The following publication R. Liu et al., "Dynamic Graph Representation Learning for Spatio-Temporal Neuroimaging Analysis," in IEEE Transactions on Cybernetics, vol. 55, no. 3, pp. 1121-1134, March 2025 is available at https://doi.org/10.1109/TCYB.2025.3531657.

Dynamic Graph Representation Learning for Spatio-Temporal Neuroimaging Analysis

Rui Liu, Zhi-An Huang, Yao Hu, Lei Huang, Ka-Chun Wong, Kay Chen Tan, *Fellow, IEEE*

***Abstract*—Neuroimaging analysis is an important research direction to reveal the information-processing mechanism of the human brain in a non-invasive way. Recently, the development of graph neural networks (GNNs) provides new insight into neuroimaging analysis due to their powerful graph representation learning ability in characterizing the non-Euclidean structure of brain networks. However, previous studies on neuroimaging analysis are largely centralized in the spatial static functional connectivity and ignore the temporal characteristics of neural dynamics observed in complex brain networks. In this study, we propose a spatio-temporal interactive graph representing network (STIGR) for dynamic neuroimaging analysis by capturing the spatio-temporal interaction from both local and global perspectives. From local perspective, a hybrid graph convolution network is developed to effectively capture complex spatio-temporal dynamics. From global perspective, a novel Transformer-based self-attention module is designed to help STIGR extract the long-term temporal correlation. To effectively model the connection relationship of dynamic graphs, the adjacency matrix of DAN-GCN is adaptively learned by contrastive learning. A cross-fusion unit is finally used to increase the interactions between local and global graph representations for extracting discriminative graph representation. To demonstrate the graph representation capability of the proposed framework, extensive experiments are conducted on neuroimaging data at three feature levels, node level, edge level, and global level, respectively. Thanks to the adopted dynamic graph attentive representation, the potent interpretability enables STIGR to detect the remarkable temporal association patterns among different brain regions based on dynamic neuroimaging signals.**

***Index Terms*—Neuroimaging analysis, graph neural networks, dynamic functional connectivity, model generalizability and interpretability, self-attention mechanism, functional magnetic resonance imaging, electroencephalography**

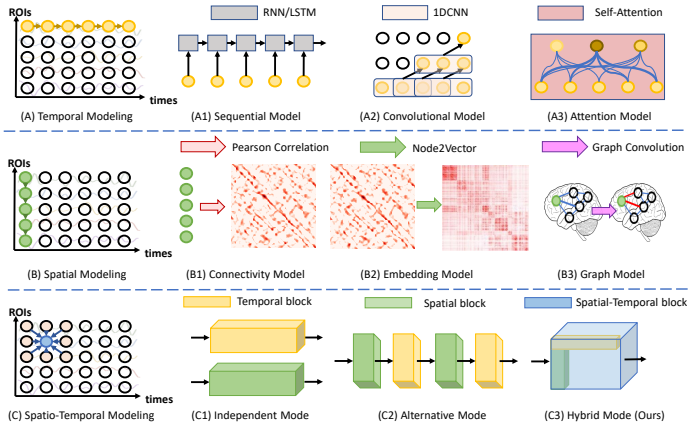


Fig. 1. Overview on different types of spatial and temporal modeling techniques. (A) is temporal modeling that attempts to infer dependencies between activities at different time steps. (A1), (A2) and (A3) illustrate previously commonly used temporal modeling techniques, including: RNN/LSTM, 1DCNN, and self-attention. (B) is spatial modeling exploring correlations between activities at different brain regions. (B1), (B2) and (B3) illustrate commonly used spatial modeling techniques, including: extracting structural connectivity via Pearson correlation coefficients, characterizing spatial similarity based on connectome embeddings, modeling adjacent correlation through graph convolution. (C) is the example of different types of spatial-temporal modeling. Specificlly, (C1) is the independent mode which seperately models spatial and temporal representation using different network blocks. (C2) is the alternative model that models the spatial and temporal representations using different network blocks one after the other (e.g. STGCN in [1]). (C3) is the hybrid mode that jointly models spatial and temporal representations using only one network block (e.g., our STIGR).

I. Introduction

The human brain is a complex neurobiological system that coordinate human behavior and cognition. Analyzing and comprehending the human brain network has become an enthralling endeavor for researchers with a variety of purposes, including mental disease diagnosis, brain-computer interface, and neuromorphic computing. Thanks to the recent advances, the activity of the human brain network can be captured through various neuroimaging techniques such as Magnetic-Resonance Imaging (MRI), Electroencephalography (EEG), Functional Near-Infrared Spectroscopy (fNIRS), etc. These techniques record brain activities by investigating different physiological parameters: 1) MRI detects the variations of blood-oxygen-level-dependent signals through utilizing a static magnetic field; 2) EEG records the electrogram of the spontaneous electrical activity of the brain; 3) fNIRS measures the blood flow signals through light sources and detectors placed on top of our head. Although these neuroimaging techniques can record brain activities at a fine spatial and temporal resolution [2], it lacks a general method that can effectively analyze the varying spatial-temporal dynamics of the underlying brain network. Especially for the development of computer-aided analysis systems on neuroimaging, different analysis models need to be established separately for different neuroimaging data. The technique heterogeneity makes it challenging to integrate them into a cross-platform analysis system, which greatly increases the training cost for neuroimaging analysts.

R. Liu, Y. Hu, and Ka-Chun Wong are with the Department of Computer Science, City University of Hong Kong, Kowloon Tong, Hong Kong SAR and also with the City University of Hong Kong Shenzhen Research Institute. (e-mail:{rliu38-c, yaohu4-c}@my.cityu.edu.hk, kc.w@cityu.edu.hk).

Z.-A. Huang is with the Center for Computer Science and Technology, City University of Hong Kong Dongguan Research Institute, Dongguan 523000, China. (e-mail:huang.za@cityu.edu.cn).

Kay Chen Tan is with the Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR (e-mail:kctan@polyu.edu.hk).

Graph neural networks (GNNs) is an optimizable transformation on all graph attributes (nodes, edges, and global context), which is robust to preserve the symmetries property of graph-structured data and permutation invariances [3]. The inherent graph-structure nature of the brain contributes to the prevalence of learning the representation of the brain network with GNN [4] so as to decode traits or states from human brain signal measurements such as fMRI, EEG, and fNIRS. Therefore, GNNs-based graph representation learning is expected to become a general model for analyzing these neuroimaging measurements. Some findings in recent studies could also demonstrate the potential of GNNs to address heterogeneity in different neuroimaging measures. For example, when GNN models were applied to diagnose depression based on fMRI, EEG, and fNIRS data, they all focused on functional changes occurring in prefrontal cortex regions [5], [6], [7]. Although these studies have shown potential strengths for learning the network representation of the brain, most of them still learn graph representation in a static manner. In neuroimaging analysis, static graph representation learning focuses on learning representations by viewing the entire spatio-temporal signal as a single fixed graph, while dynamic graph representation learning focuses on learning representations from viewing the entire spatio-temporal signal as multiple dynamic subgraphs that change over time. There is no absolute advantage or disadvantage between static and dynamic graph representation learning. Dynamic manner contains more information and has a higher upper limit in model accuracy, while static manner contains less redundant information, which is more conducive to model learning. However, growing studies suggest that distinct functional brain activity can vary within a very short period involving different activated brain regions, which can provide important insights in the field of neuroimaging analysis [8], [9]. Therefore, it is necessary to extend the GNNs by incorporating the dynamic feature of the brain network to explore spatial and temporal correlation in neuroimaging analysis flexibly.

Spatial temporal graph neural networks (STGNNs) is the recent advance in representation learning for time-series data, which can serve as one potential solution to capture the dynamic changes that occur in the brain over time from neuroimaging data [1], [10]. However, these methods of directly applying STGNNs to brain network analysis have some inherent limitations. First, current STGNNs typically model spatial and temporal dependencies independently or alternatively, which fails to capture the inter-relationship between spatial and temporal dynamics [11]. That is to say, these models remain restricted in their ability to comprehensively investigate how spatial and temporal elements interact with one another in dynamic brain network. Second, STGNNs usually apply convolution over the temporal scale of the input dynamic graph, they can not globally model spatio-temporal long-range dependencies that extend beyond the receptive field due to the restriction of the convolution kernel size [12]. Third, a fixed adjacency matrix is used to describe the topological information of the brain network in traditional STGNNs. Due to the complexity of brain activity, the connectivity of brain regions is not static but evolves over time [13]. Representing brain topological information directly with a fixed adjacency matrix makes it challenging to capture the temporal dependence of dynamic graphs accurately.

In this study, we propose the spatio-temporal interactive graph representing network (STIGR) for dynamic brain network analysis to learn discriminative spatio-temporal graph representations. The STIGR framework comprises three key components: dynamic adaptative-neighbor graph convolution network (DAN-GCN), spatio-temporal dual-attention network (ST-DAN), and contrastive learning-based adjacent matrix learning (CL-AM), which address aforementioned limitations of existing STGNNs. DAN-GCN is proposed to capture the complex inter-relationship between spatial and temporal dynamics by learning the intrinsic local dynamic graph representation between adjacent time periods. ST-DAN is designed based on the Transformer architecture to complement long-range dependencies via capturing the global attentive spatio-temporal representations. CL-AM is developed through utilizing contrastive learning to adaptively learn the adjacency matrix of the DAN-GCN to accurately model the connection relationship of dynamic graphs. Finally, we use a cross-fusion unit to increase the interactions between these spatio-temporal representations, allowing for the extraction of more discriminative graph representations for different downstream tasks (e.g., classification, prediction, or interpretation). The major contributions of this work are summarized as follows:

- We present for the first time to construct a general cross-platform and cross-task neuroimaging analysis framework from both spatial and temporal perspectives, which addresses previous difficulty in revealing complex spatial-temporal patterns of neuroimaging data.
- We proposed two novel spatial-temporal modeling components, DAN-GCN and ST-DAN, which utilize graph convolution and attention mechanisms to learn discriminative spatial-temporal representations from local and global perspectives, respectively. By exploiting contrastive learning, we adaptively learned the adjacency matrix of DAN-GCN guided by ST-DAN, which improves the interaction between local and global spatial-temporal representation learning processes.
- Extensive experiments demonstrate the successful performance of our proposed STIGR in cross-platform/task neuroimaging analysis across various types of neuroimaging data, including MRI, EEG, and fNIRS. Specifically, the proposed framework outperforms state-of-the-art (SOTA) neuroimaging analysis methods in tasks such as mental disorder diagnosis, motor imagery classification, and brain age prediction. Moreover, the learned graph representation facilitates interpretable deep learning, allowing for the exploration of salient neuroimaging patterns from node, edge, and global perspectives.

The rest of the paper is organized as follows. In Section II, we briefly review the related work regarding graph learning for brain network analysis and spatio-temporal prediction in the brain network. In Sections III, we introduce the detail of the proposed framework. Section IV presents the comprehensive experimental studies. Finally, this paper is concluded in

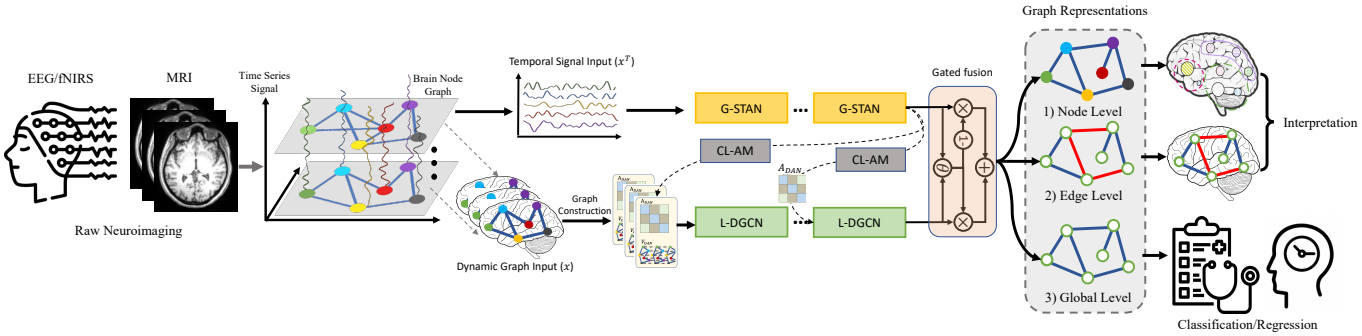


Fig. 2. An illustration of the general training and analysis framework for the proposed STIGR. For each preprocessed neuroimaging data from temporal perspective x^T (temporal signal), a stack of ST-DANs is first employed to extract the spatial-temporal representations globally. For each preprocessed neuroimaging data from spatial perspective x (dynamic graph), a stack of DAN-GCNs (along with the CL-AM module that adaptively learns adjacency matrices based on contrastive learning) is also utilized to capture graph spatial-temporal representations locally. The learned graph representations from two perspectives (temporal signal and dynamic graph) are fused via gated fusion module to obtain discriminative graph representations with local and global interactions. Based on the learned representations, a general cross-platform/task neuroimaging analysis pipeline is constructed to analyze different kinds of neuroimaging data including MRI, EEG, and fNIRS, from three representation levels: node level, edge level, and global level.

Section V.

II. Related Works

In this section, we first review the related works on graph representation learning for brain network analysis. Then, some previous works on modeling spatio-temporal dynamics in brain networks are introduced.

A. Graph Representation Learning for Neuroimaging Analysis

Graphs are a kind of data structure that describes a complicated system containing a set of objects and relationships [14]. Due to the majority of data in practical applications coming in the form of graphs (e.g., social network, knowledge graph, brain network, etc.), research on graph representation learning has raised widespread interest in the machine learning community. The intuition behind graph representation learning is to embed graphs into low-dimensional spaces while preserving graph topology and node attributes. Given that most neuroimaging problems can be modeled as graph representation learning tasks, it has recently played an important role in advancing deep learning applications in neuroimaging. For example, Parisot et al. [15] applied the vanilla GCN to learn discriminative topological graph representation of static functional connectivity (FC) brain networks for identifying in ASD and Alzheimer's disease. In [16], hypergraph learning was used to model the high-order relations for edge-level representation detection among multiple regions of interest (ROIs), which was used to calculate a unified hypergraph similarity matrix to estimate the learning ability of individuals. Additionally, for emotion recognition, [17] proposed a regularized GNN network to represent local and global inter-channel relationships in EEG signals while solving the challenge of model robustness to cross-subject EEG variations and noisy labeling. With the above static graph representation learning approaches, brain network analysis can benefit from convenient operations such as storing a fixed graph data structure and performing graph transformations. Since the topological property is consistently revealed, the reliability of graph measures in static brain connectivity can be easily studied.

Recently, dynamic graph representation learning has been demonstrated to effectively characterize the time-evolving brain connectivity at a system level. For predicting the age and gender of healthy individuals, Gadgil et al. [1] introduced a spatio-temporal GCN (STGCN) to model the non-stationary nature of short sub-sequences of fMRI time series. Following his work, Kim et al. [4] incorporated two attention-based modules in STGCN to learn dynamic graph representation of the brain connectome. A novel readout function and Transformer encoder are then employed for temporal attention statistical interpretation. Apart from the applications in fMRI data, [18] also designed a jumping knowledge graph convolution based on STGCN to explore intrinsic connections and relationships among multi-channel EEG signals for sleep stage classification. Impressively, these works provide evidence that the dynamics of brain connectivity can be highly reproducible across repeated scanning sessions. Compared with static graph representation learning methods, they can better learn the data-driven graph topology information and effectively capture the dynamics of brain activity.

B. Modeling Spatio-Temporal Dynamics in Neuroimaging Analysis

Most previous work on neuroimaging analysis has focused on exploring correlations between different signal channels in spatial domain. For example, the FC measure in fMRI data, which captures the inter-channel relations between different ROIs, is always served as the "fingerprints" to extract valuable feature representation for different neuroimaging analysis tasks in [19], [20], [21]. Common spatial patterns are also a well-known spatial filtering algorithm to effectively analyze multi-channel EEG/fNIRS signals from a spatial perspective [22], [23]. In addition to spatial correlation, these neuroimaging data also possess high temporal resolution to record salient temporal properties such as scan times and rest intervals,

which play an important role in investigating rapid activity and dynamic oscillation in different brain states. With advances in deep learning, especially sequence transduction models such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformers [24], several studies began to analyze neuroimaging from a temporal perspective and achieved decent performance in corresponding applications. For example, LSTM is successfully adopted in [25] and [26] to conduct EEG-based emotion recognition and fMRI-based ADHD classification, respectively. Through sequence modeling, dynamic connectivity features can be characterized as potential temporal biomarkers [27]. Nevertheless, all the methods discussed above focus only on the spatial or temporal dependencies of neuroimaging signals, failing to jointly study spatial and temporal dependencies, which have been shown to be one of the key approaches to understanding how the brain works [28].

Fortunately, there were a few early attempts to combine spatial and temporal features for neuroimaging analysis. Hartvig et al. [29] first proposed two statistical inference models to separately estimate the spatial and temporal activation patterns of fMRI data. Based on this work, Derado et al. [30] developed a two-stage autoregressive model to account for the spatial dependencies between voxels and the temporal dependencies between scanning sessions. Jrad et al. [31] also apply a two-stage method to address the EEG-based visual feedback error-related potentials task by extracting temporal features and spatial features by using global field power and SVM, respectively. Yet, these methods rarely explored the important spatio-temporal interaction, i.e., ignoring how the connectome-scale brain network temporally evolves, possibly due to neuroimaging data scarcity in the past. More recently, deep learning models have been capable of automatically extracting the hierarchical non-linear hidden features with different levels of complexity. Zhao et al. [32] proposed a 3D convolutional neural network (CNN) architecture to fully utilize the features on spatial and temporal dimensions for EEG-Based motor imagery classification. Mao et al. [33] further extend the 3D CNN to 4D architecture to simultaneously generate both spatial and temporal characteristics from targeted networks in fMRI data. Considering the fact that the brain network is a natural fit for graph theory, several STGNNs were presented to measure intra-subject temporal dynamics and inter-regional spatial associations [1], [4]. Among these representative STGNNs, STGCN is an efficient variant by applying convolution over the time-varying dynamic graphs. In this way, the spatial and temporal nature of neuroimaging data can be exploited by STGCN, thus can be more promise to learning discriminative graph representation.

III. Methodology

A. Problem Definition

The primary goal of our work is to learn the effective graph representation of neuroimaging data for downstream auxiliary classification/prediction/interpretation tasks. To capture the dynamic property of neuroimaging data, we model the time-series signal into a dynamic graph network $G(X) = \{\mathcal{G}(x_1), ..., \mathcal{G}(x_T)\}$ with T timestamps, where $\mathcal{G}(x_t)$ represents the graph representation of input x at the t-th timestamp. Specifically, the dynamic graph at the t-th timestamp $\mathcal{G}(x_t)$ contains a set of vertices $V(t) \in \mathbb{R}^{N\times T}$ (i.e., N involved ROIs in fMRI and channels in EEG) as well as an adjacency matrix $A \in \mathbb{R}^{N\times N}$ corresponding to the connections between $V(t)$, i.e., $\mathcal{G}(t) = \{V(t), A(t)\}$. Typically, in neuroimaging analysis, the adjacency matrix A is measured by the Pearson correlation coefficient between different v_t as,

$$A_t(i,j) = \frac{\sum_{k=1}^{S}(v_t(i,k) - \overline{v_t(i)})(v_t(j,k) - \overline{v_t(j)})}{\sqrt{\sum_{k=1}^{S}(v_t(i,k) - \overline{v_t(i)})^2 (v_t(j,k) - \overline{v_t(j)})^2}} \tag{1}$$

where i and j represent the *i*th and *j*th nodes, respectively. Therefore, our task is to learn a mapping function f that computes the correspondences between the dynamic spatio-temporal graph representation of input neuroimaging signal $X \in \mathbb{R}^{N\times T}$ and the ground-truth label Y. The optimization problem can be reached as follows:

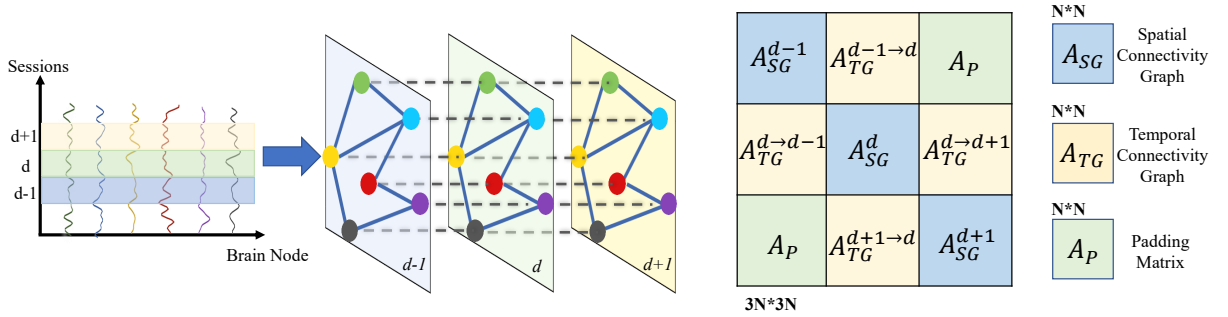
$$f^* = \arg\min_f \mathbb{E}_{X,Y}[\mathcal{L}(f(G(X)), Y)] \tag{2}$$

where $\mathbb{E}$ is the expectation of the loss function $\mathcal{L}$ over space of (X, Y).

B. Model Design

In this section, we illustrate the graph representation learning framework of STIGR, as shown in Figure 2. The STIGR mainly consists of three components including 1) DAN-GCN, 2) ST-DAN, and 3) CL-AM. Two novel modules DAN-GCN and ST-DAN are performed for jointly learning the spatial-temporal graph representation of neuroimaging data at local and global levels, respectively. The CL-AM is introduced to adaptively learn the adjacency matrix of DAN-GCN to model the dynamic graph node connections. Finally, late adding fusion is leveraged to combine the local and global spatio-temporal representations for extracting discriminative graph representations with local and global interactions. Due to different downstream tasks, the obtained graph representation is fed into different head networks for further classification/prediction/interpretation.

1) Dynamic Adaptative-Neighbor GCN: As we can see in Figure 3, the main task of DAN-GCN is to capture the spatio-temporal inter-dependences between adjacent scanning sessions based on the graph construction of dynamic adaptative-neighbor ($\mathcal{G}_{DAN} \in \{V_{DAN}, A_{DAN}\}$). In other words, our proposed dynamic adaptative-neighbor graph can connect individual spatial graph of adjacent time sessions into one graph to simultaneously capture the complicated spatio-temporal inter-dependences. It is worth noting that we divided the T timestamps signal into D number of scanning sessions with length S (e.g., $D = \lfloor T/S \rfloor$) to explore the short-term dynamics of the brain network. The scale of $\mathcal{G}_{DAN}$ is determined by the number of adjacent scanning sessions, e.g., we set it to 1 as shown in Figure 3, resulting in $V_{DAN} = [V(d-1), V(d), V(d+1)] \in \mathbb{R}^{3N\times S}$ (as shown in Figure 3(a)) and A_{DAN} with the size of $3N \times 3N$ (as



(a) Input of Dynamic Adaptative-Neighbour Graph (b) Adjacent Matrix of Dynamic Adaptative-Neighbour Graph

Fig. 3. Dynamic adaptative-neighbor graph construction. (a) is an example of input of the dynamic adaptative-neighbor graph, which would be generated along the time axis. (b) is the adjacency matrix of the dynamic adaptative-neighbor graph in (a).

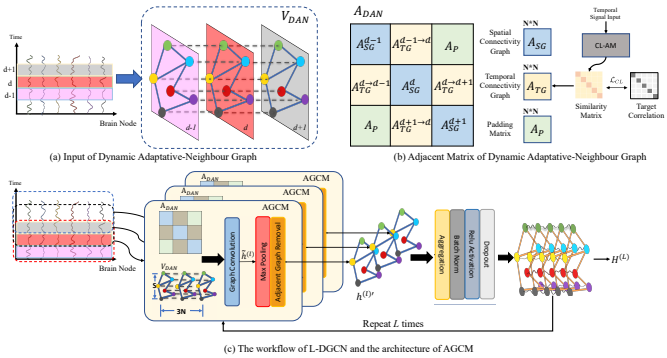


Fig. 4. (a) Workflow process of DAN-GCM at t-th timestamp. (b) The layered architecture of DAN-GCN.

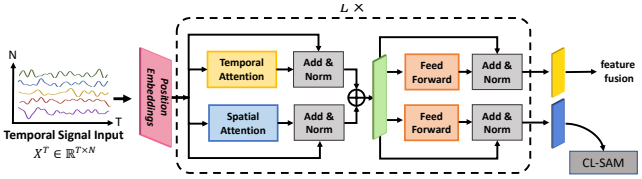


Fig. 5. The structure of the proposed ST-DAN.

shown in Figure 3(b)). Inspired by the recent advance in [11], we leverage A_{DAN} to record three kinds of adjacency matrix $\in N \times N$, i.e., spatial connectivity graph, temporal connectivity graph, and padding matrix as follows:

- Spatial connectivity graph matrix (A_{SG}), is placed on the diagonal of A_{DAN}, with the aim of representing the dynamic graph in each adjacent scanning session based on the 10% top remarkable spatial correlation computed by Eq. (2).
- Temporal connectivity graph matrix (A_{TG}) is the self-adpative learned matrix through CL-AM module, which will be elaborated in the following subsection.
- Padding matrix (A_P) is placed to fill the remaining empty space for padding operations by setting all its members to 0.

In this way, $\mathcal{G}_{DAN}$ can be constructed to extract the local spatio-temporal correlations of dynamic graph matrices by the following dynamic adaptative-neighbor graph convolution module (DAN-GCM).

By expanding such a dynamic graph correlation estimation to the whole time course, we utilize DAN-GCM to capture hybrid spatio-temporal relations among the whole dynamic brain network as shown in Figure 4. Based on the constructed $\mathcal{G}_{DAN}$, graph convolution is performed to aggregate the hybrid graph representation of the central node with its neighbors within adjacent scanning sessions. Given the l-th input graph representation as $h_{in}^{(l)} \in \mathbb{R}^{3N \times S}$, graph convolution can be formulated as follow:

$$\tilde{h}^{(l)} = \sigma(A_{DAN}(h_{in}^{(l)} W + b)) \tag{3}$$

where the variables $W \in \mathbb{R}^{S \times S}$ and $b \in \mathbb{R}^{S}$ are learnable parameters and σ represents the activation function of rectified linear unit ($ReLU$) in this work. As we can see in Figure 4(a), after graph convolution, we perform the max pooling and then reserve only the graph representation matrix $h^{(l)'}$ of $N \times S$ at the current timestamp. As such, each scanning session of spatio-temporal graph input $H^{(l)}$ is parallelly processed by DAN-GCM to construct a DAN-GCN layer to generate the local graph representations $\{h_d^{(l)'}\}_{d=1}^{D}$ as shown in Figure 4(b). Note that we need to perform zero-padding to assist the kernel of DAN-GCM in processing the first and last scanning sessions. Then, an aggregation layer is employed to concatenate $h^{(l)}$ as follows,

$$H^{(l)'} = \left\langle h_1^{(l)'}, ..., h_D^{(l)'} \right\rangle \in \mathbb{R}^{D \times N \times S} \tag{4}$$

where $\langle . \rangle$ denotes the concatenation operator. In the tail of DAN-GCN, the layers of batch normalization, $ReLU$ activation, and dropout are cascaded to output the extracted local spatio-temporal graph representation $H^{(l+1)}$. The input of the first DAN-GCN layer $H^{(0)}$ is set as the dynamic graph input signal X in this paper, i.e., $H^{(0)} = X \in \mathbb{R}^{N \times T}$. The final output after L layers DAN-GCN is represented as $H^{(L)}$ for further convenience. It is worth mentioning that the length of the scanning session S increases with the number of layers, thereby expanding the receptive field of our DAN-GCN in the temporal dimension.

2) Spatio-Temporal Dual-Attention Network: In addition to modeling the short-term dynamic behavior in DAN-GCN, characterizing global temporal dependency is also significant in estimating the time-varying architecture of FC network across the scanning time. The Transformer [24] is an effective framework to capture the global self-attention by calculating the dependency of temporal tokens/patterns on others. However, the vanilla Transformer was designed for one-dimensional sequence data and did not take into account the spatial dependencies of the inherent spatio-temporal properties of our brain network analysis tasks. To this end, we develop ST-DAN based on the Transformer encoder, which learns global spatio-temporal representation by decoupling attention in temporal and spatial dimensions.

As shown in Fig. 5, the input time-series sequences $X^T \in \mathbb{R}^{T\times N}$ is first fed into the position embedding to obtain the hidden feature representation $E \in \mathbb{R}^{T\times N}$ that encoded with position information. Then, a dual self-attention structure (stacked L layers), including temporal attention block and spatial attention block, is designed to extract the attention focus from temporal and spatial dimensions over the input embedded feature sequence E. For each input time-series signal, we defined the attention at different time points on the same node (e.g. ROI of fMRI and channel of EEG) as temporal attention, while the attention of different nodes at the same time point as spatial attention. In the temporal attention block, three matrix representations Q, K, and V ($queries$, $keys$, and $values$) is first learned given the input embedded representation E as,

$$Q = f_Q(E), K = f_K(E), V = f_V(E) \tag{5}$$

where f_Q, f_K, and f_V are the corresponding projection functions of $queries$, $keys$ and $values$. Then, the $queries$ is compared against key-$value$ pairs through dot-product similarity to obtain the attention distributions on $values$. If the $queries$ and the $keys$ are similar (i.e., high attention weight) means that the corresponding $values$ are assumed to be more related. The resulting weighted $values$ matrix forms the output of the attention block. Following [24], the multi-head attention (MHA) is adopted to extract the temporal-level attention, which projects the $queries$, $keys$, and $values$ h times with different linear projection heads. The output representation of temporal attention block $E_T^{'}$ can be formulated as follows:

$$\tilde{E}_T = MHA(E) = \langle head_1, ..., head_h \rangle W^o \tag{6}$$

$$head_j = Att_j(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{N}})V \tag{7}$$

where $\langle . \rangle$ is a concatenation operator. The $W^o \in \mathbb{R}^{N\times N}$ represents the learned linear transformation matrices of the final head concatenation. Correspondingly, the construction of the spatial attention block is the same as the temporal attention block. However, the spatial-level Q, K, and V matrix are learned over the transpose of the input embedding feature (i.e., $E^T \in \mathbb{R}^{N\times T}$) to model the attention from all nodes at the same timestamp. The output representation of the spatial attention block is denoted as $\tilde{E}_S$. Since the temporal and spatial attention blocks run in parallel to calculate $\tilde{E}_T$ and $\tilde{E}_S$ respectively, they fail to bridge the inter-dependencies between spatial and temporal representation. Therefore, we fused the temporal and spatial attention representation as $\tilde{E} = \tilde{E}_T \oplus \tilde{E}_S$ (where $\oplus$ denotes the element-wise addition) and passed to two specific feed-forward networks. In this way, our network can well extract the spatio-temporal inter-dependencies while keeping exploring the modality-specific inner-dependencies. To facilitate the model optimization, residual connections with layer normalization (shorted as $Add\&Norm$) are respectively applied to our dual attention structure as shown in Fig. 5. In this paper, we stack L spatio-temporal attention layers to successively update the fused embeddings. Finally, our ST-DAN outputs temporal and spatial representations $E_T^{(L)}$ and $E_S^{(L)}$ for further global-local feature fusion and CL-based adjacent matrix learning, respectively.

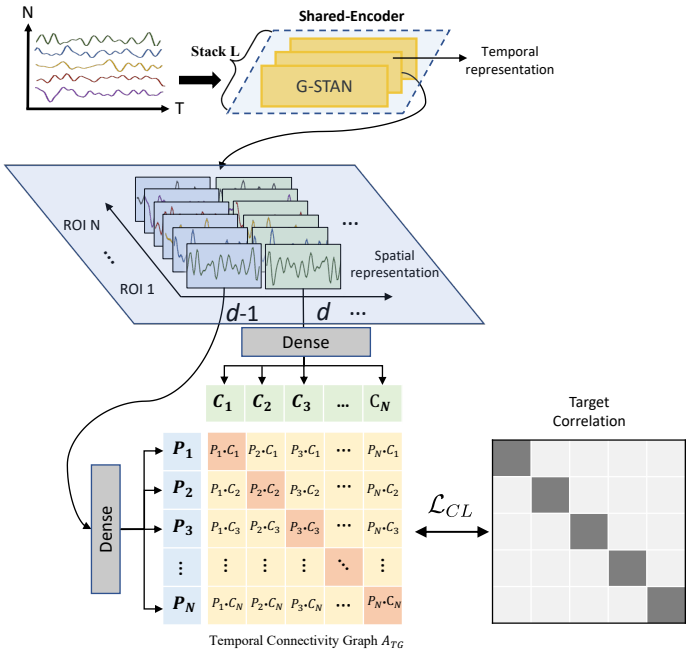


Fig. 6. The structure of the CL-AM.

3) Contrastive Learning-based Adjacent Matrix Learning: Since we expect the proposed DAN-GCN to dynamically capture the hybrid graph representation of each central node and its neighbors in adjacent scanning sessions, it is necessary to simultaneously model the node's connections within the same session and between different sessions. As we illustrated above, the node connections within the same session are modeled by the spatial connectivity graph matrix A_{SG} using the Pearson correlation coefficient. However, node connections between various sessions are challenging to obtain due to their sophisticated cross-temporal connectivity characteristic. To this end, we proposed the CL-AM module, which employs contrastive learning to address the cross-temporal connectivity problem.

The intuition behind contrastive learning is to group 'positive' samples closer and diverse 'negative' samples far from

Algorithm 1 Pseudo-code of the STIGR

Require: Preprocessed data X, and label Y
Ensure: Predicted probabilities of testing set Y_{pred}^{te}
$[X^{tr}, X^{te}, Y^{tr}, Y^{te}] \leftarrow$ Split(X, Y)
Initialize STIGR.
for $e = 1, \ldots, epochs$ **do** //e: # of training epoch
  // Spatio-temporal Dual-Attention Network
  E = PositionEmbedding (X^{trT})
  for $l = 0, \ldots, L$ **do** // l: # of attention layers
    // $E_T^{(0)} = E; E_S^{(0)} = E^T$
    $\tilde{E}^{(l)}$ = MHA$(E_S^{(l)}) \oplus MHA(\mathrm{E}_S^{(l)})$
    $E_T^{(l+1)}$ = AddNorm(FFN$(\tilde{E}^{(l)})$, $\tilde{E}^{(l)}$)
    $E_S^{(l+1)}$ = AddNorm(FFN$(\tilde{E}^{(l)})^T$, $\tilde{E}^{(l)T}$)
  // Dynamic Adaptative-Neighbor GCN
  for $l = 0, \ldots, L$ **do** // l: # of DAN-GCN layers
    $A_{DAN}^{(l)}$ = CL-AM$(E_S^{(L)})$
    $H^{(l+1)}$ = DAN-GCN$(H^{(l)}, A_{DAN}^{(l)})$ // $(H^{(0)} = X^{tr})$
  end for
  $R = E_T^{(L)} \oplus H^{(L)}$ // final dynamic graph representation
  // Connect to different predict heads for different tasks
  $output \leftarrow$ Predictor(R)
  $Loss \leftarrow \alpha\mathcal{L}(output, Y^{tr}) + (1-\alpha)\mathcal{L}_{CL}$
  STIGR.update$(Loss)$
end for
$Y_{pred}^{te} \leftarrow$ STIGR.predict (X^{te})

each other by measuring the similarity metric of two embeddings. Unlike common contrastive learning methods [34], [35], in this paper, we defined the scanning sessions at the same node as the 'positive' samples and the scanning sessions at different nodes as the 'negative' samples. In this way, we can learn the temporal connectivity graph A_{TG} by measuring the similarity metric between the embeddings of different scanning sessions based on contrastive learning. Since the representations of scanning sessions are already available from ST-GCN's spatial stream, the ST-DAN can be considered as a shared encoder for the CL-AM module to encode the underlying share information between different scanning sessions. As we can see from Figure 6, the spatial-level representation $\bar{E}_S$ is divided into D sequences to learn the pair-wise temporal connectivity based on contrastive learning. Take the learning of the $A_{TG}^{d-1\to d}$ as an example. The feature representations of scanning sessions $d-1$ and d are first encoded by the spatial-level attention network of ST-DAN. Then, the representations of two scanning sessions are fed into two different linear projection heads to map the representation to latent space for better similarity calculation [36]. Two groups of embeddings at the previous session $d-1$ and current session d are therefore obtained and denoted as P and C, respectively. Taking the embeddings of N nodes at two sessions as samples for contrastive learning, our CL-AM is trained to predict which of the $N \times N$ possible (previous session, current session) pairings actually occurred. Therefore, the CL-AM maximizes the similarity of the previous and current sessions embeddings of the N 'positive' pairs while minimizing the similarity of the embeddings of the $N^2 - N$ ' negative' pairings by optimizing the contrastive loss function L_{CL}^m as follows,

$$\mathcal{L}_{CL}^m = \sum_i (1 - SC_{ii})^2 + \lambda \sum_i \sum_{i \neq j} SC_{ij}^2, \tag{8}$$

where m indicates the m-th pair-wised temporal connectivity. The $\mathcal{L}_{CL}^m$ consists of two loss terms to simultaneously guarantee to attract of "positive" pairs while repelling "negative" pairs. The λ is a positive constant trading of the importance of the first and second terms of the loss. The notation SC is the similarity correlation matrix computed between the embeddings of the two scanning sessions along the node dimension:

$$SC_{ij} = \frac{P_i C_j}{\|P_i\|_2 \cdot \|C_i\|_2} \tag{9}$$

where $\|\cdot\|_2$ is l2-norm. Finally, by optimizing the L_{CL}^m loss, the temporal connectivity graph A_{TG} can be well modeled by the learned similarity matrix S for further dynamic graph convolution operation. Since the number of pair-wise temporal connectivity varies from different tasks, the overall contrastive loss is defined as $\mathcal{L}_{CL} = \sum_m \mathcal{L}_{CL}^m$ for final optimization. Except for the feature similarity perspective, constrative learning in our work can also be interpreted from the view of contrastive predicting coding, i.e., by considering the embedding of the previous session P as a prediction to the current session embedding C. The interpretation of contrastive learning in the predictive view is consistent with observations in neuroscience that the brain predicts various levels of abstraction [37].

4) Architecture: In this paper, due to the limited neuroimaging data, the stacked layer number L of ST-DAN and DAN-GCN are both set to 2. To expand the receptive field of DAN-GCN on the temporal level, different scanning session lengths S are adapted in different DAN-GCN layers (e.g., $S = 30$ in the first DAN-GCN and $S = 15$ in the second DAN-GCN). Therefore, adjacent matrices A_{DAN} for different DAN-GCN layers are learned independently through different CL-AM modules. After we obtained the final outputs $E_T^{(L)}$ and $H^{(L)}$, a simple adding fusion is leveraged to combine the local and global spatio-temporal representations for extracting discriminative graph representations with local and global interactions as $R = E_T^{(L)} \oplus H^{(L)}$. Finally, the fused feature is fed into different heads for different downstream tasks. For our classification task, the classification head network is composed of graph global average pooling [38] (used to represent the entire graph at global-level for further binary classification) and two dense layers. Since contrastive learning is incorporated into our model, the overall optimization function in (1) needs to be revised as follows:

$$f^* = \arg\min_f \mathbb{E}_{X,Y}[\alpha\mathcal{L}(f(G(X)), Y) + (1-\alpha)\mathcal{L}_{CL}] \tag{10}$$

where α is a learnable parameter to balance the loss weights automatically. To ease the understanding of our whole learning framework, the pseudo-code is shown in Algorithm 1.

TABLE I
MODEL PARAMETER SETTINGS FOR EACH DATASET ON CLASSIFICATION TASK

For fMRI data					
# of graph node	190	# of temporal point	90	Batch size	128
Max position embeddings	512	Length of S (l1,l2)	[15,30]	Training epochs	50
Classification head network	GlobalAveragePooling-64(Relu)-Dropout(0.5)-10(Relu)-1(Sigmoid)				
For BCICIV_2a					
# of graph node	22	# of temporal point	600	Batch size	32
Max position embeddings	1024	Length of S (l1,l2)	[100,200]	Training epochs	30
Classification head network	GlobalAveragePooling-Dropout(0.5)-1(Sigmoid)				
For BCIC2015					
# of graph node	56	# of temporal point	240	Batch size	128
Max position embeddings	512	Length of S (l1,l2)	[40,80]	Training epochs	50
Classification head network	GlobalAveragePooling-64(Relu)-Dropout(0.5)-10(Relu)-1(Sigmoid)				
For fNIRS-BCI					
# of graph node	52	# of temporal point	120	Batch size	16
Max position embeddings	512	Length of S (l1,l2)	[30,40]	Training epochs	100
Classification head network	GlobalAveragePooling-64(Relu)-Dropout(0.5)-10(Relu)-1(Sigmoid)				

IV. EXPERIMENTS AND RESULTS

A. Data Acquisition

We evaluate the performance of our STIGR on the three different kinds of publicly available neuroimaging datasets: 1) fMRI data on Autism Brain Imaging Data Exchange I & II (ABIDE I/II[1]) datasets and ADHD-200 Consortium (ADHD-200[2]) dataset, 2) EEG data on brain-computer interface (BCI) Competition IV 2a (BCICIV_2a[3]) and BCI Challenge 2015 (BCI2015[4]), and 3) fNIRS based BCI dataset (fNIRS-BCI [5]).

ABIDE I & II: The ABIDE I contains 1035 valid fMRI samples with 505 ASD subjects and 530 typical controls (TCs) aggregated from 17 different brain imaging sites. The ABIDE II dataset is also a multi-site dataset consisting of 19 different sites with 1113 valid subjects, including 521 ASD participants and 592 TCs. The preprocessed graph-structure time series signal according to the preprocessing pipeline Configurable Pipeline for the Analysis of Connectomes (C-PAC) can be directly downloaded from ABIDE I. Since ABIDE II only provides raw data without preprocessing, the Data Processing Assistant for Resting-State fMRI (DPARSF) [39] tool was applied to preprocess the raw signals in our work. The Craddock 200 (CC200) functional parcellation atlas was used to aggregate the preprocessed brain signals into 200 ROIs for reducing information redundancy.

ADHD-200: The ADHD-200 includes 939 valid fMRI samples from 8 international imaging sites involving 358 children and adolescents with ADHD and 581 TCs. Thanks to the data-sharing efforts of ADHD-200, the preprocessed data with CC200 parcellation is also open-sourced.

BCICIV_2a: The BCICIV_2a is an EEG-based motor imagery dataset collected from nine subjects. The brain signals of four different motor imagery tasks are recorded using 22-channel EEG, namely the imagination of movement of the left hand, right hand, both feet, and tongue, respectively. The EEG data were sampled at 250 Hz, bandpass filtered between 0.5 and 100 Hz, and a 50-Hz notch filter was applied to suppress line noise. Two sessions were provided for training and testing, respectively.

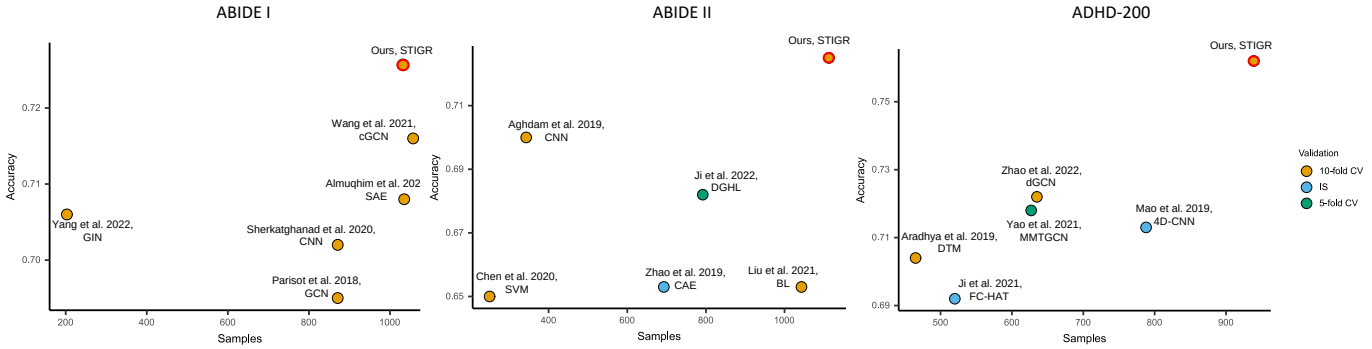
BCIC2015: The BCIC2015 EEG dataset was collected for 26 healthy subjects performing an error detection task designed based on the "P300-Speller" paradigm [40]. Each participant in this experiment was given a set of letters and numbers (36 items) to form words. Select one-word item at a time by flashing groups of screen items in random order. The goal of this challenge is to determine when the selected item is incorrect by analyzing 55 channels of EEG signals after the subject has received feedback. Their signals were sampled at 600 Hz and were bandpass filtered by a 5-th order Butterworth filter between 1 and 40 Hz.

fNIRS-BCI: The fNIRS-BCI dataset was collected from eight subjects (age 26 ± 2.8 years, three males and five females). Subjects were instructed to perform a cue-guided mental arithmetic task, i.e., subtract a one-digit number sequentially from a two-digit one for 12 seconds after the cue. The task-related period was followed by a 28-second rest period. During the mental arithmetic experiments, the changes of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) in the prefrontal cortex are recorded by a 52-channel fNIRS. To remove baseline drift, a bandpass filter with a passband of 0.002–0.018 Hz was adopted for all signals in the fNIRS-BCI dataset.

B. Model Settings

For a more reliable performance evaluation, cross-validation (CV) is conducted on ABIDE I & II, ADHD-200, BCIC2015, and fNIRS-BCI. Since BCICIV_2a sampled their data in two sessions, it is natural to train and test on the two sessions' data in an independent set (IS) manner, which is also the most common performance evaluation method on these two datasets. To evaluate with BCICIV_2a data, six combinations of binary motor imagery tasks, left-hand versus right-hand (L/R), left-hand versus feet (L/F), left-hand versus tongue (L/T), right-hand versus feet (R/F), right-hand versus tongue (R/T), and feet versus tongue (F/T), are tested. Since some subjects in these datasets have different lengths of time courses, we need to maintain the same sequence length for each subject sample to normalize the inputs for model training. Random cropping is commonly used to maintain the same sequence length while augmenting data samples, i.e., each sample is randomly cropped into a certain number of sequences with a fixed length. For a fair test, the cropped sequences of the same sample cannot be overlapped in both the training set and the testing set. The detailed model parameter settings of STIGR on the classification task of different datasets are shown in Table I. All the experiments are conducted under the same runtime environment using one intel core i7-8700K@3.70GHz, one NVIDIA GeForce RTX 2080 Ti GPU, and 64GB RAM. The experiment results are compared in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC).

[1]http://fcon_1000.projects.nitrc.org/indi/abide
[2]http://fcon_1000.projects.nitrc.org/indi/adhd200/
[3]https://www.bbci.de/competition/iv/
[4]https://www.kaggle.com/c/inria-bci-challenge
[5]http://bnci-horizon-2020.eu/database/data-sets



Note: Since both ABIDE I/II and ADHD-200 are cross-site datasets collected from different medical institutions using different scanning devices and protocols, it would cause large data heterogeneity and lead to inaccurate analysis [41]. In other words, training a model with more data from these datasets does not guarantee better model accuracy.

Fig. 7. Performance comparison with the state-of-the-art methods on ABIDE I, ABIDE II, and ADHD-200.

TABLE II
PERFORMANCE COMPARISON WITH OTHER SOTA METHODS ON BCICIV_2A DATASET

Method	task	A01	A02	A03	A04	A05	A06	A07	A08	A09	mean
Amin et al. 2019 [42]	L/R	78.10	61.10	79.88	69.08	94.09	71.49	91.70	95.85	94.10	81.71
	L/F	85.07	71.18	80.25	86.11	87.14	80.92	82.67	83.02	91.71	83.12
	L/T	88.94	79.88	87.54	85.11	86.47	78.06	87.51	92.35	98.86	86.75
	R/F	88.88	74.31	75.00	83.67	85.79	72.25	79.89	78.15	78.47	79.60
	R/T	98.24	75.71	91.31	77.04	82.24	76.03	86.48	84.36	87.52	84.33
	F/T	80.25	85.79	84.70	77.77	73.73	80.60	85.81	80.58	92.72	83.55
	mean	86.58	74.66	83.11	79.80	86.58	76.23	85.68	85.72	90.23	83.18
Bang et al. 2021 [43]	L/R	90.26	68.47	95.46	78.47	91.32	72.59	94.80	97.22	94.10	86.96
	L/F	97.91	83.31	90.31	87.16	81.92	81.22	97.91	89.29	95.14	89.35
	L/T	94.78	72.23	95.52	83.65	85.79	76.75	97.92	93.74	98.61	88.78
	R/F	95.84	77.02	93.05	89.62	88.23	78.84	97.56	89.58	84.41	88.24
	R/T	97.55	73.28	96.18	85.81	87.84	74.35	96.19	88.57	93.73	88.17
	F/T	76.07	75.05	82.67	77.76	75.71	78.82	85.11	91.00	90.28	81.39
	mean	92.07	74.89	92.20	83.74	85.14	77.09	**94.92**	91.57	**92.71**	87.15
Gaur et al. 2021 [44]	L/R	86.81	64.58	95.83	67.36	68.06	67.36	80.56	97.22	92.36	80.02
	L/F	97.22	63.89	93.06	82.64	65.97	70.83	97.92	84.2	96.53	83.64
	L/T	97.22	65.97	94.44	88.19	74.31	72.22	93.75	92.36	97.22	86.19
	R/F	97.22	80.56	93.06	89.58	70.83	64.58	93.75	88.19	84.03	84.64
	R/T	100	66.67	94.44	86.81	65.97	71.53	93.75	89.58	78.47	83.49
	T/F	69.44	73.61	69.44	62.50	68.06	70.14	79.86	82.64	85.42	72.99
	mean	91.32	69.21	90.05	79.51	68.87	69.44	89.93	88.43	89.70	81.83
Das et al. 2022 [45]	L/R	95.14	58.33	97.92	79.92	84.03	65.28	85.42	95.83	93.75	83.96
	L/F	97.92	77.78	95.14	85.42	68.06	63.19	99.31	93.75	95.14	86.19
	L/T	97.92	68.75	95.14	86.11	80.56	75.69	97.22	95.14	97.92	88.27
	R/F	99.31	80.56	95.83	85.42	77.78	55.56	100.0	90.97	85.42	85.65
	R/T	100.0	69.44	96.53	81.94	82.64	70.14	97.22	89.58	72.22	84.41
	T/F	75.69	79.86	79.86	68.75	61.11	66.67	88.19	86.81	88.19	77.24
	mean	94.33	72.45	93.40	81.26	75.70	66.08	94.56	92.01	88.77	84.28
Ours	L/R	94.44	77.08	98.26	83.33	88.27	81.25	76.39	96.53	91.32	86.42
	L/F	97.22	89.58	96.88	88.19	80.21	84.72	96.53	92.01	97.22	91.40
	L/T	97.57	81.60	94.44	89.58	82.99	82.99	94.10	93.40	98.61	90.59
	R/F	97.57	88.58	96.88	91.32	79.17	80.91	97.22	95.14	84.38	90.24
	R/T	98.96	82.64	98.96	84.38	82.99	82.99	96.53	92.01	90.63	90.01
	T/F	86.81	90.82	86.48	79.86	78.13	80.56	92.71	89.58	89.93	86.04
	mean	**95.43**	**87.43**	**95.31**	**86.11**	**80.44**	**82.24**	92.25	**93.11**	92.02	**89.28**

TABLE III
PERFOTRMANCE COMPARISON WITH SOTA METHODS ON FNIRS-BCI DATASET

Method	Validation	Acc	Sen	Spe
Panagiotis et al. 2018 [46]	CV	92.52	-	-
Lu et al 2020 [47]	5-fold CV	95.30	-	-
Wang et al. 2022 [48]	5-fold CV	92.97	93.06	93.55
Han et al. 2022 [49]	5-fold CV	95.26	96.63	95.33
Ours	5-fold CV	**97.51**	**97.63**	**96.13**

Note: [46] adopted leave-p-out cross-validation, where 70% of data were randomly chosen to be used for training and the rest (p observations) were kept for testing.

C. Performance Comparison

In this section, the proposed STIGR is compared with SOTA methods on fMRI, EEG, and fNIRS-BCI datasets in the preliminary classification (global-level representation learning) task, respectively. For fMRI datasets, our STIGR achieves competitive accuracy of 72.73% (SEN: 76.08%, SPE: 68.84%, and AUC: 79.03%), 72.53% (SEN: 73.00%, SPE: 71.56%, and AUC: 77.59%), and 76.15% (SEN: 79.47%, SPE: 72.35%, and AUC: 81.75%) on ABIDE I, ABIDE II, and ADHD-200, respectively, as shown in Figure 7. It is worth noting that both ABIDE I/II and ADHD-200 are cross-site datasets collected from different medical institutions using different scanning equipment and protocols, which can create large data heterogeneity and lead to inaccurate analysis [41]. In other words, training a model with more data from these datasets does not guarantee better model accuracy. As we can observe, deep learning techniques such as CNN, stacked autoencoder (SAE), and graph isomorphism network (GIN) have recently aroused intense scholarly interest in developing related methods for CAD of mental disorders. However, they struggle to reach an accuracy improvement when the sample size becomes larger (>800) due to the data heterogeneity issue. At this point, STIGR shows a powerful generalization capability on cross-site datasets thanks to the extracted hybrid spatio-temporal graph representation. For EEG data, our STIGR still outperforms current SOTA methods on both BCICIV_2a and BCIC2015 datasets. The proposed method achieves 89.28% mean accuracy across subjects on BCICIV_2a and 75.53% inter-subject accuracy in 4-fold CV on BCIC2015, respectively. The classification results of BCICIV_2a dataset for each subject and six binary classification tasks are shown in Table II. The highest average accuracy on the six motor imagery classification tasks is provided by our STIGR model for seven of the nine subjects. It also outperforms the SOTA method [43] by 2.13% in overall average accuracy on BCICIV_2a dataset. Since BCIC2015 dataset is used for open competition on Kaggle, we compared the performance of our STIGR with the winning algorithm. The winning algorithm without using leaky features achieves 72.94(±3.7)% accuracy[6], while our STIGR outperforms it with 75.53(±4.2)% accuracy in 4-folds

[6]https://github.com/alexandrebarachant/bci-challenge-ner-2015

TABLE IV
PERFOTRMANCE COMPARISON WITH OTHER REPRESENTATIVE STGNN ARCHITECTURES IN NEUROIMAGING FIELD

Section A: Performance comparison with STGNNs architectures on fMRI datasets												
Method	ABIDE I				ABDIE II				ADHD-200			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
STGCN	68.2	71.5	64.1	72.8	68.4	70.3	66.1	74.1	72.6	77.5	65.1	78.6
STAGIN	69.5	74.6	64.1	75.6	68.8	71.3	64.9	75.1	73.2	75.0	66.0	77.2
BrainGNN	68.7	69.8	65.8	73.4	67.6	70.7	64.1	73.7	72.9	78.9	61.3	77.8
STIGR(ours)	**72.7**†	**76.1**	**68.8**†	**79.0**†	**72.5**†	**73.0**	**71.6**†	**77.6**†	**76.2**†	**79.5**	**72.4**†	**81.8**†

Section B: Performance comparison with STGNNs architectures on EEG & fNIRS-BCI datasets												
Method	BCICIV_2a						BCIC2015			fNIRS-BCI		
	L/R	L/F	L/T	R/F	R/T	T/F	ACC	SEN	SPE	ACC	SEN	SPE
STGCN	78.6	80.9	82.4	79.4	82.4	79.5	70.9	54.9	79.8	94.3	94.0	95.1
STAGIN	79.4	81.7	83.6	80.6	83.5	79.6	72.1	51.9	80.9	94.7	93.9	95.8
BrainGNN	81.8	83.6	86.2	82.4	83.8	80.9	70.1	55.2	77.5	95.0	94.7	97.0
STIGR(ours)	**86.4**†	**91.4**†	**90.6**	**90.2**†	**90.0**†	**86.0**†	**75.5**†	**57.8**	**86.8**†	**97.5**†	**96.1**	**97.6**

Based on the Wilcoxon rank-sum test with Holm p-value correction (α = 0.05), the † indicates the marked method is significantly better than the compared methods.

CV scheme. For the fNIRS data, as we see in the table III, previous studies have achieved an average accuracy of 92.52% to 95.30% over eight subjects on the fNIRS-BCI dataset. Our STIGR also achieved the highest average accuracy, sensitivity, and specificity of 97.51%, 97.63%, and 96.13%, respectively. Furthermore, compared with the study [48] that also applies Transformer in its classification framework, our STIGR exhibits a significant accuracy improvement of 4.54%. Generally speaking, the proposed model achieves competitive classification performance while maintaining strong robustness and versatility generalization on different types of neuroimaging datasets (including fMRI, EEG, and fNIRS).

Apart from the comparison with SOTA methods, we attempt to evaluate the effectiveness of STIGR in comparison with other STGNN architectures. To the best of our knowledge, no STGNN-based methods have been proposed on these neuroimaging datasets. Therefore, several representative STGNN architectures proposed in neuroimaging domain are applied for comparison, including STGCN [1], STAGIN [4]), and BrainGNN [50]. The comparison results are summarized in Table IV. As we can see, the overall performance of STIGR consistently maintains its lead among the competitors by dominating 100% (12/12) of evaluation metrics, where nine of them show statistical significance by the Wilcoxon rank-sum test at $p \leq 0.05$ on fMRI datasets. On the EEG and fNIRS datasets, our STIGR exhibits strong generalization, which also dominates 100% (12/12) of the evaluation metrics and achieves statistical significance on eight metrics. Since STGCN and STAGIN were originally proposed for fMRI data, their performance on EEG datasets, especially for BICCIV_2a, has a large gap with our proposed method. As for the fNIRS dataset, all three SOTA STGNNs achieve comparable accuracy of around 95% but are also 2.5% $\sim$ 3.2% less accurate than our proposed STIGR. These findings further demonstrate the robustness and generalizability of our STIGR on different types of neuroimaging datasets. Overall, we can conclude that our STIGR achieves the highest accuracy on all datasets with an appropriate trade-off between specificity and sensitivity compared to STGNNs SOTA methods.

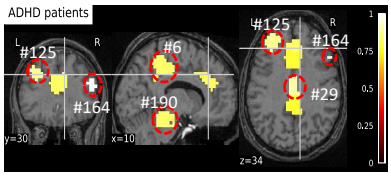
TABLE V
ABLATION STUDY OF OUR STIGR MODEL.

Method	Component			ADHD-200		BCICIV_2a	
	DAN-GCN	ST-DAN	CL-AM	Accuracy	Δ	Accuracy	Δ
(a)	–	–	–	68.4	–	80.5	–
(b)	✓	–	–	73.2	+4.8	85.2	+4.6
(c)	✓	✓	–	74.9	+6.5	87.2	+6.7
(d)	✓	✓	✓	**76.2**	**+7.8**	**89.3**	**+8.8**

Δ refers to the relative performance improvement over the baseline model. ✓ represents the given function participates in the training process.

D. Ablation Study

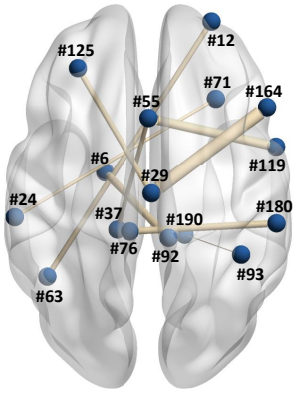
Since the proposed modules (i.e., DAN-GCN, ST-DAN, and CL-AM) account for the performance improvement of STIGR, we conduct ablation studies to evaluate their effectiveness contributions. Experiments are carried out on ADHD-200 and BCICIV_2a datasets and evaluated on 10-fold CV and IS schemes, respectively. Accordingly, the experimental results are tabulated in Table V. In particular, we choose STGCN [1] (case (a)) as the baseline model because it is the first spatio-temporal graph convolutional network proposed in the field of neuroimaging, has a relatively simple architecture, and can be flexibly extended. Model (b) replaced the STGCN in baseline with our proposed DAN-GCN, and model (c) is designed by introducing the ST-DAN module based on model (b). The proposed STIGR with all components is termed as (d) in Table V. To our surprise, our model with only DAN-GCN component (case (b)) improves the baseline by 4.8% and 4.6% mean accuracy on ADHD-200 and BCICIV_2a datasets, respectively. This result fully demonstrates the effectiveness of DAN-GCN in capturing hybrid spatio-temporal relations as well as the importance of both spatial and temporal feature extraction in dynamic neuroimaging data. That is to say, DAN-GCN can serve as a potent backbone model to extract the dynamic FC correlation within adjacent scanning sessions. In



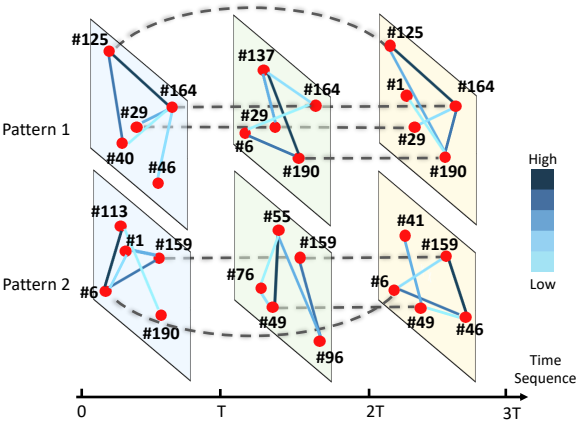
Five top predicted ADHD-specific ROIs

ID	Region	PMID	Score
164	Right Inferior Frontal Gyrus	30513395	1
29	Cingulate Gyrus	20008822	0.909
125	Left Middle Frontal Gyrus	23800705	0.891
190	Right Cerebelum	22922163	0.846
6	Right Cingulate Gyrus	27143897	0.844

(a) Remarkable ROIs in Node-level Representation



(b) Remarkable FC in Edge-level Representation



(c) Significant activities patterns in dynamic brain graph

Fig. 8. (a) and (b) are the visualization of discriminative ROIs and FCs obtained from node-level and edeg-level representation, respectively. (c) Identification of remarkable ROIs and dynamic FC patterns specific to ADHD. The dash lines indicate the involved ROIs are repeatedly highlighted throughout the whole time sequence. The darker color of the link between ROIs we observe, the more highly correlated the FC it represents. Here $T = 30$ timestamps.

addition, by introducing the ST-DAN module on top of model (a), we observe a significant improvement in the accuracy of ADHD-200 and BCICIV_2a by 1.7% and 2.1% in model (b). The performance improvement can be attributed to our ST-DAN module supplementing the missing global spatio-temporal representation of DAN-GCN through the globally self-attention mechanism. Thus, our model is developed to obtain the dynamic graph representation through DAN-GCN in the short-term (adjacent graph) while applying ST-DAN to extract the spatio-temporal dependency in the long-term (global attention). Since the CL-AM module is proposed based on DAN-GCN and ST-DAN, we cannot evaluate its effectiveness by eliminating the other two components. We evaluate the effectiveness of CL-AM module by directly comparing the performance of model (c) and model (d), which achieve 1.3% and 2.1% accuracy improvement on ADHD-200 and BCICIV_2a, respectively. This observation indicates the importance of adjacent matrix construction in graph convolution while demonstrating the great potential of contrastive learning in modeling representation similarity. The overall model (d) containing all three components (DAN-GCN, ST-DAN, and CL-AM) surpasses the baseline by 7.8% and 8.8% accuracy on two evaluated datasets. From the above analysis, we can conclude that all three components are demonstrated to make clear contributions to the performance improvement of our STIGR.

E. Exploration of Learned Graph Representation

In previous experiments, we validated the graph representing the ability of our STIGR to predict binary global properties of graph subjects. However, except for global property, each graph representation still contains node and edge information which are also important to neuroimaging analysis. As shown in Figure 2, an ideal graph representation learning framework can utilize the learned graph representation to address different graph tasks, including node-level, edge-level, and global-level. In this section, we explored the potential of our STIGR in learning node-level and edge-level representation for model interpretation. Additionally, experiments on regression prediction tasks are also performed to verify its global-level representation ability further.

1) Node-level and Edge-level: Since exploring task-related salient ROIs and functional connectivity is crucial in neuroimaging analysis, learning node-level and edge-level representations is necessary, especially for dynamic neuroimaging graph representation learning. Thanks to the graph neural network, the node-level and edge-level representation can be directly obtained by our STGIR (i.e., the final learned graph matrix R and adjacent matrix A). Since our neuroimaging data contain both node- and edge-level information, message passing algorithm (MPA) [51] is leveraged to exchange the information from neighboring nodes or edges for yielding graph representation (in node or edge-level) of increasing expressiveness and power. Therefore, our final node-leve and edge-level representation (R^N and A^E) can be obtained as $(R^N, A^E) = \text{MPA}(R, A)$. Accordingly, the final learned node-level and edge-level representations of STIGR can directly exhibit remarkable ROIs and reproducible FC patterns, revealing the unique task-related neurological activities. To validate the practical potential of our learned node-level and edge-level representations in model interpretation, we take the ADHD dataset as a case study in this section.

Figure 8(a) shows discriminative ADHD-related ROIs inferred by our learned node-level representation. Through manual investigation, all five top predicted ADHD-specific ROIs, i.e., *Right Inferior Frontal Cortex* (164), *Cingulate Gyrus* (29), *Left Middle Frontal Gyrus* (125), *Right Cerebelum* (190), and *Right Cingulate Gyrus* (6), have been validated to have associations with the neurological manifestations of ADHD by previous literature. For example, [52] found that the right inferior frontal cortex region (164), a key cognitive control hub region, is consistently dysfunctional in ADHD subjects. A significant volumetric decrease in the cingulate gyrus (29) was also founded among the treated group with ADHD [53]. In addition, we visualized top-10 FC connections with the highest scores based on learned edge-level representation. As shown

in Fig. 8(b), we can observe the strongest connectivity in the right inferior frontal cortex region (164) and the cingulate gyrus (29). This finding is consistent with a recent study [52] showing that these two regions exhibit increased positive functional connectivity in adolescents with ADHD. Based on Ahmadi's study [54], significant FC changes were found between the cingulate cortex (55) and superior temporal gyrus (119) in ADHD patients compared to TCs, which is also found in our learned edge-level representation. It is interesting to note that all five top ADHD-specific ROIs in the node-level representation were observed in the top 10 discriminative FC-connected regions in the edge-level representation, showing the great consistency between the two perspective representations. These observations suggest that node- and edge-level representations learned from our STIGR framework can be directly used for clinical interpretation analysis without additional visualization algorithms, demonstrating the feasibility of our learned graph representation in solving node-level and edge-level downstream tasks.

Furthermore, by simultaneously utilizing the learned node- and edge-level representation, significant activity patterns in dynamic neuroimaging graphs can be easily obtained. In this case study, we attempt to detect the reproducible dynamic FC patterns specific to ADHD only in light of these ADHD-related ROIs. That is, they have no significant change in BOLD signals among the TC group. We showcase the two eligible dynamic FC patterns in Figure 8(c). It is efficient to investigate how these dynamic interaction brain networks evolve along the time axis and identify the reproducible dynamic FC patterns as neuroimaging biomarkers. Through the statistical analysis, 66.9% of ADHD subjects (242/362) on ADHD-200 are found to exhibit similar dynamic FC patterns shown in Figure 8(c). This result suggests that the dynamic activity patterns identified by STIGR have great potential in advancing the understanding of brain activity mechanisms (e.g., analyzing the dynamic variation of neural stimulation between salient brain regions).

2) Global-level: In this section, to further validate the representing ability of learned graph representations at the global-level, we design experiments on the brain age prediction (regression task) based on the ABIDE I dataset. Brain age prediction is a common neuroimaging analysis problem, especially in the field of diagnosis of psychiatric disorders. It is essential to help identify novel biomarkers and develop a comprehensive system for early diagnosis of mental disorders [55]. An increasing number of neuroimaging studies in deep learning perform age prediction and classification tasks simultaneously to demonstrate the generalization ability of their models across different tasks [56], [57]. However, during the model training process, they need to completely separate the two tasks and train each task from scratch with some model architecture modifications. Given the strong representing ability of our STGIR model, we attempt to address this limitation by directly exploiting global-level graph representations learned from the previous classification task. Specifically, we use our STIGR trained on classification tasks as a feature extractor to extract high-quality representations (learned global-level graph representations). After that, these representations are put into a three-layer fully-connected network for training and finally perform age prediction. In this way, only a simple regression head needs to be trained instead of training the entire model in previous studies. As we can see from Fig. 9, compared with recent research of brain age prediction research on the ABIDE I dataset [57]–[61], our method achieved comparable performance with a mean absolute error (MAE) of 3.05($\pm$0.28) years. It is worth noting that this is a rough comparison since these methods employ various data volumes and validation methodologies. As we expected, our method outperforms the SOTA method [60] with an MAE decrease of 0.18 years. That is, by leveraging the learned graph representations, we save numerous training time for regression tasks while achieving SOTA performance. Combining the observations analyzed in the previous subsection, we can conclude that our STIGR is able to learn powerful graph representations for solving different graph tasks of nodes, edges, and global perspectives.

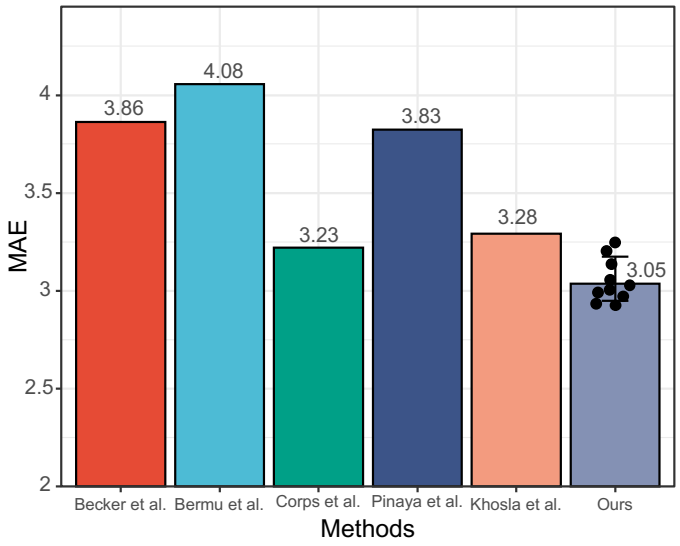


Fig. 9. Performance comparison of brain age prediction on the ABIDE I dataset.

V. Conclusion

In this paper, we propose an efficient graph representation learning framework, called STIGR, for capturing discriminative spatio-temporal graph representations of neuroimaging data. Our framework formulates the spatio-temporal representations with local and global patterns from two major components DAN-GCN and ST-DAN, respectively. Specifically, DAN-GCN is a novel graph convolution operation proposed based on the intuition of simultaneously exploring spatio-temporal dependencies by considering the temporal connection of dynamic graphs. In order to supplement the missing global spatio-temporal representation, ST-DAN is developed to learn the global spatio-temporal representation based on the self-attention mechanism and fuses it with DAN-GCN. As such, both local and global spatio-temporal dependencies are leveraged to improve the dynamic graph representation ability of our proposed model. In addition, we also designed CL-AM

module based on contrastive learning to adaptively learn the adjacent matrix of DAN-GCN for capturing the graph node connections in a dynamic manner. To evaluate the effectiveness and generalization ability of STIGR, comprehensive experiments are performed on three different kinds of neuroimaging datasets, including fMRI, EEG, and fNIRS. Evaluation results show that our STIGR achieves superior performance and outperforms the current state-of-the-art on all comparative datasets. Ablation studies are also conducted to verify the contribution of the proposed three key components. Thanks to the powerful graph representation capability of STIGR, the learned graph representation can be well utilized to explain brain activity from node and edge level (i.e., ROI analysis and connectivity analysis), respectively. The global-level graph representation learned from STIGR is also demonstrated to have great potential in different downstream tasks of classification and regression. In the future, we expect to combine our STIGR framework with multi-modality learning to explore the feasibility of learning multimodal graph representations.

REFERENCES

[1] S. Gadgil, Q. Zhao, A. Pfefferbaum, E. V. Sullivan, E. Adeli, and K. M. Pohl, "Spatio-temporal graph convolution for resting-state fmri analysis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 528–538.

[2] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, "Graph-based deep learning for medical diagnosis and analysis: past, present and future," *Sensors*, vol. 21, no. 14, p. 4758, 2021.

[3] B. Sanchez-Lengeling, E. Reif, A. Pearce, and A. B. Wiltschko, "A gentle introduction to graph neural networks," *Distill*, vol. 6, no. 9, p. e33, 2021.

[4] B.-H. Kim, J. C. Ye, and J.-J. Kim, "Learning dynamic graph representation of brain connectome with spatio-temporal attention," *Advances in Neural Information Processing Systems*, vol. 34, pp. 4314–4327, 2021.

[5] Y. Kong, S. Niu, H. Gao, Y. Yue, H. Shu, C. Xie, Z. Zhang, and Y. Yuan, "Multi-stage graph fusion networks for major depressive disorder diagnosis," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 1917–1928, 2022.

[6] Q. Yu, R. Wang, J. Liu, L. Hu, M. Chen, and Z. Liu, "Gnn-based depression recognition using spatio-temporal information: A fnirs study," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 10, pp. 4925–4935, 2022.

[7] T. Chen, Y. Guo, S. Hao, and R. Hong, "Exploring self-attention graph pooling with eeg-based topological structure and soft label for depression detection," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2106–2118, 2022.

[8] Y. Du, G. D. Pearlson, Q. Yu, H. He, D. Lin, J. Sui, L. Wu, and V. D. Calhoun, "Interaction among subsystems within default mode network diminished in schizophrenia patients: a dynamic connectivity approach," *Schizophrenia Research*, vol. 170, no. 1, pp. 55–65, 2016.

[9] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo *et al.*, "Dynamic functional connectivity: promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, 2013.

[10] S. Wein, A. Schüller, A. M. Tomé, W. M. Malloni, M. W. Greenlee, and E. W. Lang, "Forecasting brain activity based on models of spatiotemporal brain dynamics: A comparison of graph neural network architectures," *Network Neuroscience*, vol. 6, no. 3, pp. 665–701, 2022.

[11] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, 2020, pp. 914–921.

[12] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" in *ICML*, vol. 2, no. 3, 2021, p. 4.

[13] J. Cabral, M. L. Kringelbach, and G. Deco, "Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms," *NeuroImage*, vol. 160, pp. 84–96, 2017.

[14] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.

[15] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert, "Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease," *Medical Image Analysis*, vol. 48, pp. 117–130, 2018.

[16] L. Xiao, J. Wang, P. H. Kassani, Y. Zhang, Y. Bai, J. M. Stephen, T. W. Wilson, V. D. Calhoun, and Y.-P. Wang, "Multi-hypergraph learning-based brain functional connectivity analysis in fmri data," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1746–1758, 2019.

[17] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1290–1301, 2020.

[18] X. Ji, Y. Li, and P. Wen, "Jumping knowledge based spatial-temporal graph convolutional networks for automatic sleep stage classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1464–1472, 2022.

[19] K. Smitha, K. Akhil Raja, K. Arun, P. Rajesh, B. Thomas, T. Kapilamoorthy, and C. Kesavadas, "Resting state fmri: A review on methods in resting state connectivity analysis and resting state networks," *The Neuroradiology Journal*, vol. 30, no. 4, pp. 305–317, 2017.

[20] G. Marrelec and A. Giron, "Automated extraction of mutual independence patterns using bayesian comparison of partition models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2299–2313, 2020.

[21] Z.-A. Huang, Z. Zhu, C. H. Yau, and K. C. Tan, "Identifying autism spectrum disorder from resting-state fmri using deep belief network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 2847–2861, 2020.

[22] W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown, and S. Gao, "Probabilistic common spatial patterns for multichannel eeg analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 639–653, 2014.

[23] S. Zhang, Y. Zheng, D. Wang, L. Wang, J. Ma, J. Zhang, W. Xu, D. Li, and D. Zhang, "Application of a common spatial pattern-based algorithm for an fnirs-based motor imagery brain-computer interface," *Neuroscience Letters*, vol. 655, pp. 35–40, 2017.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[25] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on eeg using lstm recurrent neural network," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.

[26] N. C. Dvornek, P. Ventola, K. A. Pelphrey, and J. S. Duncan, "Identifying autism from resting-state fmri using long short-term memory networks," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2017, pp. 362–370.

[27] W. Gao, H. Zhu, K. Giovanello, and W. Lin, "Multivariate network-level approach to detect interactions between large-scale functional systems," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2010, pp. 298–305.

[28] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature Neuroscience*, vol. 20, no. 3, pp. 353–364, 2017.

[29] N. V. Hartvig, "A stochastic geometry model for functional magnetic resonance images," *Scandinavian Journal of Statistics*, vol. 29, no. 3, pp. 333–353, 2002.

[30] G. Derado, F. D. Bowman, and C. D. Kilts, "Modeling the spatial and temporal dependence in fmri data," *Biometrics*, vol. 66, no. 3, pp. 949–957, 2010.

[31] N. Jrad and M. Congedo, "Identification of spatial and temporal features of eeg," *Neurocomputing*, vol. 90, pp. 66–71, 2012.

[32] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3d convolutional neural network for eeg-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 10, pp. 2164–2177, 2019.

[33] Z. Mao, Y. Su, G. Xu, X. Wang, Y. Huang, W. Yue, L. Sun, and N. Xiong, "Spatio-temporal deep learning method for adhd fmri classification," *Information Sciences*, vol. 499, pp. 1–11, 2019.

[34] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8748–8763.

[35] H. Xu, X. Zhang, H. Li, L. Xie, W. Dai, H. Xiong, and Q. Tian, "Seed the views: Hierarchical semantic alignment for contrastive representation learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3753–3767, 2022.

[36] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1597–1607.

[37] R. P. Rao and D. H. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects," *Nature Neuroscience*, vol. 2, no. 1, pp. 79–87, 1999.

[38] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[39] C. Yan and Y. Zang, "Dparsf: a matlab toolbox for" pipeline" data analysis of resting-state fmri," *Frontiers in Systems Neuroscience*, p. 13, 2010.

[40] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced p300 speller performance," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 15–21, 2008.

[41] P. Khan, P. Ranjan, and S. Kumar, "Data heterogeneity mitigation in healthcare robotic systems leveraging the nelder–mead method," in *Artificial Intelligence for Future Generation Robotics*. Elsevier, 2021, pp. 71–82.

[42] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for eeg motor imagery classification," *IEEE Access*, vol. 7, pp. 18 940–18 950, 2019.

[43] J.-S. Bang, M.-H. Lee, S. Fazli, C. Guan, and S.-W. Lee, "Spatio-spectral feature representation for motor imagery classification using convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 3038–3049, 2022.

[44] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in eeg-bci," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[45] K. Das and R. B. Pachori, "Electroencephalogram based motor imagery brain computer interface using multivariate iterative filtering and spatial filtering," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 3, pp. 1408–1418, 2023.

[46] P. C. Petrantonakis and I. Kompatsiaris, "Single-trial nirs data classification for brain–computer interfaces using graph signal processing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 9, pp. 1700–1709, 2018.

[47] J. Lu, H. Yan, C. Chang, and N. Wang, "Comparison of machine learning and deep learning approaches for decoding brain computer interface: an fnirs study," in *International Conference on Intelligent Information Processing*. Springer, 2020, pp. 192–201.

[48] Z. Wang, J. Zhang, X. Zhang, P. Chen, and B. Wang, "Transformer model for functional near-infrared spectroscopy classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2559–2569, 2022.

[49] J. Han, J. Lu, J. Lin, S. Zhang, and N. Yu, "A functional region decomposition method to enhance fnirs classification of mental states," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5674–5683, 2022.

[50] U. Mahmood, Z. Fu, V. Calhoun, and S. Plis, "Attend to connect: end-to-end brain functional connectivity estimation," in *ICLR 2021 Workshop on Geometrical and Topological Representation Learning*, 2021.

[51] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1263–1272.

[52] K. Rubia, M. Criaud, M. Wulff, A. Alegria, H. Brinson, G. Barker, D. Stahl, and V. Giampietro, "Functional connectivity changes associated with fmri neurofeedback of right inferior frontal cortex in adolescents with adhd," *NeuroImage*, vol. 188, pp. 43–58, 2019.

[53] N. Makris, L. J. Seidman, E. M. Valera, J. Biederman, M. C. Monuteaux, D. N. Kennedy, V. S. Caviness Jr, G. Bush, K. Crum, A. B. Brown *et al.*, "Anterior cingulate volumetric alterations in treatment-naive adults with adhd: a pilot study," *Journal of Attention Disorders*, vol. 13, no. 4, pp. 407–413, 2010.

[54] M. Ahmadi, K. Kazemi, K. Kuc, A. Cybulska-Klosowicz, M. S. Helfroush, and A. Aarabi, "Resting state dynamic functional connectivity in children with attention deficit/hyperactivity disorder," *Journal of Neural Engineering*, vol. 18, no. 4, p. 0460d1, 2021.

[55] B. A. Jónsson, G. Bjornsdottir, T. Thorgeirsson, L. M. Ellingsen, G. B. Walters, D. Gudbjartsson, H. Stefansson, K. Stefansson, and M. Ulfarsson, "Brain age prediction using deep learning uncovers associated sequence variants," *Nature Communications*, vol. 10, no. 1, pp. 1–10, 2019.

[56] C. Wachinger, A. Rieckmann, S. Pölsterl, A. D. N. Initiative *et al.*, "Detect and correct bias in multi-site neuroimaging datasets," *Medical Image Analysis*, vol. 67, p. 101879, 2021.

[57] M. Khosla, K. Jamison, A. Kuceyeski, and M. R. Sabuncu, "Ensemble learning with 3d convolutional neural networks for functional connectome-based prediction," *NeuroImage*, vol. 199, pp. 651–662, 2019.

[58] B. G. Becker, T. Klein, C. Wachinger, A. D. N. Initiative *et al.*, "Gaussian process uncertainty in age estimation as a measure of brain abnormality," *NeuroImage*, vol. 175, pp. 246–258, 2018.

[59] C. Bermudez, A. J. Plassard, S. Chaganti, Y. Huo, K. S. Aboud, L. E. Cutting, S. M. Resnick, and B. A. Landman, "Anatomical context improves deep learning on the brain age estimation task," *Magnetic Resonance Imaging*, vol. 62, pp. 70–77, 2019.

[60] J. Corps and I. Rekik, "Morphological brain age prediction using multi-view brain networks derived from cortical morphology in healthy and disordered participants," *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019.

[61] W. H. Pinaya, A. Mechelli, and J. R. Sato, "Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study," *Human Brain Mapping*, vol. 40, no. 3, pp. 944–954, 2019.