Multi-Center and Multi-Channel Pooling GCN for Early AD Diagnosis Based on Dual-Modality Fused Brain Network

Xuegang Song, Feng Zhou, Alejandro F Frangi, *Fellow, IEEE*, Jiuwen Cao, *Senior Member, IEEE*, Xiaohua Xiao, Yi Lei, Tianfu Wang, Baiying Lei, *Senior Member, IEEE*

***Abstract*—For significant memory concern (SMC) and mild cognitive impairment (MCI), their classification performance is limited by confounding features, diverse imaging protocols, and limited sample size. To address the above limitations, we introduce a dual-modality fused brain connectivity network combining resting-state functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI), and propose three mechanisms in the current graph convolutional network (GCN) to improve classifier performance. First, we introduce a DTI-strength penalty term for constructing functional connectivity networks. Stronger structural connectivity and bigger structural strength diversity between groups provide a higher opportunity for retaining connectivity information. Second, a multi-center attention graph with each node representing a subject is proposed to consider the influence of data source, gender, acquisition equipment, and disease status of those training samples in GCN. The attention mechanism captures their different impacts on edge weights. Third, we propose a multi-channel mechanism to improve filter performance, assigning different filters to features based on feature statistics. Applying those nodes with low-quality features to perform convolution would also deteriorate filter performance. Therefore, we further propose a pooling mechanism, which introduces the disease status information of those training samples to evaluate the quality of nodes. Finally, we obtain the final classification results by inputting the multi-center attention graph into the multi-channel pooling GCN. The proposed method is tested on three datasets (i.e., an ADNI 2 dataset, an ADNI 3 dataset, and an in-house dataset). Experimental results indicate that the proposed method is effective and superior to other related algorithms, with a mean classification accuracy of 93.05% in our binary classification tasks. Our code is available at: https://github.com/Xuegang-S.**

***Index Terms*—Multi-center, multi-channel pooling, graph convolutional network, early Alzheimer's disease, dual-modality fusion.**

I. INTRODUCTION

Alzheimer's disease (AD) is a severe neurodegenerative disease, which is the most common type of dementia. In the United States, 5.8 million Americans lived with AD in 2020, estimated to reach 14 million by 2060 [1]. That makes AD the 5th leading cause of death in the United States for those adults aged 65 years or older [2]. Although the disease is incurable for now, it can be delayed or mitigated in its earlier stages (e.g., mild cognitive impairment (MCI) and MCI earlier stage (e.g., significant memory concern, SMC)) by using specific cognitive training and pharmacological treatments [3, 4]. Therefore, it is important to study the diagnosis of MCI and SMC. The intelligent diagnosis attracts growing attention and has been shown to perform well in neuroimaging [5, 6]. However, a few limitations remain, including the confounding neuroimaging features, multi-center data sources and limited sample size.

The popularly used neuroimaging modalities for brain disease intelligent diagnosis include magnetic resonance imaging (MRI) [7, 8], resting-state functional magnetic resonance imaging (fMRI) [9-11], and diffusion tensor imaging (DTI) [12, 13]. Nevertheless, most current methods utilize single-modality imaging data for this study. Their performance is thus limited for MCI and SMC diagnosis due to confounding neuroimaging features [14-18]. Therefore, multi-modality data fusion is studied and validated to effectively construct a brain connectivity network from multiple disease cues [9, 12, 13]. In our earlier works [9, 12, 13], we constructed fMRI and DTI networks, concatenated the dual-modality features [12] or

This work was supported partly by National Natural Science Foundation of China (Nos. 62101338, U1902209, 61871274, and 61801305), National Key R&D Program of China (No. 2022YFC2009900/2022YFC2009903), China Postdoctoral Science Foundation (No. 2019M653014), and Shenzhen Key Basic Research Project (Nos. JCYJ20180507184647636, JCYJ20170818142347251, JCYJ20170818094109846 and JCYJ20170413152804728), Royal Academy of Engineering Chair in Emerging Technologies Scheme (CiET1819/19), Pengcheng Visiting Scholars Program from the Shenzhen Government. (Corresponding authors: Baiying Lei).

Xuegang Song, Tianfu Wang, Baiying Lei are with School of Biomedical Engineering, Shenzhen University, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen, China. (sxg315@yahoo.com, tfwang@szu.edu.cn, leiby@szu.edu.cn)

Feng Zhou is with Department of Industrial and Manufacturing, Systems Engineering, The University of Michigan, Dearborn, MI, USA. (fezhou@umich.edu)

Alejandro F Frangi is with CISTIB Centre for Computational Imaging & Simulation Technologies in Biomedicine, School of Computing, University of Leeds, Leeds LS2 9LU, United Kingdom. (a.frangi@leeds.ac.uk)

Jiuwen Cao is with Artificial Intelligence Institute, Hangzhou Dianzi University, Zhejiang, China. (jwcao@hdu.edu.cn)

Xiaohua Xiao and Yi Lei are with First Affiliated Hospital of Shenzhen University, Health Science Center, Shenzhen University, Shenzhen, China. (tu_xi8888@163.com, leiyisz@2011@163.com)

designed dual-modality classifiers [13] to boost diagnosis performance by exploring the promising complementarity between functional and structural information. However, concatenating dual-modality features or developing dual-modality classifiers increase network parameters, which results in performance fluctuation in few-shot learning tasks. In view that the strong structural connectivity among ROIs typically implies a high opportunity for functional connectivity [19, 20], we propose to utilize DTI connectivity strength to assist brain functional network construction by introducing a sparse regularization term in the sparse representation (SR) method, which finally forms our fused brain connectivity networks. Based on the different distributions of brain connection strength between different groups on the same pair of brain regions, Wang *et al*. [21] proposed distribution-guided network pruning to determine thresholds for connections in functional networks. Inspired by it, we further propose a regulatory factor in sparse regularization term, which is constructed by computing the distribution diversity of DTI connectivity strength between groups.

Due to the widely spread of MCI and SMC, neuroimaging data is usually acquired from multiple medical centers, which causes diverse imaging conditions [22, 23]. For example, there are different acquisition protocols in the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, including ADNI 1, ADNI 2, ADNI GO, and ADNI 3, where they also use different equipment types (e.g., SIEMENS, GE, Philips, etc.) for data collection. Ignoring the above heterogeneity probably limits the model's ability to yield robust and general representations [24, 25]. Most MCI studies ignore the heterogeneity, which degrades the diagnosis performance. Because graph convolutional network (GCN) integrates gender and equipment type information effectively in diagnosis tasks [13, 24, 26], we extend this approach to integrate multi-center information. Besides, current GCN methods treat the non-image information equally by assigning them the same amplification coefficients on edge weights, which ignores their different impacts on features and thus limits performance. Therefore, we propose an attention mechanism to capture their different impacts on edge weights.

Although the disease is widely spread, the size of acquired dual-modality neuroimaging data is usually limited. In view that the intelligent diagnosis task with a small dataset usually has a higher request on classifiers, it is essential to study further GCN based on the characteristics of MCI and SMC. For GCN, a graph node is usually represented by the neuroimaging features of a subject. The edge between nodes is usually acquired by computing feature similarity where the impact of phenotypic information is also considered [13, 24-27]. The key to GCN lies in its convolutional filter, where edge weights correspond to the convolutional coefficients. The existing GCN methods apply the same convolutional coefficients on all features for filtering with ignorance of the difference between features, which deteriorates the filter performance. To address the shortcoming, we propose a multi-channel mechanism in GCN, which designs different filters for features based on feature statistics. Applying those nodes with low-quality features to perform convolution will also deteriorate convolutional performance. The pooling mechanism is thus studied by reducing the number of those nodes and edges. Most existing pooling methods are based on top-k selection methods to pool nodes and edges [10, 28], while others use learned cluster assignments matrix to generate the pooled graph topology [29, 30]. These pooling methods utilize trainable or learned networks to realize pooling, increase network complexity and result in performance degradation for the task with a small dataset. To address it, we propose a novel pooling mechanism, which evaluates nodes and edges by introducing the disease status of training samples in similarity computing.

The main contributions of this paper are threefold:

1) We introduce a DTI-strength penalty term in the sparse representation method to fuse fMRI functional information and DTI structural information for brain connectivity network construction. DTI structural network and the strength diversity between subject groups are used to construct the penalty term.

2) We propose a multi-center attention graph in GCN to consider non-image information (multi-center source, gender, equipment type, and disease status of those training samples), and capture their different impacts on edge weights via an attention mechanism.

3) We propose a multi-channel mechanism in GCN by designing different filters to improve filter performance. In addition, we propose a novel pooling mechanism, which introduces the disease status of those training samples into the pooling operation to boost performance.

The proposed method is validated using two public datasets from ADNI and one in-house dataset. Experimental results indicate that our method achieves remarkable performance for MCI and SMC diagnosis.

II. Background Of GCN

A. Problem Formation

Fig. 1 shows the flowchart of early AD diagnosis. The diagnosis task is treated as a binary classification task and can be divided into three parts. Specifically, a fused brain network for every subject is constructed based on the imaging data. Then, by extracting the upper triangular matrix elements from fused brain networks, each subject can be represented by a high-dimensional feature vector. Feature selection is then used for feature dimension reduction to fit in the few-shot learning task. Last, the final diagnosis results are outputted based on low-dimensional feature vectors and a designed classifier.

The construction of dual-modality brain networks and the design of the GCN classifier are the key parts, where multi-center data sources and phenotypic information are considered to improve GCN performance.

Fig. 1: Flowchart of early AD diagnosis.

B. Concept of GCN

GCN was first proposed by Kipf and Welling in 2017 [27] based on graph neural networks, where spectral graph convolution theory is introduced to improve filter by applying Fourier transform and Taylor's expansion formula.

Supposing N nodes form a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{A})$, $\mathbf{V}$ is a set of nodes, $\mathbf{E}$ is a set of edges, and $\mathbf{A}$ is the adjacency matrix which is composed by edge weights. The propagation rule of GCN is denoted as:

$$\mathbf{H}^{l+1} = \sigma\left(\widehat{\mathbf{A}}\mathbf{H}^{l}\mathbf{W}^{l}\right), \tag{1}$$

where $\widehat{\mathbf{A}} = \widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}$, $\mathbf{H}^{\boldsymbol{l}}$ is the feature matrix of all nodes in the l^{th} layer, $\mathbf{W}^{l}$ is trainable weight matrix of the l^{th} layer and σ is an activation function. $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, $\widetilde{\mathbf{D}}_{ii} = \sum_j \widetilde{\mathbf{A}}_{ij}$. The graph convolution theory is applied in the construction of $\widehat{\mathbf{A}}$. In the process, $\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}$ first converts $\widetilde{\mathbf{A}}$ to the Fourier domain, then truncates it using Chebyshev polynomials, and the last converts it to the original time domain.

By using the above propagation rule, the structure of popular used two-layer GCN for classification tasks is denoted as:

$$\mathbf{Z} = \mathrm{softmax}\left(\widehat{\mathbf{A}}\mathrm{ReLU}\left(\widehat{\mathbf{A}}\mathbf{X}\mathbf{W}^{0}\right)\mathbf{W}^{1}\right). \tag{2}$$

C. Filtering Principle in Graph Theory

Compared to traditional two-layer neural networks, pre-multiplying adjacency matrix $\widehat{\mathbf{A}}$ (shown as $\widehat{\mathbf{A}}\mathbf{X}$ in Eq. (2)) in the GCN layer before propagation plays a role as a filter, and the filter's effect is the key to boosting performance.

For the feature vector of subject i, represented by $\mathbf{x}_i$, its filtered form $\hat{\mathbf{x}}_{\boldsymbol{i}}$ can be described as Eq. (3).

$$\hat{\mathbf{x}}_i = \sum_{m=1}^{N} \mathbf{A}_{im} \times \boldsymbol{x}_m. \tag{3}$$

It shows that existing GCN methods apply the same convolutional coefficients $(\mathbf{A}_{i1}, \mathbf{A}_{i2}, \ldots, \mathbf{A}_{iN})$ on all features $(\mathbf{x}_i)$ for filtering, which ignores the difference between features. To address it, a multi-channel mechanism is proposed in this paper.

D. Integration of Non-Image Information

Experiments in works [13, 24, 26] show that integrating non-image information can significantly improve classification performance. These works integrate information into edges by assigning edge weights amplification coefficients:

$$\mathbf{A}_{ij} = sim\left(\mathbf{x}_i, \mathbf{x}_j\right) \sum_{h=1}^{H} r\left(\mathrm{P}_i^h, \mathrm{P}_j^h\right), \tag{4}$$

$$r\left(\mathrm{P}_i^h, \mathrm{P}_j^h\right) = \begin{cases} 1, & \left|\mathrm{P}_i^h - \mathrm{P}_j^h\right| < \Theta \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where $sim(\cdot)$ is a similarity function between feature vectors, r represents the measure distance between non-image information, P_i^h represents h^{th} non-image information of subject i.

In the above methods, the edge weight between those subjects with the same non-image information is assigned as an amplification coefficient. These methods consider age, gender, and equipment, but they ignore the multi-center and disease status information in graph learning. Besides, they treat information equally by assigning them the same amplification coefficient on edge weights, which ignores their different impacts. Therefore, we propose a multi-center attention graph to consider further multi-center source information and disease status information of those training samples, and capture their different impacts on edge weights via an attention mechanism.

III. Materials and Methods

Fig. 2 shows an overview of the proposed framework, which is divided into three parts. First, we construct a dual-modality fused brain connectivity network for each subject, where the DTI-strength penalty term is introduced in brain connectivity network construction. Second, we construct a multi-center attention graph to include node's feature and connection information, where multi-center source, disease status information of those training samples, gender and equipment type information are considered in connection establishment. Third, a multi-channel pooling GCN is designed and it outputs the score of each subject.

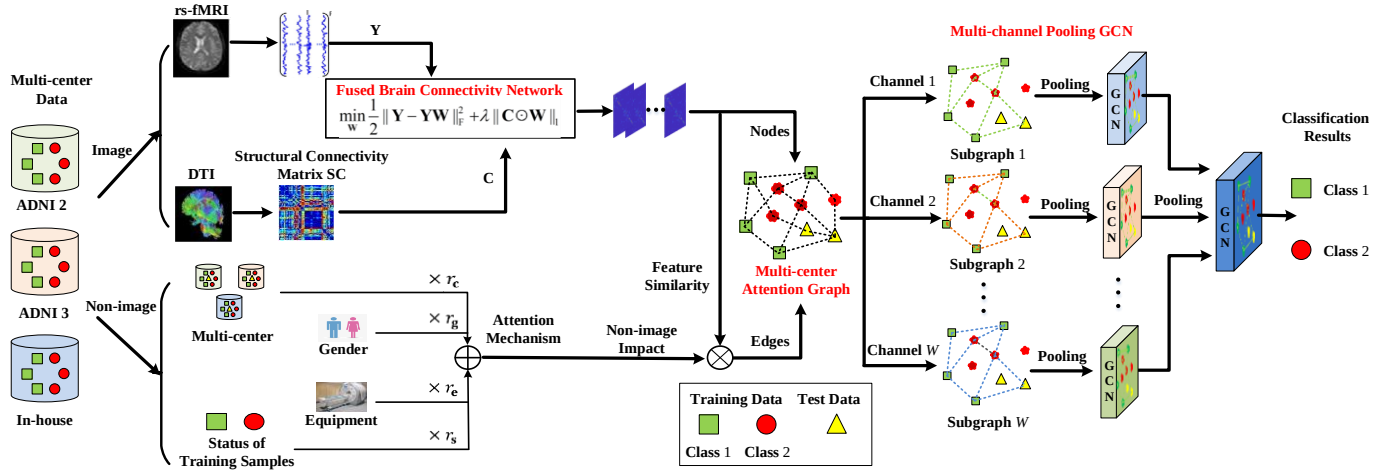

Fig. 2: Proposed framework for disease diagnosis. (a) Fused brain connectivity network construction. For each subject, its DTI structural network and the strength diversity between subject groups construct the penalty term **C**. (b) Multi-center attention graph. For N subjects with each represented by its fused connectivity network, we construct N feature vectors by selecting their discriminative features. Each feature vector is described as a node on the graph. We construct edges and compute their weights by considering the multi-center source, disease status of those training samples, gender, equipment type information, and similarity between feature vectors. (c) Multi-channel pooling GCN. Based on statistics information of features, we further divide the multi-center attention graph into several subgraphs and use a pooling mechanism to process them, then input every pooled subgraph into its corresponding single-layer GCN, then concatenate outputted features and

input them into the second-layer GCN to get the final classification results.

A. Fused Brain Network Construction

1) Dataset

In this study, three datasets with 459 subjects are collected, including datasets from ADNI 2, ADNI 3, and an in-house dataset. Every collected subject in the above three datasets includes the dual-modality data (fMRI and DTI). Generally, 163 normal control (NC), 44 SMC, 86 early MCI (EMCI), and 166 late MCI (LMCI) are included. Demographic details of the used subjects are shown in Table I.

For fMRI data, the standard preprocessing procedures are performed using the GRETNA toolbox. 1) The first ten acquired fMRI volumes are discarded, and then the remaining 170 volumes are corrected by applying mean-subtraction. 2) Head movement correction, spatial normalization with DARTEL, and smooth filtering by employing the Gaussian kernel are applied to improve data. 3) The automated anatomical labelling (AAL) is used to segment brain space into 90 regions of interest (ROIs). After the above process, we obtain the time-series of 90 ROIs for each individual.

For DTI data, PANDA toolbox is used to get the global brain deterministic fiber bundle. Then, we obtain the fractional anisotropy (FA) as feature vectors and use the AAL template on the DTI image to divide the brain space into 90 ROIs. Last, the average FA of links between network nodes is defined as the connection weight in the DTI network. After the above process, we get a 90×90 DTI connectivity network for each individual.

2) Fused brain connectivity network

We use DTI structural information to restrict the construction of functional connectivity networks. Then the restricted functional connectivity networks are treated as fused connectivity networks. Traditional SR-based methods generate the brain functional connectivity network $\mathbf{W}$ by adding a l_1-norm or l_2-norm regularization to control network sparsity. Inspired by the works in [10, 15], which weighted the sparse regularization by using functional connectivity strength, we introduce a DTI-strength penalty term in the SR method to improve its sparse regularization using:

$$\min_{\mathbf{W}} \frac{1}{2}||\mathbf{Y}-\mathbf{YW}||_{\mathrm{F}}^2+\lambda\,||\mathbf{C}\odot\mathbf{W}||_1, \tag{6}$$

where $\mathbf{Y}\in\mathbb{R}^{180\times 90}$ is the fMRI data matrix, and $\mathbf{W}\in\mathbb{R}^{90\times 90}$ is the constructed brain connectivity network. $\mathbf{C}\in\mathbb{R}^{90\times 90}$ is the constructed structural connectivity penalty matrix. $\odot$ denotes the element-wise product, σ is a positive parameter, and λ is a parameter to control sparsity.

Constructing $\mathbf{C}$ is the key to boosting performance. We consider the structural connectivity network and connectivity strength diversity between subject groups. Given T training subjects and their corresponding disease statuses, we divide all training subjects into two groups based on their labels. Here, we also consider the diversity between multi-center datasets. Then the two groups with DTI connectivity strength information are denoted as $\mathbf{SC}^{+}=[\mathbf{SC}^{1+},\mathbf{SC}^{2+},\dots,\mathbf{SC}^{T_1+}]$ and $\mathbf{SC}^{-}=[\mathbf{SC}^{1-},\mathbf{SC}^{2-},\dots,\mathbf{SC}^{T_2-}]$, respectively. Here T_1 and T_2 are the number of subjects in two groups ($T_1+T_2=T$). Therefore, the strength diversity matrix $\mathbf{SC}^{\#}$ is defined as:

$$\mathbf{SC}^{\#}=\left|\frac{\frac{1}{T_1}\sum_{i=1}^{T_1}\mathbf{SC}^{i+}-\frac{1}{T_2}\sum_{j=1}^{T_2}\mathbf{SC}^{j-}}{\frac{1}{T_1}\sum_{i=1}^{T_1}\mathbf{SC}^{i+}+\frac{1}{T_2}\sum_{j=1}^{T_2}\mathbf{SC}^{j-}}\right|, \tag{7}$$

where $\mathbf{SC}^{\#}\in\mathbb{R}^{90\times 90}$ represents DTI strength diversity between two groups, $\mathbf{SC}\in\mathbb{R}^{90\times 90}$ represents DTI connectivity network, $\mathbf{SC}^{i+}$ and $\mathbf{SC}^{j-}$ represent the DTI connection strength matrices of subjects i and j, where subjects i and j come from different groups.

For every subject, by considering its structural connectivity matrix $\mathbf{SC}$ and its corresponding strength diversity matrix $\mathbf{SC}^{\#}$, the structural connectivity penalty matrix $\mathbf{C}$ is defined and its element $\mathbf{C}_{ij}$ is denoted as:

$$\mathbf{C}_{ij}=\exp\left(\frac{-\mathbf{SC}_{ij}{}^2}{\sigma_1}\right)\times\left(1+\exp\left(\frac{-\mathbf{SC}_{ij}^{\#}{}^2}{\sigma_2}\right)\right), \tag{8}$$

where σ_1 and σ_2 are set as the mean value of the standard variation of all subjects' structural connectivity matrix $\mathbf{SC}$ and strength diversity matrix $\mathbf{SC}^{\#}$.

B. Multi-Center Attention Graph

A multi-center attention graph is proposed in this subsection. Edge connections between each pair of nodes retain more useful information. Edge weights are adaptively computed by considering feature similarity and pieces of non-image information (medical center, disease status, gender, equipment type). The attention mechanism considers different impacts of every piece of information, and assigns different amplification coefficients on edge weights. Compared to our earlier work [13], more edges are retained and different impacts of non-image information are considered.

Supposing there are N subjects with feature matrix $\mathbf{X}\in\mathbb{R}^{N\times M}$ representing their features. Adjacency matrix $\mathbf{A}\in\mathbb{R}^{N\times N}$ includes all edge weights, and its element $\mathbf{A}_{ij}$ represents the edge weight between subjects i and j, $sim(\cdot)$ is the feature similarity function, r_{g} represents the gender distance, r_{e} represents the equipment type distance, r_{c} represents the medical center (source) distance, and r_{s} represents the disease status distance. For subject i, we define $\mathbf{x}_i$, g_i, e_i, c_i, and s_i to represent feature vector, gender, equipment type, medical center, and disease status, respectively. The corresponding edge weight on the multi-center attention graph is calculated as:

$$\mathbf{A}_{ij}=sim(\mathbf{x}_i,\mathbf{x}_j)\times\Big(a_0+a_{\mathrm{g}}\times r_{\mathrm{g}}(g_i,g_j)+a_{\mathrm{e}}\times r_{\mathrm{e}}(e_i,e_j)+a_{\mathrm{c}}\times r_{\mathrm{c}}(c_i,c_j)+a_{\mathrm{s}}\times r_{\mathrm{s}}(s_i,s_j)\Big), \tag{9}$$

$$s.t.\ a_0+a_{\mathrm{g}}+a_{\mathrm{e}}+a_{\mathrm{c}}+a_{\mathrm{s}}=1,\ a_0,a_{\mathrm{g}},a_{\mathrm{e}},a_{\mathrm{c}},a_{\mathrm{s}}\in(0,1).$$

where similarity function is defined as:

$$sim(\mathbf{x}_i,\mathbf{x}_j)=exp\left(-\frac{[\rho(\mathbf{x}_i,\mathbf{x}_j)]^2}{2\sigma^2}\right), \tag{10}$$

where $\rho(\cdot)$ is the correlation distance function and σ is the width of the kernel. r_{g}, r_{e}, r_{c}, and r_{s} are defined as:

$$r_{\mathrm{g}}(g_i,g_j)=\begin{cases}1, & g_i=g_j\\ 0, & g_i\neq g_j\end{cases},$$
$$r_{\mathrm{e}}(e_i,e_j)=\begin{cases}1, & e_i=e_j\\ 0, & e_i\neq e_j\end{cases}, \tag{11}$$

$$r_c(c_i, c_j) = \begin{cases} 1, & c_i = c_j \\ 0, & c_i \neq c_j \end{cases},$$

$$r_s(s_i, s_j) = \begin{cases} 1, & s_i = s_j \quad \text{(training data)} \\ 0, & s_i \neq s_j \quad \text{(training data)} \\ 0, & s_i \text{ or } s_j \text{ is unkown (test data)} \end{cases}.$$

The amplification coefficients a_0, a_g, a_e, a_c, and a_s are acquired from the network training. Specifically, we realize our attention mechanism as follows. First, we construct a similarity matrix $\mathbf{Sim} \in \mathbb{R}^{N\times N}$, identity matrix $\mathbf{I} \in \mathbb{R}^{N\times N}$, gender distance matrix $\mathbf{A}_g \in \mathbb{R}^{N\times N}$, equipment distance matrix $\mathbf{A}_e \in \mathbb{R}^{N\times N}$, center distance matrix $\mathbf{A}_c \in \mathbb{R}^{N\times N}$, and status distance matrix $\mathbf{A}_s \in \mathbb{R}^{N\times N}$ according to Eqs. (10) and (11). Then Eq. (9) can also be described as: $\mathbf{A} = \mathbf{Sim} \odot (a_0 \times \mathbf{I} + a_g \times \mathbf{A}_g + a_e \times \mathbf{A}_e + a_c \times \mathbf{A}_c + a_s \times \mathbf{A}_s)$, as shown in Fig. 3. Second, the above five matrices (i.e., $\mathbf{I}, \mathbf{A}_g, \mathbf{A}_e, \mathbf{A}_c$, and $\mathbf{A}_s$) are combined by using Conv2d (5, 1, 1, 1) in python, and the five attention coefficients (i.e., a_0, a_g, a_e, a_c, and a_s) correspond to the convolution coefficients in Conv2d function. Moreover, attention coefficients (i.e., a_0, a_g, a_e, a_c, and a_s) subject to conditions (ranging from 0 to 1 and summed up to 1).

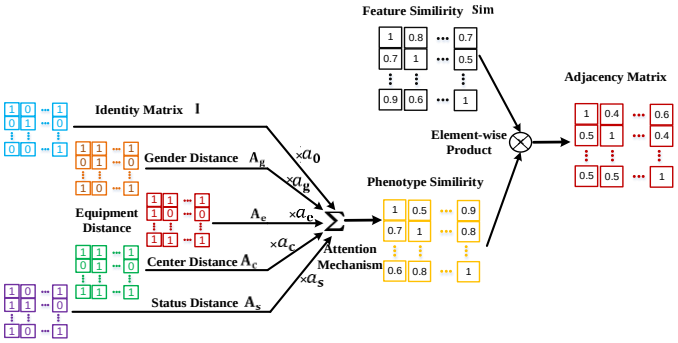

Fig. 3. Construction of adjacency matrix in multi-center attention graph.

C. Multi-Channel Pooling GCN

We propose a multi-channel mechanism to improve filter by assigning different filters to features based on their characteristics. We propose a pooling mechanism to pool edges by introducing the disease status of training samples.

1) Multi-channel mechanism

The procedure of the proposed multi-channel mechanism is described as follows. First, we rank the extracted features based on their statistics. Second, we split ranked features into several parts, and then design several adjacency matrices as filters. We can get several subgraphs by combing the split features and their corresponding adjacency matrices. Third, we get a feature matrix for every subgraph by inputting subgraphs into their corresponding GCN. Then we concatenate them in one feature matrix, and combine it with the initial adjacency matrix to form a graph. Last, we feed the graph into the second-layer GCN, and then output the final classification results. The multi-channel mechanism is shown in Fig. 4.

Inspired by the work [31] that utilized the feature mean value and feature standard deviation for feature selection, we propose our feature ranking method and its evaluation criteria are denoted as:

$$\frac{\|Mean(\mathbf{X}^+) - Mean(\mathbf{X}^-)\|}{Std(\mathbf{X}^+) + Std(\mathbf{X}^-)}, \tag{12}$$

where supposing there are total N subjects on graph with T samples in the training set and each subject has M features. Here T_1 and T_2 are the number of subjects in two groups ($T_1 + T_2 = T$). $\mathbf{X}^+ \in \mathbb{R}^{T_1\times M}$ and $\mathbf{X}^- \in \mathbb{R}^{T_2\times M}$ represent the feature matrices of samples in the two groups. $Mean(\cdot)$ represents the mean value function. $Std(\cdot)$ represents the standard deviation function.

After ranking features based on the criterion as Eq. (12), we divide the feature matrix $\mathbf{X} \in \mathbb{R}^{N\times M}$ into W parts ($\mathbf{X}_i \in \mathbb{R}^{N\times(M/W)}$), as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots \mathbf{X}_W]$, where $\mathbf{X}_1$ represents those high-ranking features and $\mathbf{X}_W$ represents those low-ranking features. Based on feature matrix $\mathbf{X}$ and non-image information, we compute the initial adjacency matrix $\mathbf{A}_0 \in \mathbb{R}^{N\times N}$ according to Eqs.(9)-(11). Similarly, we compute the multi-channel adjacency matrices $\mathbf{A}_1, \mathbf{A}_2, \cdots \mathbf{A}_W$ by using feature matrices $\mathbf{X}_1, \mathbf{X}_2, \cdots \mathbf{X}_W$ and non-image information, respectively. Finally, multi-channel subgraphs are formed, which are represented by $\mathbf{A}_1\mathbf{X}_1^1, \mathbf{A}_2\mathbf{X}_2^1, \cdots \mathbf{A}_W\mathbf{X}_W^1$. $\mathbf{X}_i^1$ represents the feature matrix $\mathbf{X}_i$ in the first GCN layer, and $\mathbf{X}_i^2$ represents the feature matrix $\mathbf{X}_i$ in the second GCN layer. To enhance robustness, we use an initial adjacency matrix $\mathbf{A}_0$ to multiply the multi-channel adjacency matrix in Eq. (13). $\odot$ represents Hadamard product.

$$\mathbf{A}_i = \mathbf{A}_0 \odot \mathbf{A}_i \,. \tag{13}$$

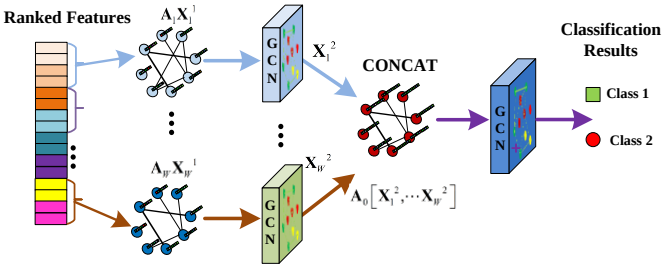

Fig. 4. Overview of the proposed multi-channel mechanism.

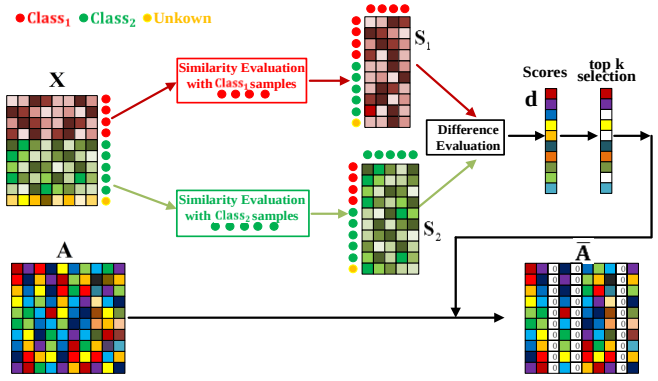

Fig. 5. Overview of the proposed pooling mechanism. All samples are included on the graph, and those training samples are divided into two groups according to their labels (denoted as Class_1 and Class_2). The test samples with unknown labels are the target of the classification task.

2) Pooling mechanism

Supposing there are total of N subjects on a graph with T_1 training samples in Class_1 group and T_2 training samples in Class_2 group. First, we establish a similarity matrix $\mathbf{S}_1 \in \mathbb{R}^{N\times T_1}$ to evaluate the similarity between every sample and those

Class_1 samples, and establish a similarity matrix $\mathbf{S}_2 \in \mathbb{R}^{N\times T_2}$ to evaluate the similarity between every sample and those Class_2 samples. Second, we evaluate the difference between similarity matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ as Eq. (14), denoted as $\mathbf{d} \in \mathbb{R}^{N\times 1}$. Third, we select those nodes with high scores based on the top-k selection method and record their indices. Last, we use the indices to form a new graph by establishing a new adjacency matrix $\overline{\mathbf{A}}$ and keeping feature map invariant as $\overline{\mathbf{X}} = \mathbf{X}$, where $\overline{\mathbf{A}}$ and $\overline{\mathbf{X}}$ are pooled $\mathbf{A}$ and $\mathbf{X}$. Our pooling mechanism is shown in Fig. 5.

$$\mathbf{d} = \|Mean(\mathbf{S}_1) - Mean(\mathbf{S}_2)\|. \quad (14)$$

IV. Experiments And Results

We divide this section into three parts to evaluate the effectiveness of our methods, including performance evaluation of dual-modality fused brain connectivity network, performance evaluation of the proposed methods for improving GCN classifier, and performance comparison of our method with the other related methods. Based on the 10-fold cross-validation strategy, prediction accuracy (ACC), sensitivity (SEN), specificity (SPE) and area under the curve (AUC) are used as evaluation criteria. The GCN parameters of all strategies in this paper are fixed and chosen according to previous work [24]. Parameter details are set as below: dropout rate is set as 0.1, regularization rate is set as 5×10^{-4}, the learning rate is set as 0.005, the number of epochs is set as 200, the default polynomial order is set as 3, the number of neurons per layer is set as 32, and the number of selected features by using RFE is set as 200.

A. Performance of Fused Brain Connectivity Network

The SR method uses fMRI signals to construct a functional connectivity network, and we introduce the DTI-strength penalty term to induce its sparse regularization. Here, we evaluate the effectiveness of our fusion method by comparing it with the other three fusion methods (shown in Fig. 6), and the comparison results are summarized in Table II. In Table II, the classification results based on single-modality data are also provided. All comparison experiments in this subsection adopt our multi-center and multi-channel pooling GCN as the classifier.

Based on fMRI data, the mean ACC of all six binary classification tasks is 83.9%, whereas the mean ACC is 82.8% by using DTI data. Compared to DTI data, using fMRI data for classification shows a 1.1% ACC improvement. This result conflicts with our earlier work [13], where DTI data shows a better performance with ACC improved by 5.5%. The inconsistency is caused by the difference between multi-center data sources, where the data used in this paper comes from three datasets (including ADNI 2, ADNI 3, and in-house) while the data in work [13] comes from one dataset (ADNI 3).

To evaluate the effectiveness of our dual-modality fusion method, we compare it with three popular fusion methods [12], [13]. The mean ACC of our method is 93.1%, whereas the mean ACC of the other three fusion methods is 86.8%, 86.4% and 87.9%. This result shows that our method achieves at least 5.1% performance improvement. Compared to the performance of using single-modality data, it shows 9.1% and 10.3% improvements, respectively. SEN, SPE and AUC reach to 91.1%, 88.8%, and 96.4%, respectively. Comparison results show our method can achieve the best performance.

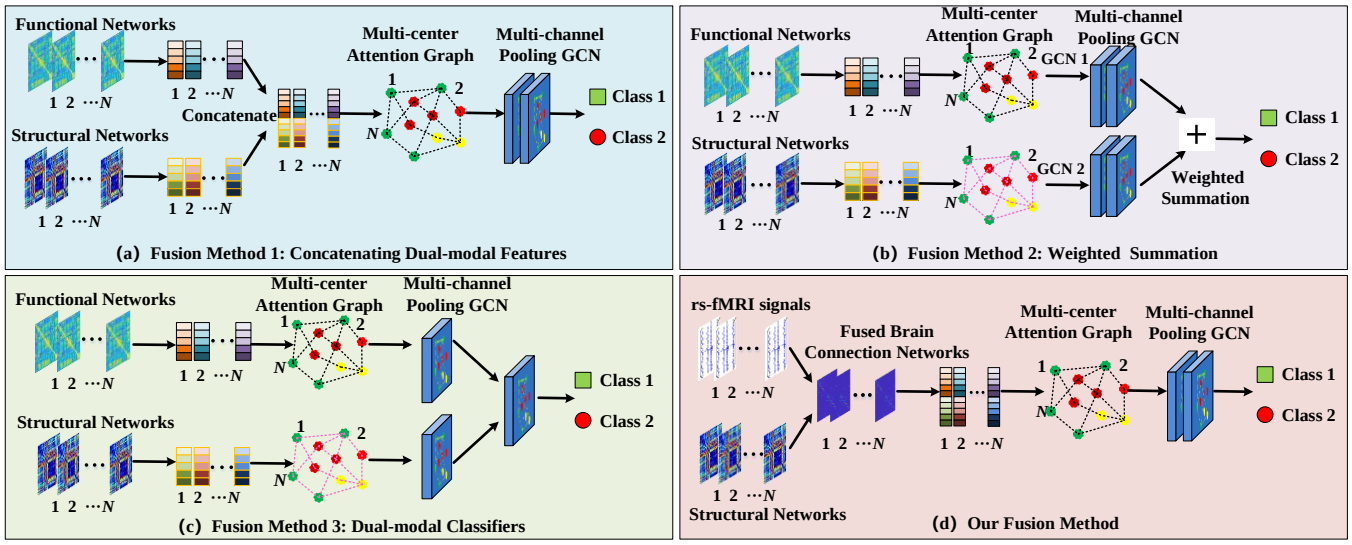

Fig. 6. Overview of the three popular used dual-modality fusion methods for comparison. Fusion method 1: After selecting features from functional and structural networks, it concatenates functional features and structural features as one feature vector for every subject. It constructs a graph and uses a two-layer GCN for final classification. Fusion method 2: After selecting features from functional and structural networks, it constructs a functional graph and a structural graph, and then uses two GCN frameworks for classification. The dual-modal classification results are combined based on the weighted summation. Fusion method 3: After constructing a functional graph and a structural graph, it uses a single-layer GCN to learn deep features for every modality. It concatenates deep functional and deep structural features as one feature vector for every subject. By using another single-layer GCN, it outputs final classification results.

TABLE I. DEMOGRAPHIC CHARACTERISTICS OF THE USED SUBJECTS.

	ADNI 2 (90)			ADNI 3 (200)				In-house (169)	
Subject	NC	EMCI	LMCI	NC	SMC	EMCI	LMCI	NC	LMCI

Number	29	34	27	64	44	52	40	70	99
Gender(M/F)	13/16	15/19	18/9	28/36	17/27	28/24	18/22	26/44	28/71
Age	75.5±5.1	73.1±5.8	72.2±7.2	75.5±6.4	76.3±5.4	76.2±6.5	75.6±8.1	64.6±5.7	65.1±6.5
SIEMENS/GE/Philips	0/0/29	0/0/34	1/0/26	36/26/2	21/21/2	35/14/3	27/12/1	*/*/*	*/*/*

TABLE II. CLASSIFICATION PERFORMANCE OF DIFFERENT MODALITIES AND FUSION METHODS.

Modality	Method	NC vs. SMC			NC vs. EMCI			NC vs. LMCI		
		ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
fMRI	MMP-GCN	82.69	45.45	92.63	80.80	62.79	90.18	83.33	94.57	71.77
DTI	MMP-GCN	84.14	56.81	86.50	78.78	63.95	86.50	87.52	97.59	77.30
fMRI + DTI	Fusion method 1	88.05	63.63	94.47	83.10	80.23	85.27	87.88	91.56	84.04
fMRI + DTI	Fusion method 2	85.64	61.36	92.02	82.40	62.79	92.63	88.18	87.34	88.95
fMRI + DTI	Fusion method 3	89.04	59.09	97.54	89.18	79.06	94.47	88.48	89.15	87.73
fMRI + DTI	Ours	93.23	77.27	97.54	91.16	82.55	95.70	94.22	92.77	95.70
Modality	Method	SMC vs. EMCI			SMC vs. LMCI			EMCI vs. LMCI		
		ACC (%) SPE(%)		SEN (%)	ACC (%)	SEN (%)	SPE(%)	ACC (%)	SEN (%)	SPE (%)
fMRI	MMP-GCN	84.62	96.51	61.36	87.14	95.78	54.54	85.4	94.57	67.44
DTI	MMP-GCN	76.92	80.23	70.45	84.76	90.96	61.36	84.68	93.97	66.27
fMRI + DTI	Fusion method 1	85.38	87.2	81.81	89.52	91.56	81.81	87.06	96.38	68.60
fMRI + DTI	Fusion method 2	87.69	87.20	88.63	87.62	95.18	59.09	87.06	92.16	76.74
fMRI + DTI	Fusion method 3	84.62	86.04	81.81	89.52	92.77	77.27	86.65	90.36	79.06
fMRI + DTI	Ours	91.53	100	75	95.71	100	79.54	92.46	93.97	89.53

TABLE III. CLASSIFICATION PERFORMANCE COMPARISON OF DIFFERENT CLASSIFIERS.

Method	NC vs. SMC			NC vs. EMCI			NC vs. LMCI		
	ACC(%)	SEN(%)	SPE(%)	ACC(%)	SEN(%)	SPE(%)	ACC(%)	SEN(%)	SPE(%)
MLP	82.60	43.18	93.25	72.69	44.18	87.73	80.24	80.72	79.75
SVM	79.24	36.37	90.79	77.15	42.08	95.70	79.63	74.56	84.89
GCN	86.47	54.54	95.09	85.54	76.74	90.18	87.53	80.72	94.47
HGNN	82.07	40.91	93.25	76.40	56.98	86.50	81.78	90.36	73
Graph U-Net	85.57	36.37	98.77	79.47	51.16	94.47	82.98	100	65.64
Graphsage	84.71	31.82	98.77	82.38	54.65	96.93	82.41	82.53	82.21
GAT	83.71	47.73	93.25	73.97	58.14	82.21	79.07	74.10	84.05
AGNN	85.57	50	95.09	86.77	69.77	95.71	86.93	89.16	84.66
AM-GCN	87.12	56.82	95.09	83.58	61.63	95.09	88.14	80.74	95.71
M-GCN	89.37	70.45	94.47	85.94	75.58	91.41	90.57	90.96	90.18
MM-GCN	91.30	63.63	98.77	90.76	83.72	94.47	90.88	91.56	90.18
MMP-GCN	93.23	77.27	97.54	91.16	82.55	95.70	94.22	92.77	95.70
Method	SMC vs. EMCI			SMC vs. LMCI			EMCI vs. LMCI		
	ACC(%)	SEN(%)	SPE(%)	ACC(%)	SEN(%)	SPE(%)	ACC(%)	SEN(%)	SPE(%)
MLP	73.07	90.69	38.63	83.33	96.38	34.09	67.85	80.12	44.18
SVM	67.69	86.04	31.81	82.86	95.18	36.36	65.95	81.32	36.04
GCN	84.61	93.02	68.18	88.57	94.57	65.90	85.71	90.36	76.74
HGNN	70.77	83.72	45.45	75.71	84.33	43.18	71.06	89.76	34.88
Graph U-Net	89.23	100	68.18	83.33	92.77	47.72	88.26	96.98	70.93
Graphsage	84.62	95.35	63.64	86.67	100	36.36	84.71	98.19	58.14
GAT	78.46	100	36.36	82.38	81.93	84.09	79.74	91.57	56.98
AGNN	84.62	91.86	70.45	88.57	96.99	56.82	87.74	94.58	74.42
AM-GCN	85.38	98.84	59.09	89.52	91.57	81.82	87.02	94.58	72.09
M-GCN	86.15	90.69	77.27	90.95	95.78	72.72	88.09	89.75	84.88
MM-GCN	89.23	96.51	75.00	92.38	99.39	65.90	88.49	93.37	79.06
MMP-GCN	91.53	100	75.00	95.71	100	79.54	92.46	93.97	89.53

TABLE IV. ALGORITHM COMPARISON WITH RELATED WORKS.

Ref.	Modality	Subject	Method	Task	ACC(%)	SEN(%)	SPE(%)
[10]	fMRI	50 MCI, 49 NC	WGraphSR, SVM	MCI vs. NC	88.8	88.0	89.8
[32]	fMRI	29 EMCI, 30 NC	Fused Multiple Graphical Lasso, SVM	EMCI vs. NC	79.6	75.8	70.0
[33]	fMRI	50 EMCI, 48 NC	Wck-CNN Framework	EMCI vs. NC	84.6	83.3	85.7
[34]	fMRI	191 MCI, 179 NC	PCC, Multiple Triplet GCN	MCI vs. NC	86	86.9	85.1
[35]	fMRI	65 MCI, 51 NC	Multiview Feature Learning with Multiatlas-based FC Networks, SVM	MCI vs. NC	85.5	/	/
[36]	fMRI	145 LMCI, 165 EMCI	Spatial-temporal Convolutional Recurrent Neural Network	LMCI vs. EMCI	79.3	80.9	77.0
[37]	fMRI	53 LMCI, 72 EMCI, 64 NC	Dynamic Effective Connectivity, Deep Feature Aggregation Network	NC vs. EMCI EMCI vs. LMCI	90.9 89.8	90.4 87.6	91.4 91.4
[9]	rs-fMRI+ ASL-fMRI	28 MCI, 33 NC	Multi-modality Hyper-connectivity Networks, SVM	MCI vs. NC	86.9	82.1	90.9
[11]	fMRI+DTI	36 MCI, 37 NC	Adaptive Dynamic Functional Connectivity, SVM	MCI vs. NC	87.7	88.9	86.5
[12]	fMRI+DTI	38 LMCI, 44 EMCI, 44 SMC, 44 NC		NC vs. SMC NC vs. EMCI	82.9 85.2	88.6 86.3	77.2 84.1

			Low-Rank Self-calibrated Brain Network, Joint Non-Convex Multi-Task Learning, SVM	NC vs. LMCI	87.8	84.2	90.9
				SMC vs. EMCI	84.0	81.8	86.3
				SMC vs. LMCI	90.2	89.4	90.9
				EMCI vs. LMCI	81.7	78.9	84.0
[13]	fMRI+DTI	38 LMCI, 44 EMCI, 44 SMC, 44 NC	PCC,Similarity-aware Adaptive Calibrated GCN	NC vs. SMC	84.9	88.6	79.5
				NC vs. EMCI	85.2	90.9	79.5
				NC vs. LMCI	89.0	89.4	88.6
				SMC vs. EMCI	88.6	95.4	81.8
				SMC vs. LMCI	87.8	84.2	90.9
				EMCI vs. LMCI	85.5	92.1	81.8
Ours	fMRI+DTI	166 LMCI, 86 EMCI, 44 SMC, 163 NC	Dual-modality Fused Brain Connectivity Network, Multi-center and Multi-channel Pooling GCN	NC vs. SMC	**93.2**	77.2	97.5
				NC vs. EMCI	**91.1**	82.5	95.7
				NC vs. LMCI	**94.2**	92.7	95.7
				SMC vs. EMCI	**91.5**	100	75.9
				SMC vs. LMCI	**95.7**	100	79.5
				EMCI vs. LMCI	**92.4**	93.7	89.5

B. Performance of Multi-center and Multi-channel Pooling GCN

We propose three mechanisms to improve the current GCN classifiers: a multi-center attention graph, a multi-channel mechanism, and a novel pooling mechanism. To describe our method, we use M-GCN to represent the traditional GCN with our multi-center attention graph, MM-GCN to represent the traditional GCN with our multi-center attention graph and multi-channel mechanism, MMP-GCN to represent the traditional GCN with our above three mechanisms. Other nine related popular frameworks are also implemented on our dataset, including multiple layer perception (MLP), support vector machine (SVM), GCN [24], HGNN [38], Graph U-Net [28], Graphsage [39], GAT [40], AGNN [41], and AM-GCN [42]. All experiments are based on the constructed dual-modality fused brain connectivity network. Experimental results are given in Table III. GCN parameters are set as above, MLP parameters follow the GCN implementation. SVM is based on the Scikit-learn library. For GCN variants (HGNN, Graph U-Net, Graphsage, GAT, AGNN, and AM-GCN), we get their results based on the implementations released by the corresponding source code. Their fundamental parameters (e.g., dropout rate, regularization, learning rate, number of epochs, and number of neurons per layer) are the same with GCN.

Table III shows that the traditional classifiers' performance (MLP and SVM) is poor, where the mean ACC of six tasks based on MLP and SVM are 76.6% and 75.4%, respectively. Compared with the two traditional methods, GCN achieves better performance with ACC increased by 9.7%. This result follows the previous work [24], which also validates the effectiveness of GCN. The proposed multi-center attention graph considers the multi-center source, disease status information of those training samples, gender, and equipment type. Also, it uses an attention mechanism to assign them appropriate amplification factors adaptively. By comparing M-GCN with GCN, Table III shows the mean ACC, SEN, and SPE of the six tasks increase by 2.1%, 3.8%, and 3.4%, respectively. These results indicate that our multi-center attention graph is effective. Based on the statistical information of features, we divide features into several parts and then design different filters for filtering. Using the multi-channel mechanism, experimental results show the mean ACC and SEN increase by 2.0% and 2.4%, respectively, whereas SPE decreases by 1.2%. To improve pooling operation, we introduce the disease status of training samples into it. By comparing MMP-GCN with MM-GCN, experimental results show an increase by 2.5%, 3.1%, and 4.9% in the mean ACC, SEN, SPE, respectively. The above results validate the effectiveness of our three mechanisms on GCN. Finally, using GCN with our three mechanisms, the mean ACC, SEN, and SPE of six tasks increase by 6.6%, 9.4%, and 7.1%, respectively. In our six classification tasks, the performance of SMC vs. LMCI reports the highest improvement (ACC increases by 7.1%). In contrast, NC vs. EMCI gets the lowest improvement (ACC increases by 5.6%). Finally, the mean ACC of six tasks reaches 93.1%.

For those popular GCN variants (HGNN, Graph U-Net, Graphsage, GAT, AGNN, and AM-GCN), the classification accuracy has big differences. For example, AGNN and AM-GCN show good performance, with the mean ACC of six tasks reaching 86.7%, whereas HGNN and GAT show bad performance, with the mean ACC of six tasks reaching 76.3% and 79.5%. GAT utilizes trainable or learned networks to construct an adjacency matrix, which increases network complexity and results in performance degradation for our task with a small dataset. Our method has a simpler structure with fewer parameters than these methods. More details and comparisons of the above variants are provided in the discussion section.

C. Comparison with Related Works

Table IV compares the final classification performance between the proposed method and related ones. Related works mainly focus on the study of brain network construction and classifier. Brain network construction methods include hyper-network, dynamic effective connectivity network, low-rank self-calibrated brain network, PCC method, fused multiple graphical Lasso method, WGraphSR method, etc. These traditional methods try to describe brain networks from different aspects. Although they have achieved good performance, the ACC is still limited by insignificant features of MCI. Table IV shows their ACC ranges from 72.8% to 90.9%. Various multi-modality data fusion methods are widely studied (e.g., fusion fMRI and ASL-fMRI data, fusion fMRI

and DTI data). Their experimental results show that the classification performance based on fused data is better. Unlike traditional brain network construction methods above, we introduce a DTI-strength penalty term in the sparse representation to get the dual-modality fused brain connectivity network.

Compared to the traditional classifiers (e.g., SVM and MLP), deep learning (e.g., GCN) shows better performance. Table IV shows that the mean ACC based on GCN can reach 86.8% in our earlier work [13], whereas the traditional classifiers reach 74.3%. This result shows that the filter effect of GCN can significantly improve classification performance. Based on the advantage of GCN, we propose three mechanisms and name the novel method as MMP-GCN. Table IV shows that our method can achieve the best performance.

Unlike current works, our work also studies the heterogeneity between different datasets, where we combine the data from ADNI 2, ADNI 3 and in-house. Compared to the current GCNs that set the impacts of non-image information as constants to amplify edge weights, we adopt an attention mechanism to combine non-image information. The impacts of the above information on edge weights are learned by network training. Compared to other works that mainly study the classification task of MCI vs. NC, we study more tasks. Specifically, NC vs. SMC, NC vs. EMCI, NC vs. LMCI, SMC vs. EMCI, SMC vs. LMCI, and EMCI vs. LMCI are included. Besides, compared to the limited samples in related works (ranging from 59 to 370), our work has more samples (457).

Our earlier work [12] concatenated fMRI functional features and DTI structural features. We adopted multi-task learning for feature selection, and finally used SVM to get the final classification results. In [13], we designed a dual-modality GCN (combining fMRI and DTI data) for classification. In this work, a structural connectivity penalty term is proposed in constructing functional connectivity networks and then using the fused network (i.e., functional connectivity network) to realize classification. Compared to the related works, our MMP-GCN achieves state-of-art performance. Compared to dual-modality GCN framework, it has more stable discriminative features and a simpler classifier structure. Moreover, the multi-center information, the difference between features, and the pooling effect are further considered.

V. Discussion

A. Analysis of Regularized Parameter

There is only one regularized parameter, named λ, in constructing our fused brain connectivity network to control the network sparsity. To study its effect on classification performance, we implement an experiment with λ varying from 2^{-10} to 2^{0} based on our MMP-GCN classifier. The experimental results are shown in Fig. 7.

The ACC varies by setting different λ values, and the highest ACC is usually acquired with λ set as 2^{-6} - 2^{-4}. Specifically, SMC vs. EMCI and SMC vs. LMCI get their highest ACC with λ set as 2^{-4}, NC vs. SMC, NC vs. EMCI, and NC vs. LMCI get their highest ACC with λ set as 2^{-5}, EMCI vs. LMCI gets its highest ACC with λ set as 2^{-6}. Generally, the mean ACC reaches the highest value with λ set as 2^{-5}. The parameter λ can result in a mean 6.2% fluctuation on ACC, which indicates that our proposed method plays an important role in brain connectivity network construction. Compared to other brain network construction methods, our method has a simple structure.

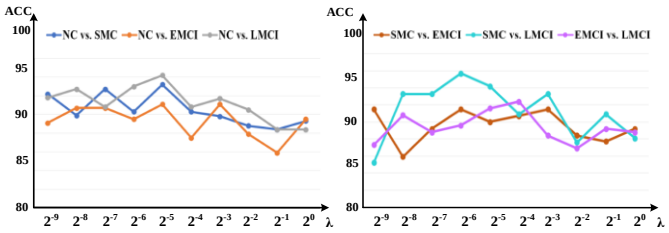

Fig. 7. Effect of the regularized parameter λ on ACC of six tasks.

B. Analysis of Non-image Information

In earlier GCN studies, non-image information is an important factor to affect classification performance, where the edge weight between the nodes with the same information is assigned a large value. For example, the experimental results in work [24] showed that considering the gender and equipment type in GCN can cause 3% ACC improvement for AD and ASD prediction. The results in work [13] indicated that they can cause a mean 7.1% improvement on ACC for SMC and MCI prediction. Besides, different data sources (multi-center studies) also affect the performance of ASD prediction [43, 44]. In this subsection, we evaluate the effect of non-image information on ACC. Their effect is shown in Fig. 8 and attention coefficients in experiments (i.e., a_0, a_g, a_e, a_c, and a_s in Eq. (9)) are listed in Table V.

TABLE V: ATTENTION COEFFICIENTS IN DIFFERENT TASKS.

Tasks/Atten coef	a_s	a_g	a_e	a_c	a_0
NC vs. SMC	0.202	0.194	0.209	0.211	0.182
NC vs. EMCI	0.203	0.191	0.213	0.208	0.179
NC vs. LMCI	0.209	0.192	0.206	0.205	0.183
SMC vs. EMCI	0.211	0.187	0.199	0.217	0.181
SMC vs. LMCI	0.204	0.185	0.210	0.211	0.184
EMCI vs. LMCI	0.196	0.189	0.212	0.214	0.184

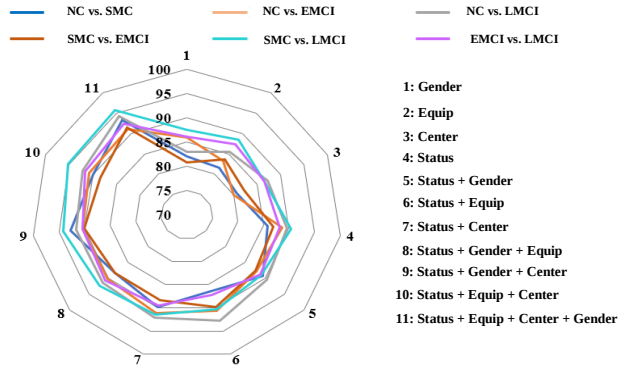

Fig. 8. Effect of non-image information on ACC of six tasks.

For our six classification tasks, experimental results show that integrating single information (gender, equipment type, multi-center information, and disease status information) can achieve mean ACC of 84.2%, 84.9%, 83.8%, and 88.2%, respectively. It indicates that disease status information has

much more influence on performance improvement. By integrating all the above information, experimental results show that the mean ACC can reach 93.1%. Table V shows we can get different attention coefficients in the training process and there is a little difference between different tasks. The above experimental results indicate that the above non-image information are important factors affecting classification performance, and integrating them all on a graph can achieve the best performance.

C. Analysis of the Number of Channels

As illustrated in our earlier work [13], the selected features have different noise characteristics. This study presents a multi-channel mechanism to improve the filtering process, which splits the selected features into multiple parts based on their statistics and then designs the corresponding filters to adapt to their characteristics. We evaluate the effect of the number of channels on the performance, and the experimental results are shown in Fig. 9.

The highest ACC is usually achieved with the number of channels set as 4 or 5. For example, NC vs. SMC, NC vs. EMCI and SMC vs. LMCI get their highest ACC with the number set as 4. NC vs. LMCI, SMC vs. EMCI and EMCI vs. LMCI get their highest ACC with the number set as 5. The mean ACC of the six tasks gets the highest value with the number set as 5. Setting the number of channels from 1 to 9 shows a mean 4.5% ACC change between the best and worst performance in the six tasks. The above results validate the effectiveness of our multi-channel mechanism and indicate that considering the difference between features in the filter plays an important role in performance improvement. The number of trainable parameters is also an important factor, which is used to evaluate the network complexity. Supposing that there are N subjects with each having M features and Q1 neurons per layer. Hence, the total parameters' number in a traditional single-layer GCN is $M \times Q1$. In contrast, the total parameters' number in a single-layer multi-channel GCN (Q2 channels) is $((M/Q2) \times Q1) \times Q2 = M \times Q1$. This show no increase in the number of total parameters by using our multi-channel mechanism.

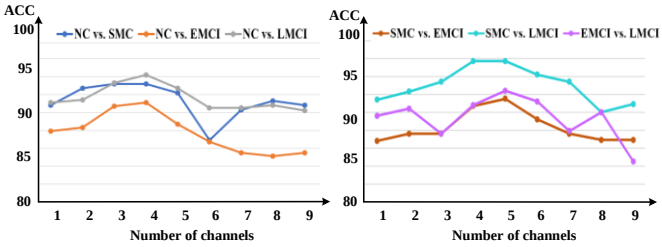

Fig. 9. Effect of the number of channels on ACC of six tasks.

D. Analysis of Pooling Strategy

We propose a simple and untrained selection method for edge pooling to reduce the number of edges on the graph to boost filter robustness. Based on the proposed MMP-GCN, we adjust the number of retained edges by pooling 0-30% of them to test its effect, and the effect on performance is shown in Fig. 10.

Experimental results report that the best pooling performance is usually obtained while setting the pooling rate as 10% or 15%. Specifically, NC vs. EMCI and SMC vs. EMCI get their highest ACC by setting the pooling rate as 10%. NC vs. SMC, NC vs. EMCI, NC vs. LMCI and SMC vs. LMCI get their highest ACC with setting the pooling rate as 15%. The mean ACC of six tasks gets the highest value with setting the pooling rate as 15%. Using our pooling mechanism by setting the best pooling rate, the mean ACC gets up to 2.3% improvement, validating our pooling mechanism's effectiveness.

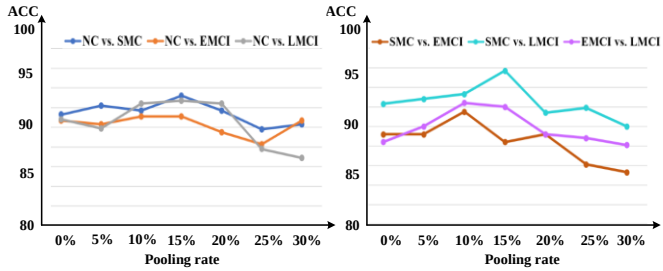

Fig. 10. Effect of the pooling rate on ACC of six tasks.

E. Most Discriminative Connectivity Features and Related ROIs

When constructing the dual-modality brain connectivity network, well-known and highly-correlated ROIs in AD/MCI disease are used [45-48]. For example, the inferior temporal gyrus (ITG.R), insula (INS.R), olfactory cortex (OLF.L), angular gyrus (ANG.L), amygdala (AMYG.R) and precuneus (PCUN.R). Specifically, the ITG.R shows the most important region in our tasks related to advanced cognitive functions [45]. Table VI lists the top 10 most related ROIs for our six classification tasks.

Based on fMRI or DTI data, the top 10 related ROIs are partly related to early AD. Whereas most of the top 10 related ROIs extracted from our dual-modality fused connectivity network are associated with early AD [46-50]. Specifically, the hippocampus (HIP.L), the extracted putamen lentiform (PUT.L), amygdala (AMYG.L) and middle occipital gyrus (MOG.R) in the top 10 ROIs based on fMRI data are related to early AD disease. The extracted INS.R, posterior central gyrus (PoCG.R), PCUN.R and supplementary motor area (SMA.R) in the top 10 ROIs based on DTI data are associated with early AD disease. The above results indicate that the proposed dual-modality fused brain network construction method is more efficient in finding discriminative features. For different classification tasks, we show the top 20 most discriminative connectivity features and the top 10 ROIs in Fig. 11. It shows there exist differences among different classification tasks.

TABLE VI: TOP 10 MOST DISCRIMINATIVE ROIS.

DTI		fMRI		Ours (Fusion)	
ROI index	ROI abbr	ROI index	ROI abbr	ROI index	ROI abbr
86	MTG.R	24	SFGmed.R	90	ITG.R
87	TPOmid.L	37	HIP.L	30	INS.R
59	SPG.L	71	CAU.L	21	OLF.L
30	INS.R	61	IPL.L	61	IPL.L
78	THA.R	73	PUT.L	65	ANG.L
58	PoCG.R	41	AMYG.L	42	AMYG.R
68	PCUN.R	84	TPOsup.R	68	PCUN.R
75	PAL.L	87	TPOmid.L	40	PHG.R
56	FFG.R	36	PCG.R	74	PUT.R
20	SMA.R	52	MOG.R	78	THA.R

TABLE VII: CROSS-VALIDATION RESULTS.

Training	Test	Model	ACC(%)	SEN(%)	SPE(%)	AUC
ADNI 2 and ADNI 3	In-house	MLP	63.31	69.69	54.28	0.6271
		SVM	65.08	66.66	62.85	0.7101
		GCN	67.45	67.67	67.14	0.6391
		Ours	74.55	87.87	55.71	0.7251
ADNI 2 and In-house	ANDI 3	MLP	64.70	60.52	67.18	0.6978
		SVM	67.64	68.42	67.18	0.6809
		GCN	69.60	76.31	65.62	0.7257
		Ours	74.50	52.63	87.50	0.7405
ADNI 3 and In-house	ANDI 2	MLP	65.45	61.53	68.96	0.6618
		SVM	67.27	65.38	68.96	0.6658
		GCN	70.90	69.23	72.41	0.6883
		Ours	78.18	73.07	82.75	0.7772

F. Cross-validation on Three Datasets

To further verify the robustness of our model, we perform cross-validation on our three datasets. The used data in this subsection includes 163 NC and 166 LMCI. Specifically, 29 NC and 27 LMCI in ADNI 2, 64 NC and 40 LMCI in ADNI 3, 70 NC and 99 LMCI in in-house dataset. Three experiments are performed and their results are shown in Table VII.

As Table VII shows, there are different classification performances by setting distinct datasets for training and test. Specifically, based on our brain network construction method and MMP-GCN classifier, the ACC in the three tasks are 74.5%, 74.5% and 78.2%, respectively. Compared to the mean ACC limited to 64.4%, 66.6%, and 69.3% for MLP, SVM, and GCN, our method performs best performance with a mean ACC of 75.7%. The above results verify the effectiveness of our method in the cross-validation task.

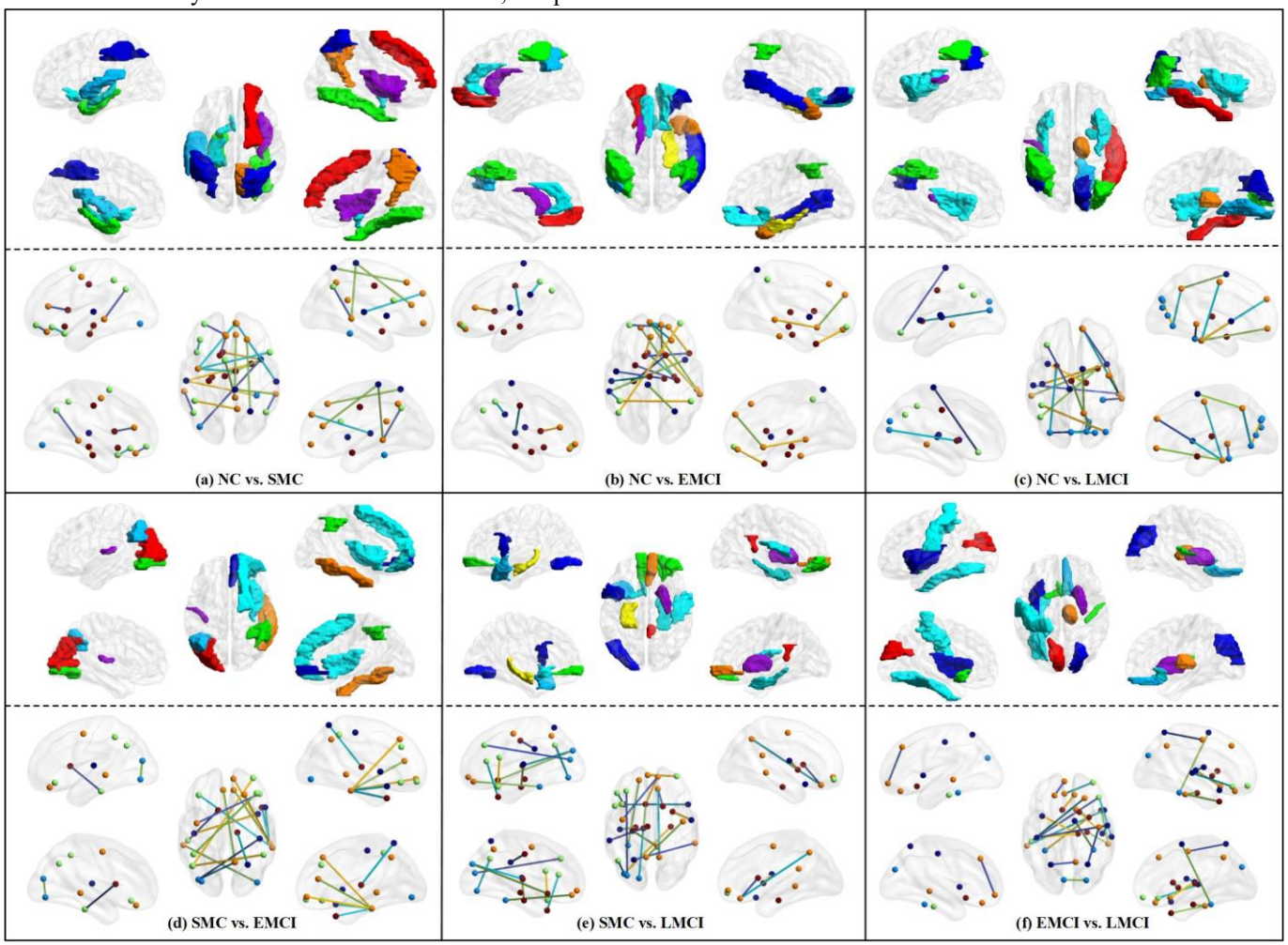

Fig. 11. Top 10 most discriminative ROIs and top 20 most discriminative connectivity features in our dual-modality fused brain connectivity network.

G. GCN related works

Many works are devoted to improving GCN. The popular GCN variants include HGNN [38], Graph U-Net [28], Graphsage [39], GAT [40], and AGNN [41]. We compare them and show their differences in this subsection.

AGNN is a simple GCN variant, which mainly updates the adjacency matrix in every propagation layer by computing feature similarity. Updating the adjacency matrix in propagation increases the difficulty of network training and results in performance fluctuation. HGNN designs a hyperedge convolution operation for learning the hidden layer representation by considering the high-order data structure. Compared to the GCN that establishes pairwise connections, a hypergraph can encode high-order data correlation using degree-free hyperedges. HGNN has a simple structure without increasing network parameters. Graph U-Net proposes graph pooling (gPool) and unpooling (gUnpool) operations. The gPool layer adaptively selects some nodes to form a smaller graph based on their scalar projection values on a trainable projection vector. The gUnpool layer (the inverse operation of the gPool layer) restores the graph into the original structure.

Graphsage trains a set of aggregator functions that aggregates feature information from a node's local neighborhood. The aggregator function is trainable and maintains high representational capacity. The aggregator function and concatenation operation increase the total number of parameters. GAT uses learnable linear transformation (i.e., a trainable weight matrix) and a single-layer feedforward neural network (i.e., a trainable weight vector) to construct an adjacency matrix, which significantly increases the network complexity. Similar to GAT, self-attention and Transformer also use learnable linear transformation (i.e., three trainable weight matrices) to realize filtering, increasing network complexity significantly. Compared to the above GCN-related methods, our attention mechanism with only five trainable parameters (i.e., a_0, a_g, a_e, a_c, a_s) and our pooling mechanism without trainable parameters is very simple, and therefore fits in the few-shot learning task.

H. Limitations and Future Work

Still several limitations need to be considered. 1) The effect of multi-center and phenotypic information on features can be further evaluated and considered in GCN. 2) The proposed multi-channel mechanism increases the complexity of the GCN classifier to some extent. 3) The proposed dual-modality fusion method ignores the condition of incomplete multi-modality neuroimages. We will further improve the GCN classifier and solve the limitation of incomplete multi-modality neuroimages in our future work.

VI. Conclusion

We propose to use structural connectivity strength to construct the functional connectivity network, which realizes the fusion of dual-modality imaging data (fMRI and DTI). Its better performance than the other popular fusion methods indicates that stronger structural connectivity among ROIs implies better discriminative functional connectivity feature in MCI and SMC classification tasks. By analyzing the effect of multi-center, disease status, gender, and equipment type information on classification performance, we find that the above information has different impacts on performance. Considering disease status information of those training samples in graph construction shows the most obvious performance improvement. The effect of multi-center information validates the feature heterogeneity among imaging data acquired from different medical centers. Our attention mechanism is proposed to combine the above information into graph construction by considering their different impacts, which shows good performance. This result also verifies the efficiency of the graph theory. By analyzing the effect of the number of channels, we observe that designing multiple filters to consider feature differences is effective. By analyzing the pooling rate in our pooling mechanism, we find it could obtain up to 2.3% improvement on ACC. This result validates that including those nodes with low-quality features to perform convolution will degrade the filter performance. Generally, the proposed MMP-GCN method mainly improves the filter performance, resulting in our good classification performance. The proposed method can be also applied to other classification tasks.

References

[1] A. A. Report, "2020 Alzheimer's disease facts and figures," *Alzheimer's and Dementia*, vol. 16, no. 3, Mar. 2020.

[2] M. Heron, "Deaths: Leading causes for 2010," *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System,* vol. 62, no. 6, pp. 1-97, Dec. 2013.

[3] R. C. Petersen *et al.*, "Practice guideline update summary: Mild cognitive impairment," *Neurology,* vol. 90, no. 3, pp. 126-135, Jan. 2018.

[4] S. Gauthier *et al.*, "Mild cognitive impairment," *Lancet,* vol. 367, no. 9518, pp. 1262-1270, Apr. 2006.

[5] P. Yang *et al.*, "Fused sparse network learning for longitudinal analysis of mild cognitive impairment," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 233-246, Sep. 2019.

[6] Z. Huang *et al.*, "Parkinson's disease classification and clinical score regression via united embedding and sparse learning from longitudinal data," *IEEE Trans. Neur. Net. Learn. Sys.*, pp. 1-15, Feb. 2021.

[7] K. Hett, V. Ta, I. Oguz, J. V. Manjón, P. Coupé, and A. D. N. Initiative, "Multi-scale graph-based grading for Alzheimer's disease prediction," *Med. Image Anal.*, vol. 67, p. 101850, Jan. 2021.

[8] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 42, no. 4, pp. 880-893, Apr. 2020.

[9] Y. Li *et al.*, "Multimodal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification," *Med. Image Anal.*, vol. 52, pp. 80-96, Feb. 2019.

[10] R. Yu, L. Qiao, M. Chen, S. Lee, X. Fei, and D. Shen, "Weighted graph regularized sparse brain network construction for MCI identification," *Pattern Recognit.,* vol. 90, pp. 220-231, Jun. 2019.

[11] Y. Li, J. Liu, Z. Tang, and B. Lei, "Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification," *IEEE Trans. Med. Imaging,* vol. 39, no. 9, pp. 2818-2830, Sep. 2020.

[12] B. Lei *et al.*, "Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease," *Med. Image Anal.*, vol. 61, p. 101652, Apr. 2020.

[13] X. Song *et al.*, "Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction," *Med. Image Anal.*, vol. 69, p. 101947, Apr. 2021.

[14] J. Gonzalez-Castillo *et al.*, "Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns," *Proc. Nat. Acad. Sci. (PNAS),* vol. 112, no. 28, pp. 8762-8767, Jun. 2015.

[15] R. Yu, H. Zhang, L. An, X. Chen, Z. Wei, and D. Shen, "Connectivity strength-weighted sparse group representation-based brain network construction for MCI classification," *Hum. Brain Mapp.*, vol. 38, no. 5, pp. 2370-2383, May 2017.

[16] Y. Zhang *et al.*, "Strength and similarity guided group-level brain functional network construction for MCI diagnosis," *Pattern Recognit.*, vol. 88, pp. 421-430, Apr. 2019.

[17] L. Qiao, H. Zhang, M. Kim, S. Teng, L. Zhang, and D. Shen, "Estimating functional brain networks by incorporating a modularity prior," *NeuroImage,* vol. 141, pp. 399-407, Nov. 2016.

[18] Z. Zhou *et al.*, "A toolbox for brain network construction and classification (BrainNetClass)," *Hum. Brain Mapp.*, vol. 41, no. 10, pp. 2808-2826, Mar. 2020.

[19] J. S. Damoiseaux and M. D. Greicius, "Greater than the sum of its parts: A review of studies combining structural connectivity and resting-state functional connectivity," *Brain Struct. Funct.*, vol. 213, no. 6, pp. 525-533, Oct. 2009.

[20] C. J. Honey, J. P. Thivierge, and O. Sporns, "Can structure predict function in the human brain?," *NeuroImage,* vol. 52, no. 3, pp. 766-776, Sep. 2010.

[21] Z. Wang *et al.*, "Distribution-guided network thresholding for functional connectivity analysis in fMRI-based brain disorder identification," *IEEE J. Biomedical Health Informat.*, vol. 26, no. 4, pp. 1602-1613, Apr. 2021.

[22] H. Guan, E. Yang, P. T. Yap, D. Shen, and M. Liu, "Attention-guided deep domain adaptation for brain dementia identification with multi-site neuroimaging data," in *Med. Image Comput. Comput. Assisted Intervention (MICCAI)*, Lima, Peru (Virtual), 2020, vol. 12444, pp. 31-40.

[23] H. Guan, Y. Liu, E. Yang, P. T. Yap, D. Shen, and M. Liu, "Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification," *Med. Image Anal.*, vol. 71, p. 102076, Jul. 2021.
[24] 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.
[25] Y. Zhang, L. Zhan, W. Cai, P. Thompson, and H. Huang, "Integrating heterogeneous brain networks for predicting brain disease conditions," in *Med. Image Comput. Comput. Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, vol. 11767, pp. 214-222.
[26] S. I. Ktena *et al.*, "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431-442, Apr. 2018.
[27] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, 2017.
[28] H. Gao and S. Ji, "Graph u-nets," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, United states, 2019, pp. 2083-2092.
[29] R. Ying, C. Morris, W. L. Hamilton, J. You, X. Ren, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," in *Adv. Neural Inf. Proces. Syst. (NeurIPS)*, Montreal, QC, Canada, 2018, vol. 2018, pp. 4800-4810.
[30] Y. Ma, S. Wang, C. C. Aggarwal, and J. Tang, "Graph convolutional networks with eigenpooling," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Anchorage, AK, United states, 2019, pp. 723-731.
[31] Z. Huang, Z. Zhu, C. H. Yau, and K. C. Tan, "Identifying autism spectrum disorder from resting-state fMRI using deep belief network," *IEEE Trans. Neur. Net. Learn. Sys.* , vol. 32, no. 7, pp. 2847-2861, Jul. 2020.
[32] C. Y. Wee, S. Yang, P. T. Yap, and D. Shen, "Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification," *Brain Imaging Behav.*, vol. 10, no. 2, pp. 342-356, Jun. 2016.
[33] B. Jie, M. Liu, C. Lian, F. Shi, and D. Shen, "Designing weighted correlation kernels in convolutional neural networks for functional connectivity based brain disease diagnosis," *Med. Image Anal.*, vol. 63, p. 101709, Jul. 2020.
[34] 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. Imaging*, vol. 40, no. 4, pp. 1279-1289, Apr. 2021.
[35] Y. Zhang, H. Zhang, E. Adeli, X. Chen, M. Liu, and D. Shen, "Multiview feature learning with multiatlas-based functional connectivity networks for MCI diagnosis," *IEEE Trans. Cybern.* , pp. 1-12, Dec. 2020.
[36] M. Wang, C. Lian, D. Yao, D. Zhang, M. Liu, and D. Shen, "Spatial-temporal dependency modeling and network hub detection for functional MRI analysis via convolutional-recurrent network," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 8, pp. 2241-2252, Aug. 2020.
[37] Y. Li, J. Liu, Y. Jiang, Y. Liu, and B. Lei, "Virtual adversarial training based deep feature aggregation network from dynamic effective connectivity for MCI identification," *IEEE Trans. Med. Imaging*, vol. 41, no. 1, pp. 237-251, Jan. 2022.
[38] J. Jiang, Y. Wei, Y. Feng, J. Cao, and Y. Gao, "Dynamic hypergraph neural networks," in *Proc. 28th Inter. Joint Conf. Artif. Intell. (IJCAI)*, Macao, China, 2019, pp. 2635-264.
[39] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Adv. Neural Inf. Proces. Syst. (NeurIPS)*, Long Beach, CA, United states, 2017, vol. 30, pp. 1025-1035.
[40] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, 2018, pp. 1-12.
[41] K. K. Thekumparampil, C. Wang, S. Oh, and L. Li, "Attention-based graph neural network for semi-supervised learning," in *Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, 2018.
[42] X. Song *et al.*, "Augmented multi-center graph convolutional network for COVID-19 diagnosis," *IEEE Trans. Ind. Inform.*, vol. 17, no. 9, pp. 6499-6509, Sep. 2021.
[43] M. Wang, D. Zhang, J. Huang, P. T. Yap, D. Shen, and M. Liu, "Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation," *IEEE Trans. Med. Imaging*, vol. 39, no. 3, pp. 644-655, Mar. 2020.
[44] J. Wang *et al.*, "Multi-class ASD classification based on functional connectivity and functional correlation tensor via multi-source domain adaptation and multi-view sparse representation," *IEEE Trans. Med. Imaging*, vol. 39, no. 10, pp. 3137-3147, Oct. 2020.
[45] H. Chertkow and D. Bub, "Semantic memory loss in dementia of Alzheimer's type: What do various measures measure?," *Brain*, vol. 113, no. 2, pp. 397-417, Apr. 1990.
[46] F. Lin *et al.*, "Insula and inferior frontal gyrus' activities protect memory performance against Alzheimer's disease pathology in old age," *J. Alzheimer's Dis.*, vol. 55, no. 2, pp. 669-678, Jan. 2017.
[47] L. Xu *et al.*, "Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers," *J. Alzheimer's Dis.*, vol. 51, no. 4, pp. 1045-1056, Apr. 2016.
[48] Z. Qi *et al.*, "Impairment and compensation coexist in amnestic MCI default mode network," *NeuroImage*, vol. 50, no. 1, pp. 48-55, Mar. 2010.
[49] Y. Li *et al.*, "Abnormal resting-state functional connectivity strength in mild cognitive impairment and its conversion to Alzheimer's disease," *Neural Plast.*, vol. 2016, p. 4680972, Jan. 2016.
[50] S. E. Rose *et al.*, "Diffusion indices on magnetic resonance imaging and neuropsychological performance in amnestic mild cognitive impairment," *J. Neurol., Neurosurg. Psychiatry*, vol. 77, no. 10, pp. 1122-1128, Oct. 2006.