Classification of Brain Disorders in rs-fMRI via Local-to-Global Graph Neural Networks

Hao Zhang, *Graduate Student Member, IEEE*, Ran Song, *Member, IEEE*, Liping Wang, Lin Zhang, *Graduate Student Member, IEEE*, Dawei Wang, Cong Wang, and Wei Zhang

***Abstract*— Recently, functional brain network has been used for the classification of brain disorders, such as Autism Spectrum Disorder (ASD) and Alzheimer's disease (AD). Existing methods either ignore the non-imaging information associated with the subjects and the relationship between the subjects, or cannot identify and analyze disease-related local brain regions and biomarkers, leading to inaccurate classification results. This paper proposes a local-to-global graph neural network (LG-GNN) to address this issue. A local ROI-GNN is designed to learn feature embeddings of local brain regions and identify biomarkers, and a global Subject-GNN is then established to learn the relationship between the subjects with the embeddings generated by the local ROI-GNN and the non-imaging information. The local ROI-GNN contains a self-attention based pooling module to preserve the embeddings most important for the classification. The global Subject-GNN contains an adaptive weight aggregation block to generate the multi-scale feature embedding corresponding to each subject. The proposed LG-GNN is thoroughly validated using two public datasets for ASD and AD classification. The experimental results demonstrated that it achieves the state-of-the-art performance in terms of various evaluation metrics.**

***Index Terms*— Brain disorders, classification, non-imaing information, GNN, attention.**

Manuscript received 10 August 2022; revised 5 October 2022; accepted 28 October 2022. Date of publication 3 November 2022; date of current version 2 February 2023. This work was supported in part by the National Key Research and Development Plan of China under Grant 2018AAA0102504, in part by the National Natural Science Foundation of China under Grant 61991411 and Grant 62076148, in part by the Young Taishan Scholars Program of Shandong Province under Grant tsqn201909029, in part by the Natural Science Foundation of Shandong Province under Grant ZR2021MH236, and in part by the China Postdoctoral Science Foundation under Grant 2021M691935. *(Corresponding authors: Ran Song; Dawei Wang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Local Institutional Review Board of ABIDE and the Institutional Data Access/Ethics Committee of ADNI.

Hao Zhang, Lin Zhang, and Cong Wang are with the School of Control Science and Engineering, Shandong University, Jinan 250100, China (e-mail: hao_zhang12345@163.com; zl935546110@gmail.com; wangcong@sdu.edu.cn).

Ran Song and Wei Zhang are with the School of Control Science and Engineering, Institute of Brain and Brain-Inspired Science, Shandong University, Jinan 250100, China (e-mail: ransong@sdu.edu.cn; davidzhang@sdu.edu.cn).

Liping Wang is with the School of Information Science and Engineering, Shandong Normal University, Jinan 250012, China (e-mail: wangliping19872011@gmail.com).

Dawei Wang is with the Department of Radiology, and the Department of Epidemiology and Health Statistics, School of Public Health, Institute of Brain and Brain-Inspired Science, Qilu Hospital, Shandong University, Jinan 250100, China (e-mail: daweiwangtj@126.com).

Digital Object Identifier 10.1109/TMI.2022.3219260

I. Introduction

BRAIN disorders, such as Autism Spectrum Disorder (ASD) and Alzheimer's disease (AD), are irreversible neurodevelopmental and neurodegenerative diseases, and their prevalence rates are increasing year by year [1]. The etiological bases and neural substrates of ASD and AD have not been fully understood, and the mechanism of how brain disorders affecting human mental health and behavioral characteristics is still less studied [2]. At present, the diagnosis of ASD and AD is based primarily on simple symptomatic observations and clinicians' experience, which are easy to cause misdiagnosis. It is thus appealing to establish an automatic method for the accurate diagnosis of such diseases.

Brain disorders can be represented as aberrant structures within the functional brain networks. As visualized in Fig. 1 generated by the Harvard-Oxford atlas [3], in such a network, the nodes represent specific brain regions while the edges convey the temporal relationships between them. As a commonly used non-invasive imaging technique, resting-state functional magnetic resonance imaging (rs-fMRI) provides a neuropathological way to study the organisation of the brain and its cognitive functions by measuring hemodynamic changes and aids in simulating the functional and structural mechanism of the brain. Compared to other modality of images, such as structural MRI and positron emission tomography (PET), rs-fMRI provides brain functional connectivity between different regions of the brain. The functional connectivity is highly useful in the study of AD as the decreases in functional connectivity might be reflective of the decreases in the integrity of the neural network. Recent findings [4] show that human mental health and behavioral characteristics are closely related to the changes in functional connectivity measured by rs-fMRI.

A large portion of methods [5], [6], [7], [8], [9] using rs-fMRI mainly consist of 1) feature selection from fMRI such as clustering-based [5] and decomposition-based methods [6] and 2) disease classification based on the selected features such as ridge classifier [7], logistic regression [8] and support vector machine (SVM) [9]. However, such handcrafted methods have two intrinsic shortcomings. On one hand, unreliable results in the stage of feature selection may lead to significant errors in the later stage of disease classification. On the other hand, important topological information of the brain network that links different brain regions is largely ignored. Due to the complexity of brain functional connections and topology

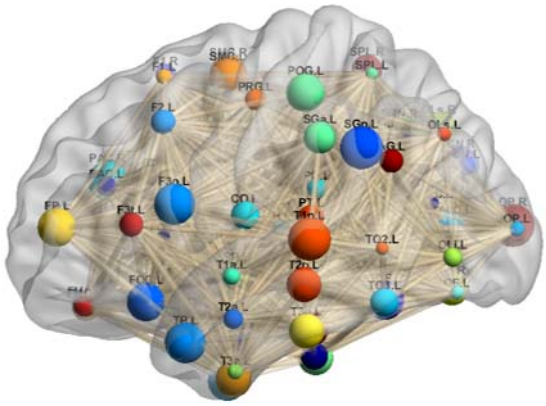

Fig. 1. Visualization of the functional brain network. Different nodes represent different brain regions generated by the Harvard-Oxford atlas.

variations, as shown in Fig. 1, it is challenging to obtain accurate brain disorder classification.

Recently, deep learning-based methods have demonstrated a good capability of feature extraction in medical image analysis [10], [11], [12]. In particular, graph neural networks (GNNs) combine deep neural networks with graph representation by explicitly capturing the topological information of brain network and is hence well-suited for the analysis of the relationship between various brain regions.

Generally, GNN-based methods for disease diagnosis are built upon two types of graphs: regional brain graph and subject graph. The former type of methods [13], [14], [15] are good at identifying and analyzing disease-related local brain regions and biomarkers. This is because they learn to classify subjects based on a graph where each node denotes a local brain region, i.e. a region of interest (ROI) in the brain of one subject. However, the relationship between subjects is ignored and non-imaging information, such as gender, age and acquisition site, is not sufficiently exploited. The latter type of GNN-based methods [16], [17], [18] overcome the above issues as they learn to classify subjects based on a global graph where each node corresponds to the brain of an individual subject and each edge encodes the relationship between two subjects. However, existing methods based on subject graphs cannot identify and analyze disease-related local brain regions and biomarkers. In addition, it is also challenging to learn a precise embedding for the representation of the brain functional connectivity network [12].

Therefore, this paper proposes a local-to-global graph neural network (LG-GNN) to combine the merits of the above two types of methods for the accurate classification of brain disorders in an end-to-end fashion. Specifically, a local ROI-GNN based on a regional brain graph is proposed to learn and analyze the features of local brain regions and identify biomarkers. And a global Subject-GNN is then established to learn the relationship between subjects with the feature embeddings generated by the local ROI-GNN and encode non-imaging information of multiple subjects for classifying the brain disorders.

Such a combination poses a requirement for the high quality of the feature embeddings learned through the local ROI-GNN. In particular, even a small number of feature embeddings damaging to the classification of brain disorders may have a significant overall impact due to the message passing within the global Subject-GNN. On one hand, this is usually the case as in a regional brain graph, not all brain regions benefit the prediction of diseases [15]. In other words, using all node features as the embeddings fed into the Subject-GNN may lead to feature redundancy. On the other hand, some ROIs are more indicative of predicting brain disorders than others. And more attention should be paid to the node features corresponding to the more indicative ROIs in the regional brain graph. We thus propose a self-attention based pooling module embedded in the local ROI-GNN to retain some important nodes only and discard redundant nodes.

Furthermore, in order to obtain the multi-scale representation for a node feature, we propose a multi-scale residual GNN as the global Subject-GNN for the classification of brain disorders, which compensates GNN with the capacity of modeling hierarchical and contextual information of global brain graph by designing multiple GNNs with a multi-scale architecture.

Overall, the contributions of our work are summarized as below:

- A novel GNN framework is proposed for classifying brain disorders with rs-fMRI in an end-to-end fashion. It employs a local ROI-GNN to learn good brain graph embeddings and identify biomarkers, and a global Subject-GNN to incorporate the non-imaging information and the relationships between subjects into the framework.
- A pooling strategy based on an attention mechanism is proposed to select the most discriminative feature embeddings generated by the local ROI-GNN. It significantly benefits the following global Subject-GNN and the overall performance of our method.
- The experimental results on two public medical datasets for ASD and AD classification demonstrate that the proposed LG-GNN achieves state-of-the-art performance.

II. Related Work

In this section, we first briefly review the handcrafted methods for classifying brain disorders and then focus more on the GNN-based approaches.

A. Handcrafted Methods

A handcrafted method usually constructs a functional connection network for each subject by calculating the statistical dependence (such as Pearson correlation coefficient) between the time series of neurophysiological signals measured from functional MRI data. Such a connection network usually facilitates the extraction of the measured values for the subsequent classification task. For example, Wee et al. [5] proposed to construct a functional connection network by extracting weighted local clustering coefficients from the brain connectivity network from rs-fMRI, and then employed multiple-kernel based-SVM algorithm for the subsequent MCI classification. Tang et al. [6] used SVM to classify AD from normal control (NC) after PCA and Student's t-test for dimension reduction

based on shape and diffusion tensor imaging. Plitt et al. [19] studied the impact of different classifier and ROI settings on the classification accuracy of rs-fMRI applied to ASD and found that linear SVM among those methods classifies ASD more effectively. However, existing handcrafted methods generally treat feature selection and classifier training as two separate steps, and thus unreliable results in the first stage can lead to significant errors in the second one. Besides, they cannot extract the high-level topological features from the brain connectivity networks. Both issues may lead to sub-optimal performance of brain disorder classification.

B. GNN-Based Methods

As a popular deep neural network, GNN in previous work has demonstrated its potential for medical image analysis. In general, GNN-based methods in the diagnosis of brain disorders can be divided into two groups in accordance with the graphs that they are based on: regional brain graph and subject graph.

1) GNNs Based on Regional Brain Graphs: This type of GNN-based methods are built upon brain graphs where a node denotes a brain region. For instance, Xing et al. [13] proposed the dynamic spectral graph convolution networks (DS-GCNs) for early MCI diagnosis using functional MRI. Ktena et al. [14] proposed a siamese graph convolution network (s-GCN) which employed the polynomial filters to learn a similarity metric between brain connectivity graphs for ASD classification in rs-fMRI data. Yao et al. [4] proposed a mutual multi-scale triplet GCN for classifying brain disorders based on functional and structural connectivity. Li et al. [15] proposed an interpretable GNN for fMRI analysis and biomarkers identification. Wen et al. [20] proposed a multiview graph convolutional neural network (MVS-GCN) which combines the graph structure learning with multi-task graph embedding learning to improve the classification performance of the brain disorders and identify the potential functional subnetworks in rs-fMRI. Although disease-related brain regions and biomarkers can be identified and analyzed by the above methods, the relationship between the subjects are ignored and the non-imaging information of individuals are not used effectively.

2) GNNs Based on Subject Graphs: This type of GNN-based methods learn to classify the subjects using the graphs where a node denotes the brain graph of a subject. As one of the earliest works in this area, Parisot et al. [16] proposed to use GCN for brain analysis in fMRI and T1 MRI data, where a group of subjects was represented as a sparse graph. The graph nodes were associated with image-based individual feature vectors and the phenotypic information was encoded as the edge. Kazi et al. [17] proposed another network based on the subject graph, namely InceptionGCN, to capture the local and global context of heterogeneous graph structures by the "inception modules" for predicting ASD and AD. Huang et al. [18] designed a well-regularized spectral GCN for disease prediction combining imaging data and non-imaging data to separate NC from ASD patients. Lostar et al. [21] designed a hypergraph of subjects where a brain graph of a patient represented a node. Song et al. [22] proposed a similarity-aware adaptive calibrated GCN (SAC-GCN) to combine functional and structural information to predict SMC and MCI. A limitation of the methods based on subject graphs is that they used handcrafted features as node descriptors. Note that the main advantage of deep learning is that a deep network can learn itself an optimal feature representation of the raw input data for the task at hand. Hence, we argue that the above methods do not make good use of the powerful feature learning capability of deep neural networks in an end-to-end manner, which means that there is still a large room for improvement in such kind methods.

Jiang et al. [23] proposed a hierarchical GCN (Hi-GCN) model which is capable of jointly learning the graph embedding from both the brain functional network and the population network in rs-fMRI images. Building on the above work, Li et al. [24] developed a transfer learning scheme enabling GCN to learn generic graph structural features by leveraging the commonality in two related domains in rs-fMRI images. However, it is still difficult for them to retain important brain ROIs and identify biomarkers.

III. Proposed Method

An overview of the proposed method is illustrated in Fig. 2. It mainly consists of a local ROI-GNN and a global Subject-GNN. For the input of m subjects composed of imaging (e.g. resting-state fMRI) and non-imaging data (e.g. gender, age and site), the local ROI-GNN is first employed to produce the brain graph embeddings based on the regional brain graphs constructed from the rs-fMRI images. Then, the global Subject-GNN is introduced for node classification to obtain the disease state (e.g. normal or diseased) of a subject where the embedding of each subject forms a node. We elaborate the local ROI-GNN and the global Subject-GNN in the following subsections.

A. Local ROI-GNN

1) Regional Brain Graph Construction: The construction of the regional brain graph is vital for brain graph embedding learning as improperly constructed graphs (i.e. graphs that cannot properly correspond to the similarities between ROIs) cannot sufficiently exploit the topological information within the brain connectivity network through the GNN. For each subject, we construct the regional brain graph based on the rs-fMRI image of the subject. First, the brain of the subject depicted by the rs-fMRI image is parcellated into n ROIs with the Harvard-Oxford atlas and the brain graph $G_{local} = \{\mathbf{V}, \mathbf{A}\}$ is constructed where each node corresponds to an ROI, i.e. $\mathbf{V} = \{v_1, \ldots, v_n\}$. The node features in the brain graph G_{local} can be represented as a matrix $\mathbf{X} = [x_1, \ldots, x_n]^\top$, where x_i is the feature representation associated with node v_i. $\mathbf{A}$ is the adjacency matrix of $\mathbf{V}$, defined as

$$\mathbf{A}_{ij} = \begin{cases} \dfrac{Cov\left(v_i, v_j\right)}{\sigma_{v_i}\sigma_{v_j}}, & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

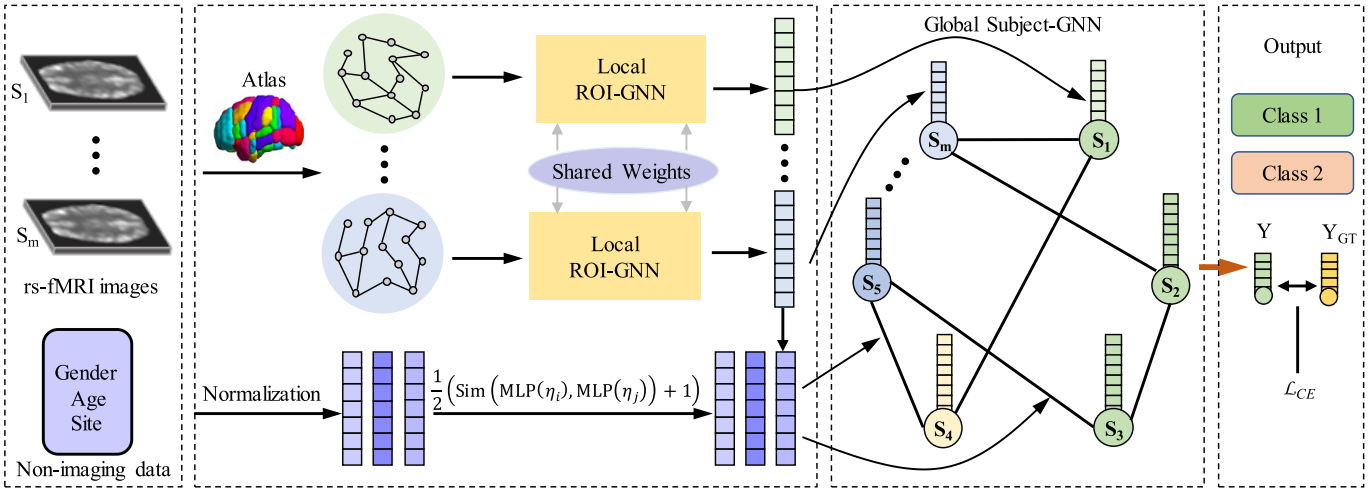

Fig. 2. Overall framework of the proposed local-to-global GNN for classifying brain disorders with rs-fMRI. The local ROI-GNN is designed to produce the brain graph embeddings based on the ROIs in the brain of each subject. The global Subject-GNN is then introduced for node classification where the embedding of each subject forms a node.

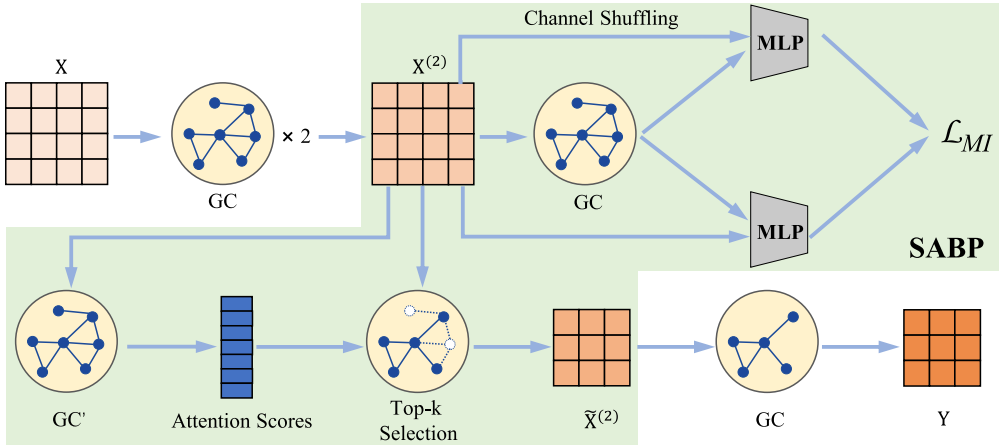

Fig. 3. Framework of the proposed local ROI-GNN comprising three GC layers and a self-attention based pooling module.

where $Cov\left(v_i, v_j\right)$ is the cross covariance between the ROIs corresponding to v_i and v_j. Therefore, for two adjacent nodes v_i and v_j, $\mathbf{A}_{ij}$ is equal to the Pearson correlation coefficient between them.

2) Local ROI-GNN Model: As shown in Fig. 3, the local ROI-GNN consists of three graph convolution (GC) layers and the Self-Attention Based Pooling (SABP) module. The local ROI-GNN contains a total of five GC layers, and two of them are included in the SABP. The GC layers in the local ROI-GNN are used to extract the topological feature for each node in a regional brain graph. The SABP is used to learn the node importance and select the discriminative features by retaining some important nodes only. Specifically, the t-th ($t \in \{1, 2\}$) GC layer with ReLU activation is defined referring to [25] as:

$$\begin{aligned}\mathbf{X}^{(t)} &= \mathrm{ReLU}(\mathrm{GC}(\mathbf{X}, \mathbf{A}) \\ &= \mathrm{ReLU}\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\mathbf{X}^{(t-1)}\mathbf{W}^{(t)}\right)\end{aligned} \tag{2}$$

where $\mathbf{D} = \mathrm{diag}\left(\sum_j \mathbf{A}_{1j}, \sum_j \mathbf{A}_{2j}, \ldots, \sum_j \mathbf{A}_{nj}\right)$ is the degree matrix of $\mathbf{A}$ and $\mathbf{W}^{(t)}$ denotes a trainable weight matrix of the t-th layer. $\mathbf{X}^{(t)} = \left[x_1^{(t)}, x_2^{(t)}, \ldots, x_n^{(t)}\right]^\top$ denotes the node representation of the t-th GC layer. The local ROI-GNN extracts features from each node of the regional brain graph and then aggregates them based on the spatial structure of the graph through localized graph filters to deliver the graph embedding.

The receptive field in one GC layer is restricted to one-hop neighborhoods, which limits their ability to explore more complex functional interactions from broader node features [26]. Two-hop (adjacent neighbors of one-hop nodes) approaches can present an influence in the center ROI by information transfer to one-hop regions [27]. We thus stack two GC layers to capture two-hop information. Then, we obtain a set of representations $\mathbf{X}^{(2)} \in \mathbb{R}^{n \times c}$ for the graph nodes, where c represents the number of channels of a node feature. Next, we elaborate SABP which learns node importance and how to encode it into a unified graph representation.

3) Self-Attention Based Pooling: Recent findings have shown that some ROIs are more important than the others in the prediction of brain disorders [15]. Thus, it is crucial to use a node pooling layer to reduce the size of the graph and only preserve some important nodes. Moreover, the pooling layers

can reduce the size of network parameters and thus avoid overfitting. As such, the node representations on coarsened graphs can deliver higher graph-level representations. To this end, we propose the Self-Attention Based Pooling (SABP) module to extract the discriminative node features and select the important nodes.

Given the node feature $\mathbf{X}^{(2)} \in \mathbb{R}^{n\times c}$, we first use a GC layer with one output channel (expressed as GC′ in Eq. 3) to estimate the self-attention scores $\mathbf{z} \in \mathbb{R}^{n\times 1}$ as it better encodes the graph topology than a Sigmoid or a Softmax layer. Inspired by [28], the top-k node selection strategy is adopted to retain a portion of nodes of the regional brain graph, where k is the number of nodes selected to establish the new graph. In the selection of the k nodes, the operation of the node ranking outputs the index idx of the k-largest values based on $\mathbf{z}$. The self-attention scores of the selected nodes $\tilde{\mathbf{z}}$ can then be obtained by applying the tanh function to the k-largest elements of $\mathbf{z}$. Then the pooled node feature $\tilde{\mathbf{X}}^{(2)}$ can be captured by element-wise multiplication on the indexed feature $\hat{\mathbf{X}}^{(2)}$ and $\tilde{\mathbf{z}}$. $\mathbf{A}(\text{idx}, \text{idx})$ performs the row and column extraction to form the new adjacency matrix $\tilde{\mathbf{A}}$. To sum up, the operation of SABP is expressed below:

$$\begin{aligned}
\mathbf{z} &= \text{GC}'(\mathbf{X}^{(2)}, \mathbf{A}),\\
\tilde{\mathbf{z}} &= \tanh(\mathbf{z}),\\
\text{idx} &= \text{topk}(\tilde{\mathbf{z}}, k),\\
\hat{\mathbf{X}}^{(2)} &= \mathbf{X}^{(2)}(\text{idx}, :),\\
\tilde{\mathbf{X}}^{(2)} &= \hat{\mathbf{X}}^{(2)} \odot \tilde{\mathbf{z}},\\
\tilde{\mathbf{A}} &= \mathbf{A}(\text{idx}, \text{idx}).
\end{aligned} \tag{3}$$

Since SABP aims to reliably retain the important nodes, we propose a mutual information loss $\mathcal{L}_{MI}$ to ensure that the weights of the k selected ROIs differ significantly from those of the unselected nodes. The weights of the selected ROIs are expected to be close to 1, while the weights of the unselected ROIs are supposed to be close to 0. Thus, $\mathcal{L}_{MI}$ is defined as:

$$\begin{aligned}
\mathcal{L}_{MI} = &\frac{1}{n}\sum_{i=1}^{n} \text{MLP}(\text{Concat}(\text{GC}(\mathbf{X}^{(2)}, \mathbf{A}), \mathbf{X}^{(2)})\\
&- \log \frac{1}{n}\sum_{i=1}^{n} e^{\text{MLP}(\text{Concat}(\overline{\mathbf{X}}, \mathbf{X}^{(2)}))}
\end{aligned} \tag{4}$$

where the feature $\overline{\mathbf{X}}$ is obtained by randomly shuffling the channel positions of $\mathbf{X}^{(2)}$. The MLP is the Multi-Layer Perceptron with one output channel, which is used to perform a linear transformation and reduce feature dimensions. n is the dimension of the channel features output by the MLP. $\mathcal{L}_{MI}$ defined on the local ROI-GNN essentially imposes a global graph constraint which facilitates the discovery and preservation of long-range relationships between the embeddings not directly connected. Such long-range relationships help to describe the whole graph topology and thus extract discriminative features from it. To the best of our knowledge, such relationships are often ignored or cannot be captured by existing self-attention methods, such as [28].

As shown in Fig. 3, the final output of the local ROI-GNN $\mathbf{Y} = [y_1, y_2, \ldots, y_k]^\top$ is produced by further applying a GC layer to the selected k node features:

$$\mathbf{Y} = \text{GC}(\tilde{\mathbf{X}}^{(2)}, \tilde{\mathbf{A}}). \tag{5}$$

The SABP module essentially distills the most important knowledge from the regional brain graph and the features associated with the graph nodes by taking into account the topological relationships among such features. It also facilitates reducing the size of the output embedding of the local ROI-GNN, which significantly decreases the total number of trainable parameters of the proposed LG-GNN as the following global Subject-GNN takes as input multiple embeddings produced by the local ROI-GNN.

B. Global Subject-GNN

1) Subject Graph Construction: Once the embedding of the regional brain graph of each subject is generated by the local ROI-GNN, the next step is to recognize the disease state (e.g. normal or diseased) for each subject, where the non-imaging data of the subjects are highly complementary to such embeddings based only on the rs-fMRI images. The affinity graph needs to be carefully constructed because it accurately simulate the interaction between subjects. Therefore, we propose to construct the global graph $G_{global} = \{\mathbf{V}', \mathbf{A}'\}$ for all subjects $\mathbf{S} = \{\text{S}_1, \text{S}_2, \ldots, \text{S}_m\}$ based on both imaging and non-imaging data where m is the number of subjects and thus $|\mathbf{V}'| = m$. Specifically, each graph node in the set $\mathbf{V}'$ corresponds to an embedding Y of a subject S obtained by reshaping the output $\mathbf{Y}$ of the local ROI-GNN into a single column vector. $\mathbf{A}'$ is the adjacency matrix of $\mathbf{V}'$. The adjacency matrix $\mathbf{A}'$ can be defined as $\mathbf{A}' = \mathbf{C} * \mathbf{W}$. $\mathbf{C}$ is a binarized connectivity matrix obtained by using both image and non-imaging information as proposed in [16]. First, a similarity matrix $\mathbf{S}_1$ is constructed by using the embedding features $\mathbf{Y}$: $\mathbf{S}_1 = \exp\left(-\frac{[\rho(\mathbf{Y}_i, \mathbf{Y}_j)]^2}{2\sigma^2}\right)$. ρ is the correlation distance and σ is the mean value of $[\rho(\mathbf{Y}_i, \mathbf{Y}_j)]^2$. Second, the node similarity metric $\mathbf{S}_2$ is built by utilizing non-imaging information (e.g., gender, age) to obtain more useful clinical auxiliary information. Then we fuse these two matrixes by Hadamard product: $\mathbf{C}' = \mathbf{S}_1 * \mathbf{S}_2$. Next, we obtain the binarized connection graph $\mathbf{C}$ by setting the connectivities with the weights greater than 0.4 as 1 and the others as 0. Each element of the sparse weight matrix $\mathbf{W}$ is defined by the function of the normalized non-imaging data η_i and η_j of two subjects: $\mathbf{W}_{ij} = \frac{\text{Sim}(\text{MLP}(\eta_i), \text{MLP}(\eta_j))+1}{2}$, where the two MLP layers have shared weights and Sim denotes the cosine similarity between two input vectors.

2) Global Subject-GNN Model: GNN models often suffer from the over-smoothing issue due to stacking multiple layers [29]. Besides, the representation ability of the shallow GNN structure is insufficient. To address the above issue and improve the classification performance, we propose a multi-scale residual GNN which creates a global model based on the imaging and non-imaging information of multiple subjects for brain disorder classification. It provides the capacity of modeling hierarchical and contextual information of the graph through multiple GNNs in a multi-scale architecture. The framework of the proposed global Subject-GNN is illustrated in Fig. 4.

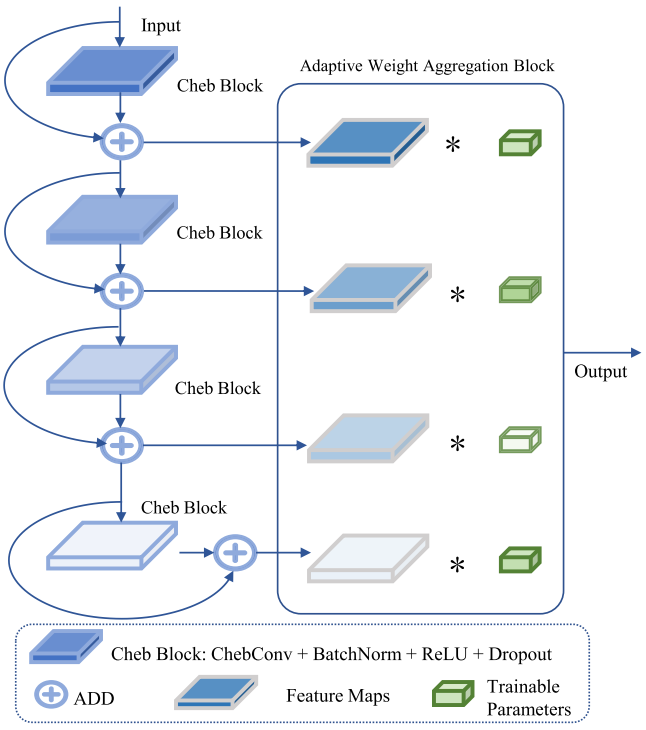

Fig. 4. Framework of the proposed global Subject-GNN comprising four Cheb blocks and one adaptive weight aggregation block.

The framework consists of four Cheb blocks and an adaptive weight aggregation block (AWAB). Each Cheb block layer takes the node features from the previous layer as input and outputs new node features for the next layer. The residual connections to each layer are applied in the forward propagation to accelerate the convergence and alleviate the over-smoothing problem. In each Cheb block, the input feature passes through a ChebConv layer, a batch normalization layer, a ReLU activation layer and a dropout layer in turn.

The ChebConv layer is a spectral-based graph convolution to combine the entire structure and the individual components, which uses the Chebyshev spectral graph convolutional operator proposed by Defferrard et al. [30]. The spectral-based graph convolution can be defined as the multiplication of a signal $h \in \mathbb{R}^m$ (a scalar for every node) with a filter $\mathbf{g}_\theta = \mathrm{diag}(\theta)$, parameterized by $\theta \in \mathbb{R}^m$:

$$\mathbf{g}_\theta \star h = \mathbf{U}\mathbf{g}_\theta(\Lambda)\mathbf{U}^\top h \tag{6}$$

where $\star$ is the convolution operator on graph. $\mathbf{U}$ is the matrix constructed by the eigenvectors of the Laplace matrix $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}} = \mathbf{U}\Lambda\mathbf{U}^\top$ where $\mathbf{I}$ is the identity matrix. Λ is the diagonal matrix composed of the eigenvalues of $\mathbf{L}$ and $\mathbf{D}$ is the degree matrix of the adjacency matrix $\mathbf{A}'$ of the graph. To reduce the computational complexity of calculating $\mathbf{U}\mathbf{g}_\theta(\Lambda)\mathbf{U}^\top$, we simplify this calculation by using the K-order Chebyshev polynomials to approximate $\mathbf{g}_\theta(\Lambda)$, expressed as:

$$\mathbf{g}_\theta \star h \approx \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}})h \tag{7}$$

where $\tilde{\mathbf{L}}$ is the rescaled graph Laplacian, θ_k are the trainable parameters. $T_k(\tilde{\mathbf{L}}) = 2\tilde{\mathbf{L}}T_{k-1}(\tilde{\mathbf{L}}) - T_{k-2}(\tilde{\mathbf{L}})$ with $T_0(\tilde{\mathbf{L}}) = 1$ and $T_1(\tilde{\mathbf{L}}) = \tilde{\mathbf{L}}$.

In this work, given the node features $\mathbf{H} = [Y_1, Y_2, \ldots, Y_m]^\top$ of the subject graph, we can define the output feature of the l-th ChebConv as:

$$\mathbf{H}^{(l+1)} = \sum_{k=0}^{K-1} \theta_k^{(l)} T_k(\tilde{\mathbf{L}})\mathbf{H}^{(l)} \tag{8}$$

In order to obtain a multi-scale node feature embedding, we aggregate the output feature maps of different Cheb blocks based on the learned weights of the neural network rather than a simple summation or concatenation. The node embeddings $\{\mathbf{H}^{(0)}, \ldots, \mathbf{H}^{(3)}\}$ of all Cheb blocks are aggregated by the AWAB to generate the final node embedding $\mathbf{Z}$, expressed as

$$\mathbf{Z} = \sum_l w^{(l)} \odot \mathbf{H}^{(l)} \tag{9}$$

where $w^{(l)}$ ($l \in \{0, 1, 2, 3\}$) is the trainable weight which follows the softmax distribution, expressed as:

$$w^{(l)} = \mathrm{Softmax}(r^{(l)}) = \frac{\exp(r^{(l)})}{\sum_l \exp(r^{(l)})} \tag{10}$$

where $r^{(l)}$ are the learnable weights with random initialization. The AWAB can effectively use the feature map of each Cheb block to obtain a multi-scale node feature embedding, which alleviates the over-smoothing problem. Finally, the global Subject-GNN uses a fully-connected layer with two output channels to generate the classification results.

C. Total Training Loss

We adopt the cross entropy loss $\mathcal{L}_{CE}$ for constraining the global Subject-GNN:

$$\mathcal{L}_{CE} = \frac{1}{N}\sum_i -\left[q_i \cdot \log(p_i) + (1 - q_i) \cdot \log(1 - p_i)\right] \tag{11}$$

where p and q represent the predicted results and the ground-truth labels, respectively. N is the number of samples. Thus, by combining $\mathcal{L}_{CE}$ and the mutual information loss $\mathcal{L}_{MI}$ used to constrain the local ROI-GNN, the total loss of the proposed LG-GNN can be expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda\mathcal{L}_{MI} \tag{12}$$

where λ is hyper-parameter and is set to 0.1 in this work.

IV. Materials and Classification Evaluation

A. Materials

We evaluate the LG-GNN on two publicly accessible datasets: ABIDE (Autism Brain Imaging Data Exchange)[1] and ADNI (Alzheimer's Disease Neuroimaging Initiative)[2]. With the rs-fMRI and non-imaging data, we aim to distinguish the NC subjects from the ASD patients on ABIDE and perform

[1] http://preprocessed-connectomes-project.org/abide/
[2] http://adni.loni.usc.edu/

TABLE I
THE DEMOGRAPHIC STATISTICS OF THE DATASETS USED IN THIS WORK

Dataset	Subgroup	Number	Gender (M/F)	Age(mean±std.)
ABIDE	ASD	403	349/54	17.07±7.95
	NC	468	378/90	16.84±7.23
ADNI	AD	34	16/18	72.89±7.12
	MCI	100	51/49	72.50±7.69
ADNI (NC vs. MCI)	NC	100	52/48	77.22±7.56
	MCI	121	50/71	75.58±8.68
ADNI (pMCI vs. sMCI)	pMCI	41	18/23	77.61±8.28
	sMCI	80	32/48	75.09±8.91

three kinds of classification tasks on ADNI: AD vs. MCI (Mild Cognitive Impairment); NC vs. MCI; pMCI vs. sMCI. The demographic statistics of the datasets used in this work in shown in Table I.

1) ABIDE Dataset: ABIDE is collected from different acquisition sites and openly shares functional MRI and phenotypic data of 1, 112 subjects. Our experiment uses a sub-dataset of 871 samples consisting of 403 ASD samples and 468 NC samples acquired at 20 different sites. The C-PAC [31] pipeline is employed to preprocess the rs-fMRI data from this dataset, which involves skull striping, slice timing correction, motion correction, global mean intensity normalization, nuisance signal regression, band-pass filtering (0.01–0.1 Hz) and then the fMRI images are co-registered on the corresponding anatomical images with FSL BBreg, and the results are normalized to the MNI152 space with the non-linear registration from ANTS.

2) ADNI Dataset: ADNI is a large-scale dataset and contains longitudinal brain MRI, PET data and genetic and phenotypic information of over 1700 adults for Alzheimer's disease (AD) study. A standard protocol is used to preprocess the data, including motion correction, fMRI co-registration to T1-MRI, normalization to the MNI template using SPM12 [32] and Gaussian spatial smoothing with an FWHM up to 5mm. All subjects were visually inspected and excluded from the analysis if they had severe scanner artifacts or head movements with an amplitude larger than 2mm. For a fair comparison, the same set of 134 subjects as [33] was used in our experiment, comprising 96 individuals with MCI and 40 diagnosed with AD.

ADNI (NC vs. MCI) and ADNI (pMCI vs. sMCI): The rs-fMRI data of these categories were collected from ADNI website. The patients converted to AD within 36 months are labeled as pMCI patients, and those whose conditions do not deteriorate are labeled as sMCI patients. We downloaded the data from the ADNI website with the description of rs-fMRI in ADNI2 and ADNI3 cohorts. After discarding the data with NaN values, 41 pMCI and 80 sMCI subjects were available and finally used for this work. For rs-fMRI data preprocessing, we apply the standard procedures using GRETNA toolkit [34], including the removal of the first 10 volumes of each subject, slice timing correction, head motion correction, warping individual functional images to the standard MNI space by estimating their transformation to the echo planar imaging (EPI) template, spatial smoothing, regression out white matter signal and CSF Signal, removal of the linear drift, band-pass filtering (0.01–0.1 Hz). Finally, we obtain the time series by aligning the rs-fMRI images using Harvard-Oxford atlas.

B. Experimental Setting

Our LG-GNN was implemented on a PyTorch framework with a single NVIDIA RTX 3090 GPU. Adaptive moment estimation (Adam) [35] was employed for network optimization. For ABIDE, the dropout rate is set to 0.3, the learning rate is set to 0.01 and the maximum number of epochs is set to 400. For ADNI, the dropout rate is set to 0.3, the learning rate is set to 0.01 and the maximum number of epochs is set to 300. The Chebyshev polynomial order K is set to 3. The 10-fold stratified cross-validation was used for the evaluation on each dataset, where each fold is used once in turn for the test while the remaining 9 folds form the training set. The training times on ABIDE, ADNI, ADNI (NC vs. MCI), and ADNI (pMCI vs. sMCI) are about 8.3 hours, 0.6 hours, 1.5 hours, and 0.6 hours respectively. The number of trainable parameters of the model is 0.51M. The results of classification are averaged over the 10 times of test. Four metrics are used to evaluate the classification performance, including classification accuracy (Acc), area under the curve (AUC), sensitivity (Sen) and F1-score.

The types of non-image information and the reasons for their selection are expressed below. For the ABIDE dataset, since the data were acquired at different sites using diverse imaging protocols and scanners, the acquisition site is selected as a non-imaging information measure. In addition, we also use gender as the measure since different genders lead to different probabilities of ASD as observed in several studies [36], [37]. For the ADNI dataset, since females are at a greater risk of developing AD and there are age-related group differences in functional connectivity [38], we selected gender and age for specifying the data.

The dimension of each embedding generated by the local ROI-GNN is 2000, 2140, 2020, and 2020 for ABIDE, ADNI, ADNI (NC vs. MCI) and ADNI (pMCI vs. sMCI), respectively. The dimension of non-imaging data input to the LG-GNN is (871,2), (134,2), (221,2) and (121,2) for ABIDE, ADNI, ADNI (NC vs. MCI) and ADNI (pMCI vs. sMCI), respectively.

V. EXPERIMENTAL RESULTS

A. Classification Results on Different Datasets

1) Classification Results on ABIDE: We compare the proposed LG-GNN with both handcrafted and GNN-based methods to evaluate its classification performance. The handcrafted methods include Ridge Classifier with half of the input features selected [7], SVM classifier with the RBF kernel used [39] and Random Forest classifier with 100 estimators [40], which were all implemented using the scikit-learn library [43]. GNN-based methods include GCN [25], GAT [42], BrainGNN [15], MVS-GCN [20], PopulationGNN [16], InceptionGCN [17], EV-GCN [18] and Hi-GCN [23]. We also compare LG-GNN with a competing method based on a deep neural network (DNN) which consists of two autoencoders and one MLP [41].

TABLE II
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON THE ABIDE DATASET. NI INDICATES WHETHER NON-IMAGING INFORMATION IS USED: ✓ INDICATES YES AND × INDICATES NO

Method	NI	Acc	Sen	Spe	AUC	F1
Ridge Classifier [7]	×	63.04 (0.29)	65.93 (0.23)	69.82 (0.25)	68.23 (0.57)	65.61 (0.65)
SVM [39]	×	65.60 (0.25)	70.66 (0.47)	64.82 (0.39)	68.32 (0.37)	68.66 (0.51)
Random Forest [40]	×	61.39 (0.27)	68.37 (0.49)	65.81 (0.70)	67.59 (0.48)	68.46 (0.21)
DNN [41]	×	67.25 (0.22)	71.20 (0.72)	63.70 (0.87)	68.41 (0.69)	68.10 (0.72)
GCN [25]	×	67.83 (1.57)	71.40 (0.64)	72.01 (0.46)	65.94 (0.46)	71.28 (0.29)
GAT [42]	×	69.02 (0.57)	72.91 (1.61)	70.12 (1.49)	70.23 (0.88)	65.08 (0.97)
BrainGNN [15]	×	70.60 (0.47)	61.30 (0.91)	71.89 (1.56)	64.44 (0.61)	65.36 (1.05)
MVS-GCN [20]	×	68.61 (0.50)	68.73 (0.23)	64.76 (0.26)	67.38 (0.57)	69.39 (0.82)
PolulationGCN [16]	✓	71.41 (0.17)	77.55 (0.75)	64.01 (0.94)	71.76 (0.34)	74.29 (0.21)
InceptionGCN [17]	✓	70.94 (0.30)	78.44 (0.63)	62.37 (1.47)	71.52 (0.44)	74.63 (0.22)
EV-GCN [18]	✓	80.25 (1.23)	80.74 (1.37)	**82.47 (1.41)**	82.04 (1.43)	81.81 (1.10)
Hi-GCN [23]	✓	72.81 (0.32)	72.04 (0.69)	73.91 (1.05)	77.12 (1.37)	75.72 (0.91)
LG-GNN (Ours)	✓	**81.75 (1.10)**	**83.22 (1.84)**	80.99 (1.17)	**85.22 (1.01)**	**82.96 (0.94)**

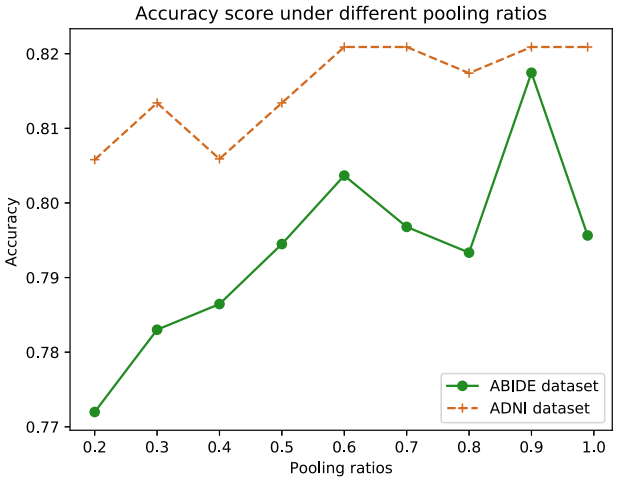

Fig. 5. Effect of different pooling ratios used by the SABP module.

The mean and variance values are reported in all comparative experiments.

Quantitative results are reported in Table II. In the case of using imaging data only, it can be seen that the handcrafted methods (i.e. Ridge Classifier, SVM and Random Forest) are outperformed by those based on DNN and GNN. And the classification accuracy of the handcrafted methods is lower than 66%, which shows that due to the inherent complexity of the rs-fMRI data, it is difficult to extract representative features from them for classification through hand engineering. Moreover, GNN-based methods perform better than DNN, which demonstrates that the high-level topological information of the brain connectivity extracted by GNN is useful for classifying brain disorders. Besides, the methods based on both imaging and non-imaging data are superior to those based only on imaging data. This is because the non-imaging data provide the clinical prior information when little medical expertise is available and such information that models the interaction and similarity between subjects is usually useful for the classification. On the ABIDE dataset, the proposed LG-GNN is the top performing method compared with other methods. In particular, our LG-GNN outperforms all of the competing GNN-based methods, which demonstrates the proposed idea of combing the advantages of both regional brain graph and subject graph. In the case of using both imaging and non-imaging data, PopulationGCN, InceptionGCN and EV-GCN are based on the recursive elimination of a fixed number of features through the ridge classifier. In comparison, our method uses local ROI-GNN to generate the most discriminative features for the global Subject-GNN, which provides a promising framework for exploiting the multi-modal data more sufficiently.

2) Classification Results on ADNI: Quantitative results on the ADNI dataset are reported in Table III. The proposed LG-GNN achieves the best classification performance compared with other methods. It can be seen that similar to ABIDE, the GNN-based methods are still the best-performing ones on ANDI compared to other types of methods. In addition, it can be observed that for almost all competing methods, the performance on ADNI is better than that on ABIDE. This is because ABIDE is highly heterogeneous with data acquired at different sites using diverse imaging protocols and scanners. Overall, such results further demonstrate the advantage of a hierarchical learning framework with both local and global GNN models.

To verify the effectiveness of the proposed method on NC vs. MCI and pMCI vs. sMCI datasets, we conducted comparative experiments with GNN-based methods that use non-imaging information. The classification results of MCI vs. NC on ADNI are illustrated in Table IV. It can be seen that our method outperforms other GNN-based methods, which demonstrates the powerful feature learning capability of LG-GNN. Moreover, the performance of classifying pMCI and sMCI is verified by the experimental results in Table V. Our method achieves the best scores: 79.23%, 87.50%, 67.00% and 84.37% in terms of Acc, Sen, Spe and F1, respectively. Hence, it shows superior performance in distinguishing pMCI from sMCI, which is critical for the identification of patients who are at-risk for AD. Notably, the recent hierarchical GNN method, Hi-GCN, performs poorly compared to our method on these two datasets. The reason that contributes to this situation may be that the feature redundancy caused by inputting all node features into global GNN are yet unconsidered by

TABLE III
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON THE ADNI DATASET. NI INDICATES WHETHER NON-IMAGING INFORMATION IS USED: ✓ INDICATES YES AND × INDICATES NO

Method	NI	Acc	Sen	Spe	AUC	F1
Ridge Classifier [7]	×	73.87 (1.02)	73.13 (2.25)	49.10 (1.47)	54.91 (3.03)	77.54 (2.84)
SVM [39]	×	73.90 (0.73)	80.04 (0.91)	50.24 (2.16)	58.21 (1.82)	84.72 (0.87)
Random Forest [40]	×	74.64 (1.62)	77.43 (3.72)	47.19 (3.12)	53.45 (4.65)	75.12 (2.76)
DNN [41]	×	74.56 (4.50)	78.57 (2.80)	52.70 (3.90)	62.01 (4.22)	85.85 (2.25)
GCN [25]	×	74.62 (2.63)	78.67 (3.18)	50.90 (4.45)	55.29 (2.87)	83.56 (1.76)
GAT [42]	×	76.86 (1.34)	79.99 (3.73)	51.92 (2.20)	61.11 (2.70)	84.91 (1.02)
BrainGNN [15]	×	77.61 (4.30)	81.81 (3.09)	54.83 (3.67)	53.07 (5.65)	84.77 (2.01)
MVS-GCN [20]	×	69.23 (4.32)	72.61 (3.26)	60.21 (2.74)	59.66 (5.43)	73.71 (2.69)
PolulationGCN [16]	✓	79.85 (1.87)	79.54 (1.34)	58.10 (3.78)	51.83 (2.02)	88.09 (1.14)
InceptionGCN [17]	✓	80.60 (1.42)	80.26 (2.84)	54.19 (3.49)	64.50 (4.22)	88.79 (1.01)
EV-GCN [18]	✓	80.59 (1.53)	80.81 (0.34)	51.21 (3.90)	62.16 (4.21)	88.63 (1.45)
Hi-GCN [23]	✓	71.60 (2.77)	78.00 (2.16)	**62.81 (4.86)**	62.60 (4.87)	76.94 (2.20)
LG-GNN (Ours)	✓	**82.09 (1.40)**	**82.96 (0.91)**	51.66 (4.84)	**68.33 (3.47)**	**88.91 (1.12)**

TABLE IV
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON THE ADNI (NC VS. MCI) DATASET

Method	Acc	Sen	Spe	AUC	F1
PolulationGCN [16]	67.41 (0.82)	75.96 (1.33)	57.00 (1.41)	62.34 (1.58)	71.65 (0.76)
InceptionGCN [17]	71.04 (1.12)	72.69 (2.66)	69.00 (1.29)	64.07 (1.50)	72.58 (1.57)
EV-GCN [18]	71.06 (0.81)	**78.71 (1.91)**	62.00 (6.36)	66.19 (1.49)	74.82 (0.54)
Hi-GCN [23]	71.08 (0.26)	76.85 (2.47)	64.00 (3.84)	65.83 (0.38)	73.90 (0.29)
LG-GNN (Ours)	**76.48 (0.44)**	77.88 (1.20)	**75.00 (2.45)**	**68.85 (1.33)**	**78.23 (0.37)**

TABLE V
CLASSIFICATION RESULTS OF DIFFERENT METHODS ON THE ADNI (PMCI VS. SMCI) DATASET

Method	Acc	Sen	Spe	AUC	F1
PolulationGCN [16]	76.02 (0.62)	81.25 (2.27)	**66.00 (3.79)**	65.06 (2.00)	81.19 (0.50)
InceptionGCN [17]	73.46 (6.67)	81.25 (8.20)	58.00 (5.26)	66.31 (7.01)	77.85 (7.13)
EV-GCN [18]	63.65 (6.25)	67.50 (9.13)	56.50 (5.30)	54.18 (5.50)	67.94 (7.18)
Hi-GCN [23]	70.12 (0.76)	78.75 (1.58)	53.50 (0.60)	63.31 (1.72)	77.33 (0.59)
LG-GNN (Ours)	**79.23 (0.58)**	**87.50 (2.19)**	63.00 (2.96)	**67.00 (2.56)**	**84.37 (1.16)**

Hi-GCN, which may cause suboptimal classification performance.

B. Ablation Studies

Our LG-GNN is mainly composed of local ROI-GNN and global Subject-GNN components. In order to verify the effectiveness of the two components, we conducted ablation studies on both ABIDE and ADNI.

1) Ablation Study for Local ROI-GNN: In order to investigate the effectiveness of the local ROI-GNN, we compare it with several state-of-the-art GNN models including GAT [42], GIN [44], GraphSAGE [45], ChebNet [30] and GCN [25]. In the experiment, we replace the local ROI-GNN with the above models in the LG-GNN framework. Table VI shows the comparative results. It can be seen that for both ADIDE and ADNI, our local ROI-GNN achieves the highest classification performance in terms of all metrics, which indicates that the local ROI-GNN best facilitates the following global Subject-GNN in the LG-GNN framework for the classification task. In particular, the local ROI-GNN outperforms GCN [25] equivalent with the local ROI-GNN without the SABP module, which demonstrates that the proposed SABP module is able to extract an optimal feature embedding of the regional brain graph. Specifically, the pooling ratio used in the SABP module is set to 0.9, which means that 90% of the nodes input to the local ROI-GNN are retained. To evaluate the performance of SABP under different pooling ratios, we conduct experiments on both ABIDE and ADNI. The number of nodes in a coarsened graph has a significant impact on classification performance, and therefore the level of coarsening in a brain network is an important factor in graph structure learning. Specifically, the topological information of the coarsened graph is condensed when the pooling ratio is smaller, and this information is enriched when the pooling ratio is larger. Brain networks with smaller pooling ratios are less connected but have more precise topological relationships between brain regions. However, some potentially important connections may inevitably be lost. In contrast, brain networks with larger pooling ratios may contain a large number of noisy connections, resulting in poorer graph embedding of the brain network. As shown in Fig. 5, the scores of classification accuracy are computed based on different pooling ratios. When the pooling ratio is 0.9, the proposed method achieves the

TABLE VI
RESULTS OF REPLACING THE LOCAL ROI-GNN WITH OTHER MODELS IN LG-GNN ON ABIDE AND ADNI DATASETS

Model	ABIDE				ADNI			
	Acc	Sen	AUC	F1	Acc	Sen	AUC	F1
GAT [42]	79.56 (1.02)	80.35 (1.23)	81.47 (0.60)	81.31 (1.35)	77.61 (1.60)	80.83 (1.06)	57.75 (3.88)	86.15 (1.26)
GIN [44]	71.98 (2.34)	77.76 (2.01)	76.15 (1.73)	72.44 (1.81)	79.59 (1.03)	81.03 (0.80)	62.25 (3.22)	87.07 (1.35)
GraphSAGE [45]	77.26 (0.84)	80.99 (1.70)	79.40 (1.14)	78.37 (1.23)	80.83 (1.37)	80.52 (1.83)	64.50 (3.54)	88.79 (1.66)
ChebNet [30]	76.80 (2.32)	80.09 (2.01)	75.24 (1.73)	79.17 (1.65)	80.34 (2.12)	80.09 (1.11)	58.16 (4.87)	87.95 (1.67)
GCN [25]	80.25 (1.15)	80.88 (2.41)	82.04 (1.44)	81.81 (1.05)	79.10 (1.71)	81.02 (1.06)	65.00 (4.26)	87.47 (1.24)
Local ROI-GNN	**81.75 (1.10)**	**83.22 (1.84)**	**85.22 (1.01)**	**82.96 (0.94)**	**82.09 (1.40)**	**82.96 (0.91)**	**68.33 (3.47)**	**88.91 (1.12)**

TABLE VII
RESULTS OF REPLACING THE GLOBAL SUBJECT-GNN WITH OTHER MODELS IN LG-GNN ON ABIDE AND ADNI DATASETS

Model	ABIDE				ADNI			
	Acc	Sen	AUC	F1	Acc	Sen	AUC	F1
GATConv [42]	80.48 (1.39)	80.05 (1.76)	83.04 (1.77)	82.74 (1.56)	76.86 (2.07)	78.89 (1.23)	61.11 (3.99)	86.37 (1.43)
GINConv [44]	79.79 (1.23)	81.03 (1.41)	81.79 (0.90)	80.06 (1.02)	79.10 (1.27)	80.13 (1.66)	60.33 (4.18)	87.18 (0.90)
GraphSAGEConv [45]	78.30 (1.67)	80.50 (1.62)	79.61 (1.72)	80.78 (1.66)	78.35 (1.56)	82.02 (0.75)	62.16 (4.67)	86.33 (1.69)
GCNConv [25]	80.59 (1.33)	82.02 (1.78)	81.82 (1.05)	82.32 (1.38)	80.59 (1.44)	81.55 (1.60)	60.66 (3.41)	87.91 (1.77)
LG-GNN w/o AWAB	80.36 (1.22)	82.70 (1.64)	82.51 (2.03)	81.11 (1.42)	81.34 (1.49)	81.00 (1.20)	64.66 (3.92)	88.73 (1.38)
LG-GNN (Ours)	**81.75 (1.10)**	**83.22 (1.84)**	**85.22 (1.01)**	**82.96 (0.94)**	**82.09 (1.40)**	**82.96 (0.91)**	**68.33 (3.47)**	**88.91 (1.12)**

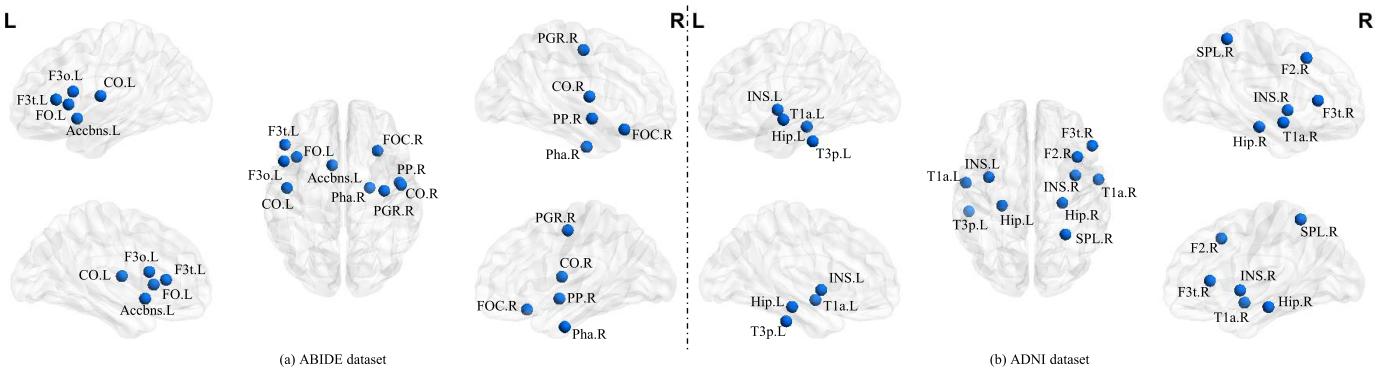

Fig. 6. 10 brain regions most important for ASD classification on ABIDE (a) and AD classification on ADNI (b) according to the learned weights output by the SABP module of the LG-GNN.

best performance on both ABIDE and ADNI. As such, the potentially important connections are retained and the noisy connections are mostly removed.

2) Ablation Study for Global Subject-GNN: To evaluate the effectiveness of the ChebConv operator in the global Subject-GNN, we replace it with various graph convolution operators in the LG-GNN framework, including GATConv [42], GINConv [44], GraphSAGEConv [45] and GCNConv [25]. Except for the graph convolution operator, all other experimental settings are the same. The comparative results in Table VII show that the proposed global Subject-GNN achieves the best classification performance in terms of all metrics on both ABIDE and ADNI. Such results demonstrate the ability of the ChebConv operator to incorporate the local individual embeddings of the regional brain graphs into the global structure.

Furthermore, to demonstrate the effectiveness of the AWAB, we attempt to directly use the output of the last Cheb block as the input of the final fully-connected layer, so as to avoid using AWAB. As shown in Table VII, the classification accuracy of the scheme without AWAB is lower than that of the global Subject-GNN, which constitutes the proof that AWAB benefits the modeling of the hierarchical and contextual information within the global subject graph.

C. Biomarker Detection

To better understand which brain regions contribute most to the ASD and the AD diagnosis, we investigate the model weights output by the SABP module, which are important for delivering discriminative features. The basis for guaranteeing that the top-k important brain regions are reasonable is stated as follows. First, we obtain the weight of each node of the brain graph which represents a brain ROI through the SABP module of the local ROI-GNN on each test set. Then we statistically analyze the top 10 brain regions with the largest weights, and the top 10 brain regions with the highest frequency are selected as biomarkers. we empirically set $k = 10$ as a safe choice so that the brain regions corresponding biomarkers can all be included. To guarantee the accuracy of the selected 10 nodes, we propose a mutual information loss $\mathcal{L}_{MI}$ to ensure that the weights of the k selected ROIs differ

TABLE VIII
INFLUENCE OF NON-IMAGING INFORMATION ON THE CLASSIFICATION RESULTS ON ABIDE DATASET

ABIDE		
Method	Acc	AUC
LG-GNN w/o site	68.16 (1.21)	66.19 (2.12)
LG-GNN w/o gender	80.90 (1.50)	83.97 (1.61)
LG-GNN	81.75 (1.10)	85.22 (1.01)

TABLE IX
INFLUENCE OF NON-IMAGING INFORMATION ON THE CLASSIFICATION RESULTS ON ADNI DATASET

ADNI		
Method	Acc	AUC
LG-GNN w/o age	81.63 (0.76)	67.58 (3.51)
LG-GNN w/o gender	80.65 (0.33)	54.75 (2.34)
LG-GNN	82.09 (0.40)	68.33 (3.47)

significantly from those of the unselected nodes. The weights of the selected ROIs are expected to be close to 1, while the weights of the unselected ROIs are supposed to be close to 0.

The 10 brain regions most important for ASD and AD classification identified by the LG-GNN are shown in Fig. 6, generated by using the BrainNet Viewer toolbox [46]. For ASD classidication, the 10 brain regions with the largest weights are mainly Inferior Frontal Gyrus, Precentral Gyrus, Frontal Orbital Cortex, Parahippocampal Gyrus, Frontal Operculum Cortex, Central Opercular Cortex, Planum Polare, Accumbens, etc. These brain regions are related to the neurological functions of social communication, perception and execution [47], [48], which are clearly deficient in ASD.

It is worth mentioning that the Parahippocampal Gyrus area discovered by our model was also highlighted by a recent study [49]. For AD classification, the 10 brain regions with the largest weights are mainly hippocampus, Superior Parietal Lobule, Temporal Gyrus, Inferior Frontal Gyrus, Insular Cortex, etc., which have been reported to be responsible for short-term memory and early stages of AD [50], [51].

D. Discussion

In the proposed method, we have introduced different non-imaging information to construct the brain graph. For the ABIDE dataset, we use site and gender as the measures. For the ADNI dataset, we use age and gender as the measures. In order to verify how non-imaging information contributes to diagnosis, we conduct ablation studies on ABIDE dataset and ADNI dataset. The results are reported in Table VIII and IX. For the ABIDE dataset, the experimental results with only site information are significantly higher than those with only gender information, which indicates that ABIDE is highly heterogeneous with data acquired at different sites using diverse imaging protocols and scanners and site information can be considered the most important factor. After adding gender information to the site information, the experimental results were improved, indicating that gender information is helpful for ASD classification. For the ADNI dataset, the experimental results with only gender information are higher than those with only age information, and those using both age and gender information are better than those using only one type of information. Although our LG-GNN has achieved promising results by using site, gender, and age non-imaging information, our model may be improved by using other non-imaging information, such as genetic information, intelligence quotient information et al. We will explore other non-imaging information in the future work.

VI. CONCLUSION

To deliver a reliable detection of the biomarkers and an accurate classification of ASD and AD diseases, this paper proposes a new GNN framework, namely LG-GNN, for classifying brain disorders. The LG-GNN is mainly composed of two GNNs: 1) a local ROI-GNN based on a regional brain graph which learns and analyzes the features of the local brain regions to identify biomarkers and 2) a global Subject-GNN which encodes non-imaging information to learn the relationship between multiple subjects with the feature embeddings generated by the local ROI-GNN. The local ROI-GNN contains a particularly designed SABP module to enable a good combination of the two GNNs. The experimental results over two public datasets demonstrate that the LG-GNN that takes advantage of the two types of GNNs achieves the state-of-the-art performance. The promising results imply that the proposed method has great potential in the application of computer-aided diagnosis based on medical images and the automatic interpretation of biological images.

REFERENCES

[1] N. Yahata et al., "A small number of abnormal brain connections predicts adult autism spectrum disorder," *Nature Commun.*, vol. 7, no. 1, pp. 1–12, 2016.
[2] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun, "Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls," *NeuroImage*, vol. 145, pp. 137–165, Mar. 2016.
[3] R. S. Desikan et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, 2006.
[4] D. Yao et al., "A mutual multi-scale triplet graph convolutional network for classification of brain disorders using functional or structural connectivity," *IEEE Trans. Med. Imag.*, vol. 40, no. 4, pp. 1279–1289, Apr. 2021.
[5] C.-Y. Wee et al., "Identification of MCI individuals using structural and functional connectivity networks," *NeuroImage*, vol. 59, no. 3, pp. 2045–2056, 2012.
[6] X. Tang, Y. Qin, J. Wu, M. Zhang, W. Zhu, and M. I. Miller, "Shape and diffusion tensor imaging based integrative analysis of the hippocampus and the amygdala in Alzheimer's disease," *Magn. Reson. Imag.*, vol. 34, no. 8, pp. 1087–1099, Oct. 2016.
[7] A. Abraham et al., "Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example," *NeuroImage*, vol. 147, pp. 736–745, Feb. 2017.
[8] S. H. Lee, D. Yu, A. H. Bachman, J. Lim, and B. A. Ardekani, "Application of fused lasso logistic regression to the study of corpus callosum thickness in early Alzheimer's disease," *J. Neurosci. Methods*, vol. 221, pp. 78–84, Jan. 2014.
[9] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson, "Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 30–43, Jan. 2010.
[10] H. Zhang et al., "Cerebrovascular segmentation in MRA via reverse edge attention network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 66–75.
[11] L. Xia et al., "A nested parallel multiscale convolution for cerebrovascular segmentation," *Med. Phys.*, vol. 48, no. 12, pp. 7971–7983, 2021.

[12] Y. Liu, L. He, B. Cao, S. Y. Philip, A. B. Ragin, and A. D. Leow, "Multi-view multi-graph embedding for brain network clustering analysis," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.
[13] X. Xing et al., "Dynamic spectral graph convolution networks with assistant task training for early MCI diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 639–646.
[14] S. I. Ktena et al., "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, Apr. 2018.
[15] X. Li et al., "BrainGNN: Interpretable brain graph neural network for fMRI analysis," *Med. Image Anal.*, vol. 74, Dec. 2021, Art. no. 102233.
[16] S. Parisot et al., "Spectral graph convolutions for population-based disease prediction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 177–185.
[17] A. Kazi et al., "InceptionGCN: Receptive field aware graph convolutional network for disease prediction," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2019, pp. 73–85.
[18] Y. Huang and A. C. Chung, "Edge-variational graph convolutional networks for uncertainty-aware disease prediction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 562–572.
[19] M. Plitt, K. A. Barnes, and A. Martin, "Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards," *NeuroImage, Clin.*, vol. 7, pp. 359–366, Jan. 2015.
[20] G. Wen, P. Cao, H. Bao, W. Yang, T. Zheng, and O. Zaiane, "MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis," *Comput. Biol. Med.*, vol. 142, Mar. 2022, Art. no. 105239.
[21] M. Lostar and I. Rekik, "Deep hypergraph U-Net for brain graph embedding and classification," 2020, *arXiv:2008.13118*.
[22] 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.* Cham, Switzerland: Springer, 2020, pp. 124–133.
[23] 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.
[24] L. Li et al., "TE-HI-GCN: An ensemble of transfer hierarchical graph convolutional networks for disorder diagnosis," *Neuroinformatics*, vol. 20, pp. 353–375, Nov. 2021.
[25] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
[26] H. Xue, X.-K. Sun, and W.-X. Sun, "Multi-hop hierarchical graph neural networks," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2020, pp. 82–89.
[27] C. Seguin, M. P. van den Heuvel, and A. Zalesky, "Navigation of brain networks," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 24, pp. 6297–6302, Jun. 2018.
[28] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3734–3743.
[29] 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.*, 2020, vol. 34, no. 4, pp. 3438–3445.
[30] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.
[31] C. Craddock et al., "Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (C-PAC)," *Front Neuroinform*, vol. 42, p. 3389, Jul. 2013.
[32] J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38, no. 1, pp. 95–113, 2007.
[33] K. Dadi et al., "Benchmarking functional connectome-based predictive models for resting-state fMRI," *NeuroImage*, vol. 192, pp. 115–134, May 2019.
[34] J. Wang, X. Wang, M. Xia, X. Liao, A. Evans, and Y. He, "GRETNA: A graph theoretical network analysis toolbox for imaging connectomics," *Frontiers Hum. Neurosci.*, vol. 9, p. 386, Jun. 2015.
[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[36] D. M. Werling and D. H. Geschwind, "Sex differences in autism spectrum disorders," *Current Opinion Neurol.*, vol. 26, no. 2, p. 146, 2013.
[37] R. K. Kana, L. Q. Uddin, T. Kenet, D. Chugani, and R.-A. Müller, "Brain connectivity in autism," *Frontiers Hum. Neurosci.*, vol. 8, p. 349, Jun. 2014.
[38] J. L. Podcasy and C. N. Epperson, "Considering sex and gender in Alzheimer disease and other dementias," *Dialogues Clin. Neurosci.*, vol. 18, no. 4, pp. 437–446, Dec. 2016.
[39] V. Vapnik, "The support vector method of function estimation," in *Nonlinear Modeling*. Boston, MA, USA: Springer, 1998, pp. 55–85.
[40] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, vol. 1, Aug. 1995, pp. 278–282.
[41] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," *NeuroImage, Clin.*, vol. 17, pp. 16–23, Jan. 2018.
[42] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *Stat*, vol. 1050, p. 20, Oct. 2017.
[43] F. Pedregosa, S. Varoquaux, A. Gramfort, V. Michel, and B. Thirion, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Dec. 2011.
[44] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" 2018, *arXiv:1810.00826*.
[45] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[46] M. Xia, J. Wang, and Y. He, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS ONE*, vol. 8, no. 7, Jul. 2013, Art. no. e68910.
[47] D. Li, H.-O. Karnath, and X. Xu, "Candidate biomarkers in children with autism spectrum disorder: A review of MRI studies," *Neurosci. Bull.*, vol. 33, no. 2, pp. 219–237, Apr. 2017.
[48] C. Press, N. Weiskopf, and J. M. Kilner, "Dissociable roles of human inferior frontal gyrus during action execution and observation," *NeuroImage*, vol. 60, no. 3, pp. 1671–1677, Apr. 2012.
[49] L. E. Libero, M. Schaer, D. D. Li, D. G. Amaral, and C. W. Nordahl, "A longitudinal study of local gyrification index in young boys with autism spectrum disorder," *Cerebral Cortex*, vol. 29, no. 6, pp. 2575–2587, Jun. 2019.
[50] J. J. Gaugler, "2017 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 13, no. 4, pp. 325–373, 2017.
[51] M. Laakso et al., "Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia: An MRI study," *Neurology*, vol. 46, no. 3, pp. 678–681, 1996.