IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 August 2024, accepted 21 September 2024, date of publication 26 September 2024,
date of current version 11 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3468439

RESEARCH ARTICLE

Hierarchical Hyperedge Graph Transformer: Toward Dynamic Interactions of Brain Networks for Neurodevelopmental Disease Diagnosis

JIUJIANG GUO[1], MONAN WANG[1], AND XIAOJING GUO[2]
[1]School of Mechanical and Power Engineering, Harbin University of Science and Technology, Harbin 150080, China
[2]School of Graduate, Heilongjiang University of Chinese Medicine, Harbin 150040, China

Corresponding author: Monan Wang (qqwmonan@163.com)

ABSTRACT The diagnosis of neurodevelopmental disorders has been a challenge due to the heterogeneity in traits and high variations in observations. Functional Magnetic Resonance Imaging (fMRI) and its Functional Connectivity (FC) analysis have become instrumental in studying these disorders by accessing underlying abnormal neural function and communications. Recently, graph neural networks (GNNs) have shifted the analysis of brain networks by capturing the structural information contained within Regions of Interest (ROIs) interactions. Nonetheless, most existing studies are limited by oversimplifying complex brain connections and the nuanced relationships among brain regions. Additionally, the underlying interactions at multiple scales and semantics are often ignored. In this regard, we propose the Hierarchical Hyperedge Graph Transformer network (HH-GraphFormer), which builds hierarchical interactions of the brain and generates hyperedge graphs to detail relationships at various scales. Based on hyperedges formed from the doubly stochastic matrices of self-attention maps of the Transformer, multiple sub-graphs are constructed and used to depict diverse inter-ROI dependencies. Moreover, to aggregate the topological properties of multiple sub-graphs, we present a cluster-based aggregation module to gather sub-graph embeddings. Experimental results on the ABIDE and ADHD-200 datasets demonstrated the superior performance of HH-GraphFormer in diagnosing neurodevelopmental disorders and interpreting complex interactions in the brain.

INDEX TERMS Brain networks, graph neural network, neurodevelopmental disorders, fMRI.

I. INTRODUCTION

Neurodevelopmental disorders encompass a diverse range of conditions characterized by abnormalities in the central nervous system, as highlighted in several key studies [1], [2], [3]. Notably, autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) are two prominent examples of these disorders, both of which are prevalent and significantly impact individuals' lives [4], [5]. Currently, the diagnosis of neurodevelopmental disorders, such as ASD, primarily relies on the observation of symptoms and the clinical experience of practitioners, which can easily lead to misdiagnosis [6], [7]. Therefore, it is crucial to develop accurate, automated diagnostic methods for these conditions.

Functional Magnetic Resonance Imaging (fMRI) has gained significant attractions in neuroimaging over the past decades. Leveraging the blood-oxygen level-dependent (BOLD) effect, fMRI has become an invaluable tool for probing brain dysfunctions. This imaging technique enables the decoding of functional activities and neural communications within the brain. Functional connectivity (FC) analysis, which quantifies the interactions among spatially distinct brain regions, is widely adopted in studying neurodevelopmental disorders [8], [9], [10]. This approach conceptualizes the brain as a network, which is crucial for understanding the intricate links between brain organization and behavioral phenotypes.

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhua Sheng.

In recent years, deep learning has significantly advanced brain network analysis, leading to remarkable improvements in diagnostic performance. Among these advancements, graph neural networks (GNNs) have emerged as state-of-the-art tools for analyzing graph-structured brain networks [3], [11], [12], [13], [14]. GNNs are particularly adept at capturing the structural information embedded in feature interactions within the graph domain, generating substantial interest in neuroscience. For instance, BrainGNN introduced ROI-aware graph convolutional layers and ROI-selection pooling layers for neurological biomarker prediction at both group and individual levels [15]. Additionally, Ktena et al. proposed learning a graph similarity metric using a siamese graph convolutional neural network [16]. The InceptionGCN was developed to capture both intra- and inter-graph structural heterogeneity during convolutions [17]. Recent studies have also proposed a similarity-aware adaptive calibrated GCN that considers disease status [13], [14].

Despite significant advancements, graph neural network (GNN) models still face several challenges in accurately distinguishing neurodevelopmental disorders. One major issue is the ambiguity of inter-ROI (Region of Interest) relationships in the graph-like brain networks [18], [19]. Most studies use functional connectivity as the adjacency matrix, which is obtained by measuring the linear correlation between brain regions. However, these matrices fail to capture the non-linear and complex inter-region dependencies in the brain. In GNNs, adjacency matrices play crucial roles in message passing on the graph, determining how information is aggregated, which can significantly impact the model's performance [15], [20]. Inaccuracies in these matrices can introduce unwanted noise and degrade performance. This lack of understanding the interconnectedness of different brain regions poses a substantial challenge in effectively modeling and interpreting neural communications. Additionally, there remain complicated interactions in the brain that are often overlooked. As shown in Figure 1, there are distinct ROIs at various scales, corresponding to different brain regions and subnetworks. In a cognitive task, multiple types of brain interactions play crucial roles in functional connectivity, representing various levels of neural communication. However, most existing studies tend to overlook interactions across multiple scales in the brain, such as inter-subnetwork dependencies.

To address the aforementioned challenges, we introduce a Hierarchical Hyperedge Graph Transformer (HH-GraphFormer) that dynamically and hierarchically models the complex inter-ROI (i.e., inter-region and inter-subnetwork) relationships in the brain at different scales. Our approach involves a hierarchical encoder to map brain region and subnetwork feature embeddings, which allows for the characterization of multi-scale and intricate neural communications in the brain network. The framework leverages the self-attention mechanism of the Transformer model to construct dynamic hyperedge graphs, resulting in more informative and representative graphs. Specifically, we construct

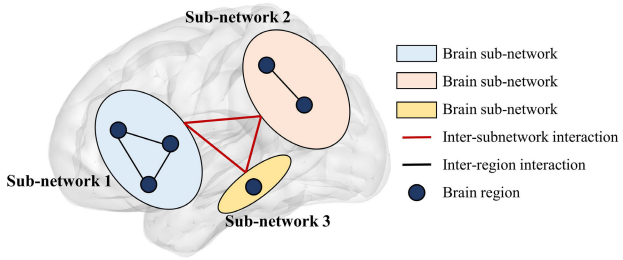

FIGURE 1. Illustration of hierarchical structures in the brain. Each subnetwork includes multiple brain regions. There remain various types of interactions in the brain, including the inter-region (black lines) and inter-subnetwork (red lines) interactions. For a cognitive task, both these two hierarchical types of dependencies play key roles in functional interactions, representing different levels of neural communications.

the hypergraph by using the output features of the Transformer as nodes and the doubly stochastic matrices, which are normalized from the self-attention maps, as hyperedges. This approach enables a nuanced depiction of various inter-ROI relationships.

Additionally, to address the potential similarity in inter-ROI dependencies yielding comparable subgraph embeddings, we introduce a cluster-based aggregation module (CAM) to integrate subgraph embeddings and their topological characteristics. We leverage hyperedge graph convolution to obtain multiple latent graph embeddings and map these graph embeddings into several clusters, with each cluster assigned a learnable weight. Finally, the graph embeddings are combined through weighted summation to avoid over-smoothing aggregation.

Through extensive evaluations on the ABIDE and ADHD-200 datasets, HH-GraphFormer has demonstrated its effectiveness and superiority in diagnosing neurodevelopmental disorders (i.e., ASD and ADHD). Additionally, statistical analysis was performed to provide biological explainability for neurodevelopmental biomarkers. In summary, our contributions are as follows:

- We propose the HH-GraphFormer model for mapping dynamic interactions within brain networks, producing hyperedge graphs that provide a nuanced depiction of inter-region and inter-subnetwork dependencies.
- We introduce a cluster-based aggregation module (CAM) to attentively integrate subgraph embeddings and aggregate topological properties of various subgraphs.
- Extensive evaluations on two real-world datasets demonstrate the superiority of our method in both diagnosing neurodevelopmental disorders and explaining the interactions within the brain.

In the following sections, we review related studies in terms of various methods for brain network studies in Section II, and then introduce the detail of the proposed HH-GraphFormer in Section III. Section IV showcases the experimental settings and the implementation details. The results are showed and discussed in Section V.

II. RELATED WORKS

Understanding the functional neural communication within the brain helps uncover the relationships between brain networks and disorders. Over the past decade, significant progress has been made using neuroimaging techniques to pinpoint brain network alterations associated with various disorders. Recent connectome-wide association studies (CWAS) have effectively identified functional dysconnectivity in neurodegenerative and psychiatric conditions. Using the key biomarkers identified through feature engineering, machine learning models such as Support Vector Machines (SVM) have been employed for classification. More recently, deep learning approaches have been widely adopted in CWAS studies, enabling the modeling of high-order intrinsic brain network embeddings and nonlinear inter-ROI dependencies. These works can be categorized into convolutional neural networks (CNNs), graph neural networks (GNNs), and Transformer networks.

A. CONVOLUTIONAL NEURAL NETWORKS FOR CWAS

BrainNetCNN [21], the first convolutional neural network specifically designed for brain network analysis, leverages the powerful nonlinear processing capabilities of convolutional neural networks to analyze complex brain network data as grid-like structures. Further research has expanded the BrainNetCNN model structure [22], proposing a convolutional neural network based on static and dynamic weighting, which enhances the hierarchical feature learning of the brain network. Another study proposed a Deep Convolutional Autoencoder (DCAE) network [23], which assigns different weights to connections between brain regions. Wang et al. [24] proposed implementing a two-layer convolution on the fMRI and DTI data simultaneously to perform multimodal brain network feature mapping. Huang et al. [25] introduced a static-dynamic convolutional neural network to capture both the static topology of the brain network and the time-varying changes in functional connectivity patterns. Additionally, some kernel-based convolutional methods have been proposed to improve the capture of complex interaction information in functional connectivity data [26], [27], [28], [29].

B. GRAPH NEURAL NETWORKS FOR CWAS

In addition to CNN models, Graph Neural Networks (GNNs) effectively capture information embedded in graph domain features and their topological structures, highlighting the correlations between structural neighborhoods [30], [31], [32]. GNNs have gained significant attention and application in the field of neuroscience. Recently, the applications of GNNs in neuroscience can be divided into two categories: transductive learning methods, such as population graphs, and inductive learning methods.

1) TRANSDUCTIVE LEARNING METHODS

These studies model each individual as a node in the graph and leverage semi-supervised learning and node classification for disease diagnosis. Parisot et al. [33] first propose to use Graph Convolutional Network (GCN) models to analyze connectome data, introducing the Population-GCN model. This model explores brain network feature similarities and performs aggregation to capture and analyze the complex interactions and connectivity patterns among individuals. Building on the Population-GCN model, the InceptionGCN method [17] is proposed to address structural heterogeneity in the data by incorporating functions with different kernel sizes and various feature selection strategies. Song et al. [14] propose an adaptive correction method to mitigate performance degradation caused by inaccuracies in the adjacency matrix. Additionally, Jiang and Li et al. [34], [35] enhance the characterization of graph topology relationships by employing different kernel functions and sparsely layered representations of connectome connectivity information. However, transductive models like population graphs are built on a specific graph of all participants and have limited predictive capabilities for new graph structures. This makes them poor adaptability to large-scale graph data or new nodes.

2) INDUCTIVE LEARNING METHODS

Inductive GNN models represent each brain network as an independent graph, with each node representing a brain region. The graph classification task is performed for brain disease diagnosis. BrainGNN [15] introduces ROI-aware graph convolutional layers and ROI selection pooling layers to effectively identify disease-related biomarkers at both group and individual levels. Additionally, Ktena et al. [16] employ a siamese graph convolutional neural network to learn graph similarity metrics, assessing the similarity between pairs of graphs and extending traditional convolution to irregular graphs through spectral graph convolution. However, such studies consistently face challenges in graph construction, particularly for the complex brain, where accurately measuring and estimating the intricate relationships between different brain regions remains a significant hurdle. To overcome these limitations, a series of studies have proposed strategies using dynamic graphs, such as Dynamic Graph Neural Networks (DGNN) and Dynamic Hyperedge Graph Neural Networks (DHGNN) [19], [36]. These methods employ dynamic strategies to generate new adjacency matrices through data fitting, effectively reducing the impact of inaccurate adjacency matrices on node representations. This approach has led to significant performance improvements in disease prediction tasks [19], [37]. In this study, inspired by the promising dynamic mechanism, we implement the dynamic graph for the attentive mapping of inter-ROI dependencies.

C. TRANSFORMER NETWORKS FOR CWAS

Recently, there has been growing interest in the Transformer architecture due to its remarkable abilities in learning graph representations [38], [39]. BrainNetTransformer [40] incorporates distinctive cluster-aware embeddings to discern

similar behavioral patterns, achieving superior performance compared to many established studies. Cai et al. [41] propose a graph transformer geometric learning framework to model the multimodal brain network by learning cross-modal interaction and fusion. Qu et al. [42] introduce a gated graph transformer model that attentively maps multi-view node feature embeddings and dynamically distributes propagation weights. Zhao et al. [43] propose a spatio-temporal graph Transformer to integrate multimodal brain networks in both the spatial and temporal domains. Zhu et al. [44] introduce a Transformer for mapping cortical structural brain networks to capture long-distance dependencies on graphs. However, most of these studies leverage Transformers for feature embeddings and fail to interpret non-linear and complex inter-ROI dependencies, which are crucial for understanding the associations between brain network organization and brain disorders. In this regard, in this study, we aim to explore the potential of Transformer for relating attentive inter-ROI dependencies and their application to graph-like brain networks.

III. METHOD

As illustrated in Fig. 2, our proposed HH-GraphFormer is constructed with a hierarchical feature parsing module, which includes A) region feature embedding and B) sub-network feature embedding. Based on the embedded features and attention maps from Multi-Head Self-Attention (MHSA), we build C) hyperedge graphs to represent the interactions of the brain network. Hyperedge Graph Convolution Networks (HGCN) are implemented for message passing and aggregation on multiple sub-graphs. The embeddings obtained from multiple graphs are then fed into the Cluster-based Aggregation Module (CAM) for attentive combination. Finally, a Multi-Layer Perceptron (MLP) is used for classification.

A. PROBLEM DEFINITION

In this study, the brain networks are derived by transforming pre-processed fMRI data into a template with M Regions of Interest (ROIs) and then calculating the Pearson Correlation Coefficient between them. These processed brain networks are represented as matrices $X \in \mathbb{R}^{M\times M}$. Each row $X_{i,:}$ or column $X_{:,j}$ of the matrix signifies the connectivity profile of each respective ROI, i or j. Our objective is to develop a learning model that effectively maps this brain network data, X, to predict the associated disease phenotype, y, formalized as a function $f : X \to y$.

To elucidate the ambiguous dependencies between different ROIs, we treat the brain networks X as the connection profile and employ a dynamic graph mechanism that attentively models these inter-ROI relationships. This is achieved by learning multiple mappings denoted as $f_V : X \to E$. In our study, we acquire this mapping through the self-attention mechanism by the Transformer model. This method allows for a more nuanced and adaptive understanding of inter-ROI relationships, dynamically reflecting the complex and variable nature of brain connectivity.

B. HIERARCHICAL ENCODER

In this study, we propose a hierarchical encoder for brain network embedding, consisting of a region feature embedding module and a sub-network feature embedding module. These modules encode the brain network at different scales, capturing both regional and sub-network features.

1) REGION FEATURE EMBEDDING

The region feature embedding module consists of a region convolution layer, followed by a layer normalization layer and a multi-head self-attention layer. The outputs are concatenated using a residual connection. Specifically, the region embedding layer is implemented with a two-layer MLP, which embeds the input $X \in \mathbb{R}^{M\times M}$ into $X_{\text{reg}} \in \mathbb{R}^{M\times D}$ with D hidden features. The output $O_{\text{reg}} \in \mathbb{R}^{M\times D}$ of the region feature embedding module has the same dimensions with X_{reg}.

2) SUB-NETWORK FEATURE EMBEDDING

This module maps the sub-network dependencies within the brain network. First, the brain regions are segmented into several ROIs, and the brain networks are divided into $\{X_1 \in \mathbb{R}^{M_1\times M_1}, X_2 \in \mathbb{R}^{M_2\times M_2}, \ldots, X_R \in \mathbb{R}^{M_R\times M_R}\}$, where $M_1 + M_2 + \ldots + M_R = M$. Each element is then flattened into a vector and embedded into an output with D features via a two-layer MLP. This process yields an output feature vector $\{O_1 \in \mathbb{R}^{1\times D}, O_2 \in \mathbb{R}^{1\times D}, \ldots, O_R \in \mathbb{R}^{1\times D}\}$. Finally, these features are concatenated into the embedded output feature $O_{\text{sub}} \in \mathbb{R}^{R\times D}$. In this study, we use the Yeo atlas [45] to segment the brain network into several sub-networks: the default mode network (DMN), the visual network (VN), the salience/ventral attention network (SAN), the dorsal attention network (DAN), the control network (CN), the somatomotor network (SMN), and the limbic network (LN). Accordingly, the brain networks are divided into $R = 7$ sub-networks.

3) MULTI-HEAD SELF-ATTENTION

In this study, we leverage Multi-Head Self-Attention (MHSA) to capture inter-region and inter-subnetwork dependencies, generating more expressive brain features:

$$\hat{X}^l_{\text{reg}} = \text{MHSA}(\text{LayerNorm}(X_{\text{reg}})) \in \mathbb{R}^{M\times D} \quad (1)$$

$$\hat{X}^l_{\text{sub}} = \text{MHSA}(\text{LayerNorm}(X_{\text{sub}})) \in \mathbb{R}^{R\times D} \quad (2)$$

For each layer l in MHSA, we first calculate the query $Q^{l,c}$, key $K^{l,c}$, and value $V^{l,c}$ for the c-th head through linear projection as:

$$Q^{l,c} = H^{l-1}W^{l,c}_q \quad (3)$$

$$K^{l,c} = H^{l-1}W^{l,c}_k \quad (4)$$

$$V^{l,c} = H^{l-1}W^{l,c}_v \quad (5)$$

where H^{l-1} is the output of the l-th layer, $H^0 = X_{\text{reg}}$, or X_{sub}, and $W^{l,c}_q, W^{l,c}_k, W^{l,c}_v$ are learnable parameters. c is in the range of $\{1, 2, \ldots, C\}$, and C denotes the number of self-attention heads. The output for each head $A^{l,c}$ is computed

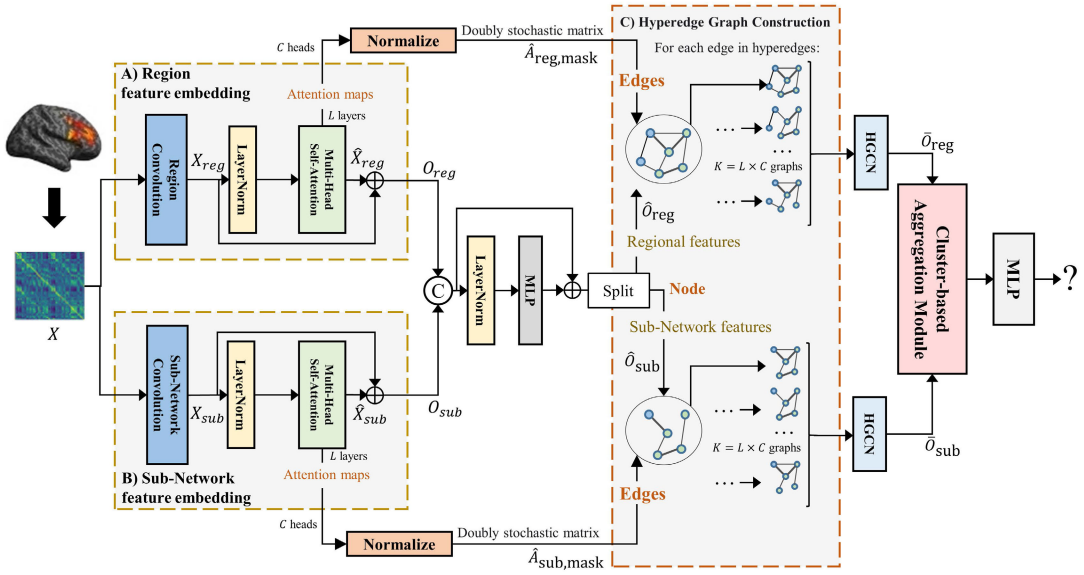

FIGURE 2. Illustration of the proposed HH-GraphFormer method. The hierarchical encoder includes A) region and B) sub-network feature embedding layers, which enable the construction of multi-scale interactions. Based on these attentive interactions, we build hyperedge graphs in C). Finally, the cluster-based aggregation module integrates the multiple graph embeddings for classification.

as a scaled dot-product as:

$$A^{l,c} = \text{Softmax}(\frac{Q^{l,c}(K^{l,c})^T}{\sqrt{d}}) \tag{6}$$

where d is the first dimension of $W^{l,c}$. To avoid self-ROI relations for brain networks and give more attention on inter-ROI relations, we formulate the diagonal forces 0 on diagonal components of $A^{l,c}$. The proposed diagonal masking is defined by

$$A^{l,c}_{\text{mask}}(i,j) = \begin{cases} A^{l,c}(i,j), i \neq j \\ 0, i = j \end{cases} \tag{7}$$

where $A_{i,j}$ indicates each component of the masked similarity matrix. The encoded features of MHSA of two branches are further obtained by corresponding adjacency matrices $A^{l,c}_{\text{reg,mask}}$ and $A^{l,c}_{\text{sub,mask}}$:

$$\hat{X}_{\text{reg}} = H^{L,C}_{\text{reg}} = A^{L,C}_{\text{reg,mask}} V^{L,C}_{\text{reg}} \tag{8}$$

$$\hat{X}_{\text{sub}} = H^{L,C}_{\text{sub}} = A^{L,C}_{\text{sub,mask}} V^{L,C}_{\text{sub}} \tag{9}$$

Finally, the output O_{reg} and O_{sub} is obtained by residual connections:

$$O_{\text{reg}} = H^{L,C}_{\text{reg}} \bar{W}_{\text{reg},O} + X_{\text{reg}} \tag{10}$$

$$O_{\text{sub}} = H^{L,C}_{\text{sub}} \bar{W}_{\text{sub},O} + X_{\text{sub}} \tag{11}$$

$\bar{W}_{\text{reg},O}$ and $\bar{W}_{\text{sub},O}$ are learnable model parameters for node feature embedding.

C. HYPEREDGE GRAPH CONSTRUCTION

The MHSA can generate multiple attention maps that capture inter-region and inter-subnetwork dependencies. These attention maps are then used to construct dynamic graphs representing the interactions within brain networks. Specifically, these graphs are dynamic, with their node features and adjacency matrices obtained through attention mechanisms.

1) EDGE

Based on the Transformer architecture with L layers and C heads, we generate $K = L \times C$ distinct attention maps, denoted as $A^{l,c}_{\text{mask}}$, where $l \in [1, L]$ and $c \in [1, C]$. To note that, the number of layers and attention heads determines the number of hyperedges. Since these two parameters are hyperparameters and difficult to determine, we use grid search to select their values. The parameter C is chosen from the set $[2, 4, 8, 16, 32]$, while L is searched within the range $[1, 5]$. Furthermore, these attention maps are then normalized into doubly stochastic matrices, represented as $\hat{A}^{l,c}\text{mask} \in \mathbb{R}^{M \times M}$. The normalization process ensures that each row and column contributes uniformly by adjusting the entries accordingly. We interpret these doubly stochastic matrices as adjacency matrices for hyperedge graphs, which model inter-ROI relationships. In the region and sub-network embedding modules, the corresponding adjacency matrices are denoted as $\hat{A}^{l,c}\text{reg,mask} \in \mathbb{R}^{M \times M}$ and $\hat{A}^{l,c}_{\text{sub,mask}} \in \mathbb{R}^{R \times R}$, respectively.

2) NODE

The encoded features O_{reg} and O_{sub} are concatenated and passed through a layer normalization layer followed by an FFN layer. The outputs are then split into region features $\hat{O}\text{reg} \in \mathbb{R}^{M \times D}$ and subnetwork features $\hat{O}\text{sub} \in \mathbb{R}^{R \times D}$.

D. HYPEREDGE GRAPH CONVOLUTION

After constructing various dynamic graphs, we use the graph convolution to aggregate inter-ROI messages by performing spatial hyperedge graph convolution. The spatial graph convolution is obtained for each graph as:

$$\bar{O}^{l,c}_{\text{reg}} = \sigma(A^{l,c}_{\text{reg}} \hat{O}_{\text{reg}} W^{l,c}_{\text{reg}}) \tag{12}$$

$$\bar{O}^{l,c}_{\text{sub}} = \sigma(A^{l,c}_{\text{sub}} \hat{O}_{\text{sub}} W^{l,c}_{\text{sub}}) \tag{13}$$

where $W^{l,c}_{reg}$ and $W^{l,c}_{sub}$ represent the parameters for the l-th layer and the c-th head in the inter-region and inter-subnetwork Transformer layer. These parameters are used to tune attention maps and map features from a D-dimensional space to a D'-dimensional space. To note that, in this study, we use the spatial graph convolution to reduce computational cost instead of spectral graph convolution, due to the constructed dynamic graphs. Compared to spectral graph convolution, spatial graph convolution offers a reduction in computational cost while maintaining similar performance levels.

Finally, considering a Transformer layer includes multiple heads and would generate multiple graphs, we leverage a readout operation to concatenate them into Z, where $Z = \|_{l=1}^{L}(\|_{c=1}^{C}\bar{O}^{l,c}) \in \mathbb{R}^{K\times M\times D'}$. $\|$ represents the concatenation operation and $K = L \times C$. This operation is repeated for each participant, capturing the nuanced, multi-dimensional relationships across the brain's network.

E. CLUSTER-BASED AGGREGATION MODULE

After performing spatial hyperedge graph convolution, the features are still in a high-dimensional space. The Transformer layers facilitate the generation of multiple edge types, leading to a diverse array of dynamic graphs that embody various topological characteristics and locality-specific details. In this regard, there is a risk that these sub-graph embeddings may exhibit similar properties. A straightforward combination might lead to over-smoothing aggregation.

To address this issue, we propose to attentively aggregate these embeddings, where sub-graph embeddings are assembled adaptively. The outputs of the hyperedge graph convolution from two branches are concatenated $\hat{Z} = \|\{Z_{\text{reg}}, Z_{\text{sub}}\} \in \mathbf{R}^{(M+R)\times D'}$ and further fed into the CAM. We first map these encoded representations from $(M+R)$ features into Q clusters using Softmax($\hat{Z}$). Each cluster is assigned a learnable weight w_c, $c \in [1, Q]$. Finally, the aggregated embedded features are obtained by:

$$\tilde{O} = \sum_{c=1}^{M+R} w_p \hat{Z}_c, p = argmax_c\{Softmax(\hat{Z}_c)\} \tag{14}$$

F. OPTIMIZATION

The final aggregated features obtained from CAM is further fed into a three-layer MLP classifier for classification. We leverage the cross-entropy loss function for optimization.

IV. EXPERIMENTAL SETTINGS

A. DATASET

Table 1 summarizes the demographic details of the two datasets. We utilized resting-state functional MR images (rsfMRI) from two large multi-center datasets:

1) ABIDE-I DATASET [46]

The ABIDE-I dataset consists of 3D-T1 weighted images as well as resting-state functional MRIs. The images are collected from 17 data centers, including 505 individuals diagnosed with ASD and 530 normal controls (NC). To exclude the effects of phenotypes like age and gender, we selected 502 ASD participants and 520 NCs, matching them by sex and age.

2) ADHD-200 DATASET [47]

The ADHD-200 dataset was released by the ADHD-200 global competition. We used the training set in our study, which includes 488 normal controls and 280 patients with ADHD. For sex and age matching, we selected 311 NCs and 261 ADHD patients to form our cohort.

TABLE 1. The demographic details on the ABIDE and ADHD-200 datasets.

	Group	ABIDE	ADHD-200
Gender (M/F)	NC	445/75	230/81
	ASD	443/59	209/52
	F-Statistics (p-value)	1.264 (0.206)	1.728 (0.085)
Age (Mean±Std)	NC	17.21 (7.93)	11.50 (2.37)
	ADHD	17.06 (8.34)	11.18 (2.42)
	F-Statistics (p-value)	0.303 (0.706)	-1.509 (0.112)

B. PRE-PROCESSING

1) THE ABIDE DATASET

The preprocessing of all images in the ABIDE dataset was carried out using the C-PAC pipeline. This process included skull stripping, slice timing correction, motion correction, global mean intensity normalization, regression of nuisance signals using 24 motion parameters, and band-pass filtering within the 0.01-0.08 Hz range. To enable comparisons across participants, the functional images were registered to the MNI152 standard anatomical space. The regression of nuisance variables was modeled with 24 motion parameters.

2) THE ADHD-200 DATASET

The Athena pipeline was used for preprocessing the images, incorporating AFNI and FSL neuroimaging tools. The preprocessing steps included removing the first four volumes, slice timing correction, motion correction through realignment, and linear transformation between the mean functional volume and the corresponding structural MRI. The functional images were transformed into MNI152 space using T1-weighted MRI for the MNI nonlinear warp. To remove noise and head drifts in the time series, nuisance regression

TABLE 2. Diagnosis performance (Mean-Std) on the ABIDE and ADHD-200 datasets in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC). The best results are highlighted in bold.

Dataset	ABIDE				ADHD-200			
Task	NC vs. ASD				NC vs. ADHD			
Metric	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SVM	65.6-5.8	70.0-7.9	61.0-11.3	67.2-7.9	54.0-7.0	56.1-14.6	59.7-15.6	49.6-15.6
MLP	68.4-5.3	64.0-18.2	72.2-14.7	69.3-4.2	60.1-2.9	62.3-4.5	59.9-6.1	61.2-3.3
BrainNetCNN [21]	68.1-2.0	67.6-7.0	69.7-4.4	72.0-3.1	62.8-3.8	67.9-12.8	60.4-13.6	64.0-5.1
DHGNN [36]	64.3-1.5	63.8-5.0	65.0-2.6	66.2-3.4	59.8-2.0	53.5-2.7	61.7-2.5	59.0-1.4
BrainGNN [15]	69.3-3.0	68.2-11.1	68.6-13.8	69.5-3.5	61.0-2.6	54.6-4.1	64.1-2.9	64.0-3.9
PopGCN [33]	69.1-3.2	67.1-4.3	71.2-3.2	72.6-3.2	61.0-3.6	56.7-2.2	66.4-3.0	55.8-7.4
HI-GCN [34]	69.3-5.1	**73.0-10.6**	70.3-4.6	72.0-6.2	63.2-4.4	63.1-11.9	66.2-9.0	61.0-4.4
Transformer	69.0-1.1	65.0-5.2	72.5-4.0	72.4-7.1	61.6-1.1	63.2-5.2	61.8-1.7	67.51-3.2
BrainNetTF [40]	70.5-2.6	72.3-4.7	70.2-3.8	71.2-3.8	63.4-2.9	63.5-2.5	62.9-2.3	65.8-4.8
HH-GraphFormer (Ours)	**71.7-4.9**	**73.0-6.0**	**73.6-6.2***	**73.2-5.4***	**65.9-2.6***	**68.1-4.6***	**67.1-8.3***	**68.1-2.5***

* indicates significant outperforming (p-value < 0.05) all the alternative methods.

models were applied. Finally, the denoised time series underwent band-pass filtering (0.009-0.08 Hz).

3) BRAIN NETWORK CONSTRUCTION

The mean time series for a set of regions defined by the Schaefer template [45] were obtained and normalized to have zero mean and unit variance. Functional connectivity was measured using Pearson's Correlation Coefficient. In our study, we utilized the Schaefer atlas with 100 brain regions for evaluation. Each region was associated with one of the seven network parcellations: the default mode network, the visual network, the somatomotor network, the dorsal attention network, the salience/ventral attention network, the limbic network, and the control network.

C. IMPLEMENTATION DETAILS

In this study, all experiments were conducted on the PyTorch 1.9 platform using an NVIDIA V100 with 32 GB of memory. We utilized the Adam optimizer for training, with the learning rate set to 1×10^{-4}. The settings for the number of heads and layers in the MHSA are discussed in the results section. The models were trained for 80 epochs, with an early stopping strategy applied to avoid overfitting. We used the Schaefer-100-7 atlas for brain network construction, which maps the brain into 100 regions and 7 subnetworks, resulting in $M = 100$ and $R = 7$. Each ROI belongs to a subnetwork and a region. For example, the brain region LH.PCC.DMN belongs to the default mode network subnetwork among the 7 subnetworks, and it is also a region among all 100 brain regions.

For a fair comparison, we trained and tested all models using 10-fold stratified cross-validation, a method adopted in most related studies. For each fold, 10% of samples were randomly selected for testing. Disease diagnosis accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC) were used for evaluation.

D. COMPETITIVE METHODS

In this study, to evaluate the superiority of our proposed HH-GraphFormer, we compared it with several baselines, including:

1) MACHINE LEARNING METHODS

We included conventional machine learning methods such as SVM and MLP. The brain networks were flattened into feature vectors and fed into the classifiers for classification. The number of MLP layers was determined within a search range of 1 to 4.

2) TRANSDUCTIVE METHODS

We compared our model with Population-GCN (PopGCN) [33] and HI-GCN [34]. These methods construct a population graph where each node represents the feature vector of a participant. The adjacency matrix is built based on the similarity of brain network features and phenotypic values such as site ID, sex, and age.

3) INDUCTIVE METHODS

We also compared BrainNetCNN [21], DHGNN [36], BrainGNN [15], Transformer, and BrainNetTransformer (BrainNetTF) [40]. These models were implemented based on their original architectures, with hyperparameters determined by a grid search.

V. RESULTS

A. DISEASE DIAGNOSIS PERFORMANCE

Table 2 shows the diagnostic performance on two datasets, with the average and standard deviation results displayed. The results demonstrate that deep learning approaches outperform traditional machine learning methods like SVM and MLP. Despite the limited sample size of brain networks, deep learning methods achieve better performance. Additionally, the Transformer model achieves comparable performance with graph-based methods, including both transductive and inductive methods. This indicates traditional Transformer models,

TABLE 3. Ablation studies on the interactions of the brain network including the inter-region and inter-subnetwork dependencies.

Dataset		ABIDE				ADHD-200			
Task		NC vs. ASD				NC vs. ADHD			
Inter-region	Inter-subnetwork	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
✓		70.6-5.1	72.8-6.3	70.7-6.9	71.1-4.6	64.3-2.8	66.1-5.1	62.8-7.1	66.6-4.0
	✓	69.9-3.9	72.5-5.1	69.7-5.5	70.6-4.5	63.5-2.5	66.3-3.3	61.7-4.8	65.4-3.5
✓	✓	**71.7-4.9**	**73.0-6.0**	**73.6-6.2***	**73.2-5.4***	**65.9-2.6**	**68.1-4.6***	**67.1-8.3***	**68.1-2.5***

* indicates significant outperforming (p-value < 0.05) all the alternative methods.

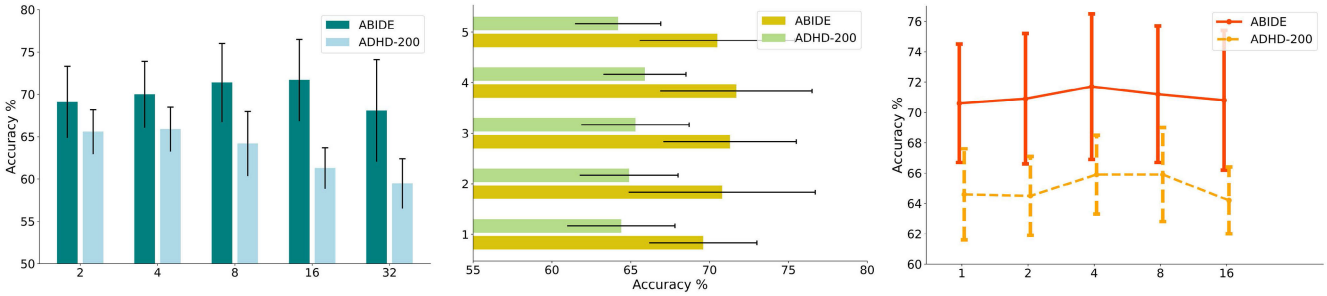

FIGURE 3. Sensitivity analysis on the settings of A) heads and B) depth of Transformer layers, and C) the cluster number of CAM.

as proposed in other domains, may not be fully suitable for brain network studies without specific modifications. In this regard, BrainNetTF outperforms both the graph-based models and the vanilla Transformer model. Finally, our proposed HH-GraphFormer achieves the best performance across both datasets (accuracy: 71.7% and 65.9%, sensitivity: 73.0% and 68.1%, specificity: 73.6% and 67.1%, AUC: 73.2% and 68.1%) for distinguishing ASD and ADHD from NC. This superior performance can be attributed to the improved dynamic inter-region and inter-subnetwork interactions enabled by the hierarchical encoder and hyperedge graphs.

B. ABLATION STUDIES

Moreover, to assess the effect of the hierarchical encoder, we conducted ablation studies on the region embedding module and the subnetwork embedding module. These two layers are referred to as inter-node and inter-subnetwork in Table 3. The results show that the inter-region embedding (region embedding module) consistently achieves better performance than the inter-subnetwork embedding (subnetwork embedding module) on both datasets. This can be attributed to the fact that brain region features contain more diverse and representative information than subnetwork features. Additionally, when both embedding modules are integrated, the hierarchical encoder facilitates the extraction of more representative features. The interactions between brain regions and subnetworks can produce disease-specific and informative representations. However, most existing studies fail to address inter-subnetwork dependencies. In this regard, our proposed HH-GraphFormer effectively relates hierarchical interactions in the brain, providing valuable insights into brain network studies.

C. SENSITIVITY ANALYSIS

Moreover, we evaluated the impact of model hyperparameter settings on performance across two datasets. The results, displayed in Figure 3, include: A) the number of attention heads, B) the number of Transformer layers, and C) the number of clusters in CAM.

1) THE EFFECT OF VARIOUS NUMBERS OF ATTENTION HEADS

As is shown in Figure 3 A), with the number of attention heads increases, performance on both datasets initially improves and then declines. The optimal settings are obtained by $C = 16$ for distinguishing ASD and $C = 4$ for distinguishing ADHD. This suggests that an increase in attention heads could contribute to more diverse and dynamic inter-ROI dependencies, thereby improving our understanding of brain interactions and enhancing performance. However, an excessive number of redundant attention heads may cause over-smoothing in feature combinations, which can negatively impact performance.

2) THE EFFECT OF VARIOUS NUMBERS OF TRANSFORMER LAYERS

Figure 3 B) displays the results on various Transformer layers, which is similar to the trend of various hyperedge graphs. For both datasets, the best results are achieved with $L = 4$ consistently. Unlike graph-based models, which are typically constructed with fewer than three layers,

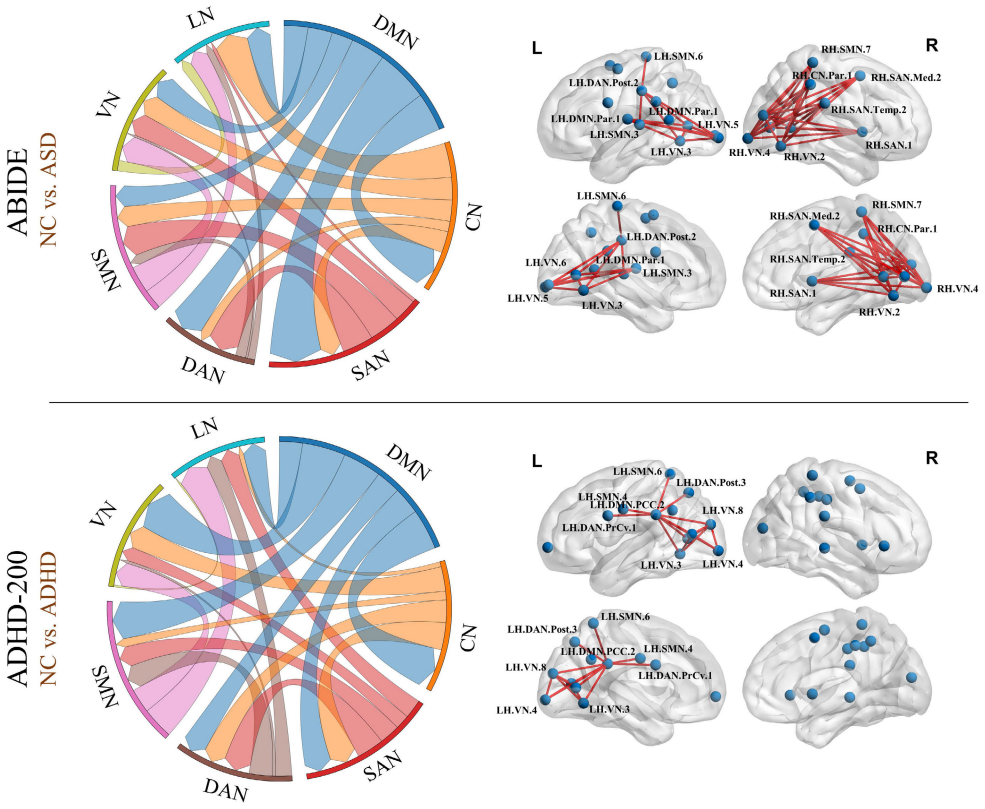

FIGURE 4. Biological explanation on the inter-subnetwork and inter-region interactions in the brain on two datasets by group-wise comparison. The inter-subnetwork dependencies are shown on the left, where the widths of lines indicate the interaction strength between subnetworks. The key inter-region connections (p-value < 0.05) after the correction are displayed. LH: the left hemisphere. RH: the right hemisphere. DMN: the default mode network. VN: the visual network. DAN: the dorsal attention network. Par: the parietal. PCC: the post cingulate cortex.

Transformer models can have more than three layers without performance degradation. Our proposed HH-GraphFormer is feasible to produce meaningful and disease-specific representations while avoiding the issue of over-smoothing in the graph neural networks [48], [49].

3) THE EFFECT OF VARIOUS NUMBERS OF CLUSTERS IN CAM

As shown in Figure 3 C), the performance for the ABIDE dataset is comparable across various cluster settings, indicating that the classification of ASD is not sensitive to the number of clusters. In contrast, the ADHD-200 dataset results are more sensitive to cluster settings, with the optimal range being between 4 and 8 clusters.

D. BIOLOGICAL EXPLANATION

In this study, our proposed HH-GraphFormer facilitates inter-region and inter-subnetwork dependencies through its hierarchical encoder. We extracted the learned interactions of the brain regions and subnetworks by averaging the corresponding attention maps of all Transformer layers and MHSA.

On the left of Figure 4, we display the interactions of subnetworks across two datasets, where the widths of the lines indicate the values of inter-subnetwork dependencies. In both datasets, the most important interactions involve the DMN and other subnetworks. For example, in the ABIDE dataset, the interaction between the DMN and SAN has the highest value, while in the ADHD-200 dataset, the interaction between the DMN and CN is the highest. This underscores the key role of the DMN in these two neurodevelopmental diseases. Additionally, we further showcase the key connections in the brain on the right of Figure 4. We implemented multivariate analysis to identify key connections with significant differences between groups (p-value < 0.05). These key connections mainly involve interactions between the DMN (ABIDE: LH.DMN.Par, ADHD-200: LH.DMN.PCC) and other subnetworks (including VN, SAN, and DAN), similar to the findings of the inter-subnetwork dependencies.

Previous studies have indicated that DMN dysfunction is related to specific components of social cognition, implicated in social-cognitive deficits in ASD and ADHD [50], [51]. Aberrancies in key nodes of the DMN and their dynamic functional interactions contribute to atypical integration of information about the self in relation to others, as well as

impairments in the ability to flexibly attend to socially relevant stimuli [52]. Our findings provide evidence to these studies and highlight the key role of the interactions between the DMN and other subnetworks.

VI. DISCUSSION

In this study, our proposed HH-GraphFormer facilitates the integration of multiple levels of interactions underlying brain activity. Built on a graph neural network and Transformer architecture, HH-GraphFormer effectively captures long-range inter-ROI dependencies while aggregating neighborhood information to identify topological properties. Experimental results on two large public neurodevelopmental datasets demonstrate that HH-GraphFormer outperforms other baselines in terms of accuracy, sensitivity, specificity, and AUC. Notably, significant improvements are observed in specificity and AUC across both datasets.

One of the key advantages of HH-GraphFormer is its ability to capture multi-scale interactions in the brain. Currently, most studies on brain networks leverage pre-defined brain templates to map fMRI data into over 100 brain regions using different atlases, such as AAL, Harvard-Oxford, and CC200. However, these studies often ignore the potential dependencies between sub-networks. As shown on the left side of Figure 4, DMN, a key sub-network in neurodevelopmental diseases related to attention/inhibition, working memory, and information processing, is difficult to detect when focusing only on inter-region interactions (depicted on the right of Figure 4). For example, in the ABIDE dataset, connections with the VN and DAN are identified as key biomarkers. While these biomarkers have been reported in previous studies, the more significant region, the DMN, is not as prominently detected as in prior research. From this perspective, our proposed HH-GraphFormer can associate key biomarkers at different scales and uncover key connections between the DMN and other sub-networks.

In this study, we have also found that the DMN emerges as a key biomarker in both ASD and ADHD. And the connection weights between the DMN and CN are significant in both conditions. Previous studies indicate that the DMN-CN inter-network is critical for the development of executive functions in children [53]. Dysfunctions between the DMN and CN may contribute to decreased performance in working memory, social cognition, and other cognitive functions [54]. Additionally, different inter-subnetwork connection patterns were observed between the two disorders. For instance, in the ABIDE dataset, the connection between the DMN and SAN is emphasized, whereas in the ADHD-200 dataset, the connection between the DMN and VN is more prominent. Studies have reported that DMN-SAN connectivity contributes to aberrant social processing in youth at high risk for psychosis [55]. The connection between the DMN and SAN is also associated with social deficits in ASD [56]. Moreover, ADHD shows significant reductions in the VN, which may be linked to the relatively frequent transient attention lapses in ADHD patients. Overall, our findings are consistent with these studies, and our HH-GraphFormer effectively identifies differences across various neurodevelopmental diseases at multiple scales, underscoring the importance of hierarchical encoding in brain networks.

Our study also has its limitations. The first is that our method is evaluated solely on neurodevelopmental diseases in this study, where the dynamic graph and hyperedges are more sensitive to these conditions due to the high variability in patients' connectome features. In this regard, one of our future directions is to assess the method's adaptability to other diseases such as Parkinson's Disease and Alzheimer's Disease. Additionally, the methods discussed in this study, including HH-GraphFormer, are still trained on a specific and single dataset, which limits their generalizability. We aim to explore pre-trained general models based on HH-GraphFormer to enhance its clinical applicability.

VII. CONCLUSION

In this study, we propose the HH-GraphFormer network for brain network analysis, which implements a hierarchical encoder for capturing multi-scale interactions among brain regions and subnetworks. The proposed HH-GraphFormer leverages dynamic graph modeling using the self-attention mechanism to exploit dynamic and adaptive interactions in the brain. Experimental results demonstrate the superiority of HH-GraphFormer in distinguishing neurodevelopmental diseases and highlight its importance in modeling these dynamic interactions. Additionally, our proposed HH-GraphFormer effectively identifies discriminative key interactions in the brain, providing potential for interpreting the complex dependencies among regions and subnetworks.

REFERENCES

[1] B. Candiri, A. N. Arıkan, and S. C. Çolak, "Balance and physical activity in children with neurodevelopmental disorders," in *Current Researches Health Sciences-III*. Russia: Özgür Yayın-Dağıtım Co. Ltd., 2023, p. 117.

[2] D. Bishop and M. Rutter, "Neurodevelopmental disorders: Conceptual issues," in *Rutter's Child Adolescent Psychiatry*. U.K.: Blackwell Publishing Ltd., 2008, pp. 32–41.

[3] Y. Yang, C. Ye, and T. Ma, "A deep connectome learning network using graph convolution for connectome-disease association study," *Neural Netw.*, vol. 164, pp. 91–104, Jul. 2023.

[4] A. Thapar, M. Cooper, and M. Rutter, "Neurodevelopmental disorders," *Lancet Psychiatry*, vol. 4, no. 4, pp. 339–346, 2017.

[5] C. F. Norbury and A. Sparks, "Difference or disorder? Cultural issues in understanding neurodevelopmental disorders," *Develop. Psychol.*, vol. 49, no. 1, pp. 45–58, 2013.

[6] D. J. Morris-Rosendahl and M.-A. Crocq, "Neurodevelopmental disorders—The history and future of a diagnostic concept," *Dialogues Clin. Neurosci.*, vol. 22, no. 1, pp. 65–72, 2020.

[7] S. E. Soden, C. J. Saunders, L. K. Willig, E. G. Farrow, L. D. Smith, J. E. Petrikin, J.-B. LePichon, N. A. Miller, I. Thiffault, and D. L. Dinwiddie, "Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders," *Sci. Transl. Med.*, vol. 6, no. 265, Dec. 2014, Art. no. 265ra168.

[8] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers Syst. Neurosci.*, vol. 9, p. 175, Jan. 2016.

[9] Y. Yang, C. Ye, J. Sun, L. Liang, H. Lv, L. Gao, J. Fang, T. Ma, and T. Wu, "Alteration of brain structural connectivity in progression of Parkinson's disease: A connectome-wide network analysis," *NeuroImage, Clin.*, vol. 31, Jan. 2021, Art. no. 102715.

[10] A. A. Fingelkurts, A. A. Fingelkurts, and S. Kähkönen, "Functional connectivity in the brain-is it an elusive concept?" *Neurosci. Biobehavioral Rev.*, vol. 28, no. 8, pp. 827–836, 2005.

[11] N. S. Dsouza, M. B. Nebel, D. Crocetti, J. Robinson, S. Mostofsky, and A. Venkataraman, "M-GCN: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations," in *Proc. 4th Conf. Med. Imag. With Deep Learn.*, 2021, pp. 119–130.

[12] Y. Yang, C. Ye, X. Guo, T. Wu, Y. Xiang, and T. Ma, "Mapping multimodal brain connectome for brain disorder diagnosis via cross-modal mutual learning," *IEEE Trans. Med. Imag.*, vol. 43, no. 1, pp. 108–121, Jan. 2024.

[13] X. Song, F. Zhou, A. F. Frangi, J. Cao, X. Xiao, Y. Lei, T. Wang, and B. Lei, "Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101947.

[14] X. Song, A. Frangi, X. Xiao, J. Cao, T. Wang, and B. Lei, "Integrating similarity awareness and adaptive calibration in graph convolution network to predict disease," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Lima, Peru. Cham, Switzerland: Springer, Oct. 2020, pp. 124–133.

[15] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, and J. S. Duncan, "BrainGNN: Interpretable brain graph neural network for fMRI analysis," *Med. Image Anal.*, vol. 74, Dec. 2021, Art. no. 102233.

[16] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, Apr. 2018.

[17] A. Kazi, "InceptionGCN: Receptive field aware graph convolutional network for disease prediction," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, Hong Kong. Cham, Switzerland: Springer, Jun. 2019, pp. 73–85.

[18] Y. Zhang and H. Huang, "New graph-blind convolutional network for brain connectome data analysis," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2019, pp. 669–681.

[19] K. Zhao, B. Duka, H. Xie, D. J. Oathes, V. Calhoun, and Y. Zhang, "A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD," *NeuroImage*, vol. 246, Feb. 2022, Art. no. 118774.

[20] Y. Zhang, H. Ren, J. Ye, X. Gao, Y. Wang, K. Ye, and C.-Z. Xu, "AOAM: Automatic optimization of adjacency matrix for graph convolutional network," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 5130–5136.

[21] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017.

[22] T.-E. Kam, H. Zhang, Z. Jiao, and D. Shen, "Deep learning of static and dynamic brain functional networks for early MCI detection," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 478–487, Feb. 2020.

[23] B. Jie, M. Liu, C. Lian, F. Shi, and D. Shen, "Designing weighted correlation kernels in convolutional neural networks for functional connectivity based brain disease diagnosis," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101709.

[24] Y. Wang, Y. Yang, X. Guo, C. Ye, N. Gao, Y. Fang, and H. T. Ma, "A novel multimodal MRI analysis for Alzheimer's disease based on convolutional neural network," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 754–757.

[25] J. Huang, M. Wang, H. Ju, Z. Shi, W. Ding, and D. Zhang, "SD-CNN: A static-dynamic convolutional neural network for functional brain networks," *Med. Image Anal.*, vol. 83, Jan. 2023, Art. no. 102679.

[26] J. Ji, X. Xing, Y. Yao, J. Li, and X. Zhang, "Convolutional kernels with an element-wise weighting mechanism for identifying abnormal brain connectivity patterns," *Pattern Recognit.*, vol. 109, Jan. 2021, Art. no. 107570.

[27] C. Hua, H. Wang, J. Chen, T. Zhang, Q. Wang, and W. Chang, "Novel functional brain network methods based on CNN with an application in proficiency evaluation," *Neurocomputing*, vol. 359, pp. 153–162, Sep. 2019.

[28] L. Meng and J. Xiang, "Brain network analysis and classification based on convolutional neural network," *Frontiers Comput. Neurosci.*, vol. 12, p. 95, Dec. 2018.

[29] J. Ji and Y. Yao, "A novel CNN framework to extract multi-level modular features for the classification of brain networks," *Appl. Intell.*, vol. 52, no. 6, pp. 6835–6852, Apr. 2022.

[30] S. Kan, Y. Cen, Y. Li, M. Vladimir, and Z. He, "Local semantic correlation modeling over graph neural networks for deep feature embedding and image retrieval," *IEEE Trans. Image Process.*, vol. 31, pp. 2988–3003, 2022.

[31] Y. Yang, X. Guo, Z. Chang, C. Ye, Y. Xiang, and T. Ma, "Multi-modal dynamic graph network: Coupling structural and functional connectome for disease diagnosis and classification," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2022, pp. 1343–1349.

[32] Y. Yang, X. Guo, G. Cai, C. Ye, and T. Ma, "Topology-regularized self-knowledge distillation for transductive-inductive learning of brain disorder diagnosis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 4045–4049.

[33] 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," *Med. Image Anal.*, vol. 48, pp. 117–130, Aug. 2018.

[34] H. Jiang, P. Cao, M. Xu, J. Yang, and O. Zaiane, "Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction," *Comput. Biol. Med.*, vol. 127, Dec. 2020, Art. no. 104096.

[35] L. Li, H. Jiang, G. Wen, P. Cao, M. Xu, X. Liu, J. Yang, and O. Zaiane, "TE-HI-GCN: An ensemble of transfer hierarchical graph convolutional networks for disorder diagnosis," *Neuroinformatics*, vol. 20, no. 2, pp. 353–375, Apr. 2022.

[36] J. Jiang, Y. Wei, Y. Feng, J. Cao, and Y. Gao, "Dynamic hypergraph neural networks," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 2635–2641.

[37] X. Xing, "Dynamic spectral graph convolution networks with assistant task training for early MCI diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Shenzhen, China. Cham, Switzerland: Springer, Oct. 2019, pp. 639–646.

[38] Y. Yang, C. Ye, G. Su, Z. Zhang, Z. Chang, H. Chen, P. Chan, Y. Yu, and T. Ma, "BrainMass: Advancing brain network analysis for diagnosis with large-scale self-supervised learning," *IEEE Trans. Med. Imag.*, early access, 2024, doi: 10.1109/TMI.2024.3414476.

[39] L. Zhao, Z. Wu, H. Dai, Z. Liu, T. Zhang, D. Zhu, and T. Liu, "Embedding human brain function via transformer," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2022, pp. 366–375.

[40] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, "Brain network transformer," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 25586–25599.

[41] H. Cai, Y. Gao, and M. Liu, "Graph transformer geometric learning of brain networks using multimodal MR images for brain age estimation," *IEEE Trans. Med. Imag.*, vol. 42, no. 2, pp. 456–466, Feb. 2023.

[42] G. Qu, A. Orlichenko, J. Wang, G. Zhang, L. Xiao, K. Zhang, T. W. Wilson, J. M. Stephen, V. D. Calhoun, and Y.-P. Wang, "Interpretable cognitive ability prediction: A comprehensive gated graph transformer framework for analyzing functional brain networks," *IEEE Trans. Med. Imag.*, vol. 43, no. 4, pp. 1568–1578, Apr. 2024.

[43] C. Zhao, L. Zhan, P. M. Thompson, and H. Huang, "Revealing continuous brain dynamical organization with multimodal graph transformer," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2022, pp. 346–355.

[44] Y. Zhu, Y. Zhang, Y. Li, H. Ye, Y. Kong, and Y. Yuan, "Cortical network transformer for classification of structural connectivity in cerebral cortex," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–14, 2024.

[45] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. T. Yeo, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, Sep. 2018.

[46] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, and M. Dapretto, "The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism," *Mol. Psychiatry*, vol. 19, no. 6, pp. 659–667, Jun. 2014.

[47] T. Consortium, "The ADHD-200 consortium: A model to advance the translational potential of neuroimaging in clinical neuroscience," *Frontiers Syst. Neurosci.*, vol. 6, p. 62, Jan. 2012.

[48] D. Chen, Y. Lin, W. Li, P. Li, J. Zhou, and X. Sun, "Measuring and relieving the over-smoothing problem for graph neural networks from the topological view," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 4, pp. 3438–3445.
[49] T. K. Rusch, M. M. Bronstein, and S. Mishra, "A survey on oversmoothing in graph neural networks," 2023, *arXiv:2303.10993*.
[50] A. Harikumar, D. W. Evans, C. C. Dougherty, K. L. H. Carpenter, and A. M. Michael, "A review of the default mode network in autism spectrum disorders and attention deficit hyperactivity disorder," *Brain Connectivity*, vol. 11, no. 4, pp. 253–263, May 2021.
[51] C. He, Y. Chen, T. Jian, H. Chen, X. Guo, J. Wang, L. Wu, H. Chen, and X. Duan, "Dynamic functional connectivity analysis reveals decreased variability of the default-mode network in developing autistic brain," *Autism Res.*, vol. 11, no. 11, pp. 1479–1493, Nov. 2018.
[52] A. Padmanabhan, C. J. Lynch, M. Schaer, and V. Menon, "The default mode network in autism," *Biol. Psychiatry, Cognit. Neurosci. Neuroimaging*, vol. 2, no. 6, pp. 476–486, 2017.
[53] M. Chen, Y. He, L. Hao, J. Xu, T. Tian, S. Peng, G. Zhao, J. Lu, Y. Zhao, H. Zhao, M. Jiang, J.-H. Gao, S. Tan, Y. He, C. Liu, S. Tao, L. Q. Uddin, Q. Dong, and S. Qin, "Default mode network scaffolds immature frontoparietal network in cognitive development," *Cerebral Cortex*, vol. 33, no. 9, pp. 5251–5263, Apr. 2023.
[54] D. Vatansever, D. K. Menon, and E. A. Stamatakis, "Default mode contributions to automated information processing," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 48, pp. 12821–12826, Nov. 2017.
[55] A. Pelletier-Baldelli, J. A. Bernard, and V. A. Mittal, "Intrinsic functional connectivity in salience and default mode networks and aberrant social processes in youth at ultra-high risk for psychosis," *PLoS ONE*, vol. 10, no. 8, Aug. 2015, Art. no. e0134936.
[56] H. Chen, L. Q. Uddin, X. Duan, J. Zheng, Z. Long, Y. Zhang, X. Guo, Y. Zhang, J. Zhao, and H. Chen, "Shared atypical default mode and salience network functional connectivity between autism and schizophrenia," *Autism Res.*, vol. 10, no. 11, pp. 1776–1786, Nov. 2017.

JIUJIANG GUO was born in Taiyuan, Shanxi, China. He is currently a postgraduate student in mechanical engineering with the School of Mechanical and Power Engineering, Harbin University of Science and Technology, Harbin, China. He has gained some practical experience and academic foundation by being actively involved in various projects. His research interests include medical image analysis, machine learning, and computer-aided medical treatment.

MONAN WANG received the Ph.D. degree in automation from Harbin Engineering University, Harbin, China, in 2004. From 2004 to 2006, she was a Postdoctoral Fellow with the State Key Laboratory of Robotics and System, Harbin Institute of Technology, China. Currently, she has been a Professor with the Mechatronic Engineering Department, Harbin University of Science and Technology. Her research interests include medical image analysis and computer-aided medical treatment.

XIAOJING GUO was born in Shanxi, China. She received the M.S. degree in rehabilitation of traditional Chinese medicine from Heilongjiang University of Chinese Medicine, in 2022, where she is currently pursuing the Ph.D. degree in rehabilitation of traditional Chinese medicine. Her research interests include brain function and neurorehabilitation.