

# Less Is More: Brain Functional Connectivity Empowered Generalizable Intention Classification With Task-Relevant Channel Selection

Haowei Lou<sup>1</sup>, Zesheng Ye, Lina Yao<sup>2</sup>, *Senior Member, IEEE*,  
and Yu Zhang, *Senior Member, IEEE*

**Abstract**—Electroencephalography (EEG) signals are gaining popularity in Brain-Computer Interface (BCI)-based rehabilitation and neural engineering applications thanks to their portability and availability. Inevitably, the sensory electrodes on the entire scalp would collect signals irrelevant to the particular BCI task, increasing the risks of overfitting in machine learning-based predictions. While this issue is being addressed by scaling up the EEG datasets and handcrafting the complex predictive models, this also leads to increased computation costs. Moreover, the model trained for one set of subjects cannot easily be adapted to other sets due to inter-subject variability, which creates even higher over-fitting risks. Meanwhile, despite previous studies using either convolutional neural networks (CNNs) or graph neural networks (GNNs) to determine spatial correlations between brain regions, they fail to capture brain functional connectivity beyond physical proximity. To this end, we propose 1) removing task-irrelevant noises instead of merely complicating models; 2) extracting subject-invariant discriminative EEG encodings, by taking functional connectivity into account. Specifically, we construct a task-adaptive graph representation of the brain network based on topological functional connectivity rather than distance-based connections. Further, non-contributory EEG channels are excluded by selecting only functional regions relevant to the corresponding intention. We empirically show that the proposed approach outperforms the state-of-the-art, with around 1% and 11% improvements over CNN-based and GNN-based models, on performing motor imagery predictions. Also, the task-adaptive channel selection demonstrates similar predictive performance with only 20% of raw EEG data, suggesting a possible shift in direction for future works other than simply scaling up the model.

**Index Terms**—Motor imagery recognition, graph neural network, channel selection.

## I. INTRODUCTION

ELECTROENCEPHALOGRAPHY (EEG)-based brain-computer interaction (BCI) systems have enabled a variety of neurological tasks, such as motion intention recognition [1], emotion analysis [2] and brain disease detection [3]. The EEG-based BCI uses non-invasive scalp electrodes to record and further analyze electrical fluctuations that occur as a result of brain activity. A task that is of interest to this study is recognising Motor Imagery (MI), a cognitive process in which subjects imagine moving different parts of their bodies. There have been years of research into algorithms for detecting a particular MI, as well as implications for individuals with disabilities in a range of applications, including brain typing [4], mind-controlled wheelchairs [5] and prosthetic arm [6].

### A. NN-Based MI Recognition

Recent MI recognition research takes advantage of deep neural networks (NN) to extract discriminative representations from enormous amounts of data without the need for carefully-designed features. For instance, convolutional neural networks (CNNs) are widely adopted to extract spatial correlations between different sensory channels, whilst recurrent neural networks (RNNs) are promising for capturing temporal dependencies from raw EEG signals. A single CNN architecture consisting of two convolution blocks, dubbed EEGNet, was proposed by [7] for EEG classification across different paradigms. Zhang et al. [4] design a hybrid model using CNN and RNN to extract spatial and temporal features from EEG signals, paired with an Auto-encoder to eliminate artifacts. Zhang et al. [8] additionally transform raw EEG signals into a matrix-like form to explore correlations between physically adjacent sensory channels.

### B. Subject-Independent MI Recognition

While achieving success in subject-dependent settings, they still suffer from subject-independent evaluation, in which training and test data are collected from different subjects.

Manuscript received 8 December 2022; revised 20 February 2023; accepted 1 March 2023. Date of publication 6 March 2023; date of current version 31 March 2023. (Corresponding author: Lina Yao.)

Haowei Lou and Zesheng Ye are with the School of Computer Science and Engineering, University of New South Wales, Sydney NSW 2052, Australia (e-mail: haowei.lou@unsw.edu.au; zesheng.ye@unsw.edu.au).

Lina Yao is with CSIRO's Data61 and the School of Computer Science and Engineering, University of New South Wales, Sydney NSW 2052, Australia (e-mail: lina.yao@data61.csiro.au).

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

Digital Object Identifier 10.1109/TNSRE.2023.3252610

In fact, EEG signals may manifest different patterns even when subjects are performing the same cognitive task. It is necessary either to adapt to new subjects or to model subject-invariant features to address this issue. Fahimi et al. [9] combine general mental states with new subject’s data for personalized modeling. Chen et al. [10] reduce the discrepancy between two subjects with adversarial training. In [11], the mean covariance matrix of each user is transformed into an identity matrix for the alignment of their EEG trials. However, each new subject must be adapted once, and negative transfer may also occur in view of noisy EEG signals [12]. Alternatively, another line of work seeks to identify patterns that are universal across many subjects. For instance, Zhang et al. [13] incorporate self-attention [14] into a convolutional-recurrent model to explore concentrated temporal periods.

### C. Topological Pattern of Brain Activity

Despite being dominant in learning spatial representations, CNNs are structured on dense and regular “grid”-like inputs, which limits their ability to identify non-Euclidean relationships. It is, however, true that EEG channels are non-Euclidean by nature, as are brain regions. Additionally, even with the same acquisition device, signals do not necessarily originate from the same location due to the variance between subjects. Zhang et al. [15], [16] represent EEG nodes as a graph to learn a topological-based positioning relationship, which appears to be less subject-dependent than conventional representations. It brings to light recent interest in using graph neural networks (GNNs) as an alternative to CNNs in obtaining non-Euclidean representations from raw EEG input signals [17]. There are concerns raised by existing graph-based EEG representations despite the topological features showing robustness to inter-subject variability.

1) *Channel Redundancy*: standard EEG acquisition devices measure brain activity from the entire scalp. A particular cognitive task will, however, not require the activation of all brain regions [18], and different tasks may even activate different regions [19]. It follows that redundancy would be included if all channels are used indiscriminately.

2) *Edge Formation*: confusion still exists regarding how inter-regional brain connections should be represented as graph edges. The graph representation of certain non-euclidean structures, such as brain networks, is shown to be more effective than using CNNs [17]. Still, the extensively studied distance-based edge formation [15], [16] does not account for dynamic functional connectivity between different MI intentions.

3) *Graph Noise*: EEG recordings are bound to contain noise from both external and internal factors, implying raw EEG signals are likely to result in poor model generalization and over-fitting risks given GNNs’ sensitivity to the quality of the input graph [20]. This makes it even more challenging to identify task-relevant patterns while invariant to subjects, as EEG datasets are practically impossible to scale up as much as other domains.

Targeting each of these concerns, our study presents a Subject-Independent MI classification model using brain

Functional Connectivity (SIFT-EEG). For *redundant channel*, functional connectivity statistically contributes to identifying the brain regions involved when performing a particular cognitive task. For *edge definition*, graph adjacency is dynamically determined by functional connectivity, which filters out weak associations and identifies task-relevant active brain regions based on their importance in the functional connectivity graph. This is followed by performing self-attentive temporal convolution to extract discriminative task-adaptive temporal embeddings that mitigate *graph noise* contained in raw EEG signals. Finally, we derive robust topological embeddings from the task-adaptive temporal graphs of different subjects using a Graph Isomorphic Network (GIN) that reliably detects equivalent graphs [21].

### D. Contributions

- We propose a subject-independent MI prediction model built upon functional topological adjacency, which further takes into account self-attentive temporal convolution and graph isomorphism, thus capturing task-adaptive but subject-invariant EEG embeddings.
- We present a data-driven channel selection algorithm based on active brain regions, which can exclude non-contributory channels and reduce the impact of task-irrelevant noises.
- We evaluate SIFT-EEG on a large-scale EEG-based MI dataset, demonstrating its effectiveness for subject-independent classification, with competitive performances against state-of-the-art even using 20% of raw data.

## II. PRELIMINARIES

### A. Graph Representation for EEG Signals

A graph  $G = \langle V, E, W \rangle$  is an abstract structure in non-euclidean space, composed of nodes  $V$ , edges  $E$ , and weights  $W$ . Thus, the brain network topology can be represented by graphs. The nodes  $\{v_i\}_{i=1:n} \in V$  refer to EEG scalp electrodes located on specific brain regions, with a total electrode count of  $n$ . Each edge  $e_{ij} \in E$  represents the inter-regional relation between electrode pairs  $(v_i, v_j)$ . Accordingly, the pairwise edge weight  $w_{ij} \in W$  indicates the strength of each relation.

Graph adjacency  $A$  takes the form of a 2D matrix  $R^{N \times N}$ , where  $N$  denotes the number of nodes. One can be either a weighted or an unweighted graph. For unweighted adjacency,  $A_{ij} = 1$  represents a pair of connected nodes while  $A_{ij} = 0$  means there is no connection between them. A weighted adjacency has an additional attribute to the strength of relationship by setting  $A_{ij} = w_{ij}$ . Degree  $d(v_i)$  measures a node’s centrality by the number of edges connecting to it, indicating the node importance within the graph, calculated by adding up the  $i$ -th row of  $A$ .

Edge definitions for EEG-based graph representations are yet to be provided. The literature mostly uses either complete graph [17], i.e., each pair of nodes are connected, or distance-based connections [15], [16], that is, two nodes connected if their physical distance between electrodes is lower than a pre-defined threshold.

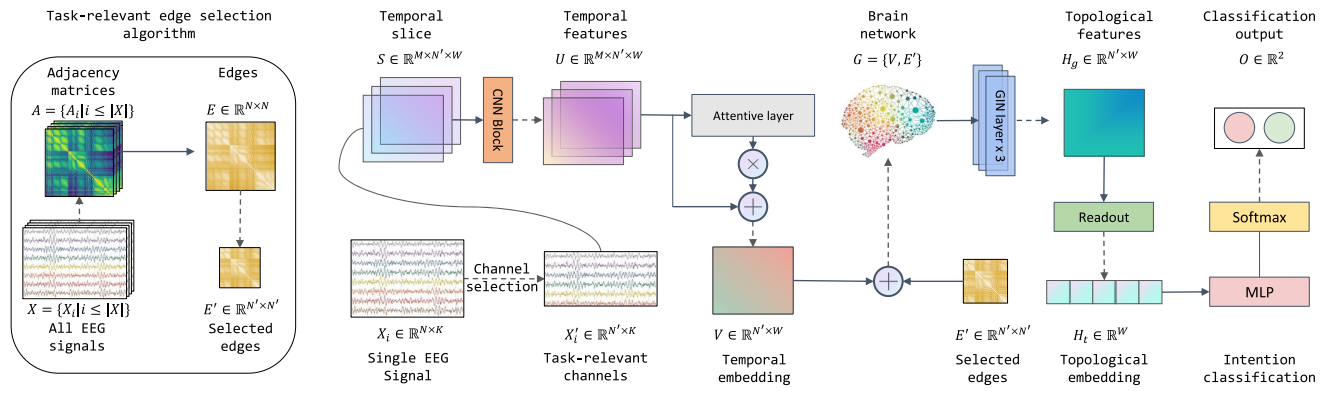


Fig. 1. Overview of the Subject-Independent MI classification model using brain Functional Connectivity (SIFT-EEG) model. We first crop raw EEG signal into a sequence of time slices using the slide window technique; adopt CNN to extract temporal features and a self-attentive module to search for the most discriminative temporal slice; then we combine nodes in the temporal embedding with functional connectivity to generate the graph representation of EEG signals; select three layers of Graph Isomorphism Network (GIN) to extract topological features and lastly, the extracted topological embedding are classified to different motion intention using a fully connected network with a softmax activation function.

## B. Graph Neural Networks

Graph Neural Networks (GNNs) are a family of generalized neural networks excelling at analyzing graph-structured data.

Graph Convolutional Network (GCN) [22] simplify spectral graph convolutions to the framework of neural networks, which defines a GCN operator to exchange node-wise information through edge connections, along with a layer-wise propagation rule that updates hidden node features. Throughout the  $K$ -layer propagation process, the nodes receive the averaged features from their neighbors via a mean aggregation, and the shape of the graph structure remains the same at the next layer.

$$h_v^{(k+1)} = \sigma \left( W \cdot \text{MEAN} \left\{ h_v^{(k)} \cup h_n^{(k)} \mid n \in \mathcal{N}(v) \right\} \right) \quad (1)$$

where  $h_v^{(k+1)}$  is node  $v$ 's hidden feature at  $k$ -th GCN layer,  $W$  are model parameters and  $\sigma$  is non-linear activation. READOUT( $\cdot$ ) function is further applied after propagating the last layer to extract the graph-level embedding,

$$h_G = \text{READOUT} \left( \left\{ h_v^{(K)} \mid v \in V \right\} \right) \quad (2)$$

GraphSAGE [23] interpret GNNs from a spatial perspective and generalize the mean-aggregator in GCN to a wider range of operators,

$$h_v^{(k+1)} = \sigma \left( W \cdot \text{AGG} \left\{ h_v^{(k)} \cup h_n^{(k)} \mid n \in \mathcal{N}(v) \right\} \right) \quad (3)$$

where  $W$  are model parameters and AGG( $\cdot$ ) refers to a permutation invariant function such as min/max/mean pooling.

The recent Graph Isomorphism Network (GIN) [21] is derived from the Weisfeiler-Lehman (WL) isomorphism test for checking if graphs are topologically identical, proving that GIN is as powerful as the WL test when AGGREGATE( $\cdot$ ) and READOUT( $\cdot$ ) are permutation invariant and injective, which produces more discriminative embeddings than other GNNs variants in graph-level classification.

Assume the function  $f : \mathcal{X} \rightarrow \mathbb{R}^n$  for any countable node feature space  $\mathcal{X}$ . Then for infinitely many choices of  $\epsilon$  including all irrational numbers, such that any function  $g$  can be decomposed into  $g(c, X) = \phi((1 + \epsilon) \cdot f(c) + \sum_{x \in X} f(x))$  for some function  $\phi$ , where  $c \in X$  and  $X \subset \mathcal{X}$ . A multilayer

perceptron (MLP) with more than one hidden layer can be used to approximate injective function [24], so the node embedding  $h_v$  becomes distinguishable, using the GINConv operator defined by

$$h_v^{(k+1)} = \text{MLP}^{(k)} \left( (1 + \epsilon^{(k)}) h_v^{(k)} + \sum_{n \in \mathcal{N}(v)} h_n^{(k)} \right) \quad (4)$$

where the AGG( $\cdot$ ) of neighbor nodes is implemented as a summation to impose injective mapping.

## III. METHODOLOGY

### A. Overview

We now describe the goal of EEG-based Motor Imagery (MI) classification and overview the key steps of our approach. Given an EEG segment  $X \in \mathbb{R}^{N \times K}$  collected while a subject was performing a MI task, we aim to estimate the associated specific intention  $Y$  by training a predictive model that performs supervised classification under the subject-independent setting, meaning that model training and evaluation are conducted on two disjoint groups of subjects.  $K = T \times f$  is the number of time points within a segment,  $N$  is the number of electrodes,  $T$  is the recording duration, and  $f$  is the sampling frequency.

Let  $X$  be an EEG segment and  $Y$  be the intention, our approach approximates the mapping  $f : X \rightarrow Y$ , parameterized by a Neural Network  $\hat{Y} = f_{\Theta}(X)$ , by involving the following five steps:

- 1) Calculating adjacency  $A \in \mathbb{R}^{N \times N}$  across the readings of all sensory channels  $X$  based on functional connectivity;
- 2) Selecting the top- $N'$  strongest channels  $X' \in \mathbb{R}^{N' \times K}$  tailored to the MI task and identifying task-relevant edges  $E' \in \mathbb{R}^{N' \times N'}$ ;
- 3) Extracting temporal node embedding  $V \in \mathbb{R}^{N' \times \omega}$ , by summarizing the features of each temporal slice  $S_m \in \mathbb{R}^{N' \times \omega}$ , where  $S = \{S_m\}_{m=1}^M$  results from sliding window applied on raw data of task-relevant channels  $X'$ ;
- 4) Generating topological graph embedding  $H \in \mathbb{R}^h$  from the brain functional network  $G = \{V, E'\}$  constructed by temporal node embeddings and task-relevant edges;

- 5) Predicting the intention  $f(X) = \hat{Y}$  of an EEG segment from the graph-level topological embedding.

### B. Functional Adjacency Matrix

We first calculate the adjacency of an EEG-based brain network with functional connectivity. We prefer this approach to previous distance-based or complete graph connections which either ignore topological relationships or are very computationally intensive if too many nodes are involved. Functional connectivity defines the statistical dependencies among temporal signals, commonly measured as the similarity between two brain regions by using the Pearson coefficient:

$$P(i, j) = \frac{\text{Cov}(i, j)}{\sqrt{\text{Var}(i)\text{Var}(j)}} \quad (5)$$

where  $\text{Cov}(i, j)$  is the covariance of measurement readings between sensory electrodes  $i$  and  $j$ ,  $\text{Var}(i)$  denotes the standard deviation of  $i$ -th channel readings throughout  $K$  timesteps.

This results in a complete functional adjacency  $A \in \mathbb{R}^{N \times N}$  with  $N$  being the number of electrodes. The following section discusses how we further reduce nodes and edges by only keeping task-relevant channels.

### C. Task-Adaptive Channel Selection

The activation of different brain regions varies with brain activity [25], indicating that some regions may not respond as actively to certain tasks as others. However, raw EEG data collected from electrodes distributed throughout the scalp will thus always contain information irrelevant to a specific MI task. The presence of such redundancy increases the computational burden and over-fitting risks.

This has led to the demand for filtering out task-irrelevant channels, yet most efforts are directed at manually solving complex optimization problems [26], which even requires extensive domain expertise [27]. The problem may, however, require an automatic solution in some cases, such as cross-subject analysis [28]. Instead, we propose a simple yet effective data-driven channel selection strategy upon node **importance** within all EEG channels, representing the activation intensity of corresponding brain regions. In particular, we suggest two metrics for measuring the nodes' importance  $\Lambda \in \mathbb{R}^N$ .

1) *Degree-Based Importance*: The degree of a node indicates its centrality within a brain network as well as the implication at the graph-level. Within this metric, we simply define the nodes' importance  $W$  as the number of incoming edges by looking up the channel-wise adjacency matrix  $A$ . Then, we apply a proportional threshold  $T^*$  to preserve channels  $N'$  adapting to a specific intention by descending nodes upon the importance  $\Lambda$ . The above steps are applied for each EEG segment  $X$  in the dataset.

2) *Strength-Based Importance*: We assume the channels with high correlation strength across different MI tasks are more active than those less correlated. Say we have EEG segments of  $C$  tasks with  $N_c$  for each, being performed by multiple subjects. We compute the absolute Pearson coefficient between the same channel's  $N_c$  readings of every

---

### Algorithm 1 Task-Relevant Channel Selection

---

**Require:** raw EEG segments set  $\mathcal{X}$ , intention set  $C$ , importance weight  $\Lambda$  and threshold  $T^*$

**Ensure:** EEG segments set  $\mathcal{X}'$  of task-relevant channels

- 1:  $\mathcal{X}' \leftarrow \emptyset$
- 2: **for**  $X \in \mathcal{X}$  **do**
- 3: Sort all channels by importance scores  $S$  in descending order, and store the indices into  $I$

$$I \leftarrow \text{argsort}(W)$$

- 4: Select the most active channels with threshold  $T^*$

$$X' \leftarrow X_i, \quad \forall i < T^* \text{ and } i \in I$$

- 5: Preserve readings of the most active channels  $X'$  and append them to the task-relevant EEG set

$$\mathcal{X}' \leftarrow \mathcal{X}' \cup X'$$

- 6: **end for**
- 

two intentions  $c_i, c_j \in C$  across all intentions. Then the task-relevant strength of each channel is obtained by averaging its  $C - 1$  coefficient values across all intention pairs. We define this strength of all channels as the importance weight  $\Lambda$  for segment  $X$ . Lastly, we apply a threshold  $T^*$  to preserve task-relevant channels  $N'$  by descending nodes upon their task-relevant strength.

### D. Temporal Embedding

There might be different concentration periods between subjects during MI, leading to different temporal properties even if two subjects were performing the same task [13]. Thus, our objective is to identify the most discriminative period within each EEG segment of different subjects, which we refer to as the subject-invariant temporal embedding.

Specifically, we first crop the EEG segment  $X' \in \mathbb{R}^{N' \times K}$  into  $M$  temporal slices  $S \in \mathbb{R}^{M \times N' \times \omega}$  using a sliding window of size  $\omega$ , after selecting the strongest  $N'$  channels. We then create initial temporal features  $U_m$  of each temporal slice  $S_m \in \mathbb{R}^{N' \times \omega}$  by applying a  $3 \times 3$  convolution kernel followed by an elu non-linear activation,

$$U_m = \text{elu}(\text{conv}(S_m)), \quad \text{for } m = 1, \dots, M \quad (6)$$

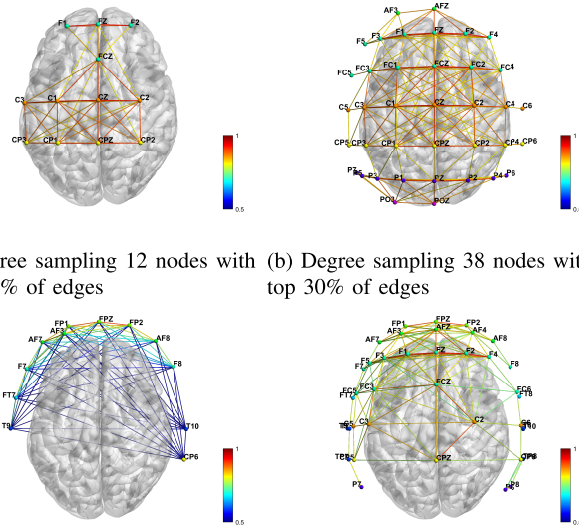
where  $U_m$  retains the same tensor shape as input slice  $S_m$ . Next, we use self-attention [14] to capture slice-wise correlations and adapt temporal weights to their temporal features, leading to a temporal embedding  $V \in \mathbb{R}^{N' \times W}$  that summarizes the EEG segment,

$$V = \sum_{m=1}^M \frac{\exp(H_m^\top W_m)}{\sum_{j=1}^M \exp(H_j^\top W_j)} U_m \quad (7)$$

with

$$H_m = W_h U_m + b_h \quad (8)$$





(a) Degree sampling 12 nodes with top 100% of edges (b) Degree sampling 38 nodes with top 30% of edges

(c) Strength sampling 12 nodes with top 100% of edges (d) Strength sampling 38 nodes with top 30% of edges

**Fig. 2.** Selected task-relevant channels and channel-wise connectivity. We present part of the edges with the highest correlation for readability purposes. The graph is visualized using BrainNet Viewer [29].

where  $W_m$ ,  $W_h$  and  $b_h$  are learnable parameters. The temporal embedding  $V$  aggregates all of  $M$  slices, taking into account each slice's importance, which derives the most discriminative representation upon input values.

### E. Topological Embedding

The brain activity is coordinated by multiple brain regions [30]. These inter-regional correlations are potentially beneficial when incorporated into predictive models. While CNN-based methods [4], [8], [13] assume Euclidean-structure of EEG electrodes, these works ignore the natural geometry of brain structure and connections between different regions beyond their immediate vicinity. As opposed, graph-based brain representations appear to reflect better the non-Euclidean nature of human's scalp [17], but also encode the subject-invariant positioning priors of electrodes [15], [16] to the model. Nevertheless, their graph constructions cannot encode dynamic functional connectivity that adapts to different MI tasks, since the edges therein are fixed in terms of the distances between nodes.

1) *Task-Adaptive Edge Formation*: In a brain network, the correctly illustrated edges could model how each brain region associates with other active regions during a MI task. Meanwhile, channel-wise correlations across MI intentions do not remain constant as aforementioned. Having identified the task-relevant channels  $N'$ , we now formulate the task-relevant edges  $E' \in \mathbb{R}^{N' \times N'}$  to leverage strong associations tailored to the task [31], in the following stages:

- 1) *Constructing*: Calculate channel-wise correlation adjacency  $A_c \in \mathbb{R}^{L \times S \times N \times N}$  of all subjects for each intention  $c \in C$ , using the Pearson coefficient, where  $L$  denotes the number of labels,  $S$  is the number of subjects, and  $N$  represents the number of channels/electrodes.

### Algorithm 2 Task-Relevant Edge Formation Algorithm

**Require:** EEG segments set  $\mathcal{X}$ , adjacency matrices  $\mathcal{A}$ , intention set  $C$ , edge threshold  $T^e$

- 1: **for**  $c$  in  $C$  **do**
- 2: Average the matrices  $A_c$  of all EEG segments within the same intention  $c$

$$A_c = \text{mean}_{X \rightarrow c}(A), \forall A \in \mathcal{A}, \forall X \in \mathcal{X}$$

- 3: Sort edges  $E_{A_c}$  in  $A_c$  based on the absolute value of weights  $\Lambda$  in descending order

$$E_c = \text{sort}(\text{abs}(E_{A_c}))$$

- 4: Derive critical edges corresponding to the strongest associations using edge threshold  $T^e$

$$E_c = \text{indices}(E_c(0 : t \times |E_c|))$$

- 5: **end for**
- 6: Merge critical edges together to produce task-relevant edges

$$E = \text{union}_{c \in C}(E_c)$$

- 2) *Generalizing*: Find the generalized connectivity across all subjects by averaging the adjacency matrices of all subjects for each intention  $c$ .
- 3) *Thresholding*: Preserve critical connections that exceed the threshold  $T^e$  for each intention  $c$ .
- 4) *Merging*: Derive task-relevant edges  $E'$  by merging the critical connections of all  $C$  intentions.

Algorithm 3 shows the pseudocode to derive a purely functional adjacency  $E'$ . Moreover, considering the temporal embedding  $V$  results from the task-relevant channels  $X'$ , we construct the task-adaptive brain network  $G = \langle V, E' \rangle$  to obtain topological embedding.

2) *Embedding Computation*: We adopt a  $L$ -layer GIN to investigate the topological embedding of a given brain network  $G$ , as it shows promise for graph-level classification [21], [32]. The graph-level topological embedding  $H \in \mathbb{R}^W$  after  $L$ -layers' propagation is produced by

$$H = \text{READOUT} \left\{ h_v^{(L)} \mid v \in V \right\} \quad (9)$$

with

$$h_v^{(l)} = \text{MLP}^{(l-1)} \left( (1 + \epsilon^{(l-1)})h_v^{(l-1)} + \sum_{n \in N(v)} h_n^{(l-1)} \right) \quad (10)$$

where  $W$  is the embedding dimension. We set MLP with 2 hidden layers,  $\epsilon = 0$ , and  $k = 3$  in practice. While READOUT( $\cdot$ ) function can be either non-injective aggregations (e.g., pooling) or injective mappings (e.g., MLP( $\cdot$ )), it is believed that the latter will yield a more discriminative graph embedding, as shown by [21]. Our empirical studies experiment with different setups to verify this argument in EEG-based applications.

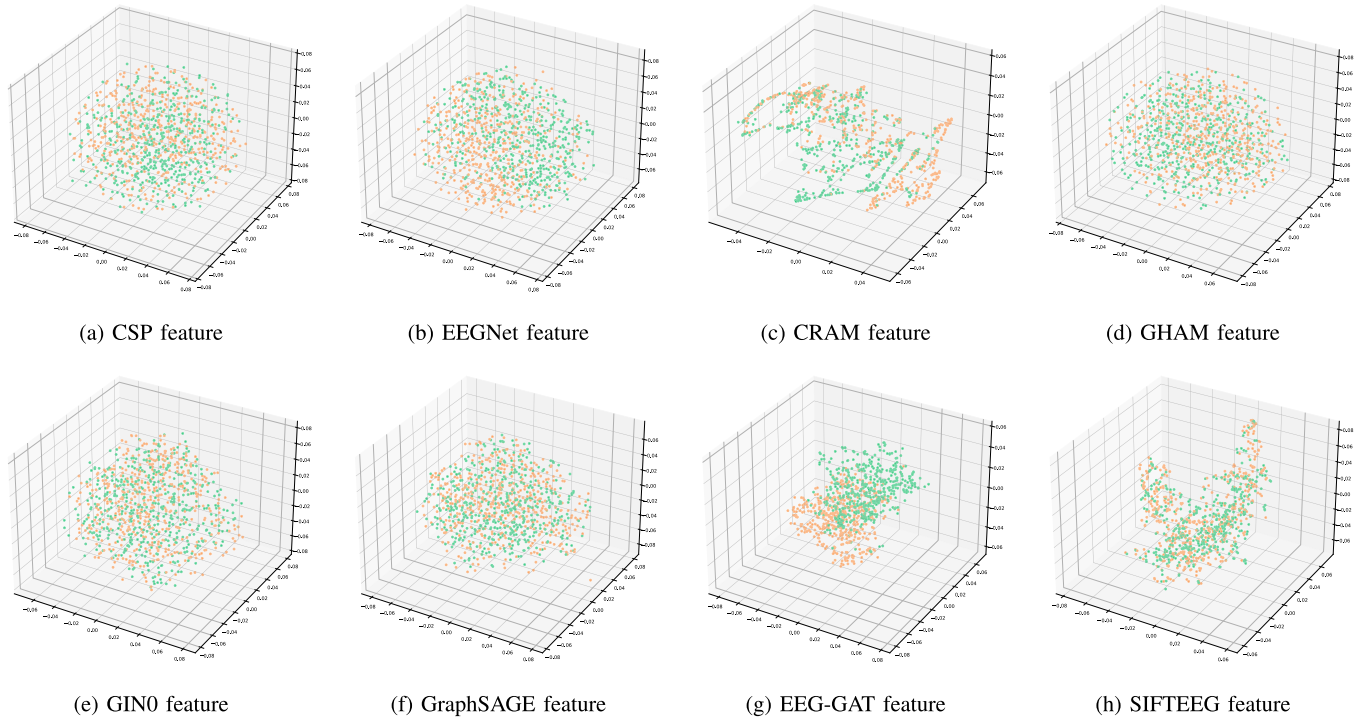


Fig. 3. The visualization of hidden features with t-SNE. Green dots refer to left-hand imaginary motion, while orange dots indicate right-hand imaginary motion.

## F. Intention Classification

We take the topological embedding  $H$  of an EEG segment  $X$  to perform intention prediction, with a softmax function to estimate the probabilities of each possible intention,

$$P_c = \text{softmax}(W_o H + b_o), \quad \text{for } c = 1, \dots, C \quad (11)$$

where  $W_o$  and  $b_o$  are trainable parameters. The predicted intention is thus given by  $\hat{Y} = \arg \max_c P_c$ .

## IV. EMPIRICAL STUDIES

### A. Experiment Setting

1) *Dataset*: In this work, we evaluate the performance of the proposed SIFT-EEG on a widely studied large-scale cross-subject EEG-based Motor Imagery dataset eegmidb (EEG motor movement/imagery database) downloaded from Physionet database [33]. The dataset was collected using BCI2000 [34] containing 64 electrodes with the sampling frequency of 160Hz from 109 healthy subjects. Following the conventions [15], [16], we remove subjects #88, #89, #92, #100 from the dataset due to consecutively resting states. Within three sessions of Motor Imagery tasks, EEG signals were recorded when subjects were executing left/right fist open and closed imagery. There are approximately fifteen 3.1-second segments of imaginary left/right-hand movement in each session; each subject performed three sessions. We randomly select 90 subjects and use their EEG segments as the training set, whereas the remaining 15 subjects are used as the test set. There are no disjoint subjects in the training and testing set, which ensures the evaluation is conducted in a subject-independent setting. We experiment with twelve different train/test splits to reduce randomness in the results.

We fix a unique random seed for each split that specifies the training and testing subjects. Noticeably, the last split includes all untested subjects from the previous eleven splits to ensure each subject is tested at least once.

2) *Preprocessing*: In line with most related works, raw EEG data is normalized with a z-score, calculated by subtracting the mean value and dividing by the standard deviation of training samples  $X$ . We do not further apply filters or data augmentation on raw EEG signals. We use temporal data as model input.

3) *Baselines*: We reproduce several baselines for subject-independent MI classification, including both traditional and DNN-based approaches. We adopt the Common spatial pattern and linear discriminant analysis (CSP+LDA) [35] as the representative traditional means. The DNN-based approaches are further divided into the CNN and GNN families.

Specifically, the CNN-based methods include EEGNet [7], CRAM [13], GHAM [16]. The GNN members GIN0 [21], GraphSAGE [23] and EEG-GAT [36] apply different GNNs under the same framework [17]. In addition, we implement a two-layer RNN as a vanilla DNN baseline to model temporal data.

4) *Implementation Detail*: Each input EEG segment  $X \in \mathbb{R}^{64 \times 496}$  contains 496 timesteps with 64 channels. For temporal embedding, we apply the slide window technique with window size of 400 and a step size is 10. Hence, the input signal contains ten temporal slices, where each slice has the shape of [64,400] (i.e.,  $N = 64, \omega = 400$ ). All the models are implemented with PyTorch<sup>1</sup> and trained and trained on an NVidia 3060-Ti GPU in a fully-supervised manner.

<sup>1</sup><https://pytorch.org>

TABLE I

OVERALL PERFORMANCES FOR SUBJECT-INDEPENDENT EVALUATION. ALL RESULTS ARE OBTAINED OVER 12 RUNS WITH MEAN  $\pm$  STD

	Method	Accuracy	ROC-AUC
Baselines	CSP+LDA	64.73 $\pm$ 1.85 <sub>(+20.79%)</sub>	0.6471 $\pm$ 0.019 <sub>(+21.62%)</sub>
	RNN	54.77 $\pm$ 3.47 <sub>(+42.76%)</sub>	0.5374 $\pm$ 0.010 <sub>(+46.45%)</sub>
	EEGNet	72.29 $\pm$ 8.10 <sub>(+8.17%)</sub>	0.7202 $\pm$ 0.082 <sub>(+9.28%)</sub>
	CRAM	75.65 $\pm$ 3.12 <sub>(+3.35%)</sub>	0.7643 $\pm$ 0.041 <sub>(+2.97%)</sub>
	GHAM	77.31 $\pm$ 3.49 <sub>(+1.13%)</sub>	0.7785 $\pm$ 0.035 <sub>(+1.09%)</sub>
	GIN0	54.47 $\pm$ 2.75 <sub>(+43.55%)</sub>	0.5449 $\pm$ 0.027 <sub>(+44.43%)</sub>
Ablation	GraphSAGE	58.42 $\pm$ 3.86 <sub>(+33.84%)</sub>	0.5845 $\pm$ 0.039 <sub>(+34.66%)</sub>
	EEG-GAT	68.18 $\pm$ 4.30 <sub>(+14.68%)</sub>	0.6817 $\pm$ 0.076 <sub>(+15.44%)</sub>
	SIFT-EEG (w/o d)	61.87 $\pm$ 1.72 <sub>(+26.38%)</sub>	0.6187 $\pm$ 0.017 <sub>(+27.21%)</sub>
	SIFT-EEG (w/o t)	64.10 $\pm$ 3.95 <sub>(+21.97%)</sub>	0.6454 $\pm$ 0.036 <sub>(+21.95%)</sub>
	SIFT-EEG (Avg)	61.51 $\pm$ 2.94 <sub>(+27.11%)</sub>	0.6089 $\pm$ 0.018 <sub>(+29.26%)</sub>
	SIFT-EEG (Sum)	62.91 $\pm$ 4.42 <sub>(+24.30%)</sub>	0.6159 $\pm$ 0.034 <sub>(+27.79%)</sub>
	SIFT-EEG (Max)	64.10 $\pm$ 6.44 <sub>(+21.99%)</sub>	0.6365 $\pm$ 0.056 <sub>(+23.64%)</sub>
	SIFT-EEG	<b>78.19 <math>\pm</math> 3.42</b>	<b>0.7870 <math>\pm</math> 0.035</b>

We use cross-entropy as the objective function and optimize model parameters using Adam with a learning rate of 0.001. The training batch size is 500. Each model is trained for 120 epochs, and the dropout probability is set to 0.5 to avoid over-fitting. We implement all GNN-related components using Torch\_Geometric.<sup>2</sup>

## B. Result & Discussion

The proposed SIFT-EEG is empirically compared with a range of baselines, focusing on three research questions.

- 1) Does the modeling of temporal dependencies benefit spatial/topological correlations and model performance?
- 2) Does task-adaptive dynamic connectivity outperform distance-based connectivity for topological learning?
- 3) Does task-relevant channel selection with topological features still achieve competitive performance?

The evaluation metrics include classification Accuracy and the Area Under ROC-Curve(ROC-AUC). All models are trained and evaluated with the same setting for fair comparison.

1) *Overall Performance:* Tab. I reports the model performance of all approaches on both metrics. Observe that SIFT-EEG outperforms all baseline models, with 1.13% and 14.68% accuracy improvements, as well as ROC-AUC gains of 1.09% and 15.44%, compared to the best performing CNN-based state-of-the-art (GHAM) and GNN-based state-of-the-art (EEG-GAT), respectively. Meanwhile, the top performers in both DNN families prove to be more accurate than CSP+LDA. In addition, all DNN baselines (except RNN) include mechanisms to represent the spatial correlation between EEG channels, providing better results than vanilla RNN. It follows that such representations should be incorporated into modeling.

Fig. 5 depicts how the training loss changes with the number of training epochs increase. GNN-based methods generally have lower training losses and faster convergence rates compared with CNN-based methods, which suggests the capability of topological features in task-specific predictions. However, their test performances are worse than CNN-based methods, showing a tendency to overfit the training data. The reason

TABLE II

COMPARISON OF THE CLASSIFICATION ACCURACY FOR TASK-ADAPTIVE AND RANDOM SELECTION OF 20% OF ORIGINAL CHANNELS

Model	Random	Task-Adaptive	All Channels
CRAM	55.11	<b>77.05</b>	75.93
GHAM	59.55	<b>76.96</b>	77.04
SIFT-EEG	53.11	<b>79.81</b>	77.93

may be that GNN-based methods extract topological features directly from raw EEG signals. Recall that our evaluations take place in a subject-independent setting. In this case, the variances in temporal patterns between subjects, i.e., graph noises, may cause them to perform inconsistently [37], [38]. Conversely, CNN-based methods explicitly handle temporal correlations before looking at the spatial domain. For instance, CRAM locates discriminative temporal features adaptively for different subjects by using attention. This eases the burden on the spatial feature extractor and generalizes the model to new subjects, albeit fitting these models takes longer. The proposed SIFT-EEG combines the merits of both, leveraging flexible topological features of the brain signal, as well as subject-adaptive temporal features that reduce task-irrelevant noise.

In addition, we perform statistically significant tests to evaluate the model performance improvements of SIFT-EEG over baselines. We use the pairwise t-test, assuming the pairwise difference is significant if  $p$ -value is less than 0.05. The results are reported in Tab. III, where statistically significant differences are **bolded**.

2) *Impact of Temporal Embedding:* Now we analyze how temporal embedding contributes to EEG-based MI classification and answer the first research question. EEGNet design a CNN block to extract temporal features, while CRAM and GHAM adopt self-attention on a recurrent network to capture long-term temporal dependency and adaptive subject-specific patterns, thus improving model performance in subject-independent experiments. In contrast, GIN0 and GraphSAGE take raw EEG signals directly as input without learning temporal features. This leads to much lower predictive results than their CNN-based counterpart. EEG-GAT, on the other hand, attempts to extract temporal information with a 2D temporal convolution operator, which derives around 32% and 24% improvement over GIN0 and GraphSAGE, even when all three approaches are developed under a similar predictive framework. We also examine the impact of temporal embedding within SIFT-EEG. Denoted as SIFT-EEG(w/o t), a SIFT-EEG variant without temporal embedding is compared with full SIFT-EEG to reveal its effect. There is an increase of 29.01% and 28.96% in the results of accuracy and ROC-AUC, respectively. It is thus clear from the comparisons of all three groups that modeling temporal dependency in conjunction with spatial/topological embedding is essential.

3) *Comparison of Topological Feature:* The second research question investigates whether the proposed task-adaptive dynamic connectivity facilitates topological learning. We first identify the graph connection of each model in comparison.

<sup>2</sup><https://pytorch-geometric.readthedocs.io>



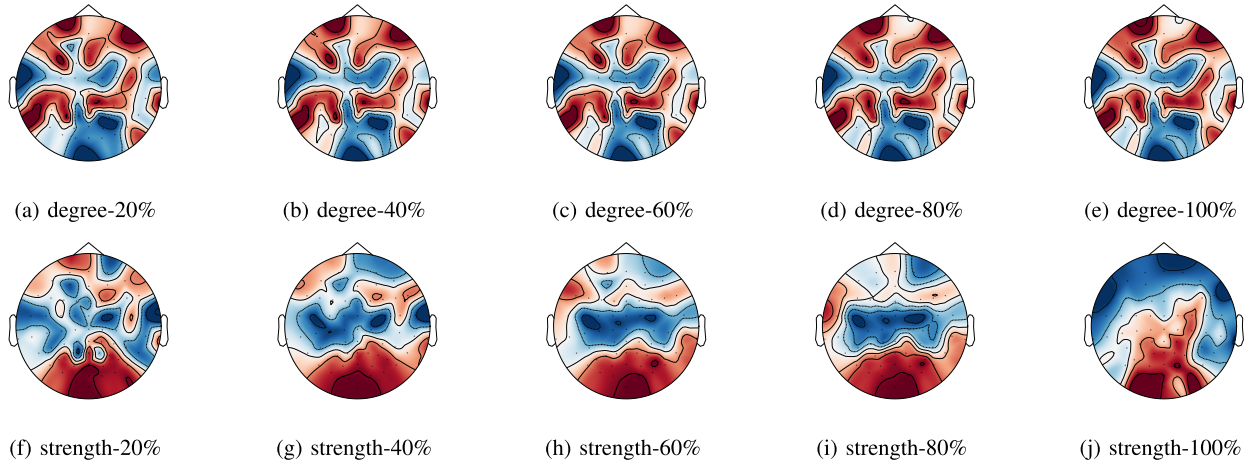


Fig. 4. The brain topology maps acquired with various importance metrics under different ratios of subjects sampled from 105 subjects.

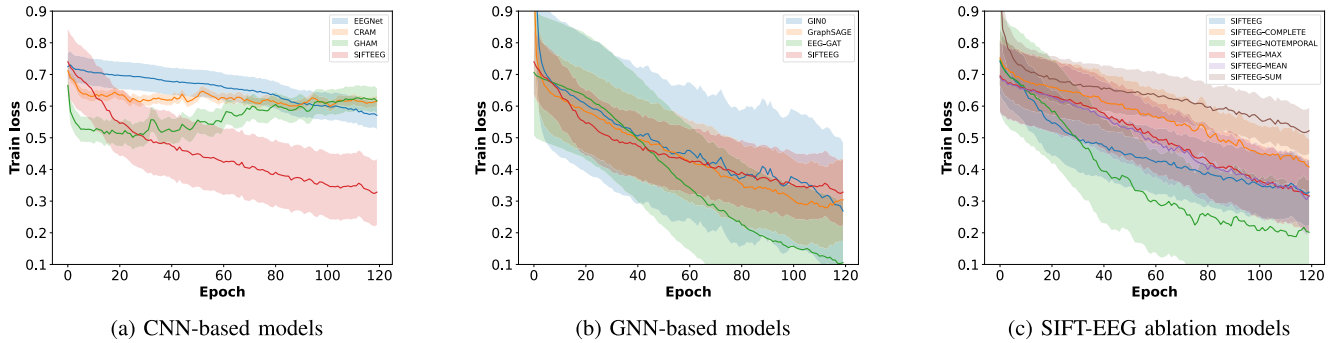


Fig. 5. Comparison for training loss change when the number of training epochs increases.

TABLE III

THE STATISTICALLY SIGNIFICANT TESTS. WE FIND THAT 20 OUT OF 24 COMPARISONS ARE SIGNIFICANT ( $p \leq 0.05$ ), WITH RESULTS SHOWN IN **BOLD**

Metric	Method					
Accuracy P-value	CSP+LDA <b>0.00</b>	EEGNet <b>0.03</b>	CRAM 0.07	HAM 0.54	GINO <b>0.00</b>	GraphSAGE <b>0.00</b>
	EEG-GAT <b>0.00</b>	SIFTEEG (Avg) <b>0.00</b>	SIFTEEG (Sum) <b>0.00</b>	SIFTEEG (Max) <b>0.00</b>	SIFTEEG (w/o d) <b>0.00</b>	SIFTEEG (w/o t) <b>0.00</b>
ROC-AUC P-value	CSP+LDA <b>0.00</b>	EEGNet <b>0.02</b>	CRAM 0.16	HAM 0.56	GINO <b>0.00</b>	GraphSAGE <b>0.00</b>
	EEG-GAT <b>0.00</b>	SIFTEEG (Avg) <b>0.00</b>	SIFTEEG (Sum) <b>0.00</b>	SIFTEEG (Max) <b>0.00</b>	SIFTEEG (w/o d) <b>0.00</b>	SIFTEEG (w/o t) <b>0.00</b>

GHAM uses distance-based graph representations of EEG channels, but not GNN for representation learning. All three GNN baselines define complete graph connections to yield the best results. SIFT-EEG chooses functional connectivity and bypasses the position limits.

GHAM extends CRAM with a graph definition of input that leads to an increase in 1.5% on accuracy and 0.7% on ROC-AUC to CRAM, implying the merit of non-euclidean assumption imposed by graph representation. Among the three GNN baselines, EEG-GAT shows better performance than the other two. A possible explanation is, GAT could benefit from well-designed attention in large and noisy graphs [39],

suggesting GAT seems to win out over GIN in a complete graph of the brain network (as per their framework).

We further replace dynamic connectivity with complete connection in SIFT-EEG and observe a clear performance drop as in SIFT-EEG (w/o d). The GIN applied to SIFT-EEG aggregates all nodes indiscriminately under complete connections. This leads to the conclusion that SIFT-EEG needs to be defined with task-adaptive edges.

4) *Injective Mapping Benefits SIFT-EEG*: We additionally test with different READOUT functions in SIFT-EEG, including three pooling-based variants SIFT-EEG (Avg), SIFT-EEG (Sum), SIFT-EEG (Max), referring to mean-, sum-,



TABLE IV

COMPARISON OF MODEL PERFORMANCE WITH DIFFERENT CHANNEL SELECTION STRATEGIES APPLIED TO SIFT-EEG

Strategy	Accuracy	#Channels	Percent
Original	$78.19 \pm 0.034$	64	100%
Random	$62.00 \pm 0.010$	39	61%
Subject-specific [41]	$75.89 \pm 0.038$	33	52%
Motor cortex region [42]	$76.60 \pm 0.019$	18	28%
Task-adaptive (Degree)	$63.83 \pm 0.013$	13	20%
Task-adaptive (Strength)	<b><math>79.81 \pm 0.045</math></b>	13	20%

and max-pooling. The performance also drops noticeably. As a reminder, SIFT-EEG emphasizes a) graph-level prediction instead of node-level prediction; and b) structural information of the graph (brain network). Our results empirically align with the theoretical findings of [21], i.e., MLP-based injective mapping increases capacity over its non-injective READOUT counterpart.

5) *Analysis of Task-Adaptive Channel Selection*: Our third research question examines task-adaptive channel selection from four perspectives. First, we visualize and compare the effects of two importance metrics. We next apply task-adaptive channel selection to CRAM and GHAM to examine its applicability. We evaluate its effectiveness against three additional channel selection approaches using SIFT-EEG. Finally, we discuss the model elapsed time results with different channel selection ratios.

a) *Qualitativeness*: Fig. 2 showcases the positions of task-relevant channels selected by two different importance metrics. The nodes resulting from degree-based importance are primarily located around the central sulcus, while those derived by strength-based importance clusters in the frontal lobe correspond to the functional area for motor control in the study of neuroscience [40].

b) *Applicability*: In addition to SIFT-EEG, the performance of two other models, CRAM and GHAM, is compared with task-adaptive and random selection by keeping 20% original channels. For comparison, CRAM makes no graph assumptions, GHAM represents the input with a hard-ruled graph definition, whilst SIFT-EEG dynamically determines the graph representation. As seen by Table II, the accuracy of all models for task-adaptive selection is similar to (with CRAM and GHAM) or even slightly higher (with SIFT-EEG) than all channels when only 20% are used, whereas random strategy returns lower performance. Hence, our task-adaptive channel selection could benefit all models in general. Still, it works best when coupled with dynamic functional graph connectivity.

c) *Effectiveness*: Moreover, we compare the proposed task-adaptive selection with two additional channel selection strategies. Shan et al. [41] identify subject-specific channels by finding channels with strong correlations to the central channel. In reproducing this strategy, we select the channels with an average correlation  $\geq 0.7$ , except for the reference channels C3/C4/Cz. Mattioli et al. [42] reduce the number of channels required by segmenting motion functional regions and producing regions of interest. In our experiments, we select channels located in the motor cortex region for this method. Table IV reports the least number of channels

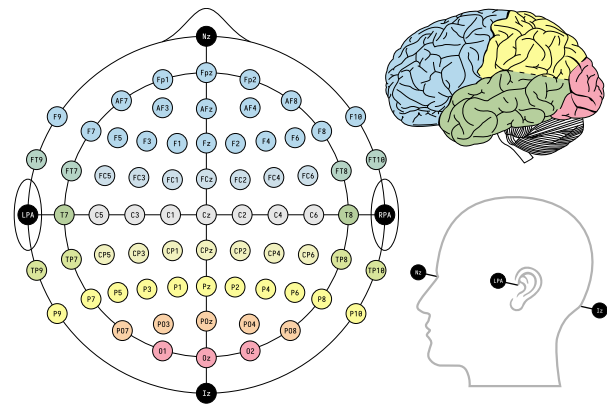


Fig. 6. Electrodes position in international 10-10 EEG signal acquisition device.

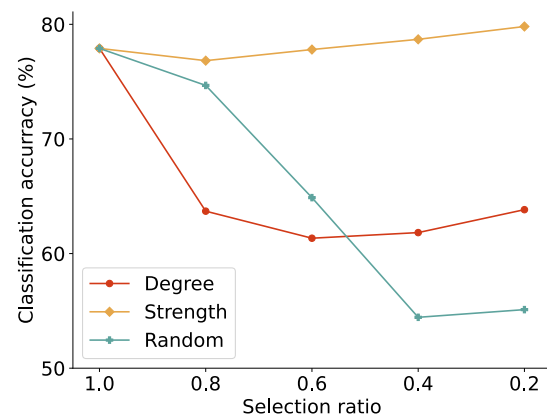


Fig. 7. Classification accuracy for top 1% task-relevant channels selected by the proposed algorithm.

required to achieve an accuracy greater than 60%, for each of these strategies applied to SIFT-EEG. The task-adaptive selection with strength-based importance is the top performer with the fewest channels required. The degree-based metric, however, appears to be relatively ineffective. It might be the case that nodes are unequally distributed throughout the scalp. According to Fig. 6, the sensory nodes near edge areas (e.g., AF7, O1 and P10) have fewer neighbors than those in central areas such as Cz, Cpz and Fcz. Meanwhile, the distance between nodes affects both the number of incoming edges and the strength of the connection. This eventually biases the estimation when using the degree-based importance metric.

On the other hand, the strength-based task-adaptive selection does not only report higher accuracy but with fewer channels, even compared to using channels from the widely-recognized motor cortex region [42]. The results indicate that task-adaptive channel selection can help encode the most discriminative EEG embedding with the fewest channels, even in the absence of extensive domain knowledge.

d) *Efficiency*: Our final step is to examine the model performance and running efficiency when varying the number of channels and edges. Fig. 7 and Fig. 8 show accuracy and relative elapsed time, respectively. The random selection consistently results in degraded performance as available channels decrease. In contrast, both task-adaptive selections perform better than random selection. The strength-based selection

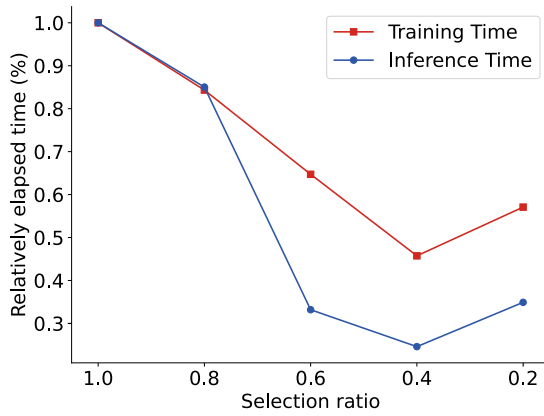
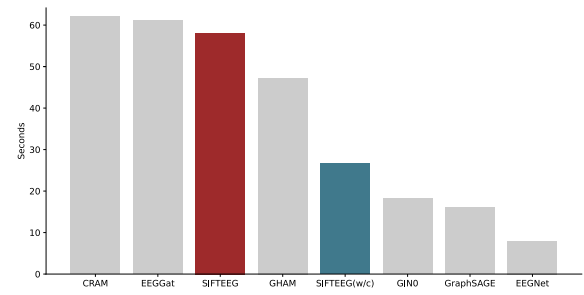


Fig. 8. Relative elapsed train and inference time for SIFT-EEG under different channel selection ratios.

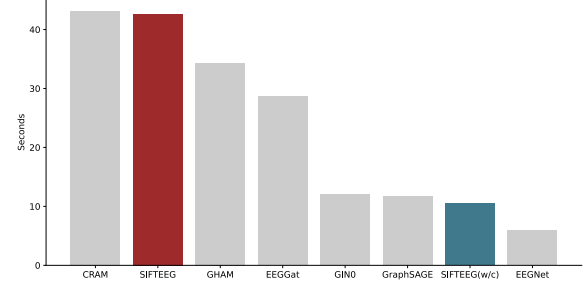
maintains the highest consistency of performance, regardless of the number of channels. Meanwhile, both training time and inference time are trending downward with fewer channels. In an interesting twist, running with 20% channels would cost slightly more time than the case of 40%, which, however, would lead to rebounded classification accuracy.

In addition, we compare the computational efficiency of SIFT-EEG with baselines when all 64 channels are used, as shown in Fig. 9. We evaluate how long it takes to handle 400 arbitrarily sampled EEG segments with 400 forward passes with batch size 100, for both training and inference. The proposed SIFT-EEG takes 3.4% less training time than EEG-GAT, the strongest GNN baseline whilst improving the predictive performance by 14.68%. Moreover, SIFT-EEG consumes 5 fewer but 10 more seconds than CRAM and GHAM, respectively. The higher recognition accuracy of these three models comes at the expense of being more costly during inference. Nevertheless, we note that SIFT-EEG can improve its efficiency with our task-relevant channel selection, denoted by SIFT-EEG (w/c) in Fig. 9. At the time it runs with top-40% task-relevant channels, which improves training and inference efficiency by more than 50% and 75%, respectively. This proves SIFT-EEG to be faster than most baselines while gaining substantial performance benefits.

6) *Visualization of Dynamic Functional Connectivity*: Having learned the task-adaptive dynamic functional connectivity topology, we compare it with a fixed functional connectivity topology originating from the pre-motor, supplementary, and primary motor area [43]. We visualize the active brain region across certain proportions of 105 subjects upon performing MI tasks, according to two important metrics, as in Fig. 4. The degree-based importance is associated with activities in the frontal and parietal lobes, whereas strength-based importance seems to target the occipital lobes mainly. There are overlaps between the dynamic active region and the fixed motor region regardless of the number of subjects used. Moreover, our dynamic brain topology shows that active regions within each importance metric are distributed similarly across subjects. Increasing the number of subjects converges the distribution to an almost identical pattern, implying that the statistical characteristics of the dynamic topology can be improved



(a) Elapsed training time for 400 iterations.



(b) Elapsed inference time for 400 iterations.

Fig. 9. Comparison of computation efficiency for total time taken for processing arbitrarily sampled 400 EEG segments.

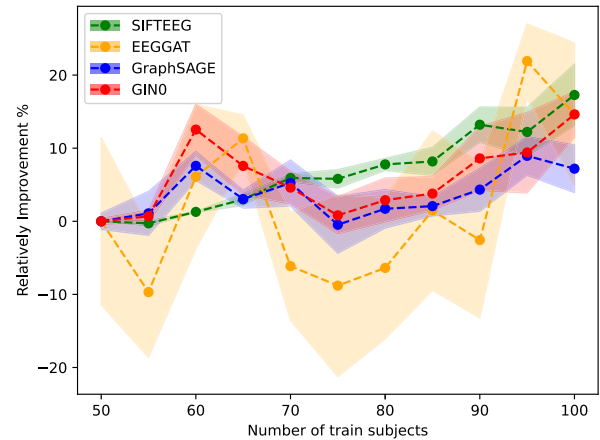


Fig. 10. Impact of varying the number of training subjects. Every subject is included in the test set at least once for each train/test ratio. The results are reported with mean  $\pm$  std values of the relatively improved classification accuracy within each method itself.

by having more samples. Still, it shows robustness despite different sample sizes.

7) *Impact of Varying Training Subjects*: Furthermore, we investigate how different numbers of training subjects affect classification accuracy. Our experiment begins with fifty subjects used for training, i.e., the training and test subjects are close to a 1:1 ratio. Following, the training set expands by five subjects per ratio record until 100 subjects have been included. For each train/test ratio, we perform cross-validation to split the train and test sets, ensuring each subject has been tested at least once. We report performance improvements across GNN family methods with varying numbers of training subjects, using fifty as the baseline and the mean and standard deviation for each ratio, as illustrated in Fig. 10.

Fig. 10 illustrates a strong correlation is observed between the number of subjects and classification accuracy. Increasing the number of train subjects from fifty to seventy almost linearly improves classification accuracy. This may suggest that SIFT-EEG learns more subject-independent features as it is trained on more subjects. However, the improvement becomes negligible as the training set size increases, when there are seventy to eighty-five subjects. Observations show that the overall performance continues to improve beyond eighty-five training subjects. The presence of more training subjects may increase the likelihood of testing subjects exhibiting similar patterns to those in the training set, as well as improved predictive performance.

We also include the performance changes of other GNN-based methods. Whereas all methods demonstrate performance improvements, SIFT-EEG benefits the most from the use of a greater number of training subjects, as this may facilitate the extraction of subject-independent patterns among the population. In contrast, other methods show fluctuating results and irregular patterns while underperforming SIFT-EEG consistently, mostly because neither of these methods bakes subject-independent features into the representation. Interesting to note that EEG-GAT reports considerably higher variances despite being the best-performing GNN member.

## V. CONCLUSION

This paper focused on task-adaptive modeling of brain networks with functional connectivity. We conducted extensive experiments with a large-scale EEG dataset to demonstrate that the proposed predictive model outperforms the state-of-the-art for Motor Imagery classification in the subject-independent setting. Moreover, our investigation indicated that task-adaptive region selection produces similar predictive performance with only 20% of raw EEG data, with a considerable reduction in computation cost during model training and deployment. Our future work may lend itself to the data scarcity problem in EEG-related research by examining other Motor Imagery datasets with fewer subjects. We will examine how to apply the proposed subject-independent channel-selection methods more efficiently in situations where training subjects are limited.

## REFERENCES

- [1] W. Chen et al., "EEG-based motion intention recognition via multi-task RNNs," in *Proc. ICDM*, 2018, pp. 279–287.
- [2] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affective Comput.*, vol. 10, no. 3, pp. 374–393, Jul./Sep. 2019.
- [3] J. Jeong, "EEG dynamics in patients with Alzheimer's disease," *Clin. Neurophysiol.*, vol. 115, no. 7, pp. 1490–1505, Jul. 2004.
- [4] X. Zhang, L. Yao, Q. Z. Sheng, S. S. Kanhere, T. Gu, and D. Zhang, "Converting your thoughts to texts: Enabling brain typing via deep feature learning of EEG signals," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2018, pp. 1–10.
- [5] I. A. Mirza et al., "Mind-controlled wheelchair using an EEG headset and Arduino microcontroller," in *Proc. Int. Conf. Technol. Sustain. Develop. (ICTSD)*, Feb. 2015, pp. 1–5.
- [6] T. Beyrouthy, S. K. Al Kork, J. A. Korbane, and A. Abdulmonem, "EEG mind controlled smart prosthetic arm," in *Proc. IEEE Int. Conf. Emerg. Technol. Innov. Bus. Practices Transformation Societies (EmergiTech)*, Aug. 2016, pp. 404–409.
- [7] 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.
- [8] D. Zhang et al., "Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 1–8.
- [9] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026007.
- [10] K. Chen, L. Yao, D. Zhang, X. Chang, G. Long, and S. Wang, "Distributionally robust semi-supervised learning for people-centric sensing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3321–3328.
- [11] H. He and D. Wu, "Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.
- [12] Y.-P. Lin and T.-P. Jung, "Improving EEG-based emotion classification using conditional transfer learning," *Frontiers Hum. Neurosci.*, vol. 11, p. 334, Jun. 2017.
- [13] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent EEG signal analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 715–719, May 2019.
- [14] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," 2018, *arXiv:1803.02155*.
- [15] D. Zhang, L. Yao, K. Chen, S. Wang, P. D. Haghighi, and C. Sullivan, "A graph-based hierarchical attention model for movement intention detection from EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 11, pp. 2247–2253, Nov. 2019.
- [16] D. Zhang, K. Chen, D. Jian, and L. Yao, "Motor imagery classification via temporal attention cues of graph embedded EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2570–2579, Sep. 2020.
- [17] A. Demir, T. Koike-Akino, Y. Wang, M. Haruna, and D. Erdogmus, "EEG-GNN: Graph neural networks for classification of electroencephalogram (EEG) signals," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 1061–1067.
- [18] J. N. Sanes and J. P. Donoghue, "Plasticity and primary motor cortex," *Annu. Rev. Neurosci.*, vol. 23, pp. 393–415, Mar. 2000.
- [19] A. Bechara, "Emotion, decision making and the orbitofrontal cortex," *Cerebral Cortex*, vol. 10, no. 3, pp. 295–307, Mar. 2000.
- [20] D. Luo et al., "Learning to drop: Robust graph neural network via topological denoising," in *Proc. 14th ACM Int. Conf. Web Search Data Mining*, Mar. 2021, pp. 779–787.
- [21] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *Proc. ICLR*, 2019, pp. 1–17.
- [22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
- [23] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [24] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, no. 2, pp. 251–257, 1991.
- [25] M. C. Tate, G. Herbet, S. Moritz-Gasser, J. E. Tate, and H. Duffau, "Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited," *Brain*, vol. 137, no. 10, pp. 2773–2782, 2014.
- [26] T. Chugh, K. Sindhya, J. Hakanen, and K. Miettinen, "A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms," *Soft Comput.*, vol. 23, no. 9, pp. 3137–3166, May 2019.
- [27] L. A. Moctezuma and M. Molinas, "Classification of low-density EEG for epileptic seizures by energy and fractal features based on EMD," *J. Biomed. Res.*, vol. 34, no. 3, p. 180, 2020.
- [28] L. A. Moctezuma and M. Molinas, "EEG channel-selection method for epileptic-seizure classification based on multi-objective optimization," *Frontiers Neurosci.*, vol. 14, p. 593, Jun. 2020.
- [29] M. Xia, J. Wang, and Y. He, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS ONE*, vol. 8, no. 7, Jul. 2013, Art. no. e68910.
- [30] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature Neurosci.*, vol. 20, no. 3, pp. 353–364, Feb. 2017.
- [31] X. Wu, W.-L. Zheng, Z. Li, and B.-L. Lu, "Investigating EEG-based functional connectivity patterns for multimodal emotion recognition," *J. Neural Eng.*, vol. 19, no. 1, Feb. 2022, Art. no. 016012.

- [32] B.-H. Kim and J. C. Ye, "Understanding graph isomorphism network for rs-fMRI functional connectivity analysis," *Frontiers Neurosci.*, vol. 14, p. 630, Jun. 2020.
- [33] A. Goldberger et al., "Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, 2000.
- [34] 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.
- [35] S.-L. Wu, C.-W. Wu, N. R. Pal, C.-Y. Chen, S.-A. Chen, and C.-T. Lin, "Common spatial pattern and linear discriminant analysis for motor imagery classification," in *Proc. IEEE Symp. Comput. Intell., Cognit. Algorithms, Mind, Brain (CCMB)*, Apr. 2013, pp. 146–151.
- [36] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.
- [37] Y. Rong, W. Huang, T. Xu, and J. Huang, "DropEdge: Towards deep graph convolutional networks on node classification," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–18. [Online]. Available: <https://openreview.net/forum?id=Hkx1qkrKPr>
- [38] E. Dai, W. Jin, H. Liu, and S. Wang, "Towards robust graph neural networks for noisy graphs with sparse labels," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, Feb. 2022, pp. 181–191.
- [39] B. Knyazev, G. W. Taylor, and M. Amer, "Understanding attention and generalization in graph neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
- [40] M. Bear, B. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain*. Burlington, MA, USA: Jones & Bartlett Learning, 2020.
- [41] H. Shan, H. Xu, S. Zhu, and B. He, "A novel channel selection method for optimal classification in different motor imagery BCI paradigms," *BioMed. Eng. OnLine*, vol. 14, no. 1, pp. 1–18, 2015.
- [42] F. Mattioli, C. Porcaro, and G. Baldassarre, "A 1D CNN for high accuracy classification and transfer learning in motor imagery EEG-based brain-computer interface," *J. Neural Eng.*, vol. 18, no. 6, Dec. 2021, Art. no. 066053.
- [43] S. Sreeja, J. Rabha, D. Samanta, P. Mitra, and M. Sarma, "Classification of motor imagery based EEG signals using sparsity approach," in *Proc. Int. Conf. Intell. Hum. Comput. Interact.* Cham, Switzerland: Springer, 2017, pp. 47–59.