achieving significantly higher classification accuracies than existing state-of-the-art methods.

Keywords: Motor imagery (MI), electroencephalogram (EEG), feature learning, graph representation, self-attention

1. Introduction

Brain-Computer Interface (BCIs) can provide an alternative communication pathway without the need for any muscle and peripheral nerve activation(Wolpaw, 2007). The BCI technology has been applied in many fields, such as, but not limited to, medical rehabilitation, medical diagnosis, military applications (Kotchetkov et al., 2010), gaming (Wang et al., 2019), intelligent applications control (Tang et al., 2018) and driving assistance (Jafarifarmand and Badamchizadeh, 2019).

* Corresponding author. Jing Jin.
E-mail address: jinjingat@gmail.com.

Motor imagery (MI) is a classic paradigm used in building BCI systems, in which the participant imagines a movement and the brain activity related to this imagined movement is identified and translated into a control action (Pfurtscheller and Neuper, 2001). MI-based BCI systems have exhibited significant performance for stroke rehabilitation and for assisting patients with movement disorders (Ang et al., 2015). When people imagine their body movements without any actual movement, two phenomena of energy change occur in the sensorimotor regions of the contralateral hemisphere and ipsilateral hemisphere in the brain (Pfurtscheller and Neuper, 1997). These phenomena are called the event-related synchronization (ERS) and event-related desynchronization (ERD) respectively (Neuper et al., 2006). Building a BCI system includes signal acquisition, preprocessing, feature extraction, and feature classification. Learning distinguishable features and selecting suitable classifiers are key steps in the development of MI-based BCI systems, and the correct extraction and classification of features can significantly improve performance.

Within MI-BCI systems the power spectral density (PSD) is, arguably, the most popular feature used for motor imagery classification(Demuru et al., 2020). In many BCI systems, extraction of the PSD features is usually combined with the channel selection method (Jin et al., 2019), because not all EEG channels can provide distinguishable information in terms of PSD features. However, the drawback of the traditional PSD feature set is that it ignores the structure and spatial information available in the EEG(Sun et al., 2021a). Consequently, the wide-spread use of this feature extraction method may neglect considerable distinguishable information available within raw EEG data (Sun et al., 2021b).

Machine learning and deep learning technologies have been wildly used in the motor imagery classification tasks. Common spatial pattern (CSP) (Pfurtscheller and Neuper, 2001) is a traditional feature extraction method, which maximize the covariance of the EEG from two kinds of motor imagery tasks. Filter bank common spatial pattern (FBCSP) is an improvement of CSP algorithm by splitting the EEG data with different frequency band and selecting feature with mutual information (Ang et al., 2008). Convolutional neural networks (CNNs) are capable of extracting features from raw EEG signals in both the temporal and frequency domains. Several CNN-based models for feature extraction, such as EEGNet (Lawhern et al., 2018), ETRCNN(Xu et al., 2020), EEG-Inception (Zhang et al., 2021), and others, have been proposed and achieved good performance in motor imagery classification tasks. However, most of these models do not consider the physical position information of EEG channels, which is also important for improving the classification accuracy.

A graph can be used to present structural information about a dataset via nodes and edges, and can be used to describe features extracted from raw EEG signals (Stefano Filho et al., 2018). When used to describe EEG data the EEG channels are defined by graph nodes, while the edges within the graph are often used to represent a measure of functional or effective connectivity between EEG channels (Xu et al., 2014). Graph neural networks (GNNs) can be used to extract features from EEG-based graphs and improve the classification performance of MI-BCI systems(Jin et al., 2021). Most of the GNNs used for BCIs only focus on the structural information of EEG graphs and ignore the properties of the nodes, but we consider both the nodes' properties and the structural characteristics of the graphs in this study. We used the PSD features extracted from individual EEG channels as the properties of the nodes and redesigned the EEG graph as the input to the graph neural networks. We focused on exploring effective feature learning framework to extract distinguish features for commonly used classifiers in the BCI field: support vector machines (SVMs) and logistic regression classifiers.

Moreover, to train a model for motor imagery classification, a substantial amount of accurately labeled training data is required. However, during the data collection process, missing labels may occur due to software or hardware malfunctions. To address this issue, we propose a feature extraction framework capable of extracting features from EEG signals in the absence of label information.

This paper makes three significant contributions. First, we introduce a novel EEG-based graph that incorporates both node properties and structural information derived from the EEG channels. Node properties are defined using PSD features, and edges are defined based on channel position. Second, we propose a feature extraction

model called the EEG Graph Self-Attention Network (EGSAN), which can extract distinguishable features from EEG-based graphs. Third, we apply an unsupervised feature extraction strategy to the networks in an experiment setting that considers the unlabeled condition.

2. Method

2.1. Graph Representation

We define a graph $\mathcal{G} = (V, E)$, where $\{v_i | v_i \in V\}$ and $\{e_{ij} | e_{ij} \in E\}$ represent the set of nodes v_i and the set of edges e_{ij} between node v_i and v_j, respectively. The adjacency matrix $a_{ij} \in A \in \mathbb{R}^{n \times n}$ can describe the topological structure of a graph with n nodes. Additionally, we define $p_{v_i} \in P \in \mathbb{R}^{n \times d}$ to represent the properties of each node v_i, where d denotes the dimension of the node attributes. Consequently, a graph $\mathcal{G} = (V, E)$ can also be described as $\mathcal{G} = (A, P)$. According to the adjacency matrix A, we can define a set $\mathcal{N}_v$ of neighbors for each node $v \in V$. Figure 1 depicts the structure of the EEG-based graph, showcasing the corresponding adjacency matrix and the properties of a simple graph consisting of five nodes. The different colors represent nodes with distinct properties. The adjacency matrix, denoted as A, has a dimension of (5×5). P represents the node properties, with a dimension of $(5 \times d)$, where d refers to the dimension of properties.

In the case of EEG, the recording electrodes can be seen as the nodes, but the edges of graph are defined by the spatial position of electrodes.

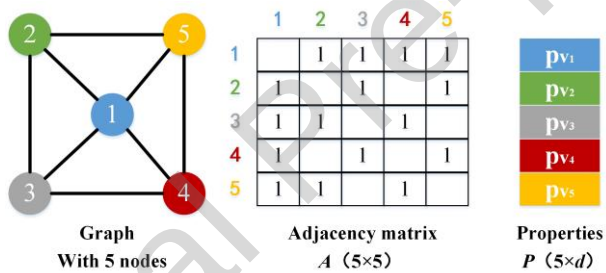


Figure 1. Graph data with adjacency matrix and properties.

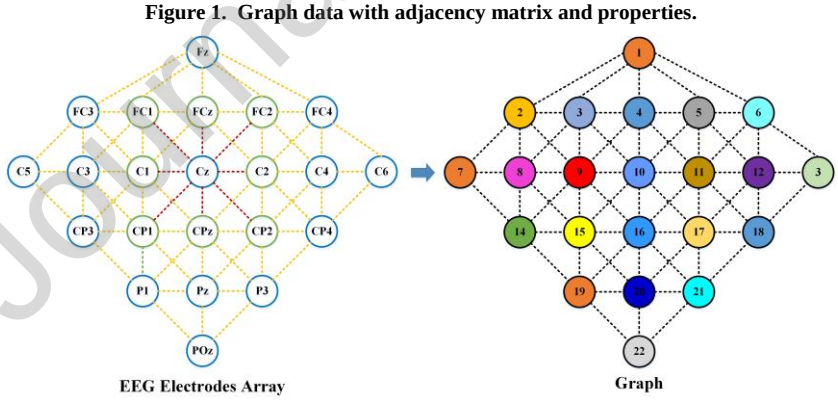


Figure 2. An example graph structure of 22 EEG electrode channels.

2.2. From EEG to Graph

In this study, we establish a brain graph to capture the characteristics of EEG signals during motor imagery tasks. The nodes of the graph correspond to EEG channels obtained from an electrode array, while the edges represent the physical distances between neighboring channel pairs. We assume that each node is connected to its immediate neighbors, defined as those channels with a one-step distance. Figure 2 illustrates the EEG electrode array and the corresponding EEG-based graph with 22 nodes. As depicted on the left side of Figure 2, each node can have a maximum of eight neighbors. For instance, electrode Cz has eight neighboring nodes (FC1, FC2, C1, C2, FCz, CPz, CP1, and CP2). On the right side of Figure 2, the 22 nodes of the EEG-based graph are color-coded to represent their distinct properties. During each trial of a motor imagery experiment, we calculate the power spectral density (PSD) of each EEG channel using the Welch spectrogram method. These resulting PSD values are then used to define the properties of the nodes. By utilizing the EEG graph, we can capture the PSD characteristics of electrodes while retaining the spatial information through the graph edges.

2.3. EEG based Graph Self-Attention Networks

Extracting suitable features plays a significant role in classification tasks. In order to classify the different motor imagery tasks, we need to extract the feature from the EEG graph, which can be sent to classifier to get the final classification result. In the process of experiment, the labels may be absent because of the equipment broken or negligence of technician, we only get some motor imagery EEG data without label. It is meaningful to design a unsupervised feature learning model to solute this situation. In this study we designed a feature learning model to extract feature from EEG based graph with unsupervised training.

Graph embedding is a popular technique used for extracting feature from graph, it can learn low-dimensional vector representations for graphs. Graph neural networks (GNNs) have become an essential method for graph embedding, which use an aggregation function and a graph pooling function to obtain graph vector representation (Ying et al., 2018). In general, the GNNs update the node feature vector by transforming and aggregating the feature vector of neighboring nodes with an aggregation function. The resulting graph embedding can be obtained after a graph pooling function. GNN based methods have excellent performance on graph feature extraction and classification tasks.

The self-attention mechanism in transformers (Vaswani et al., 2017) has been widely used across domains such as natural language processing and computer vision, and has been shown to improve the performance of graph neural network (GNN) models when used to build aggregation functions (Nguyen et al., 2022). In this study, we propose applying self-attention to calculate attention coefficients between nodes in EEG graphs, which can capture synergistic characteristics among different channels during motor imagery tasks. Our novel approach utilizes GNNs with self-attention to extract features from EEG graphs for motor imagery classification tasks. To enhance the feature extraction process, we introduce an advanced EEG graph self-attention network (EGSAN), composed of a self-attention module, a recurrent transition module, residual connections, and layer normalization modules.

Given a graph $\mathcal{G} = (V, E)$, and set of neighbors $\mathcal{N}_v$ for each node $v \in V$. The set $\{\mathcal{N}_v \cup v\}$ is the input to the graph transformer self-attention network. For each training batch, we sample a different neighbor set for node v. For the structure of the EGSAN, we consider using multiple layers stacked on top of each other to improve the node embedding. Given a node $v \in V$, the embedding of v in the k-th layer can be obtained by a novel transformer based aggregation function as:

$$p_v^{(k)} = TransAggregation\left(p_u^{(k-1)}\right) \tag{1}$$

where $u \in \{\mathcal{N}_v \cup v\}$. The $TransAggregation$ function can be divided to two parts as:

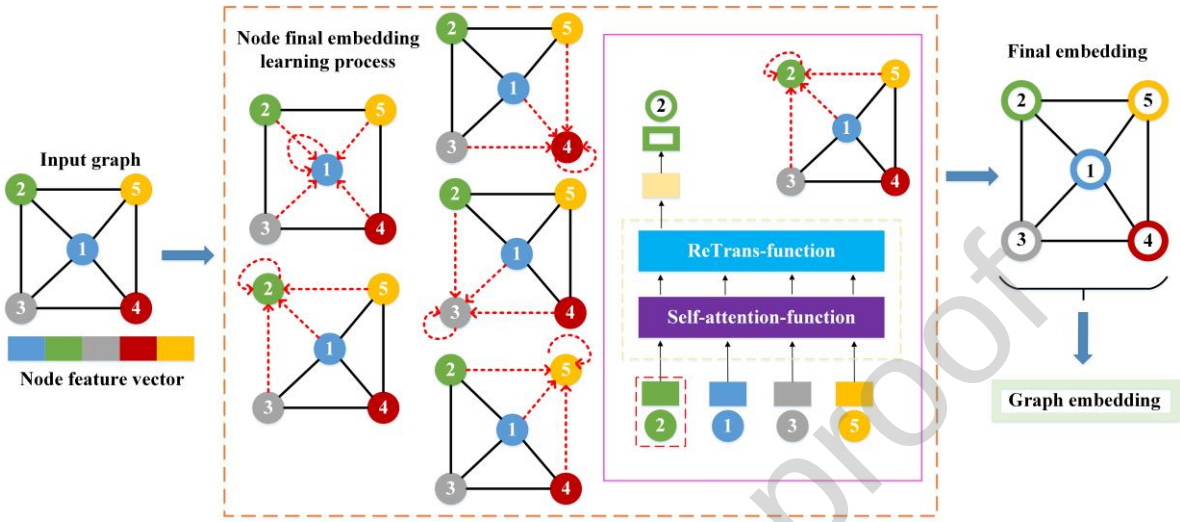


Figure 3. Example of applying the EGSAN model to a five node graph.

$$x_u^{(k)} = LayerNorm\left(p_u^{(k-1)} + Attention\left(p_u^{(k-1)}\right)\right) \quad (2)$$

$$p_v^{(k)} = LayerNorm\left(x_u^{(k)} + ReTrans\left(x_u^{(k)}\right)\right) \quad (3)$$

where $p_u^{(k)} \in \mathbb{R}^d$. $LayerNorm$ is layer normalization function, which can be used to normalize the inputs across feature dimensions to stabilize the network, to make gradients smoother, and to decrease the training time. The term $ReTrans(.)$ denotes a multiple layer perceptron network with two fully connected layers as:

$$ReTrans\left(x_u^{(k)}\right) = W_2^{(k)} Relu\left(W_1^{(k)} x_u^{(k)} + b_1^{(k)}\right) + b_2^{(k)} \quad (4)$$

where $W_1^{(k)} \in \mathbb{R}^{s \times d}$ and $W_2^{(k)} \in \mathbb{R}^{d \times s}$ are weight matrices, and $b_1^{(k)}$ and $b_2^{(k)}$ are bias parameters. The term $Attention(.)$ denotes a self-attention neural network as:

$$Attention\left(p_u^{(k-1)}\right) = \sum_{u' \epsilon \{\mathcal{N}_v \cup v\}} \alpha_{u,u'}^{(k)} \left(V^{(k)} p_{u'}^{(k-1)}\right) \quad (5)$$

where $V^{(k)} \in \mathbb{R}^{d \times d}$ is a value-projection weight matrix. The term $\alpha_{u,u'}$ is an attention coefficient calculated by the softmax function and dot products between nodes u and u'

$$\alpha_{u,u'}^{(k)} = softmax\left(\frac{\left(Q^{(k)} p_u^{(k-1)}\right)^T (K^k) p_{u'}^{(k-1)}}{\sqrt{d}}\right) \quad (6)$$

where $Q^{(k)} \in \mathbb{R}^{d \times d}$ is the query-projection matric and $K^k \in \mathbb{R}^{d \times d}$ is the key-projection matrix. The features extracted by the k-th layer of all nodes in a graph can be described as:

$$P^{(k)} = Attention_{\mathcal{N}_v \cup v}\left(P^{(k-1)}Q^{(k)}, P^{(k-1)}K^{(k)}, P^{(k-1)}V^{(k)}\right) \tag{7}$$

Specially, $p_v^{(0)} \in \mathbb{R}^d$ is the feature vector of node v. We concatenate the feature vectors across the layers to obtain the node embedding $\boldsymbol{e}_v$ of the node v as:

$$\boldsymbol{e}_v = \left[p_v^{(1)}; p_v^{(2)}; \dots; p_v^{(K)}\right] \tag{8}$$

In this paper, we train the feature learning frame without any labels. This approach is suitable for the situation where there are insufficient class labels available. We assume a final embedding $\boldsymbol{f}_v$ for each node v, and the similarity between $\boldsymbol{f}_v$ and $\boldsymbol{e}_v$ should be higher than the similarity between $\boldsymbol{e}_v$ and the final embedding of the other nodes. The aim of this setting is to let the graph neural network identify and discriminate the sub-graph structural characteristics within each graph and remember the differences of structural characteristics among the graphs. We used the sampled soft-max loss function(Jean et al., 2014) as:

$$\mathcal{L}(v) = -log\frac{exp(\boldsymbol{f}_v{}^{\top}\boldsymbol{e}_v)}{\sum_{v' \epsilon V'} exp(\boldsymbol{f}_{v'}{}^{\top}\boldsymbol{e}_v)} \tag{9}$$

where V' is a subset sampled from $\{\cup V_m\}_{m=1}^{M}$. The final embedding $\boldsymbol{f}_v$ is learned implicitly as model parameters. The graph embedding $\boldsymbol{e}_{\mathcal{G}}$ is obtained by summing all the final embeddings $\boldsymbol{f}_v$ of nodes v in the graph $\mathcal{G}$. The whole process of the graph embedding is presented in Algorithm 1. To illustrate our EGSAN model, we give an example of applying the EGSAN model to a single graph with five nodes. The detailed process of node final embedding and graph embedding can be found in the figure 3.

Algorithm 1:Graph embedding with EGSAN

for $k = 0, \dots, K-1$ **do**
  for $v \in V$ **do**
    Sample $\mathcal{N}_v$ for v
    $\forall u \in \{\mathcal{N}_v \cup v\}$
    $x_u^{(k)} \leftarrow LayerNorm\left(p_u^{(k-1)} + Attention\left(p_u^{(k-1)}\right)\right)$
    $p_v^{(k)} \leftarrow LayerNorm\left(x_u^{(k)} + ReTrans\left(x_u^{(k)}\right)\right)$
$\boldsymbol{e}_v \leftarrow \left[p_v^{(1)}; p_v^{(2)}; \dots; p_v^{(K)}\right]$
$\boldsymbol{f}_v \leftarrow \boldsymbol{e}_v$
$\boldsymbol{e}_{\mathcal{G}} \leftarrow \sum_{v \in V} \boldsymbol{f}_v$

2.4. Classification

The EGSAN model is designed solely for feature extraction, and as such, the obtained features must be classified using a separate classifier to obtain the classification results. To this end, we employed a support vector machine (SVM) classifier, a commonly used algorithm in motor imagery tasks, to classify the graph embedding $\boldsymbol{e}_{\mathcal{G}}$, which serves as the extracted feature from the graph. Our SVM classifier was built with a linear kernel and utilized a one-versus-rest (OVR) strategy for multi-class classification.

2.5. Whole experiment framework

Figure 4 depicts the comprehensive experimental framework, which encompasses several steps. First, the power spectral density of each channel is computed to derive node properties, and EEG signals are transformed into graph data. The graph data is represented by an adjacency matrix along with node properties. Next, the resulting graph data is fed into the EGSAN model to extract features and generate graph embeddings. These graph embeddings are subsequently utilized as input for a SVM classifier, facilitating the classification process.

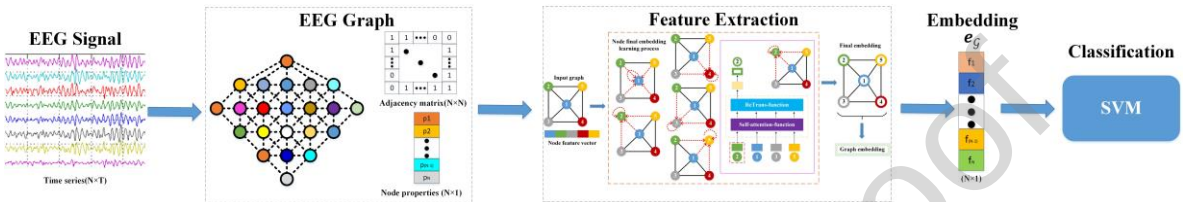


Figure 4. The entire experiment framework.

3. Experiments

In this section, we describe our experiments on two publicly available MI EEG datasets that are commonly used in EEG MI classification to evaluate the effectiveness of our proposed feature learning method.

3.1. Dataset 1

The first dataset we used to evaluate our model is the GinaDB dataset (Cho et al., 2017), which was recorded from 52 healthy participants. Among the participants, 33 are male and the rest are female. All participants gave written informed consent to collect information on brain signals and the dataset was approved by the Institutional Review Board of Gwangju Institute of Science and Technology. Every participant was required to perform two kinds of motor imagery tasks (left and right-hand motor imagery). The EEG signals were recorded by 64 Ag/AgCl active electrodes, which were placed according to the international 10-10 system. Each participant performed 100 or 120 trials of every kind of MI task, and all EEG signals were recorded with a sampling rate of 512Hz. The data set can be download from the website: http://gigadb.org/dataset/100295.

3.2. Dateset 2

The second dataset we used to evaluate our model is the BCI competition Ⅳ Dataset Ⅱa (Naeem et al., 2006), which consists of EEG recorded from nine healthy participants. The participants were right-handed, had normal or corrected-to-normal vision and were paid for participating in the experiments. The EEG signals were recorded via 22 electrodes with a sampling rate of 250Hz and band-pass filtered between 0.5Hz and 100Hz. Each participant was required to perform four kinds of motor imagery tasks according to visual cues. The four kinds of MI tasks are left-hand MI, right-hand MI, both feet MI, and tongue MI. Each participant performed 288 trials in total, and the proportion among different MI tasks is balanced. The data set can be download from the website: http://www.bbci.de/competition/iv /.

3.3. Data processing

All EEG data taken from the two datasets were band-pass filtered using a fifth-order Butterworth filter from 0.5Hz to 30Hz. According to the characteristic of the paradigm, we use different time windows for different datasets. For dataset 1, we took a 2 seconds time window from 0.5s after the cue presentation time until 2.5s after the cue presentation time, then we took the mean-subtraction process for every trial data. To get useful signal of motor imagery task and limit the size of the graph, we only used 18 channels near to the motor cortex (FC1, FC2, FC3, FC6, Cz, C1, C2, C3, C4, C5, C6, CPz, CP1, CP2, CP3, CP4, CP5, and CP6). For dataset 2, we took a 3 seconds time window from 0.5s after the cue presentation time until 3.5s after the cue presentation time, and all 22 channels were used.

The EEG signal of each trial were converted to one graph, with the number of nodes equal to the number of channels used. We used the sum of PSD calculated from α and β frequency bands from each of the EEG channels as the node properties, and the shape of nodes initial properties is $(n \times 1)$ the n represents the number of EEG channels.

3.4. Model setting

We vary the number of K of EGSAN layers in $\{1,2,3\}$. We set the number of neighbors to 6, set the hidden size in $ReTrans(.)$ to 1024, and the batch size is set to 5. We apply the Adam optimizer to train our EGSAN and select Adam initial learning rate $\mathrm{lr} = 0.005$. We run up 50 epochs to evaluate our EGSAN.

4. Results

4.1. Label limited experiment

To evaluate the feature extraction model EGSAN, we conducted experiments considering different label-lacking conditions. In this section, we assumed a scenario where motor imagery EEG data had a limited number of labels. We transformed all EEG data into graph data and applied the EGSAN model to extract features through an unsupervised feature learning process. Subsequently, we constructed an SVM classifier using a restricted amount of labeled data, varying the proportion of correctly labeled data from 10% to 90% across five different ratios. For instance, when only 10% of the data had accurate labels and 90% lacked labels, only 10% of the extracted features from the EEG data with correct labels were employed for training the SVM classifier, while the remaining 90% without labels were utilized as test data.

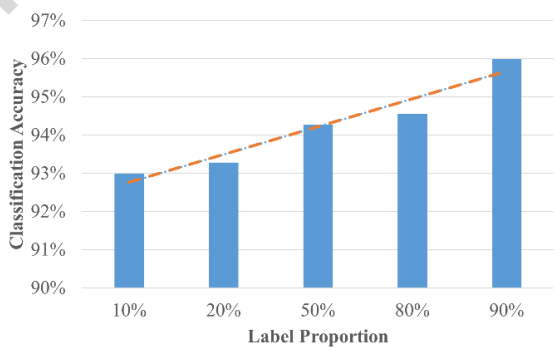


Figure 5. Classification accuracy of dataset1 with different label proportion

We evaluated our EGSAN model using classification accuracy as the evaluation metric for the binary classification task on dataset 1. As shown in figure 5, even with only 10% labeled data, the SVM classifier achieved more than 90% classification accuracy. Furthermore, the trendline shows that as the proportion of labeled data increased, the classifier's performance improved.

We evaluate the performance of our EGSAN model on the multiple classification task using dataset 2, using kappa value and classification accuracy as the evaluation metric. The advantage of kappa value in this context is that it provides a standardized measure of inter-classifier agreement that takes into account the complexities of the task, including the possibility of class imbalance and varying levels of complexity among the classes. As shown in figure 6, even with only 10% labeled data, the SVM classifier achieved more than 86% classification accuracy and 0.8 kappa value. Furthermore, as the proportion of labeled data increased, the kappa value and classification accuracy improved.

The experiments in figure 5 and 6 demonstrate that EGSAN can extract distinguishable features from motor imagery EEG signals through an unsupervised feature learning process. Therefore, EGSAN is well-suited for offline data analysis of motor imagery tasks under label-limited conditions.

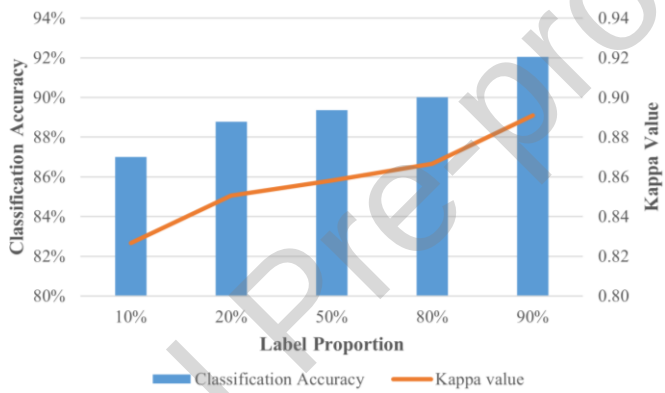


Figure 6. Classification accuracy and kappa value of dataset 2 with different label proportion

4.2. Comparable experiment

We conducted experiments to compare the effectiveness of our EGSAN model with other recently proposed models for extracting distinguishable features from EEG signals. In this section, we initially partitioned the data into a training set (90%) and a test set (10%). During the training phase, our EGSAN model was trained for 50 epochs exclusively on the training data, deviating from the label-limited experimental setup. Additionally, we trained an SVM classifier using the training data. Once both the EGSAN model and SVM classifier were fully trained, we evaluated their performance on the test data using classification accuracy and kappa value as performance metrics. To assess the performance of our model, we applied it to binary motor imagery (MI) tasks (dataset 1) and four-class MI tasks (dataset 2). We used the same datasets to compare our model with the following other models proposed in recent papers:

1) CSP+SVM: (Wang et al., 2006) A baseline algorithm that combined common spatial patterns with SVM classifier for binary motor imagery classification. CSP algorithm is the traditional feature extraction method in MI tasks.
2) OPTICAL:(Kumar et al., 2019) This is predictor model, which consists of CSP feature extraction module and a long-short memory (LSTM) neural network to classify the motor imagery tasks.
3) VMD+LR: (Sadiq et al., 2021) The approach utilizes both the Variational Mode Decomposition method and the Linear Regression (LR) feature selection method.

4) STR: (Rodrigues et al., 2019) This is a space-time recurrence-based (STR) alternative for estimating EEG brain functional connectivity used in motor imagery tasks, which takes into account the recurrence density between pairwise electrodes over a time window.
5) SICR-EEGNet: (Jeon et al., 2021) A deep neural network that learns subject-invariant and class-relevant representation via mutual information estimation among features in different levels for BCI tasks in an end to end manner. But in this paper, the SICR-EEGNet was compared in the subject specific experiment setting.
6) KCS-FCnet: (García-Murillo et al., 2023) The Kernel Cross-Spectral Functional Connectivity Network (KCS-FCnet) method employs a single 1D-convolutional neural network to extract temporal-frequency features from raw EEG data, along with a cross-spectral Gaussian kernel connectivity layer that models functional relationships among channels.
7) CNN+LSTM: (Zhang et al., 2019) This method includes a one-versus-rest FBCSP module, for pre-extracting features from MI EEG signals, and a hybrid deep neural network based on a convolutional neural network (CNN) and LSTM to learn spatial and temporal features.
8) DSCNN: (Ma et al., 2022) A shallow double-branch network, which has two different branches to extract more abundant features related to motor imagery signals.
9) ETRCNN: (Xu et al., 2020) A novel EEG topographical representation energy calculation method for learning EEG patterns of brain activities using a CNN
10) EEG-TCNet: (Ingolfsson et al., 2020) A temporal convolutional network with a low memory foot-print and low computational complexity that means it can be realized in resource-limited devices.
11) EEG-Inception: (Zhang et al., 2021) An end-to-end model built on the backbone of the inception-time network. It takes the raw EEG signals as its input.
12) AMSI-EEGNet: (Riyad et al., 2021) The input of this model can be sampled at different sampling frequencies, allowing to extract features thar are specific to each scale. Each scale are processed with an auxiliary convolutional blocks that have different hyperparameters but with the same structure. The multi-scale features can be used by the main network with aggregate function for motor imagery classification.
13) DST: (Razi et al., 2019) A novel classifier fusion module by employing Dempster-Shafer theory, which can solve the problem that the traditional CSP algorithm is not suitable for multi-class classification.
14) Mcnn: (Amin et al., 2019) A multi-layer CNNs method for fusing CNNs, which utilizes different convolutional layers to capture the spatial and temporal feature from EEG signal.
15) MKSSP: (Galindo-Noreña et al., 2020) A method named Multiple Kernel Stein Spatial Patterns maps EEG signal into low-dimensional covariance matrices preserving the nonlinear channel relationships, and stein kernel provides a parameterized similarity metric for covariance matric.

We compare these 15 models with our EGSAN+SVM method.

4.3. Binary MI Classification

In the comparable experiment, Figure 7 illustrates the classification accuracy of each participant in dataset 1 using the EGSAN+SVM model. Different colors indicate varying ranges of classification accuracy. Upon examining individual performance, it is evident that our proposed method performs exceptionally well on dataset 1, with all participants achieving over 75% accuracy, thus ensuring the feasibility of the motor imagery BCI system. Specifically, five participants achieved more than 90% classification accuracy (represented by the orange color), while 31 participants achieved classification accuracy between 85% and 90% (represented by the yellow color). Fifteen participants achieved classification accuracy ranging from 80% to 85% (represented by the blue color), and only one participant achieved below 80%. These results demonstrate the effectiveness of our proposed EGSAN model in extracting distinguishable features from graph data converted from EEG signals in binary motor imagery classification tasks.

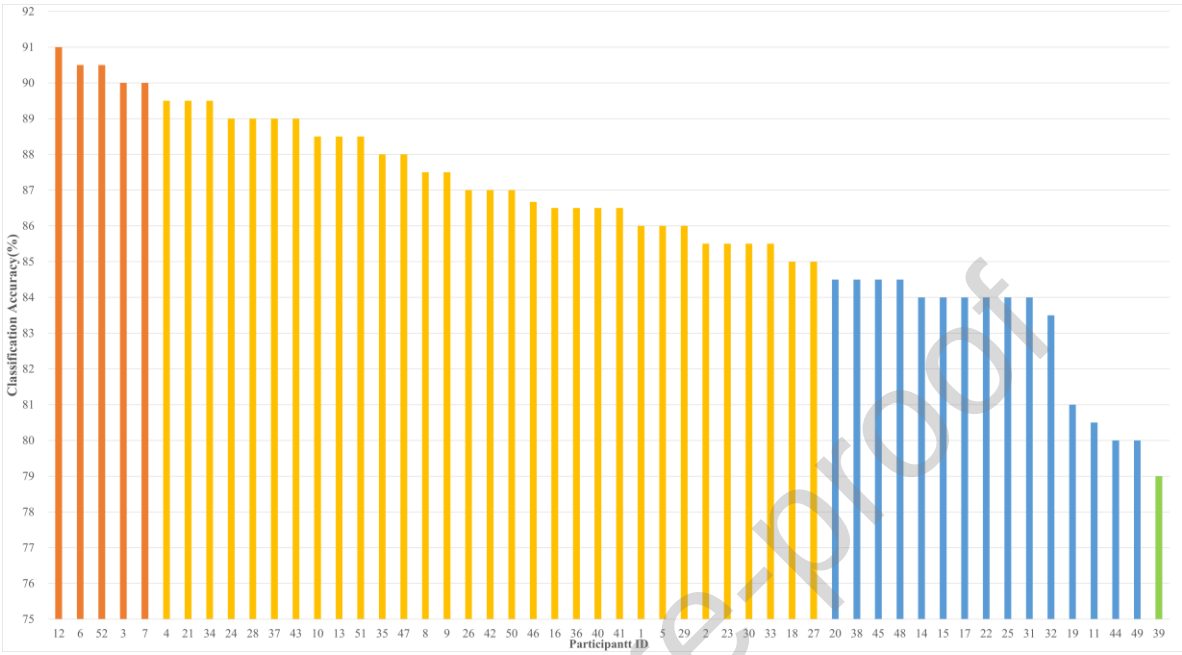


Figure 7. Classification Accuracy of dataset 1 in the comparable experiment with EGSAN+SVM.

Table 1 summarizes the mean classification accuracy of comparable traditional machine learning or deep learning methods for MI classification on dataset 1. Results indicate that the baseline CSP+SVM method achieved only 60% accuracy, while hybrid models based on CNNs, such as OPTICAL, improved accuracy to 68.89%. Both STR and KCS-FCnet extract features by calculating functional connectivity between pairwise electrodes, achieving an accuracy of 76%. The best method among the comparable ones is VMD+LR, with 85.02% classification accuracy. However, our proposed EGSAN model with an SVM classifier achieved 86.21% classification accuracy, surpassing VMD+LR by 1.19%. The last line of Table 1 shows the pair-test result, indicating that our proposed EGSAN+SVM is significantly ($p<0.05$) superior to other methods except the VMD+LR method.

Table 1 The mean classification accuracy of dataset 1 with different methods

Method	CSP+SVM	OPTICAL	STR	SICR_EEGNet	VMD+LR	KCS-FCnet	EGSAN+SVM
Mean	60.68%	68.89%	76.00%	76.60%	85.02%	76.40%	86.21%
Std	±15.74%	±9.36	±12.00%	±12.48%	±7.29%	±11.30	±2.88%
p-value	<0.05	<0.05	<0.05	<0.05	0.27	<0.05	-

Our proposed model outperforms the aforementioned excellent models in terms of classification accuracy. Additionally, the standard deviation of our proposed models is lower, indicating stronger robustness than the other models. These results demonstrate that our proposed models are highly competitive in comparison to these models.

4.4. Multi-class MI classification

To evaluate the performance of our proposed feature learning framework on multi-class MI classification tasks, we applied our models to dataset 2. Table 2 presents a comparison between the performance of our proposed models and that of competing models for four-class MI EEG signal decoding tasks, using the kappa value as the evaluation metric for multi-class classification performance. Our proposed EGSAN+SVM method achieved an excellent average kappa value of 0.862 ± 0.02 (Mean ± std), outperforming most models. Among the nine participants, four (S2, S4, S5, S9) obtained the highest kappa value with our proposed models, while the remaining five obtained the highest kappa values with various competing methods. We conducted paired t-tests for the competing methods and the p-values obtained are displayed in the last column of Table 2. The p-values for most competing methods (DST, AMSI-EEGNet, Mcnn) are less than 0.05, indicating that EGSAN+SVM significantly outperforms these methods, while the p-values for the other methods indicate that our proposed EGSAN+SVM method competes closely with them.

Table 2 Multi-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	p-value
CNN-LSTM	0.85	0.54	0.87	0.78	0.77	0.66	**0.95**	0.83	0.90	0.794±0.13	0.1493
ETRCNN	0.84	0.66	0.88	0.76	0.57	**0.89**	0.81	0.89	0.91	0.801±0.12	0.2266
DSCNN	**0.90**	0.63	0.93	0.75	0.72	0.60	0.94	0.86	0.84	0.797±0.13	0.1854
DST	0.78	0.59	0.85	0.72	0.67	0.57	0.81	0.86	0.88	0.748±0.12	0.0261
EEG-inception	0.75	0.72	0.92	0.74	0.73	0.70	0.89	0.88	0.86	0.799±0.09	0.0981
EEG-TCNet	0.86	0.63	**0.97**	0.68	0.78	0.61	0.91	0.82	0.80	0.784±0.12	0.1269
AMSI-EEGNet	0.79	0.40	0.86	0.60	0.55	0.49	0.92	0.80	0.74	0.680±0.18	0.0212
Mcnn	0.87	0.51	0.86	0.62	0.50	0.30	0.88	0.78	0.76	0.680±0.20	0.0270
CBN+SVM	0.69	0.51	0.87	0.85	0.78	0.42	0.54	**0.97**	0.45	0.676±0.20	0.0351
MKSSP	**0.90**	0.66	0.89	0.72	0.83	0.68	0.90	0.89	0.87	0.816±0.10	0.2808
EGTSN+SVM	0.82	**0.86**	0.86	**0.95**	**0.86**	0.82	0.91	0.77	**0.91**	**0.862±0.02**	-

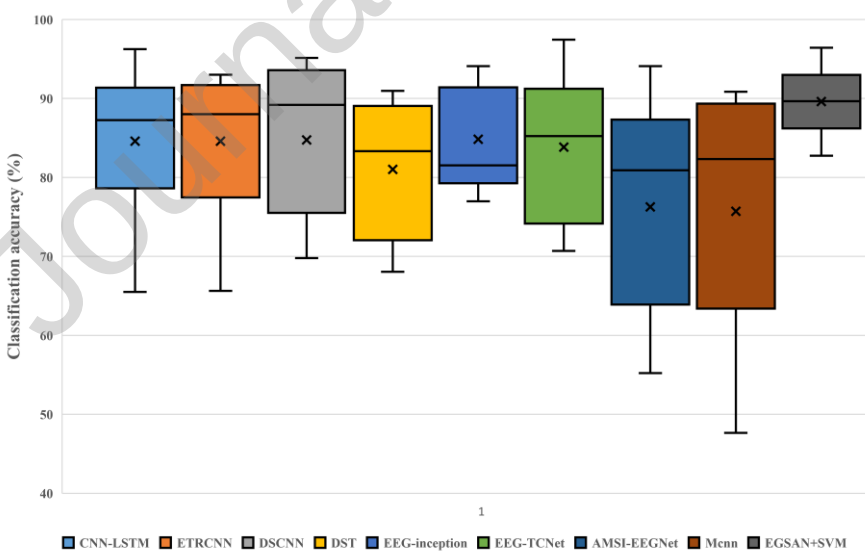


Figure 8. The classification accuracy of dataset 2 with different models.

Figure 8 presents the classification accuracies achieved by different competing methods. The symbol '×' represents the mean value, and the '—' represents the median line. Our methods perform better than ETRCNN, CNN-LSTM, DSCNN, EEG-Inception, Mcnn, AMSI-EEGNet, DST, and EEG-TCNet methods, with more than a 2% improvement in mean classification accuracy, as shown in Figure 8. Moreover, our proposed methods exhibit greater stability when decoding different participants, indicated by the smaller standard deviation in our results.

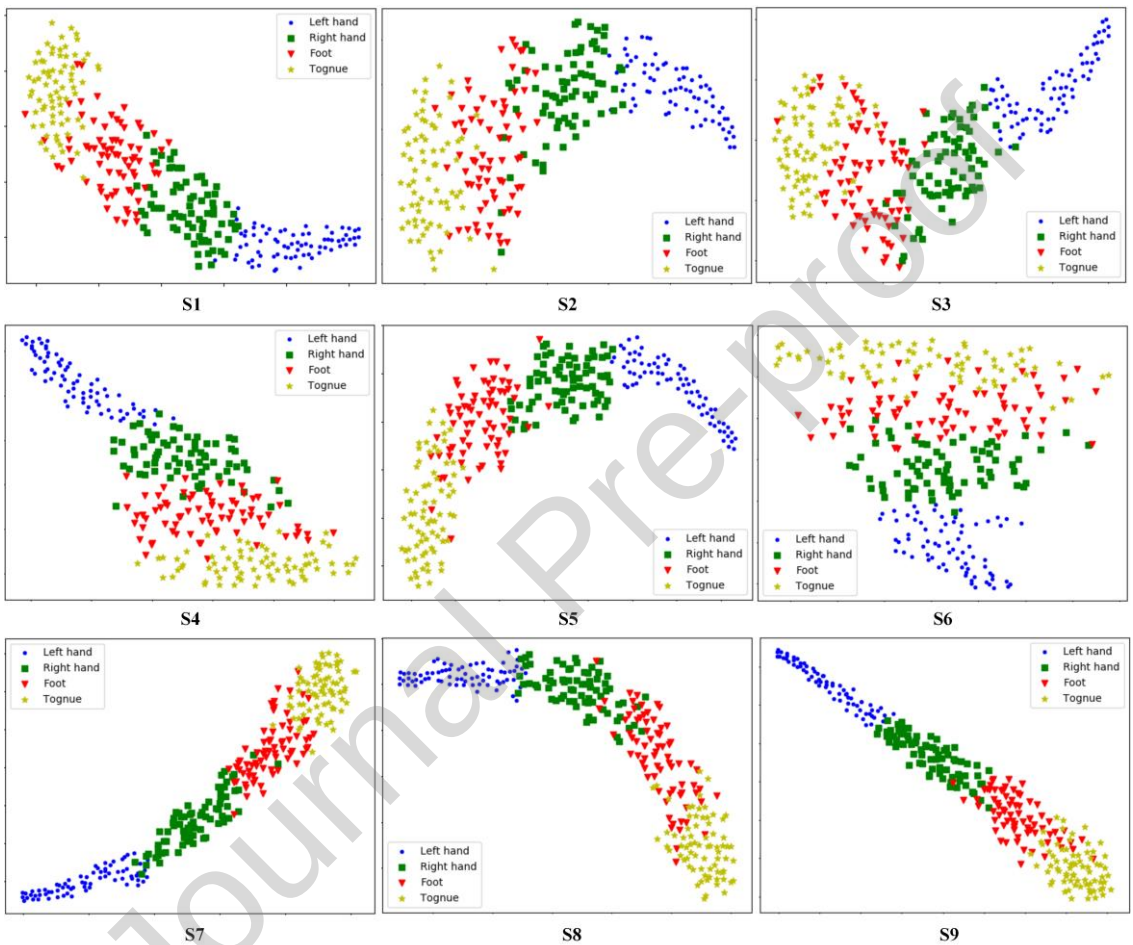


Figure 9. Feature distribution of dataset 2.

5. Discussion

In this study, we propose EGSAN, a model designed to extract features from EEG-based graphs. EEG-based graphs utilize edges to preserve spatial information by establishing connections exclusively between neighboring nodes, effectively representing the relative positions of EEG channels. During feature embedding with EGSAN, only nodes connected by edges are aggregated, while nodes without edge connections are excluded from aggregation. The loss function employed in EGSAN differs from the traditional cross-entropy loss function. We utilize a sampled softmax loss function to optimize the hyperparameters of the EGSAN model,

which does not require the labels information of samples. The goal of this loss function is to guide EGSAN to recognize and distinguish the sub-graph structural information within each graph, leading to improve the classification accuracies. Consequently, the EGSAN model can effectively extract distinguishable features from EEG-based graphs without relying on labels, making it suitable for unsupervised learning conditions.

5.1. Feature distribution

We demonstrate the effectiveness of our proposed feature learning framework by applying t-distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton, 2008) to visualize the distribution of learned features. This reduces high-dimensional feature sets into lower dimensional projections for easy visualization. We applied the t-SNE method to the final graph embedding obtained by our EGSAN method, and the resulting feature distributions are presented in Figure 9. The figure clearly shows that the features extracted by our EGSAN method are easily distinguishable in most participants. Notably, left-hand motor imagery tasks were easier to distinguish than other motor imagery tasks, due to all participants being right-handed. However, the features of foot and tongue motor imagery tasks have overlapping ratios, as indicated in Figure 9, making them harder to distinguish. Overall, the distribution of features extracted by our EGSAN method confirms its ability to extract distinctive features from the graph data converted from raw EEG signals.

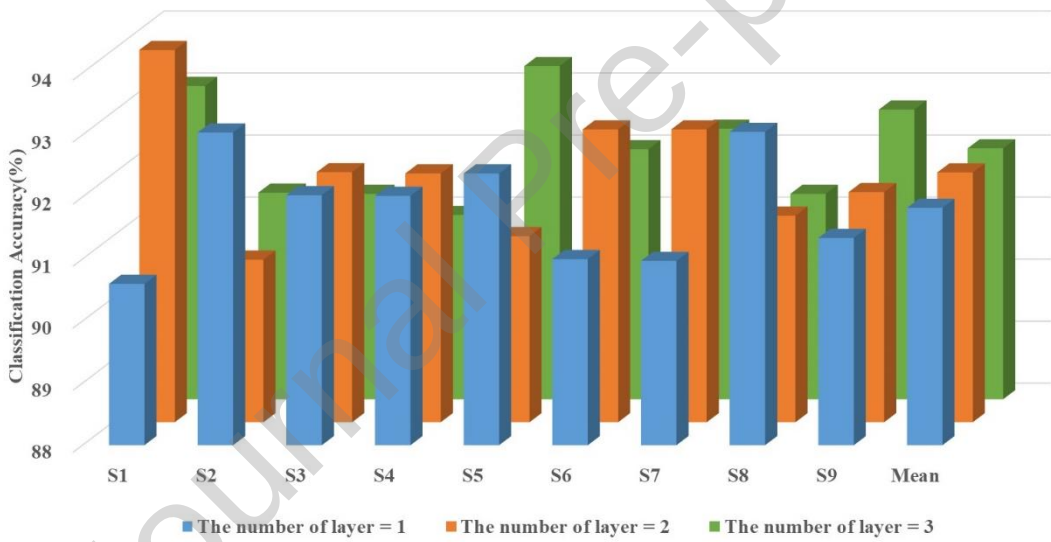


Figure 10. Classification accuracy with different number of layer.

5.2. Effect of the number of layers

The number of layers is a crucial factor that determines the behavior of our EGSAN model. To investigate this, we conducted experiments using different numbers of layers and present the results in Figure 10, where each color represents the model's performance with different layer numbers. Considering the size of the graph data, we limited the maximum number of layers to three. As shown in Figure 8, the optimal number of layers for each participant varied. For example, participant '2' achieved the highest classification accuracy with one layer, while participant '5' obtained the best result with three layers. Although participants had different performances with different numbers of layers, our proposed EGSAN model was able to extract distinguishing

features using simple structures, as illustrated in Figure 10. The mean classification accuracy increased as the number of layers increased, but there was only a slight improvement in mean classification accuracy when the number of layers increased from two to three, indicating that the EGSAN model can achieve excellent performance with a simple structure.

5.3. Future work

In this study, we only focus on the subject-dependent scenario, we will try our best to apply transfer learning on the EGSAN model and aim to achieve few-shot learning on motor imagery classification tasks. Moreover, we will explore the performance of our EGSAN model in BCI systems with other paradigms such as P300, SSVEP, and emotion detection. All the experiments in this study are performed on offline data, in the future, we will incorporate our EGSAN model into a real-time BCI system in order to evaluate its performance in this scenario.

6. Conclusion

This study proposes a novel feature learning framework named EEG graph Self-attention network (EGSAN) that aims to improve motor imagery EEG classification performance. The input of our EGSAN model is graph data derived from raw EEG signals, wherein the nodes are defined by EEG channels, and the node properties are PSD features extracted from individual channels. The structure of the graph retains the spatial information from the raw EEG channels. Our EGSAN model uses a graph embedding and self-attention mechanism to extract attentive features from EEG-based graphs. In our experiments, we consider the un-labelled scenario and apply an unsupervised feature learning strategy, the features learned by EGSAN can be classified by SVM classifiers. The experiment results demonstrate how our EGSAN model can extract distinguishable features and outperform than other state of art models. Our EGSAN model has the potential to improve the performance of BCI.

Acknowledgements

This work was supported by STI 2030-major projects 2022ZD0208900 and the Grant National Natural Science Foundation of China under Grant 62176090; in part by Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX，in part by the Program of Introducing Talents of Discipline to Universities through the 111 Project under Grant B17017; This research is also supported by National Government Guided Special Funds for Local Science and Technology Development (Shenzhen, China) (No. 2021Szvup043) and by Project of Jiangsu Province Science and Technology Plan Special Fund in 2022 (Key research and development plan industry foresight and key core technologies) under Grant BE2022064-1.

References

Amin, S.U., Alsulaiman, M., Muhammad, G., Mekhtiche, M.A., Shamim Hossain, M., 2019. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. Future Generation Computer Systems 101, 542–554. https://doi.org/10.1016/j.future.2019.06.027

Ang, K.K., Chin, Z.Y., Zhang, H., Guan, C., 2008. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). pp. 2390–2397. https://doi.org/10.1109/IJCNN.2008.4634130

Ang, K.K., Guan, C., Phua, K.S., Wang, C., Zhao, L., Teo, W.P., Chen, C., Ng, Y.S., Chew, E., 2015. Facilitating effects of transcranial direct current stimulation on motor imagery brain-computer interface with robotic feedback for stroke rehabilitation. Arch Phys Med Rehabil 96, S79–S87. https://doi.org/10.1016/j.apmr.2014.08.008

Cho, H., Ahn, M., Ahn, S., Kwon, M., Jun, S.C., 2017. EEG datasets for motor imagery brain-computer interface. Gigascience. https://doi.org/10.1093/gigascience/gix034

Demuru, M., La Cava, S.M., Pani, S.M., Fraschini, M., 2020. A comparison between power spectral density and network metrics: an EEG study. Biomed Signal Process Control 57, 101760.

Galindo-Noreña, S., Cárdenas-Peña, D., Orozco-Gutierrez, Á., 2020. Multiple kernel stein spatial patterns for the multiclass discrimination of motor imagery tasks. Applied Sciences 10, 8628.

García-Murillo, D.G., Álvarez-Meza, A.M., Castellanos-Dominguez, C.G., 2023. KCS-FCnet: Kernel Cross-Spectral Functional Connectivity Network for EEG-Based Motor Imagery Classification. Diagnostics 13, 1122.

Ingolfsson, T.M., Hersche, M., Wang, X., Kobayashi, N., Cavigelli, L., Benini, L., 2020. EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces, in: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp. 2958–2965.

Jafarifarmand, A., Badamchizadeh, M.A., 2019. EEG artifacts handling in a real practical brain–computer interface controlled vehicle. IEEE Transactions on Neural Systems and Rehabilitation Engineering 27, 1200–1208.

Jean, S., Cho, K., Memisevic, R., Bengio, Y., 2014. On using very large target vocabulary for neural machine translation. arXiv preprint arXiv:1412.2007.

Jeon, E., Ko, W., Yoon, J.S., Suk, H.-I., 2021. Mutual information-driven subject-invariant and class-relevant deep representation learning in BCI. IEEE Trans Neural Netw Learn Syst.

Jin, J., Miao, Y., Daly, I., Zuo, C., Hu, D., Cichocki, A., 2019. Correlation-based channel selection and regularized feature optimization for MI-based BCI. Neural Networks 118, 262–270.

Jin, J., Sun, H., Daly, I., Li, S., Liu, C., Wang, X., Cichocki, A., 2021. A novel classification framework using the graph representations of electroencephalogram for motor imagery based brain-computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering 30, 20–29.

Kotchetkov, I.S., Hwang, B.Y., Appelboom, G., Kellner, C.P., Connolly, E.S., 2010. Brain-computer interfaces: military, neurosurgical, and ethical perspective. Neurosurg Focus 28, E25.

Kumar, S., Sharma, A., Tsunoda, T., 2019. Brain wave classification using long short-term memory network based OPTICAL predictor. Sci Rep 9, 9153.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. J Neural Eng 15, 056013.

Ma, W., Gong, Y., Xue, H., Liu, Y., Lin, X., Zhou, G., Li, Y., 2022. A lightweight and accurate double-branch neural network for four-class motor imagery classification. Biomed Signal Process Control 75, 103582.

Naeem, M., Brunner, C., Leeb, R., Graimann, B., Pfurtscheller, G., 2006. Seperability of four-class motor imagery data using independent components analysis. J Neural Eng 3, 208.

Neuper, C., Wörtz, M., Pfurtscheller, G., 2006. ERD/ERS patterns reflecting sensorimotor activation and deactivation. Prog Brain Res 159, 211–222.

Nguyen, D.Q., Nguyen, T.D., Phung, D., 2022. Universal graph transformer self-attention networks, in: Companion Proceedings of the Web Conference 2022. pp. 193–196.

Pfurtscheller, G., Neuper, C., 2001. Motor imagery and direct brain-computer communication. Proceedings of the IEEE 89, 1123–1134.

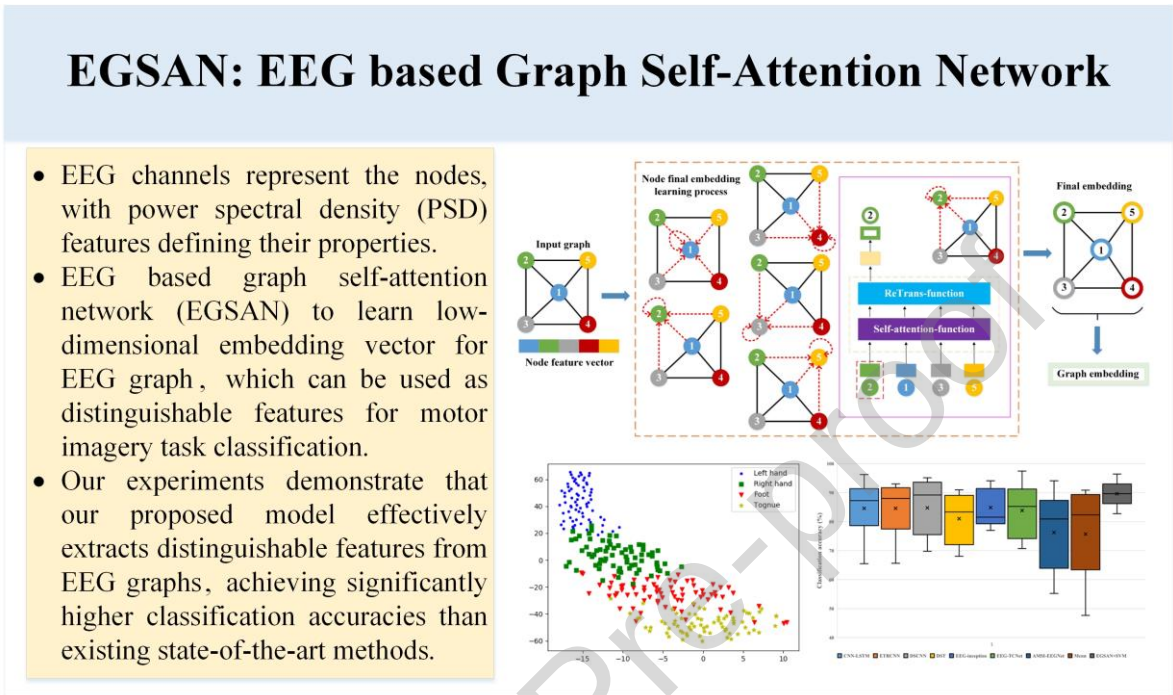
Pfurtscheller, G., Neuper, C., 1997. Motor imagery activates primary sensorimotor area in humans. Neurosci Lett 239, 65–68.

Razi, S., Mollaei, M.R.K., Ghasemi, J., 2019. A novel method for classification of BCI multi-class motor imagery task based on Dempster–Shafer theory. Inf Sci (N Y) 484, 14–26.
Riyad, M., Khalil, M., Adib, A., 2021. A novel multi-scale convolutional neural network for motor imagery classification. Biomed Signal Process Control 68, 102747.
Rodrigues, P.G., Filho, C.A.S., Attux, R., Castellano, G., Soriano, D.C., 2019. Space-time recurrences for functional connectivity evaluation and feature extraction in motor imagery brain-computer interfaces. Med Biol Eng Comput 57, 1709–1725.
Sadiq, M.T., Yu, X., Yuan, Z., Aziz, M.Z., Siuly, S., Ding, W., 2021. Toward the development of versatile brain–computer interfaces. IEEE Transactions on Artificial Intelligence 2, 314–328.
Stefano Filho, C.A., Attux, R., Castellano, G., 2018. Can graph metrics be used for EEG-BCIs based on hand motor imagery? Biomed Signal Process Control 40, 359–365.
Sun, H., Jin, J., Kong, W., Zuo, C., Li, S., Wang, X., 2021a. Novel channel selection method based on position priori weighted permutation entropy and binary gravity search algorithm. Cogn Neurodyn 15, 141–156.
Sun, H., Jin, J., Xu, R., Cichocki, A., 2021b. Feature selection combining filter and wrapper methods for motor-imagery based brain–computer interfaces. Int J Neural Syst 31, 2150040.
Tang, J., Liu, Y., Hu, D., Zhou, Z., 2018. Towards BCI-actuated smart wheelchair system. Biomed Eng Online 17, 1–22.
Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. Journal of machine learning research 9.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv Neural Inf Process Syst 30.
Wang, Y., Gao, S., Gao, X., 2006. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, pp. 5392–5395.
Wang, Z., Yu, Y., Xu, M., Liu, Y., Yin, E., Zhou, Z., 2019. Towards a hybrid BCI gaming paradigm based on motor imagery and SSVEP. Int J Hum Comput Interact 35, 197–205.
Wolpaw, J.R., 2007. Brain-computer interfaces (BCIs) for communication and control, in: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 1–2.
Xu, L., Zhang, H., Hui, M., Long, Z., Jin, Z., Liu, Y., Yao, L., 2014. Motor execution and motor imagery: a comparison of functional connectivity patterns based on graph theory. Neuroscience 261, 184–194.
Xu, M., Yao, J., Zhang, Z., Li, R., Yang, B., Li, C., Li, J., Zhang, J., 2020. Learning EEG topographical representation for classification via convolutional neural network. Pattern Recognit 105, 107390.
Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., Leskovec, J., 2018. Hierarchical graph representation learning with differentiable pooling. Adv Neural Inf Process Syst 31.
Zhang, C., Kim, Y.-K., Eskandarian, A., 2021. EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification. J Neural Eng 18, 046014.
Zhang, R., Zong, Q., Dou, L., Zhao, X., 2019. A novel hybrid deep learning scheme for four-class motor imagery classification. J Neural Eng 16, 066004.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Graphical abstract



Highlights

- A novel EEG-based graph representation contained node and structure information.
- Proposed EGSAN model to extract distinguish features from EEG-based graph.
- Unsupervised feature learning strategy to the network optimal processing.
- The EGSAN model can be applied on label limited dataset.