

fMRI-Based Brain Disease Diagnosis: A Graph Network Approach

Wei Tong¹, Yong-Xia Li, Xiao-Yan Zhao, Qi-Qi Chen, Yu-Bing Gao, Ping Li², *Member, IEEE*,
and Edmond Q. Wu³, *Senior Member, IEEE*

Abstract—As a non-invasive brain imaging technology, functional magnetic resonance imaging (fMRI) provides a basic tool for brain functional network modeling and brain disease diagnosis. Problems, such as large number of parameters, low training efficiency, and poor interpretability, are encountered in mainstream models because of the high complexity of fMRI and brain networks. To solve these problems, a novel structure feature combined graph neural network (SFC-GNN) with a low number of parameters is proposed. In particular, SFC-GNN is composed of 1) the graph convolution layer of the brain region perception and 2) the node pooling layer of the graph structure feature (GSF). It also receives the sparse brain graph modeled by each subject's fMRI as input. Especially, the GSF layer can select brain regions that are important for classification, thereby localizing all active regions related to brain disease. Moreover, a group network is constructed according to the correlation among subjects, and SFC-GNN can be extended further to a node classification model to achieve better diagnosis performance. The proposed method has been validated on the ABIDE and ADNI datasets, thereby showing the effectiveness of our proposed method in various experiments.

Index Terms—Brain functional network, functional magnetic resonance imaging, brain disease diagnosis, graph-based model.

I. INTRODUCTION

AS THE most complex organ of the human body, the brain has highly complex structural and functional characteristics. Many synapses transmit the excitation or inhibition generated by neurons to each other to coordinate each region of the brain, thereby forming a dynamic brain functional network. The research of brain functional network aims to

Manuscript received 10 January 2023; revised 8 March 2023; accepted 2 April 2023. Date of publication 26 April 2023; date of current version 18 May 2023. This article was recommended for publication by Associate Editor A. Nguyen and Editor P. Dario upon evaluation of the reviewers' comments. This work was supported by the National Natural Science Foundation of China under Grant 62171274, Grant 62273019, and Grant U2241228. (Wei Tong, Yong-Xia Li, and Qi-Qi Chen contributed equally to this work.) (Corresponding authors: Ping Li; Edmond Q. Wu.)

Wei Tong is with the School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China (e-mail: tongwei@njust.edu.cn).

Yong-Xia Li is with the Department of Emergency, Shanghai Sixth People's Hospital, Shanghai 200000, China (e-mail: liyongxia2021@163.com).

Xiao-Yan Zhao is with the Rehabilitation Medicine, Anhui No.2 Provincial People's Hospital, Hefei 230041, China (e-mail: zxy364@163.com).

Qi-Qi Chen, Yu-Bing Gao, and Edmond Q. Wu are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: cqjx@alumni.sjtu.edu.cn; gaoyubing@sjtu.edu.cn; edmondqwu@sjtu.edu.cn).

Ping Li is with the School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China (e-mail: liping@buaa.edu.cn).

Digital Object Identifier 10.1109/TMRB.2023.3270481

understand the working mechanism of the brain and the principles of brain diseases to protect the brain from the occurrence of various brain diseases. Previous studies [1], [2] have shown a certain correlation between brain functional networks and most neurological diseases, so the changes in topology and connection in the brain functional network can provide a reference for the diagnosis of brain diseases. Modern brain imaging technology provides the basis for brain network modeling. Among brain imaging technologies, functional magnetic resonance imaging (fMRI), as a representative non-invasive technology, has been widely used. fMRI has a good spatial resolution and can provide quantitative analysis of brain regions. Therefore, it becomes a useful tool for the modeling of brain functional networks.

Hemoglobin in human blood exhibits different magnetic properties under various oxygen contents. Once the oxygenated hemoglobin content increases, the magnetic resonance signal also rises. This phenomenon is called the blood oxygen level-dependent (BOLD) effect. When performing a specific task, oxygenated blood flows into the brain region responsible for that function and causes changes in the magnetic resonance signal. fMRI can detect this kind of change in magnetic resonance signal [3], which enables the visualization of brain metabolism [1], [4]. Resting-state fMRI refers to the fMRI data obtained in a quiet and resting-state of a subject without performing any specific tasks. It has a consistent and stable functional pattern [5] and can reflect the general state of the non-specific brain and has relatively clear physiological and pathological significance. Resting-state fMRI has been widely used in clinical studies, such as in cognitive neuroscience and clinical psychiatry, including disease monitoring and treatment [2], pharmacological efficacy research [6], and biomarker discovery [7]. Therefore, the resting-state fMRI is adopted to conduct brain functional network construction and disease diagnosis.

The main contributions of this paper are presented as follows:

1) Based on the graph convolution layer of brain functional region perception and the graph pooling layer of structure feature selection, an interpretable graph classification model called structure feature combined graph neural network (SFC-GNN) is designed for brain disease diagnosis. The network takes sparse brain graph as input and designs a special pooling layer to select nodes that are important for diagnosis tasks, which finds the brain regions related to disease and reduces computational time.

2) Based on the brain functional graph embedding generated by SFC-GNN, a node classification model with a group network is further designed to study the correlations among different subjects. From the experiment, our method demonstrates high advantages over other tested methods.

3) In addition, transfer learning from autism to Alzheimer's disease is also conducted to further improve the diagnostic accuracy of Alzheimer's disease and reveal the potential correlation between the two brain diseases.

The remainder of this paper is organized as follows. Section II introduces the related works for brain disease diagnosis. Section III describes the details of the proposed sparse brain functional graph, SFC-GNN, and group network. Section IV conducts extensive experiments to illustrate the effectiveness of our method. Section V summarizes this paper.

II. RELATED WORK

Many studies have used brain functional networks instead of original fMRI data to infer brain cognition and neural development or to predict and diagnose brain damage and diseases.

A. Machine Learning-Based Method

Many of the earlier brain network analysis tools were developed with machine learning models. Rosenberg et al. [8] developed a simple linear model to demonstrate that the brain functional network can provide a broadly applicable neural marker for the symptoms of attention deficit hyperactivity disorder. Ball et al. [9] adopted a random forest (RF)-based feature selection method to identify the discriminative edges of the neonatal brain functional connectivity network and then used a nonlinear support vector machine (SVM) to classify premature and term infants. Challis et al. [10] designed a Bayesian Gaussian process logistic regression model for the diagnosis of Alzheimer's disease. However, these types of disease diagnosis methods have poor model fitting and generalization ability, which is difficult to achieve ideal results in practical application.

B. Deep Learning-Based Method

Many neural network-based fMRI analysis methods have been proposed with the development of deep learning. Considering that fMRI data are embedded in the spatiotemporal dimension, recurrent neural networks are suitable for fMRI analysis. Cui et al. [11] designed a new deep recurrent network framework, which achieved better results than those of shallow models, to identify brain functional networks with multiple time scales. Dvornek et al. [12] used the long short-term memory (LSTM) structure to design a multi-task learning framework to identify autistic patients, generate meaningful functional communities, and improve the interpretability of the model. Kawahara et al. [13] proposed BrainNetCNN for the prediction of the clinical neurodevelopment of brain networks by utilizing the topology of brain networks and achieved superior performance over other methods.

C. Graph Network-Based Method

Graph network-based methods *i.e.*, graph convolution network (GCN) have unique advantages in describing the functional characteristics of the brain, which can consider the important topological properties of brain networks. Kazi et al. [14] and Parisot et al. [15] considered each subject as a node in the graph and integrated phenotypic information into edge weights to predict brain diseases. Li et al. [16] developed an interpretable GNN for fMRI-based biomarker analysis, which could determine the biomarkers that corresponded to specific tasks. BrainGNN [17] adopted a ROI-aware graph convolution kernel to extract the functional and topological information of fMRI for simultaneous learning and achieved higher classification accuracy than traditional machine learning-based and convolutional neural network-based methods. Gadgil et al. [18] proposed a graph convolution model to combine spatiotemporal information for predicting the genders and ages of adolescents and achieved higher classification accuracy than recurrent neural networks. Yao et al. [19] designed a multi-spatial-scale triple graph convolutional network to analyze brain function and structural connection and achieved better results than the single-scale GNN in the diagnosis of Alzheimer's disease. Zhang and Wang [20] proposed a graph isomorphic network for autism diagnosis, which also achieved high classification accuracy. Li et al. [21] introduced transfer learning into the diagnosis task of Alzheimer's disease and autism and proposed an integrated framework to combine hierarchical GCN and brain transfer learning to improve the diagnostic accuracy of diseases.

D. Novelty of the Work

Our SFC-GNN is designed according to the idea of BrainGNN [17], in which each subject is constructed as a graph and adopted as input to the graph classification model. However, different from BrainGNN, the designed graph pooling layer can combine the graph structure and node features to obtain better performance of important nodes. Besides, SFC-GNN is extended to a node classification model by constructing a group network, which can further improve the performance.

III. PROPOSED METHOD

A. Definition of Brain Functional Graph

Assuming that the whole brain can be divided into N brain functional regions, an undirected weighted brain functional graph can be defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The nodes in the graph represent the set of predefined brain functional regions (*i.e.*, $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$), and the edges in the graph represent the set of connectivity among brain functional regions. If $(v_i, v_j) \in \mathcal{E}$, then the two brain functional regions i and j , which corresponds to the weight of an edge $e_{ij} \in \mathbb{R}^+$ and $e_{ij} = e_{ji}$, have a connection, respectively; if $(v_i, v_j) \notin \mathcal{E}$, then $e_{ij} = 0$. Thus, an adjacency matrix $E = [e_{ij}] \in \mathbb{R}^{N \times N}$ can be obtained to represent the connection among brain functional regions. In addition, for i -th node in the graph, \mathbf{h}_i is defined as its feature vector.

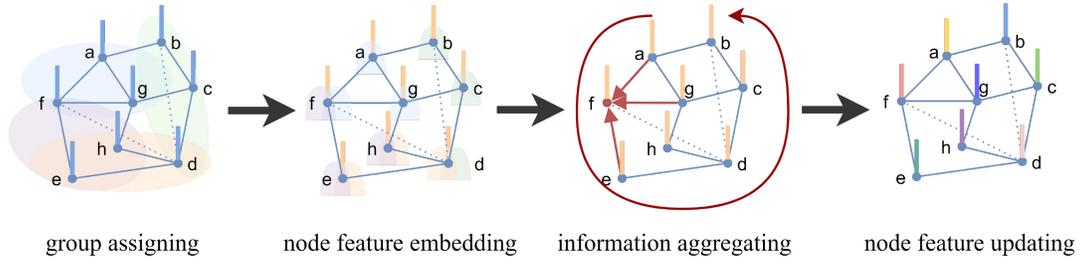


Fig. 1. Brain region perception-based graph convolution layer: RaGCN.

Specific to the construction of brain functional graph, the fMRI image in each moment is divided into N brain functional regions, and the average of all voxel values for each brain functional region is calculated to represent the brain functional region. Therefore, a corresponding average time series of BOLD signals can be obtained from each brain functional region, and a total of N average time series can be obtained from the images of the entire brain. The partial correlation coefficient matrix is used to construct the edge and adjacency matrix of the brain functional graph, and a threshold proportion is set to construct the sparse matrix. That is, the partial correlation coefficient values between pairs of brain functional regions are sorted. The brain functional regions with the highest partial correlation coefficient values are considered connected, that is, they correspond to the undirected edges in the brain functional graph. Other brain functional regions with low partial correlation coefficient values are considered to have no connection, and the corresponding edge weights are set to zero. The diagonal elements of the partial correlation matrix are all 1, but in the brain functional graph, no edge connection exists between the node with itself, that is, $(v_i, v_i) \notin \mathcal{E}$, so $e_{ii} = 0$. Thus, a symmetric sparse adjacency matrix E with all zero diagonal elements is finally obtained.

Four features are considered for the node feature \mathcal{H} in the brain functional graph: the mean and standard deviation of the BOLD signal that correspond to each node, the degree of the node, and the correlation coefficient between the nodes. Features are spliced to form the feature vector \mathbf{h}_i of the current node i , we have $\mathbf{h}_i \in \mathbb{R}^{3+N}$. Therefore, the corresponding brain functional graph of fMRI is composed of the adjacency matrix E and the node feature vector set \mathcal{H} .

B. Graph Classification-Based Disease Diagnosis Model

A graph classification-based network model of the brain functional graph is proposed. First, the graph nodes are embedded in a low-dimensional feature space and divided into different communities. Second, the state transfer and information aggregation between nodes are performed, and some nodes with larger weights are retained by node selection, which reduces calculation and improves interpretability. Finally, the information extracted by the model is aggregated into a vector and inputted into the classifier to achieve end-to-end graph classification. The proposed graph classification-based network model consists of three different network layers: graph convolution layer, node pooling layer, and readout layer. The details of these layers are discussed as follows.

1) *Graph Convolution Layer*: Message propagation mechanism-based graph network is adopted in the design of the graph convolution layer. The message transmission process can be represented by two sub-functions, namely, information aggregation function, which aggregates the information of the current node's neighbor nodes and combines them into an information vector to be transmitted to the central node; and node update function, which combines the information vector and the current node features to update the central node features [22], [23].

The structure of the graph convolution layer for brain functional region perception is shown in Fig. 1. Specifically, assuming that the nodes in the original brain graph can be divided into R different communities and the nodes from the same community have similar attributes, the brain graph contains community information in addition to the original information, and the node's community information can be mapped by its position information [17]. One-hot coding is used to represent the position information \mathbf{p}_i of each node. Thus, \mathbf{p}_i of all nodes is a N -dimensional vector with only one element of 1 and other elements of 0. The position with a value of 1 indicates the order of this node in the entire brain functional graph, and \mathbf{p}_i is the same for the same node in different brain functional graphs.

To learn the community information of nodes through position information, a two-layer perceptron is utilized to learn the kernel of feature embedding. First, the position information of N nodes in the brain functional graph is mapped to R different communities through the first layer of the perceptron. Second, the community information is mapped to the kernel embedding vector through the second layer of the perceptron, which is finally mapped to a corresponding weight matrix \mathbf{W}_i .

The two parameter matrices of the two-layer perceptron in the layer k of graph convolution are defined as $\mathbf{b}_1^{(k)}$ and $\mathbf{b}_2^{(k)}$. Therefore, we yield

$$\begin{aligned}
 \mathbf{b}_1^{(k)} &= [\alpha_1^{(k)}, \alpha_2^{(k)}, \dots, \alpha_{N^{(k)}}^{(k)}] \in \mathbb{R}^{R^{(k)} \times N^{(k)}}, \\
 \alpha_i^{(k)} &= [\alpha_{i1}^{(k)}, \alpha_{i2}^{(k)}, \dots, \alpha_{iR^{(k)}}^{(k)}]^T, \forall i \in \{1, 2, \dots, N^{(k)}\}, \\
 \mathbf{b}_2^{(k)} &= [\beta_1^{(k)}, \beta_2^{(k)}, \dots, \beta_{R^{(k)}}^{(k)}] \in \mathbb{R}^{(d^{(k+1)} \cdot d^{(k)}) \times R^{(k)}}, \\
 \beta_j^{(k)} &= [\beta_{j1}^{(k)}, \beta_{j2}^{(k)}, \dots, \beta_{j(d^{(k+1)} \cdot d^{(k)})}^{(k)}]^T, \forall j \in \{1, 2, \dots, R^{(k)}\}.
 \end{aligned} \tag{1}$$

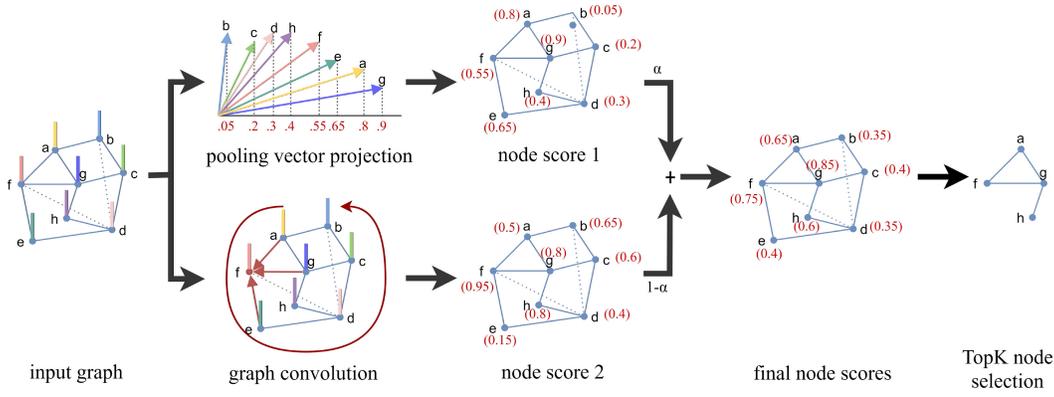


Fig. 2. The detailed explanation of GSF pooling layer based on structure and feature information.

Therefore, the two-layer perceptron can be expressed as

$$\begin{aligned}
 \text{vec}(\mathbf{W}_i^{(k)}) &= f_{MLP^{(k)}}(\mathbf{p}_i) \\
 &= \mathbf{b}_2^{(k)} \phi(\mathbf{b}_1^{(k)} \mathbf{r}_i) + \mathbf{b}_0^{(k)} \\
 &= \sum_{j=1}^{R^{(k)}} \phi(\alpha_{ij}^{(k)}) \beta_j^{(k)} + \mathbf{b}_0^{(k)}, \quad (2)
 \end{aligned}$$

where \mathbf{b}_0 is the bias, and ϕ is the activation function. PReLU [24] is adopted as the activation function of multilayer perceptron to improve model adaptability.

Therefore, the dimension of the weight vector of feature embedding obtained by the two-layer perceptron is $d^{(k+1)}d^{(k)}$, and it is mapped to obtain the weight matrix $\mathbf{W}_i^{(k)} \in \mathbb{R}^{d^{(k+1)} \times d^{(k)}}$.

The weight of the edge that connects two nodes represents the strength of connectivity between the two brain functional regions because the partial correlation matrix is introduced as the adjacency matrix E of the brain functional graph. This paper holds that the neighbor nodes with stronger connections should have a greater impact on the message propagation process of the graph convolution layer. Thus, the node feature vector should be multiplied by the edge weight during the calculation of the aggregation function. Due to the information aggregation mechanism, to control the parameter values obtained after the aggregation calculation of all neighbor nodes to a certain order of magnitude, the weights of the edges must be normalized [25]. In addition, a constant amplification ratio for the current node information is set in the node update process to balance the weight of the node's own information and the neighbor node information in the combined update. Finally, the graph convolution layer based on the message transmission mechanism can be embodied as

$$\begin{aligned}
 \mathbf{h}_i^{(k+1)} &= \phi \left((1 + \gamma^{(k)}) \mathbf{W}_i^{(k)} \mathbf{h}_i^{(k)} + \sum_{j \in \mathcal{N}(v_i)} \varphi(\mathbf{W}_j^{(k)} \mathbf{h}_j^{(k)}, e_{ij}) \right) \\
 &= \phi \left((1 + \gamma^{(k)}) \mathbf{W}_i^{(k)} \mathbf{h}_i^{(k)} + \frac{\sum_{j \in \mathcal{N}(v_i)} e_{ij} \mathbf{W}_j^{(k)} \mathbf{h}_j^{(k)}}{\sum_{j \in \mathcal{N}(v_i)} e_{ij}} \right), \quad (3)
 \end{aligned}$$

where ϕ adopts PReLU function, e_{ij} is the weight of the edge that connects the two nodes v_i and v_j in the adjacency matrix E ,

\mathbf{W} is the feature embedding to be learned in the network training process, and γ is a constant.

2) *Graph Pooling Layer*: Considering that the dimensions of nodes and features in the original graph are large, a node pooling layer between the graph convolution layers must be introduced to obtain a subgraph with fewer nodes and features [26]. Graph pooling, which can be roughly summarized as static pooling [27], hierarchical clustering pooling [28], [29], [30], [31], and node selection pooling [32], [33], [34]. Usually, only several key brain functional regions play a key role in fMRI analysis and disease diagnosis (*i.e.*, brain functional activation regions). Following this idea, the node selection pooling method is adopted, and the idea of self-attention graph pooling [33] is utilized to pool the brain functional graph. The graph convolution method is used to fuse the graph structure information to calculate the node score, and it is improved by combining the calculation method of the projection vector. Finally, the top-ranked nodes are screened by the Top-K method [32]. The improved graph pooling operation can consider the structure and feature information of the graph simultaneously. It adopts different node evaluation methods and utilizes the features of unselected nodes. As such, a new graph structure-feature (GSF)-based graph pooling layer is constructed. This graph pooling method can be applied to graphs with different sizes and structures and has a good node selection performance. The detailed structure is shown in Fig. 2.

Specifically, two-node importance evaluation methods, namely feature-based and structure-based node selection, are used in the graph pooling layer. Most of the feature information in the graph is contained in the feature vector of each node, so the score of the graph feature is calculated on the basis of these feature vectors. The formula for calculating the node score is expressed as

$$\mathbf{s}_1^{(k)} = \frac{1}{\|\mathbf{m}^{(k)}\|} \phi(H^{(k)} \mathbf{m}^{(k)}), \quad (4)$$

where $H = [\mathbf{h}_1^{(k)}, \mathbf{h}_2^{(k)}, \dots, \mathbf{h}_{N^{(k)}}^{(k)}]^T$ is the feature matrix formed by node feature vectors of the k -th layer, $\mathbf{m}^{(k)} \in \mathbb{R}^{d^{(k)}}$ is a mapping vector to be learned, $d^{(k)}$ is the dimension of the feature, $\|\cdot\|$ represents L2 regularization, ϕ is the activation

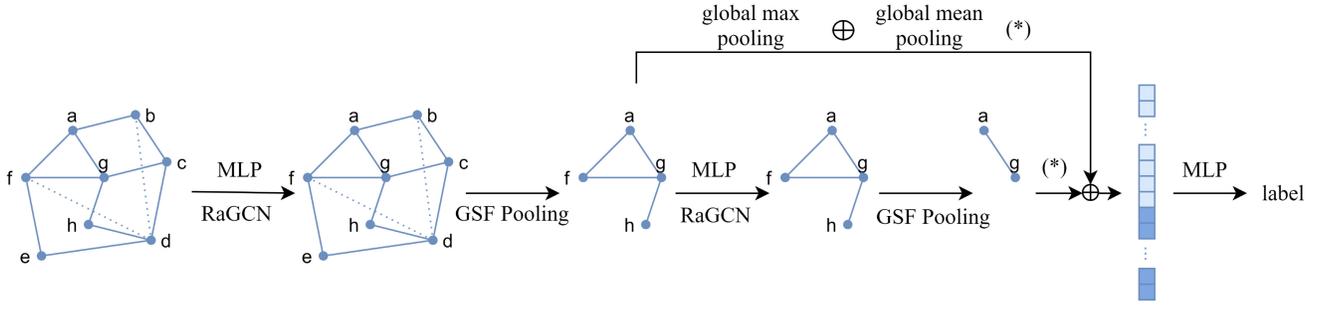


Fig. 3. The detailed network architecture of graph classification: SFC-GNN.

function, and $\mathbf{s}_1^{(k)} \in \mathbb{R}^{N^{(k)}}$ is the final calculated score of each node.

For structure-based node selection, considering the structure information in node selection is crucial because the nodes and edges in the graph and the adjacency matrix between nodes contain much structure information. Given that the graph convolution process uses the graph structure information, the graph convolution method can be used to calculate the score of the graph structure by the adjacency matrix E and the feature matrix H :

$$\begin{aligned} \mathbf{s}_2^{(k)} &= \phi\left(GNN\left(H^{(k)}, E\right)\right) \\ &= \phi\left(D^{-\frac{1}{2}}ED^{-\frac{1}{2}}H^{(k)}\mathbf{w}^{(k)} + H^{(k)}\mathbf{w}^{(k)}\right), \end{aligned} \quad (5)$$

where $D \in \mathbb{R}^{N \times N}$ is the degree matrix of the sum of the graph adjacency matrix and the identity matrix $E + I_N$, and $\mathbf{w}^{(k)} \in \mathbb{R}^{d^{(k)}}$ is the parameter vector to be learned. $\mathbf{s}_2^{(k)} \in \mathbb{R}^{N^{(k)}}$ is the final score of each node calculated by graph convolution.

When calculating \mathbf{s}_1 and \mathbf{s}_2 , the Sigmoid activation function is adopted to map the values to the same interval. The final score of each node in the brain graph is expressed as the linear sum of \mathbf{s}_1 and \mathbf{s}_2 :

$$\mathbf{s} = \alpha\mathbf{s}_1 + (1 - \alpha)\mathbf{s}_2, \quad \alpha \in [0, 1], \quad (6)$$

where α is a hyperparameter. When the value of α is 1, the calculation method of the score is Top-K pooling based only on feature information; and when α is 0, the calculation method will degenerate to the method of the graph structure information-based graph convolution calculation. After obtaining the score of each node using the aforementioned calculation method, the nodes are sorted according to score, and then the top t nodes are selected by the Top-K method as the nodes retained after pooling. The updated adjacency matrix and node features after the corresponding graph pooling layer are obtained:

$$\begin{aligned} \mathbf{i} &= \text{topk}\left(\mathbf{s}^{(k)}, t\right), \\ E^{(k+1)} &= E_{\mathbf{i}, \mathbf{i}}^{(k)}, \\ H^{(k+1)} &= \left(H^{(k)} \odot \left[\mathbf{s}^{(k)}, \mathbf{s}^{(k)}, \dots, \mathbf{s}^{(k)}\right]_{N^{(k)} \times d^{(k)}}\right)_{\mathbf{i},:}, \end{aligned} \quad (7)$$

where \mathbf{i} represents the selected node to be retained. \odot represents the Hadamard product (*i.e.*, the element-by-element multiplication of matrix). $(\cdot)_{\mathbf{i},:}$ represents the index operation of matrix (*i.e.*, select all elements specified by the row index \mathbf{i}

and the column index \mathbf{j}). Thus, through the graph pooling layer, the updated output graph $(\mathcal{V}^{(k+1)}, \mathcal{E}^{(k+1)})$ is obtained from its input graph $(\mathcal{V}^{(k)}, \mathcal{E}^{(k)})$. The GSF layer can reduce the number of parameters and select the nodes important for classification. Therefore, finding brain regions related to diseases is beneficial and makes the classification result explicable.

3) *Readout Layer*: The readout layer of the GNN aggregates all node features in the subgraph to obtain the representation vector of the entire graph. It can be implemented through a graph-level pooling operation, that is, a certain node selection strategy is used to learn graph-level features, and the original graph is mapped into a vector containing graph-level information. Different from the aforementioned pooling layer, the readout layer does not consider the hierarchical information of the graph structure and focuses on the learning of graph-level representations [35]. The readout operation obtains graph-level features via simple permutation invariance functions, such as by summing, averaging, or maximizing the hidden representations of all nodes in the subgraph. Therefore, the readout function can be expressed as:

$$\mathbf{h}_{\mathcal{G}}^{(k)} = \text{Readout}\left(\mathbf{h}_v^{(k)} \mid v \in \mathcal{G}^{(k)}\right). \quad (8)$$

The readout function in this paper simultaneously maximizes and averages all node representations of the subgraph to obtain the final graph representation vector:

$$\mathbf{h}_{\mathcal{G}}^{(k)} = \max \mathcal{H}^{(k)} \parallel \text{mean } \mathcal{H}^{(k)}, \quad (9)$$

where $\mathcal{H}^{(k)} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{N^{(k)}}\}$, and \parallel represents the splicing operation. To obtain the final graph representation vector, the operations are element-wise, and the vectors obtained in the two ways are concatenated into one vector. Finally, the graph representation vector $\mathbf{h}_{\mathcal{G}}$ obtained by the readout layer will be inputted to a classifier, such as a multilayer perceptron, to obtain the final prediction result.

4) *Network Architecture: SFC-GNN*: Using the three aforementioned modules of the graph convolution layer, graph pooling layer, and readout layer, the SFC-GNN model is constructed stacked by GNN modules, and its network architecture is shown in Fig. 3. Each GNN module consists of a brain region-aware graph convolution layer and a structure and the feature information-based graph pooling layer. The forward graph network model receives the undirected weighted brain functional graph constructed by fMRI and its node features as input. It stacks the GNN modules in sequence, and each GNN

module is followed by a separate readout layer for obtaining the graph representation vector processed by the module. Finally, all the graph representation vectors obtained by the readout layer are concatenated into a feature vector, which is inputted into the multilayer perceptron for the prediction.

5) *Loss Function*: For the classification task, the model adopts a simple cross-entropy loss function to calculate the classification error:

$$L_{ce} = \frac{1}{B} \sum_{i=1}^B L_i = -\frac{1}{B} \sum_{i=1}^B \sum_{c=1}^C y_{ic} \cdot \log(\hat{y}_{ic}), \quad (10)$$

where B is the number of samples, which is usually the batch size. C is the number of categories. y_{ic} is the sign function, which takes 1 if and only if the true label of the sample i is c , otherwise it takes 0. $\hat{y}_{ic} \in [0, 1]$ is the probability that the sample i belongs to the category c predicted by the model.

For the graph pooling operation, if the scores of each node in the graph calculated by the pooling function are similar, then the amount of useful information contained in each node is similar, and much useful information is lost through node selection. Therefore, the model expects that the score gap between the retained nodes and the unselected nodes during the node selection process should be as large as possible. In Eqs. (4) and (5), if Sigmoid is selected as the activation function, then the score value of the selected node is expected to be 1, and that of the unselected node tends to be 0. To achieve this goal, maximum mean discrepancy (MMD) is introduced [36] to define the corresponding Top-K node selection loss:

$$L_{topk} = -\left\| \frac{1}{m} \sum_{i=1}^m \varphi(x_i) - \frac{1}{n} \sum_{j=1}^n \varphi(y_j) \right\|^2. \quad (11)$$

Large MMD means better performance, and the loss function needs to be minimized. Here, L_{topk} takes its opposite.

In addition, the score vector \mathbf{s} obtained by node selection function is the result of calculating the structure and feature information of the original brain functional graph. Therefore, for different subjects, the difference of \mathbf{s} may be quite large due to the difference between their brain functional graphs. The purpose of the graph network is to classify different brain functional graphs to explore the commonalities of biological patterns among different subjects under the same neural prediction task to explore the group-level features between the brain functional graphs of subjects that belong to the same category in the classification results. \mathbf{s} can be regularized to make \mathbf{s} of different subjects in the same category consistent. Thus, group consistency loss [17] is introduced as a regularization term in the loss function:

$$\begin{aligned} L_{glc} &= \sum_{c=1}^C \sum_{i,j \in \mathcal{D}_c} \|\mathbf{s}_i - \mathbf{s}_j\|^2 \\ &= \sum_{c=1}^C [2\text{tr}((S_c)^T M_c S_c) - 2\text{tr}((S_c)^T L_c S_c)] \\ &= 2 \sum_{c=1}^C \text{tr}((S_c)^T P_c S_c), \end{aligned} \quad (12)$$

where C is the total number of categories, \mathcal{D}_c is the set of brain functional graphs that belong to this category. $S_c =$

$[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m]^T \in \mathbb{R}^{m \times N}$, where $m = |\mathcal{D}_c|$, $M_c \in \mathbb{R}^{m \times m}$ is a diagonal matrix whose diagonal elements are all m , $L_c \in \mathbb{R}^{m \times m}$ is a matrix with all elements 1, thus $P_c = M_c - L_c$ is a symmetric positive semi-definite matrix [37]. The loss function calculates the sum of the squared L_2 distances between the node score vectors of two brain functional graphs under the same category. Through training, the group consistency loss function is minimized to enable the brain graph feature under the same category to have higher consistency.

Based on the above loss function, the final loss function of the graph network is:

$$L = L_{ce} + \sum_{k=1}^K \alpha^{(k)} L_{topk}^{(k)} + \beta \cdot L_{glc}^{(1)}, \quad (13)$$

where K is the total number of layers in the network or the number of GNN modules. α and β are the hyperparameters used to adjust the weight of different loss functions. For each GNN module, L_{topk} is used to calculate the distance between the scores of the retained nodes and the unselected nodes in the graph pooling layer. Given that the score vectors of various brain functional graphs will be quite different after multiple GNN modules, L_{GLC} is only used for the first graph pooling layer.

C. Brain Functional Graph Embedded Classification Model

Through the above SFC-GNN model, the brain functional graph defined on each subject is outputted as a graph representation vector. It can be used as the embedding vector of each subject's brain functional graph. Following this idea, each subject i is regarded as a node, and the brain functional graph embedding vector obtained by SFC-GNN is used as the feature vector f_i of the node. Therefore, a group-level brain network can be constructed by combining the brain functional graphs of all subjects. Since each node of the group network represents an individual, the group network can also be learned by the graph network and regarded as a node classification task for the diagnosis of brain diseases. The overall algorithm framework is shown in Fig. 4.

The key to constructing the group network lies in the definition of the edges in the graph (*i.e.*, the definition of the adjacency matrix). Given that the brain network $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$ of each subject level is defined according to the correlation among brain functional regions, the graph kernel function is adopted to measure the topological similarity between the correlation coefficient matrices of subjects, thereby obtaining the edges of the group network. The graph kernel function can capture the similarity between the topological structures of graph data in high-dimensional space effectively.

Here, the Gaussian kernel function is adopted to calculate the correlation coefficient matrix of each subject (*i.e.*, for the brain network \mathcal{G}_i and \mathcal{G}_j), we yield

$$K(E_i, E_j) = \langle \phi(E_i), \phi(E_j) \rangle = \exp\left(-\frac{\|E_i - E_j\|_2^2}{2}\right). \quad (14)$$

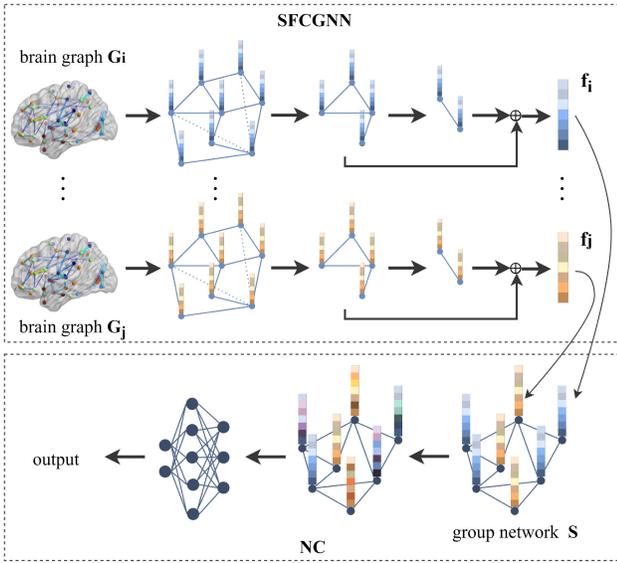


Fig. 4. The algorithm framework of group network and node classification.

Therefore, the similarity between brain networks can be defined as

$$S(\mathcal{G}_i, \mathcal{G}_j) = \frac{\sum_{i=1}^M \sum_{j=1}^M w_i w_j K(E_i, E_j)}{\sum_{i=1}^M w_i \sum_{j=1}^M w_j},$$

$$w_i = \frac{1}{\sum_{k=1}^M K(E_i, E_k)}, \quad (15)$$

where M represents the number of subjects, and E_i represents the sparse adjacency matrix that corresponds to the brain functional network \mathcal{G}_i of the subject i . The size of the obtained adjacency matrix that corresponds to the group network is $M \times M$. This group network is inputted into GCN for node classification, where only one simple graph convolution layer is used to aggregate the features between nodes:

$$F^{(1)} = \tilde{D}^{-\frac{1}{2}} \tilde{S} \tilde{D}^{-\frac{1}{2}} F^{(0)} W^{(0)}, \quad (16)$$

where $\tilde{S} = S + I_M$, $\tilde{D}_{ii} = \sum_{j=1}^M \tilde{S}_{ij}$, $F^{(0)} = [f_1, f_2, \dots, f_M]^T$ is the node feature obtained by SFC-GNN, and $W^{(0)}$ is the weight matrix to be learned. The whole graph convolution process can be regarded as Laplace smoothing on the group network structure. The feature matrix obtained through this graph convolution layer is also inputted into a two-layer perceptron to obtain the final classification vector. Then, cross-entropy is adopted to calculate the loss, and the average of all nodes is obtained.

Inductive learning is used to train and infer the node classification model. The model learns from the nodes that belong to the training set and then generalizes them to other unseen nodes. Specifically, a group network is constructed on the basis of all the subjects in the dataset [21].

IV. EXPERIMENT RESULTS

A. Dataset and Preprocessing

Two brain imaging datasets, namely, the ABIDE dataset [38] and the ADNI dataset [39] are utilized for experiments and

comparative analysis. For the ABIDE dataset, 1035 subjects are divided into autistic patients and healthy controls. For the ADNI dataset, 134 independent subjects are divided into AD patients and mild cognitive impairment patients.

B. Implementation Details

All experiments in this paper are carried out on a computer with a single RXT2080Ti graphics card, and the deep learning framework Pytorch is used for the model construction and the training and testing of the algorithm. All subjects' data are randomly shuffled, and experiments are conducted in a cross-validation manner. The Ranger optimizer [40] is used for the optimization of all models with an initial learning rate of 0.01, which is reduced to half the original value every 20 epochs and a weight decay of 0.0005. The maximum number of iteration cycles of the network is set to 100, and the batch size is set to 16. For the network model, the number of communities R is set to 8, and PReLU is used as the activation function in the two-layer perceptron. In addition, to avoid overfitting, dropout regularization is added after the graph pooling layer, and nodes are randomly dropped with a probability of 0.5 during training. Precision, Recall, F1 scores, and ROC curve are used to evaluate the performance of the model.

C. Comparative Models

To measure the performance of the proposed SFC-GNN model, a series of comparative experiments is conducted between the proposed model and nine existing classification models on the autism diagnosis task with the ABIDE dataset. The comparison models include four traditional machine learning-based models (*e.g.*, SVM, K-nearest neighbor (KNN), decision tree (DT), and RF), two typical neural network-based models for processing time-series signals (*e.g.*, LSTM [41] and TCN [42]), and three top-performing graph network-based models on fMRI (*e.g.*, GAT [43], LI_NET [16], and BrainGNN [17]).

Experiments are conducted on the ABIDE and ADNI datasets under 10 different classification models to compare the performance of various algorithms on the disease diagnosis task. The preprocessing of experimental data and the input form vary for different types of models. For machine learning-based models, the upper half of the matrix is vectorized, and the principal component analysis method, which removes redundancy and retains 99% of the original data, is used to reduce its dimension after the correlation coefficient matrix is constructed from the average time-series. The final input of the model refers to the correlation coefficient of the brain functional regions in vector form. In addition, to avoid affecting the performance of the machine learning-based model due to improper hyperparameter settings, the grid search method is used to search for the best parameter combination of the model, and then the best model is fitted automatically. For LSTM and TCN, the neural network-based models, the average time-series of all brain functional regions is directly used as the input of models. For all graph network models, the same brain functional graph is defined as the input similar

TABLE I
RESULT COMPARISON OF TEN CLASSIFICATION
MODELS ON ABIDE DATA

Model	Healthy Control Group			ASD Patients		
	Precision	Recall	F1 score	Precision	Recall	F1 score
SVM	0.6571	0.6462	0.6316	0.6664	0.6734	0.6513
KNN	0.6168	0.5038	0.5113	0.6316	0.6881	0.6129
DT	0.5076	0.5781	0.5281	0.5362	0.4644	0.4848
RF	0.5531	0.5024	0.4911	0.5656	0.6014	0.5544
LSTM	0.6923	0.4147	0.4408	0.5698	0.7086	0.6016
TCN	0.6612	0.3765	0.4157	0.5689	0.7312	0.5969
GAT	0.6587	0.7105	0.6801	0.6678	0.6156	0.6364
LI_NET	0.7057	0.6052	0.6397	0.6376	0.7219	0.6715
BrainGNN	0.6418	0.6693	0.6188	0.6571	0.5359	0.5239
SFC-GNN	0.7242	0.7744	0.7467	0.7442	0.6758	0.7041

to that in Section III-A, and the original parameter settings in the related papers are used.

First, comparative experiments are conducted on the ABIDE dataset to compare the recognition rates of autistic patients. For the ABIDE dataset, the experiment is conducted by tenfold cross-validation, and the three indicators (precision, recall and F1 score) are used to measure the performance of the model in recognizing the two categories (ASD patients and healthy control groups). Table I shows the average of the tenfold cross-validation results. All results are obtained on the validation set. The comprehensive result shows that in the ASD diagnosis task, SVM is the model with the best performance among all machine learning-based methods, and its values on three indicators are mostly higher than those of KNN, DT, and RF. The performance of the deep learning-based models is generally better because they can extract rich deep features when dealing with complex data. The comprehensive performance of the graph network-based model is the best among the three kinds of models. For the two neural network models (LSTM and TCN), regardless under which indicator, the performance of the models on the two categories of healthy control groups and ASD patients is quite different. The recall and F1 score of the two models in the ASD category are much higher than those of the healthy control group, whereas the precision of the ASD patient category is slightly lower than that of the healthy control group, indicating that the two models have good detection performance on patients with autism. However, healthy people are more likely to be misdiagnosed as patients with autism. In addition, compared with the TCN model, the performance of the LSTM model is more unstable, and the results of the 10 experiments are very scattered because these two models directly use the average time series as input. More redundancy and noise exist in the data, and the truly useful information has not been effectively extracted. The lengths of time series from different image sites are inconsistent, thereby causing the model to be disturbed and resulting in poor performance.

Among the four graph network-based models, the proposed SFC-GNN model achieves the best average results in classification precision and recall under the two categories of healthy controls and ASD patients in the ABIDE data. In the column of recall of ASD patients, TCN achieves the best results.

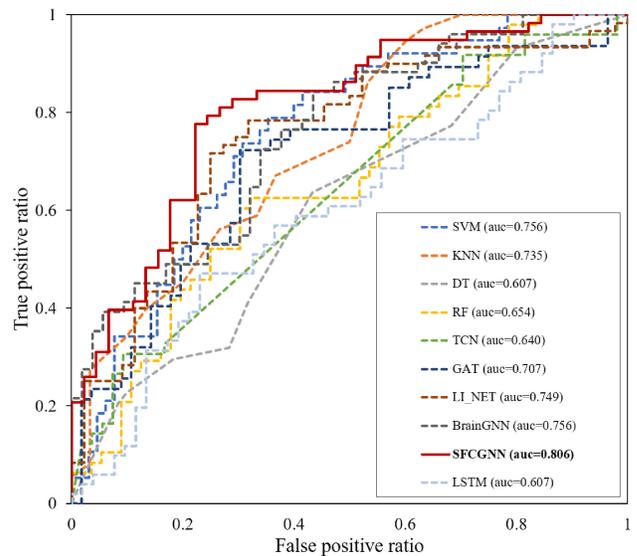


Fig. 5. The comparison of ROC curves and AUC values of ten classification models on ABIDE dataset.

However, its average recall in the healthy control group is the lowest, only 37.65%, which indicates that the model predicts a large number of healthy people as autistic patients, so the effect of TCN is not reasonable. Compared with other graph network-based models, the SFC-GNN model can achieve better performance because the graph pooling layer can select nodes that are more important for classification tasks, and the penalty term of the node score in the loss function further guides the pooling layer to retain more informative nodes. In addition, although deep learning-based method is limited by the amount of data available on ABIDE dataset, our method can achieve good results, which proves the effectiveness of the proposed architecture.

Taking ASD patients as the positive category, the receiver operating characteristic (ROC) curve is drawn for each model to further demonstrate its performance on the ASD diagnosis task. Fig. 5 shows the best result for the area under the curve (AUC) of each model. ROC curve and AUC value further illustrate that the graph network-based model has better performance than other models, and the proposed SFC-GNN model has certain advantages in the autism diagnosis task.

The same comparison experiment is performed on ADNI data. The HO atlas is used for the segmentation of the subject's fMRI data, and the sparse brain network is constructed using the top 10% as the threshold. Fivefold cross-validation is conducted on 10 classification models. Given that the ADNI data selected in this paper only contains 34 AD patient data, an imbalance of positive and negative samples is observed. Thus, AD patients are taken as the positive category. As shown in Table II, the results are compared under four indicators: weighted precision, weighted recall, weighted F1 score, and AUC value. Since weighted recall is equivalent to accuracy, no separate comparison of model accuracy is presented here.

The SFC-GNN model achieves the best average results on the evaluation metrics, which shows the effectiveness of SFC-GNN for AD diagnosis. For most classifiers, the weighted

TABLE II

COMPARISON OF RESULTS OF CLASSIFICATION MODELS ON ADNI DATA

Model	Precision	Recall	F1 score	AUC
SVM	0.7328	0.7088	0.7007	0.5562
KNN	0.5694	0.7305	0.6350	0.5085
DT	0.6670	0.6564	0.6448	0.4836
RF	0.5712	0.7453	0.6425	0.5482
LSTM	0.6073	0.7311	0.6567	0.6004
TCN	0.6237	0.7538	0.6551	0.5482
GAT	0.6831	0.6946	0.6842	0.6110
LL_NET	0.5511	0.7020	0.6165	0.6197
BrainGNN	0.5958	0.7538	0.6544	0.5834
SFC-GNN	0.7116	0.7689	0.7150	0.6219

recall is higher than the weighted precision because of the imbalance of positive and negative samples and the classifier's tendency to classify samples as negative. Only SVM, GAT, and SFC-GNN are less affected by the imbalance of samples, thereby resulting in higher F1 scores. However, because the ADNI data used for the experiment are too small, the model may overfit the training set and have poor generalization performance. When calculating the AUC value, AD patients with a small number of samples are taken as the positive category, so the AUC value of all models is low. However, graph classifiers generally obtain higher AUC values compared with other classification methods, thereby showing that the brain functional graph definition method proposed in this paper can overcome the problem of sample imbalance to a certain extent and improve the model's ability to identify small categories.

D. Group Networks and Transfer Learning

1) *Group Networks*: To further improve the performance of FC-GNN for brain disease diagnosis, a corresponding group network is constructed for 134 subjects in the ADNI dataset according to the method proposed in Section III-C. Fivefold cross-validation is performed in the experiment.

First, the feature embedding vector of each subject's brain functional graph is obtained by the output of the second graph pooling layer in the trained SFC-GNN. Then, in the node classification task, the embedding vectors of all subjects in the training set and the adjacency matrix constructed by the similarity between brain functional graphs are inputted into GCN for training. Finally, the prediction is made on the validation set. In the forward inference process, the input embedding vector and adjacency matrix include all the subjects in the training and validation sets. The feature embedding vector is also obtained from the trained SFC-GNN. In the adjacency matrix, except that the brain functional graph similarity between subjects in the validation set is 0, the other values are calculated by the Eq. (15), so that GCN can generalize the classification criteria learned through the training set to the validation set. According to the results in Table III, compared with the separate SFC-GNN model, by further constructing the group network and inductively learning the node classification-based GCN, the precision and recall of AD diagnosis on ADNI data

TABLE III

COMPARISON OF ABLATION EXPERIMENT RESULT

Model	Precision	Recall	F1 score	AUC
SFC-GNN	0.7116	0.7689	0.7150	0.6219
SFC-GNN+NC	0.7531	0.7991	0.7379	0.5911
TLGNN12+SFC-GNN	0.7102	0.7840	0.7328	0.6542
TLGNN1+SFC-GNN	0.7443	0.7989	0.7592	0.6522
TLGNN1+SFC-GNN+NC	0.7862	0.8063	0.7534	0.7170

increase by 4.15% and 3.02% respectively, but the AUC value is reduced by 3.08%. This finding suggests that the classification accuracy can be improved by constructing individual-level brain functional networks and group networks and capturing the topology information of brain functional networks and the similarity between subjects. However, it is greatly affected by the problem of sample imbalance. During node classification, the prediction of unseen nodes will be more affected by numerous types of nodes in its neighborhood.

2) *Transfer Learning*: With only 134 samples, the amount of ADNI data is relatively small. Deep learning-based models often require a large amount of data for training, so small samples may be one of the reasons for the low model accuracy. Transfer learning is often used in various medical image tasks based on deep learning. However, it is often the transfer from natural image tasks to medical image tasks. Natural and medical images have huge differences in imaging principles and intrinsic characteristics. Thus, this transfer is not necessarily applicable. Considering that some correlation may exist among different brain diseases, here we explore to transfer from ABIDE data to ADNI data to discuss whether the transfer among different brain diseases can improve the generalization ability of the model.

The brain functional network is constructed on the ABIDE data segmented by the HO atlas and then inputted to the SFC-GNN for the training to obtain a pre-trained model. The pretrained model is fine-tuned on the ADNI data. Two transfer learning approaches are considered here, one is to pretrain the entire SFC-GNN model and then fine-tune it under the brain functional network constructed from ADNI data (TLGNN12+SFC-GNN). The other is to use only the first GNN module in SFC-GNN as a pre-trained model, whereas the second GNN module under the ADNI task is still trained from random initialization (TLGNN1+SFC-GNN). The results in Table III show that the transfer learning from ASD to AD can improve the AD diagnosis performance of the model, and the AUC value is increased by 3.23%. Compared with the pre-training of the entire SFGNN, pre-training only on the first GNN module can achieve relatively better results. The above experimental results show that some inherent brain functional network topology may exist between the data of AD and ASD, and this makes the transfer learning between the two datasets contribute to the improvement of model performance. However, the activation of brain functional regions among different brain diseases (*i.e.*, the brain functional regions that are more important for the diagnosis tasks of AD and ASD are different) vary. Then, the nodes retained by the graph pooling

layer are different. Therefore, compared with transferring the entire SFC-GNN model, better AD diagnosis performance can be achieved by only transferring the first graph convolution layer.

The model that combines transfer learning and node classification of group networks (TLGNN1+SFC-GNN+NC) achieves the best AD diagnosis performance in the experiment. The AUC value is improved by 9.51% compared with the benchmark SFC-GNN model. The experiment shows that some correlation may be observed between the two brain diseases, namely, ASD and AD, and the graph network-based model for the brain disease diagnosis can be transferred across related diseases to overcome the overfitting problem in small image datasets, thereby improving the generalization performance of the disease diagnosis model.

V. CONCLUSION

A low-parameter and interpretable brain disease diagnosis model called SFC-GNN is proposed, and the construction methods of sparse brain graph and group network are defined for the high complexity of the brain functional network and its unique topological structure. The diagnosis of AD and autism are conducted on the basis of graph classification and node classification. Our model combines the convolution layer of brain functional region perception and the graph pooling layer of node selection to build a network architecture and uses the topology of brain functional graph and node features to identify important brain functional regions, thereby reducing the number of parameters and improving the interpretability of the model. In addition, SFC-GNN is further extended to a node classification model by defining the group network. The comparative experiments on the ABIDE dataset and the ADNI dataset show that the SFC-GNN model has better disease diagnosis performance. The experimental results of transfer learning between two brain diseases indicate the potential correlation among different brain diseases.

REFERENCES

- [1] J. G. Castillo, J. W. Y. Kam, C. W. Hoy, and P. A. Bandettini, "How to interpret resting-state fMRI: Ask your participants," *J. Neurosci.*, vol. 41, no. 6, pp. 1130–1141, 2021.
- [2] T. L. Richards and V. W. Berninger, "Abnormal fMRI connectivity in children with dyslexia during a phoneme task: Before but not after treatment," *J. Neurolinguistics*, vol. 21, no. 4, pp. 294–304, 2008.
- [3] G. H. Glover, "Overview of functional magnetic resonance imaging," *Neurosurg. Clin. North Amer.*, vol. 22, no. 2, pp. 133–139, 2011.
- [4] K. K. Kwong et al., "Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation," *Proc. Nat. Acad. Sci. United States Amer.*, vol. 89, no. 12, pp. 5675–5679, 1992.
- [5] E. E. O'Connor and T. A. Zeffiro, "Why is clinical fMRI in a resting state?" *Front. Neurol.*, vol. 10, p. 420, Apr. 2019.
- [6] R. G. Wise and C. Preston, "What is the value of human fMRI in CNS drug development?" *Drug Discov. Today*, vol. 15, nos. 21–22, pp. 973–980, 2010.
- [7] D. I. Kim et al., "Identification of imaging biomarkers in schizophrenia: A coefficient-constrained independent component analysis of the mind multi-site schizophrenia study," *Neuroinformatics*, vol. 8, no. 4, pp. 213–229, 2010.
- [8] M. D. Rosenberg et al., "A neuromarker of sustained attention from whole-brain functional connectivity," *Nat. Neurosci.*, vol. 19, no. 1, pp. 165–171, 2016.
- [9] G. Ball et al., "Machine-learning to characterise neonatal functional connectivity in the preterm brain," *NeuroImage*, vol. 124, pp. 267–275, Jan. 2016.
- [10] E. Challis et al., "Gaussian process classification of alzheimer's disease and mild cognitive impairment from resting-state fMRI," *NeuroImage*, vol. 112, pp. 232–243, May 2015.
- [11] Y. Cui et al., "Identifying brain networks at multiple time scales via deep recurrent neural network," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2515–2525, Nov. 2019.
- [12] N. C. Dvornek et al., "Jointly discriminative and generative recurrent neural networks for learning from fMRI," in *Proc. Int. Conf. Mach. Learn. Med. Imag.*, 2019, pp. 382–390.
- [13] J. Kawahara et al., "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017.
- [14] A. Kazi et al., "InceptionGCN: Receptive field aware graph convolutional network for disease prediction," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2019, pp. 73–85.
- [15] S. Parisot et al., "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.
- [16] X. Li et al., "Pooling regularized graph neural network for fMRI biomarker analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2020, pp. 625–635.
- [17] X. Li et al., "BrainGNN: Interpretable brain graph neural network for fMRI analysis," *Med. Image Anal.*, vol. 74, Dec. 2021, Art. no. 102233.
- [18] S. Gadgil, Q. Zhao, A. Pfefferbaum, E. V. Sullivan, E. Adeli, and K. M. Pohl, "Spatio-temporal graph convolution for resting-state fMRI analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2020, pp. 528–538.
- [19] 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.
- [20] L. Zhang and J. Wang, "Diagnosing autism spectrum disorder from functional MRI using graph isomorphic network," *J. Nanjing Univ. Nat. Sci.*, vol. 57, no. 5, pp. 801–809, 2021.
- [21] 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.
- [22] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–17.
- [23] J. You, R. Ying, and J. Leskovec, "Design space for graph neural networks," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 17009–17021.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1026–1034.
- [25] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
- [26] D. Mesquita, A. H. Souza, and S. Kaski, "Rethinking pooling in graph neural networks," in *Proc. IEEE Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 2220–2231.
- [27] E. Luzhnica, B. Day, and P. Liò, "Clique pooling for graph classification," 2019, *arXiv:1904.00374*.
- [28] R. Ying, J. You, C. Morris, X. Ren, W. L. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 4805–4815.
- [29] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 3844–3852.
- [30] H. Yuan and S. Ji, "StructPool: Structured graph pooling via conditional random fields," in *Proc. ICLR*, 2020, pp. 1–12.
- [31] Y. Ma, S. Wang, C. C. Aggarwal, and J. Tang, "Graph convolutional networks with EigenPooling," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2019, pp. 723–731.
- [32] H. Gao and S. Ji, "Graph U-nets," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2083–2092.
- [33] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3734–3743.
- [34] L. Zhang et al., "Structure-feature based graph self-adaptive pooling," in *Proc. Web Conf.*, 2020, pp. 3098–3104.
- [35] W. Liu, M. Gong, Z. Tang, A. K. Qin, K. Sheng, and M. Xu, "Locality preserving dense graph convolutional networks with graph context-aware node representations," *Neural Netw.*, vol. 143, pp. 108–120, Nov. 2021.

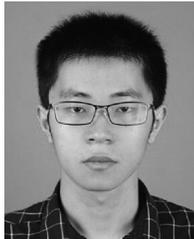
- [36] I. Tolstikhin, B. K. Sriperumbudur, and B. Schölkopf, "Minimax estimation of maximum mean discrepancy with radial kernels," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1930–1938.
- [37] U. Von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
- [38] C. Cameron et al., "The Neuro Bureau preprocessing initiative: Open sharing of preprocessed neuroimaging data and derivatives," *Front. Neuroinform.*, vol. 7, p. 41, Aug. 2013.
- [39] M. W. Weiner et al., "The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement," *Alzheimer's Dementia*, vol. 13, no. 5, pp. 561–571, 2017.
- [40] L. Wright and N. Demeure, "Ranger21: A synergistic deep learning optimizer," 2021, *arXiv:2106.13731*.
- [41] P. Malhotra, L. Vig, G. Shorff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proc. 23rd Eur. Symp. Artif. Neural Netw. Comput. Intell. Mach. Learn.*, 2015, pp. 89–94.
- [42] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.
- [43] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–12.



Qi-Qi Chen received the bachelor's degree in automation from Shanghai Jiao Tong University, Shanghai, China, in 2019, where she is currently pursuing the master's degree with the Department of Automation. Her research interests include computer vision and medical image analysis.



Yubing Gao received the bachelor's degree in mathematics from the Harbin Institute of Technology, Harbin, China, in 2020. She is currently pursuing the master's degree with the Department of Automation, Shanghai Jiao Tong University. Her research interests are machine learning.



Wei Tong is currently pursuing the Ph.D. degree with the School of Mechanical Engineering, Nanjing University of Science and Technology. He is a Research Assistant with the Department of Automation, Shanghai Jiao Tong University, China. His research interests include robot vision, human-machine interaction, 3-D reconstruction, medical image analysis, and brain cognition.



Ping Li (Member, IEEE) received the M.Sc. and Ph.D. degrees in biomedical engineering from Southeast University. She is currently a Professor with the School of Biological Science and Medical Engineering, Beihang University. Her research interests include magnetically actuated hydrogel robots, magnetolectric bifunctional tissue engineering scaffolds, and repair of nerve, bone, and skin tissue under mechanical-electrical or magnetolectric stimulation.



Yong-Xia Li received the M.Sc. degree in emergency medicine from the Shanghai Jiao Tong University School of Medicine. She is currently an Attending Physician with the Department of Emergency, Shanghai Sixth People's Hospital. Her research interests include septic encephalopathy, postoperative delirium, differential diagnosis of coma, and intensive care unit syndrome.



Edmond Q. Wu (Senior Member, IEEE) received the Ph.D. degree in controlling theory and engineering from Southeast University, Nanjing, China, in 2009. He is a Professor with the Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai Jiao Tong University, China. He is also currently with the Science and Technology on Avionics Integration Laboratory, China National Aeronautical Radio Electronics Research Institute, Shanghai, China. His research interests include deep learning, fatigue recognition, and human-machine interaction.



Xiao-Yan Zhao received the master's degree in medicine from Southeast University, Nanjing, China, in 2012. She is currently an Attending Doctor of Rehabilitation Medicine with Anhui No.2 Provincial People's Hospital, Hefei, China. Her research interests include rehabilitation of neurological diseases, especially in Alzheimer disease, and cerebrovascular disease.

He is currently an Associate Editor of IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON INTELLIGENT VEHICLES, IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS, IEEE TRANSACTIONS ON MEDICAL ROBOTICS AND BIONICS, and IEEE TRANSACTIONS ON INDUSTRIAL CYBER-PHYSICAL SYSTEMS.