Hierarchical Graph Convolutional Network Built by Multiscale Atlases for Brain Disorder Diagnosis Using Functional Connectivity

Mianxin Liu, Han Zhang, Senior Member, IEEE, Feng Shi, and Dinggang Shen, Fellow, IEEE

Abstract—Functional connectivity network (FCN) data from functional magnetic resonance imaging (fMRI) is increasingly used for the diagnosis of brain disorders. However, state-of-the-art studies used to build the FCN using a single brain parcellation atlas at a certain spatial scale, which largely neglected functional interactions across different spatial scales in hierarchical manners. In this study, we propose a novel framework to perform multiscale FCN analysis for brain disorder diagnosis. We first use a set of well-defined multiscale atlases to compute multiscale FCNs. Then, we utilize biologically meaningful brain hierarchical relationships among the regions in multiscale atlases to perform nodal pooling across multiple spatial scales, namely “Atlas-guided Pooling (AP).” Accordingly, we propose a multiscale-atlases-based hierarchical graph convolutional network (MAHGCN), built on the stacked layers of graph convolution and the AP, for a comprehensive extraction of diagnostic information from multiscale FCNs. Experiments on neuroimaging data from 1792 subjects demonstrate the effectiveness of our proposed method in the diagnoses of Alzheimer’s disease (AD), the prodromal stage of AD [i.e., mild cognitive impairment (MCI)], as well as autism spectrum disorder (ASD), with the accuracy of 88.9%, 78.6%, and 72.7%, respectively. All results show significant advantages of our proposed method over other competing methods. This study not only demonstrates the feasibility of brain disorder diagnosis using resting-state fMRI empowered by deep learning but also highlights that the functional interactions in the multiscale brain hierarchy are worth being explored and integrated into deep learning network architectures for a better understanding of the neuropathology of brain disorders. The codes for MAHGCN are publicly available at “https://github.com/MianxinLiu/MAHGCN-code.”

Index Terms—Brain disorder, brain multiscale hierarchy, functional connectivity network (FCN), graph convolutional neural network.

I. INTRODUCTION

Brain disorders, such as Alzheimer’s disease (AD) and autism spectrum disorder (ASD), propose a severe challenge to global public health. AD could induce first mild cognitive impairment (MCI) and eventually dementia in elders. Also, ASD typically affects children’s social interaction and communication abilities. As of 2020, 55 million people are living with dementia, and over 75 million people are affected by ASD worldwide according to the statistics of the World Health Organization (https://www.who.int/news-room/fact-sheets). An accurate identification as early as possible is necessary to enable effective interventions, which improve the patient’s prognosis. However, there is still a lack of clear understanding of the pathology of these brain disorders and well-established method for the diagnosis.

Recent work focuses on diagnosing human brain disorders by imaging the brain dynamics with in vivo brain functional imaging techniques. Among many of these techniques, resting-state functional magnetic resonance imaging (rs-fMRI) is a popular neuroimaging method that can probe spontaneous neural activities by measuring fluctuations in blood-oxygen-level-dependent (BOLD) signals. In the rs-fMRI studies, functional connectivity (FC) is the most prevalent measurement to quantitatively characterize spatiotemporal covariation patterns in the brain. Typical FC measures the temporal synchrony of neural activities across spatially distant regions using pairwise Pearson correlations. The FC characteristics among all pairs of regions further result in a whole-brain FC network (FCN). Many other FCN-constructing methods, such as the 3-D maps generated by seed-based correlations or independent component analyses (ICA), can also detect neural synchronizations. FCN was found to be sensitive to capture functional changes induced by brain disorders and has been used in individualized
diagnosis for various brain disorders (see systematic reviews in [1]).

In FCN-based diagnostic studies, a lasting debate is about how to efficiently extract disorder-related features from FCNs. Conventional studies typically reshaped FCNs into vectors and then inputted them to classical machine-learning-based classifiers, such as a support vector machine (SVM) [2], [3], [4]. However, such vectorizations may lose the disorder-related information carried on the topological structures of FCNs.Attributing to the rapid development of deep learning technology, graph convolution neural networks (GCNs) allow implementations of artificial neural network algorithms directly on graph structures, thus providing suitable representations for the disorder-related information of FCNs. Briefly, a typical GCN works on individual FCN and integrates nodal features along with the graph–topological neighbors based on the graph Laplacian and generates the representation for the entire FCN, based on which the brain disorders can be identified. Notably, another stream to apply GCN is to regard individual data as a node and interindividual similarity as a link. They perform nodal classification in the population-level graph to achieve the individual-level diagnosis. We call this type of GCN application as “nodal classification GCN (nc-GCN).” The GCN-based brain disorder diagnosis has earned great attention, in diagnosing ASD [5], [6], MCI, and AD [7], [8], [9].

However, so far, most of these previous studies, including those with GCNs, ignored crucial multiscale architectures of the brain networks [10], [11]. In the brain, the nodes at multiple spatial scales are often organized in a hierarchical modular structure, e.g., individual neurons often connect to each other to form a “neural mass,” many of which organize as neural circuits at a larger scale; and then, neural circuits of many functional subsystems further form large-scale functional networks through inter-subnetwork connectivities [12], [13]. Such multiscale hierarchical brain connectomes could provide much richer knowledge about high-level brain functioning. First, the multiscale brain networks are not simply duplicates of single-scale structures, but essentially different aspects of the brain system. It has been found that healthy brain connectomes at different scales exhibit different topological attributes, such as degree distribution [14], centrality [11], and small-worldness [15]. Second, multiscale hierarchical brain connectomes are significant for cognitive processing. Recent theoretical and empirical analyses have highlighted that the multiscale hierarchical brain network organization underpins the flexible switching between information integration and segregation [16], [17], and the sensitive response to external stimulus [18] for efficient brain functioning. Third, accumulating evidence revealed that brain disorders could modify FCN topology at a specific or multiple scale(s). A recent study specifically investigated modular structures of the lifespan FCNs at different spatial scales [19], indicating that the degree of segregation among network modules exhibits an interaction between age and spatial scales. The relationship or interaction “across” scales could also be affected by atypical brain states, as revealed by a recent study on “cross-scale” functional interactions (namely, “associated high-order FC”) under brain disorders [20]. Therefore, a multiscale FCN analysis is meaningful and necessary for comprehensive characterizations of brain systems and disorder statuses.

Indeed, the latest studies have started to focus on mining multiscale information from FCNs in data-driven manners [21], [22]. The advanced GCN studies have attempted to extract multiscale hierarchical topological features from the given graph, by learning various pooling operations on the nodes. Ying et al. [23] proposed a coarse-graining-based pooling “DIFFPOOL” and built a hierarchical GCN. It learns the “cluster assignments” for nodes to aggregate the nodes into a set of clusters and to update the connectivity as well. The coarsened nodes and connectivities are forwarded to the further layers of GCN and DIFFPOOL, which can thus estimate the multiscale community structures of the graph. Based on the DIFFPOOL, Xing et al. [21] generated multiscale graph representations from FCNs and detected the diagnostic brain region clusters in MCI and obsessive-compulsive disorder patients during automatic diagnoses. In addition, Gao and Ji [24] proposed a “gPOOL” operation in a “Graph-U-Net” to generate a hierarchical representation of a given graph. gPOOL learns a hierarchical downsampling scheme from a single-scale graph based on the “importance” of nodal features, learned via learnable linear projections. It thus gradually focuses on a small set of nodes and their corresponding subnetworks. Following gPOOL, “self-attention graph pooling (SAGPOOL)” is further proposed to refine the hierarchical downsampling by learning the nodal importance via GCNs on the graph [25]. Li et al. [22] adopted the idea of "gPOOL" and applied it to build the hierarchical GCN for ASD detections.

However, all the abovementioned studies can only mine multiscale features from a single-scale FCN, built by a pre-defined parcellation scheme of the atlas. On the other hand, with the development of atlas-building technology [26], [27], multiscale brain atlases have been increasingly studied, which naturally provides FCNs at different spatial scales with more solid neuroscience knowledge. Recently, Yao et al. [28] proposed a mutual multiscale triplet GCN (MMTGCN) for brain disorder classifications based on the multiscale atlas information and the intersubject sample relationships. In addition to the “triplet sampling” strategy to learn intersubject relationships, they further proposed a “mutual learning strategy” to collaboratively train multiscale encoders to generate aligned and optimal representations of the individual brain networks.

Finally, the advanced brain atlas study offers refined multiscale parcellation and the biologically validated prior brain hierarchy as the spatial relationship among multiscale atlases [29]. Inspired by it, in this work, we propose a novel multiscale-atlases-based hierarchical GCN method, to respect and utilize the prior brain network hierarchy under the GCN framework and to more effectively and sensitively capture the disorder-related intrascale and interscale topological changes in FCNs. Specifically, we first implement multiscale atlases to construct multiscale FCNs for avoiding the potential bias in single-scale analysis and to comprehensively characterize the brain network. We further invent an “Atlas-guided Pooling (AP)” to perform nodal aggregation based on the prior hierarchical relationship defined in the multiscale atlases.
Specifically, in each stage of nodal pooling, we update the corresponding graph with the empirical FCNs from multiscale atlases. We thus build a hierarchical architecture by stacking the GCNs and AP layers to extract multiscale disorder-related features for final diagnoses. We utilize three open-access datasets and perform diagnoses of AD, MCI, and ASD to demonstrate the effectiveness of our proposed method.

The main contributions of the presented study are threefold. First, to the best of our knowledge, our method is the first attempt to embed the guiding priors from multiscale brain atlases into the hierarchical graph convolutional neural networks for brain disorder diagnoses. Our method distinguishes itself from other GCN methods as a novel and biologically meaningful framework to naturally integrate domain knowledge into a data-driven learning-based method. Second, experiments demonstrate that our method can effectively detect more disorder-related features than other single-scale-based methods (including data-driven graph pooling) and other multiscale-atlases-based methods with classical fusion schemes and thus promote precise brain disorder diagnoses. Third, our method contributes to the improvement of interpretability for artificial neural networks and also revealing of pathological factors of brain disorder among multiscale brain connectomes.

A part of this work was reported at MICCAI 2021 conference [30]. In this article, we have added new variants of the proposed method by generalizing its pooling operations. We have also added new analyses to demonstrate feature interpretability. In addition, we have included more detailed descriptions of method implementations, systematic testing of the effects of parameter changing, as well as extensive experiments on additional datasets to evaluate our proposed method.

II. MATERIALS AND METHODS

A. Subjects and rs-fMRI Data Preprocessing

We use the rs-fMRI data from three datasets: 1) AD neuroimaging initiative (ADNI-2 and ADNI-GO; http://adni.loni.usc.edu/) [31]; 2) open-access series of imaging studies (OASIS-3; https://www.oasis-brains.org/) [32]; and 3) autism brain imaging data exchange (ABIDE-I; https://fcon_1000.projects.nitrc.org/indi/abide/) [33]. Table I gives the overall information of subjects in the three datasets. Chi-square test and two-sample t-test are used to confirm no significant discrepancies in gender and age across groups (i.e., p > 0.05), while there are significant differences in MMSE between HC and AD subjects. Specifically, the ADNI dataset is used as the main dataset for illustrations, while the OASIS and ABIDE datasets serve as additional independent datasets to partially evaluate the generalizability of our proposed method.

For the ADNI dataset, a total of 531 fMRI data from different subjects at baseline scans are selected (with no repeated scan for any single subject, which is the same for the OASIS and ABIDE data), which contains 364 healthy control (HC), 117 early MCI (eMCI), and 50 AD subjects. In ADNI, eMCI is defined to describe the subjects exhibiting cognitive decline but subtler than those at MCI, measured by the Wechsler memory scale neuropsychological test. Accordingly, the eMCI state can be more difficult to detect based on FCN. The fMRI data are acquired with TR = 3 s, TE = 30 ms, flip angle = 80°, resolution = 3.3 × 3.3 × 3.3 mm³, 48 axial slices, and 420 s in duration (140 volumes).

In the OASIS dataset, based on the availability, we include 207 HC and 43 AD subjects in the study. The fMRI acquisition protocols are TR = 2.2 s, TE = 27 ms, resolution = 4 × 4 × 4 mm³, 36 axial slices, and 372 s in duration (169 volumes). From the ABIDE datasets, we select scans with a duration longer than 300 s, yielding 512 HC and 499 ASD subjects. As a multisite dataset, the acquisition protocols and diagnostic criteria in ABIDE are varying across data collection sites [33]. Overall, the scanning parameters are TR = 1.5–3 s, TE = 15–33 ms, in-plane resolution 3 × 3–3.438 × 3.438 mm², slice thickness = 3–4.5 mm, 28–40 axial slices, and 304–486 s in duration (120–300 volumes).

We apply well-accepted toolboxes, AFNI [34] (for ADNI) and DPARSF [35] (for ABIDE and OASIS), to perform a standardized preprocessing procedure for fMRI data. In particular, the first five or ten volumes of each image are discarded due to potential nonequilibrium magnetization. The rigid-body transformation is performed to correct the subject’s head motion. There is no large head motion (i.e., larger than 2 mm or 2°) found in the used subjects. We do not perform scrubbing/censoring to the data as it may introduce additional artifacts [36]. The signals of white matter, cerebrospinal fluid, and head motion are regarded as nuisance covariates and are regressed out from individual data. The fMRI images are then normalized to the Montreal Neurological Institute (MNI) space and spatially smoothed with a Gaussian kernel with full-width at half-maximum (FWHM) of 6 × 6 × 6 mm³. The BOLD signals are further band-pass filtered (0.01 ≤ f ≤ 0.1 Hz) to remove the neural-irrelevant high-frequency noises and low-frequency drift from MRI. For ABIDE dataset, since volumes of scans are different among the collecting sites, we use its minimum common length, i.e., 115 volumes, from the middle of the preprocessed fMRI sequence for further processing.

B. Multiscale Atlases for Brain Parcellation and FCN Construction

Schaefer et al. [29] provided a set of atlases for multiscale brain parcellation, which is used in this article for generating

| Study | Group | Gender (M/F) | Age | MMSE |
|-------|-------|--------------|-----|------|
| ADNI  | HC    | 148/216      | 71.8±6.3 | 29.1±1.1 |
| ADNI  | eMCI  | 56/61        | 70.5±6.8 | 28.3±1.7 |
| AD    | AD    | 25/25        | 72.8±8.4 | 22.9±2.5 |
| OASIS | HC    | 107/100      | 72.3±4.5 | 28.9±1.4 |
| OASIS | AD    | 28/15        | 73.2±6.6 | 25.6±3.4 |
| ABIDE | HC    | 434/78       | 17.6±7.7 | \ |
| ABIDE | ASD   | 436/63       | 17.2±8.5 | \ |
Fig. 1. Schematic of the proposed multiscale-atlas-based hierarchical GCN (MAHGCN) workflow. For each individual, a set of predefined multiscale atlases (a) are applied to the fMRI data to obtain regional averaged BOLD signals and build FCNs at different scales (b). The colors in (a) encode different resting-state networks (RSNs). The acronyms of RSNs are attached aside the colorbar, including DMN, frontoparietal network (FP), limbic network (LIM), salience network (SAL), attention network (ATT), somatomotor network (SM), and visual network (VIS). (c) MAHGCN extracts information from the multiscale FCNs based on the AP using hierarchical relationship between neighboring-scale atlases. The extracted features from each scale are integrated by concatenation and further processed by fully connected layers to generate an individualized diagnosis.

multiscale FCNs and guiding the nodal pooling across scales. The atlases are generated by clustering voxels (or vertices) based on FC patterns by considering both global similarity and spatial proximity. Using different resolution parameters during the clustering results in a set of brain functional parcellation at multiple scales, ranging from 100 to 1000 regions of interest (ROIs). These ROIs belong to seven resting-state functional networks (RSNs) defined in [37]. It can be observed that the RSN structures are largely preserved after parcellations at all scales [Fig. 1(a)]. The relationship among the atlases at different scales thus characterizes a biologically meaningful functional hierarchy.

Given an atlas at a specific spatial scale, the ROI-level signals can be extracted by averaging voxel-level BOLD signals within each ROI. The FCN at the given scale \( R \) is then computed by the Pearson correlation among all pairs of ROI-level signals and denoted as \( \text{FCN}_R \) [Fig. 1(b)]. To demonstrate our method and reduce computational and storage costs, we use the first five scales, i.e., from 100 to 500 ROIs (note that our method can be easily extended to other scales). Therefore, \( R \) has a range from 100 to 500. The effect of increasing and decreasing the range of scales to be integrated is explored in later analyses, which could be extended to infer the situations when more scales are included.

C. Multiscale-Atlases-Based Hierarchical Graph Convolutional Network (MAHGCN)

1) Graph Convolutional Network: We implement the spectral graph convolution [38] to build GCN, which applies a series of approximation to obtain a simple formula of graph convolution operation and is the most popular version of the GCN. For a given graph with adjacency matrix \( A \) with \( R \) nodes, its corresponding symmetric normalized Laplacian \( L \) is

\[
L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}
\]

where \( I \) is the identity matrix and \( D \) is the degree matrix of \( A \), yielded by filling the diagonal elements with the degree \( \sum A_{ij} \) of \( A \) and the off-diagonal elements with 0. The eigenvectors \( V \) of graph Laplacian define the Fourier basis of the graph. With \( V \), the graph convolution operation on the inputted feature vector \( h \) (with dimensions \( R \times \text{ch}^{in} \), where \( \text{ch}^{in} \) refers to the dimension/channel size in the inputted nodal feature) can be defined in the spectral domain

\[
g_0 * h = V((V^T g_0) \times (V^T h))
\]
where \(g_\theta\) denotes the graph convolutional kernel with learnable parameters \(\theta\) and \(V^T h\) is the Fourier transform of feature vector \(h\).

Direct estimation of \(g_\theta\) can be computationally expensive. According to [38], a simplified operation can be applied where the outputted feature from GCN \(h_{\text{GC}}^{R_j} (R \times \text{ch}_{\text{out}})\) is given by the following equation:

\[
h_{\text{GC}}^{R_j} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} h\theta).
\]

2) Atlas-Guided Pooling: We further define the “AP” to bridge the GCNs at different scales and guide the fine-to-coarse processing of graph features. Multiple small ROIs in fine-scale parcellation could be merged into one ROI in coarse parcellation. The spatial overlapping of ROIs defined at different scales is thus utilized to approximately characterize the interscale hierarchy. We construct the mapping matrix between atlases by checking spatial overlapping between ROIs defined at different scales. Formally, from the atlas \(\mathcal{R}\) (with \(R\) ROIs/nodes) to the atlas \(\varphi\) (with \(P\) ROIs/nodes, where \(R > P\)), the atlas mapping matrix \(M \rightarrow \varphi\) (a \(R \times P\) matrix) is defined as follows. First, we compute how many voxels of ROI \(i\) in the finer scale atlas \(\mathcal{R}\) are spatially overlapping with ROI \(j\) in the coarser scale atlas \(\varphi\) and further calculate the overlapping ratio \(\rho\) between the number of overlapping voxels in ROI \(i\) and the total number of voxels in ROI \(i\). Second, a threshold \(T_h\) is applied to the overlapping ratio \(\rho\) to define the interscale mapping matrix in the AP

\[
M_{\mathcal{R} \rightarrow \varphi}(i, j) = \begin{cases} 1, & \rho > T_h \\ 0, & \text{Otherwise.} \end{cases}
\]

3) Stacks of GCNs and APs for Hierarchical Architecture: After defining the AP, we can then build the MAHGCN with stacks of GCNs and APs in a hierarchical manner. In this article, the MHAGCN integrating all five scales will be used as a full method for demonstrations. We extract FCNs from five scales (100, 200, . . . , 500 ROIs) at maximum, which can be used to build a five-layer MAHGCN (as shown in Fig. 1(c)). Different scale ranges are also tested in later experiments. In particular, MAHGCN starts with a graph convolution between the nodal features and the FCN from the finest parcellation [i.e., the 500-ROI atlas in the example of Fig. 1(c)]. As the initial inputted nodal features, a one-hot feature is used here, i.e., a 500 \(\times\) 500 identity matrix, rather than the features extracted from brain data. This is because the effect of choosing different nodal features is not in the scope of our study. Also, such a null configuration (one-hot feature as used in [38]) will make the tested models here focus on topological information in FCNs and thus provide an extendable baseline. Using the AP, the outputted feature vector \(h_{\text{GC}}^{R_{500}}\) (a 500 \(\times\) 1 vector) is then converted to \(h_{\text{AP}}^{R_{400}}\) (a 400 \(\times\) 1 vector), which is further used in the next-layer graph convolution with FCN_{400}. The hierarchical graph convolution continues until FCN from the coarsest scale atlas (i.e., the 100-ROI atlas) is processed.

A skip connection design, inspired by [39], is also implemented to allow outputs from all the scales to be involved in the final prediction. The outputs from all the scales are concatenated to form a fused representation for the state of the multiscale brain system and sent to two fully connected layers (FLs). In Fig. 1(c), the fused representation is \(h_{\text{GC}}^{R_{1500}}\), with its size depending on the range of integrated scale (i.e., 1500 \(\times\) 1 at most and 300 \(\times\) 1 at least). The first FL compresses the fused representation from MAHGCN into a 64 \(\times\) 1 vector (which is fixed for variable ranges of integrated scale) and the second one converts the outputs from the first FL into the predicted probability for the two classes with a Softmax function (a 2 \(\times\) 1 vector). During the implementation, the GCNs are associated with dropout (rate = 0.3), and the FLs are with the batch normalization and ReLU activation functions. These settings for GCN and FLs are kept in other methods for comparison.

D. Model Explainability Analysis

To examine the explainability of our proposed model, we apply the gradient-guided class activation map (Grad-CAM) algorithm [40] to MAHGCN. By computing the Grad-CAM from GCNs at different scales, the important brain regions, defined by multiscale atlases during the predictions, can be discovered.

Briefly, the Grad-CAM tracks the gradient of the prediction output for a certain class (i.e., the normal or disorder state) with respect to each element in the feature maps outputted from the intermediate layers in the neural network (i.e., \(h_{\text{GC}}^{R_j}\)). The gradient can be interpreted as the amount of influence of the intermediate feature elements on the outputted decision-making and thus can be regarded as the importance of each feature element. The class activation map (CAM) is computed by the product between the gradient map and the feature map of individual data. Since the elements that positively contribute to the final prediction are more focused, the negative part of the individual CAM is further filtered out with a ReLU function.

III. EXPERIMENTS

A. Experimental Settings

1) Implementation: The neural-network-based methods are implemented using the open-source framework Pytorch [41] in Python (Version 3.6.2). The training of the network is accelerated by one Nvidia GTX 3080 GPU. To address the imbalance

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
between positive and negative sample sizes, we apply weighted cross entropy as a loss function, where the weights are set as the inverse of the sample ratio in the training set. We also observe that using focal loss will not significantly change the results (see the Supplementary Material). Each model is trained with epoch = 100, learning rate = 0.001, and batch size = 30. The time cost for training MAHGCN is around 30 min for ADNI and OASIS and 70 min for ABIDE. The parameters in each model are initialized with random weights. The Adam [42] is used as an optimizer, with weight decay = 0.1 (for ADNI) or 0.01 (OASIS and ABIDE) for avoiding overfitting.

2) Validation Scheme and Evaluation Metrics: Five times of holdout cross validation are carried out to evaluate the models. The dataset is randomly assigned into the training and testing sets with 80–20 split five times. Then, each model is trained and tested in the five different training–testing assignments. The accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (area under ROC, AUC) are chosen as performance metrics. Since the sample sizes of different classes are imbalanced in ADNI and OASIS datasets, AUC can be a more informative assessment for evaluating performance there. The abovementioned performance metrics from each method are generally provided as the mean and standard deviation over five times of cross validation. Where possible, we report the confidence interval (CI) for more comprehensive statistical information.

3) Statistical Analysis: To demonstrate the significance of improvement by our proposed method, we conduct statistical comparisons for performance metrics from cross validations, with both one-sided paired t-tests and Wilcoxon signed-rank test to examine the significance. The Wilcoxon signed-rank test is a nonparametric test on the data median and has no assumption on the normality of the data distribution. Both types of tests are performed based on the built-in functions in MATLAB, i.e., ttest and signrank (R2020b, Mathworks Inc., USA).

B. Comparison Methods

First, the single-scale GCN method is adopted to provide a baseline to be compared with multiscale-based methods.

1) GCN: The GCN is based on [38] and the configurations described in the method parts. The GCN is single-layered, which is followed by two FLs to generate a prediction of the status for each subject. We report the results from GCNs based on FCNs at 500-ROI scales for a better comparison with the following experiments.

Second, several representative data-driven pooling methods are compared, including DIFFPOOL [23], gPOOL [24], and SAGPOOL [25]. To make a fair comparison, we mimic the structure of MHAGCN with five scales and let the DIFFPOOL (DP), gPOOL (GP), and SAGPOOL (SAGP) mine the 500-ROI FCN and generate 400-, 300-, 200-, and 100-ROI FCN in the same hierarchical manner. After layers of GCN and pooling, the encoded features are concatenated and fed into two FLs, where configurations are the same as the main architecture of MAHGCN as described above.

1) DP: The DP is akin to a clustering method, which learns a nodal assignment/clustering matrix based on the features extracted from the nodal and graph features by a GCN. The assignment matrix serves as the mapping matrix $M_{R \rightarrow \rho}$ in MAHGCN and can be used to generate the clustered/coarse-grained graph for further processing with GCNs. Specifically, they further proposed an auxiliary link prediction objective and entropy regularization to ensure that the learned assignment matrix is sparse and modularized.

2) GP and SAGP: They estimated the coarse-scale graph in a downsampling manner. In GP, the nodal features are inputted into FLs and a Softmax operation to estimate the nodal importance, based on which the graph is downsampled into the subgraph with the top $K$ important nodes. Here, $K$ is the parameter to control the sampling rate. In SAGP, the nodal importance is learned on the nodal and topographical features by a GCN. The downsampled graph is forwarded to the following GCNs for processing.

Third, we compare multiscale-atlases-based methods with classical fusion schemes to demonstrate the effectiveness of multiscale atlases fusion and the advantage of our proposed method, MAHGCN. In particular, we perform the following.

1) MV and WV: Decision-level fusion is first implemented. A “majority voting (MV)” among predictions from respectively trained GCN-based classifiers at different scales. The prediction from each classifier is regarded as one vote, and the final predicted label is the most voted label among all the classifiers. In addition, we also perform “weighted voting (WV),” in which the predicted probability from the pretrained GCN-based classifiers is weighted-summed with the weights learned by training.

2) MAGCN: Moreover, feature-level fusion is performed. We train a model with multiple GCNs to independently process the inputs of multiscale FCNs. The outputs from multiple GCNs are concatenated and forwarded to two FLs. We name this model “MAGCN.” Note that this model can also be viewed as MAHGCN without AP and thus can also serve as an ablation study to highlight the importance of including the prior of interscale hierarchy among GCNs at different scales.

We also conducted a comparison between our proposed method and the classical machine learning- and recurrent neural network-based methods, to cover other existing analysis streams in the fMRI-based analysis studies [4], [43]. Our proposed method also shows improved performance over these comparison methods in all diagnosis tasks. The detailed descriptions and results can be found in Supplementary Materials (Tables S1–S4).

IV. RESULTS

A. Comparison Study

In this experiment, MAHGCN is compared with three data-driven pooling methods and three multiscale-atlases-based methods. The models are trained and tested in ADNI, OASIS, and ABIDE datasets. In Table II, the performances
TABLE II
CLASSIFICATION RESULTS FROM ALL COMPARISON METHODS IN FORM OF “MEAN (STANDARD DEVIATION).” THE ASTERISK (*) BY THE METRIC FROM MAHGNCN INDICATES SIGNIFICANT IMPROVEMENT AT $p < 0.05$ LEVEL WHEN COMPARED TO THE SECOND HIGHEST PERFORMANCE, USING EITHER ONE-SIDED PAIRED $t$-TEST OR WILCOXON SIGNED-RANK TEST. THE DIAMOND (♢) INDICATES THAT THE PERFORMANCE FROM MAHGNCN IS SIGNIFICANTLY BETTER THAN THE ONE FROM MAGCN (I.E., MAHGNCN WITHOUT AP).

| Metric | ADNI: eMCI | ADNI: AD | OASIS | ABIDE |
|--------|------------|----------|--------|--------|
|        | GCN | DP | GP | SAGP | MV | WV | MAGCN | MAHGNCN | GCN | DP | GP | SAGP | MV | WV | MAGCN | MAHGNCN | GCN | DP | GP | SAGP | MV | WV | MAGCN | MAHGNCN | GCN | DP | GP | SAGP | MV | WV | MAGCN | MAHGNCN |
| ACC    | 0.728(0.035) | 0.753(0.018) | 0.757(0.266) | 0.732(0.031) | 0.780(0.047) | 0.777(0.049) | 0.722(0.028) | 0.786(0.037) | 0.839(0.042) | 0.843(0.073) | 0.858(0.037) | 0.848(0.045) | 0.884(0.045) | 0.730(0.084) | 0.843(0.055) | 0.889(0.031) | 0.704(0.080) | 0.748(0.032) | 0.760(0.108) | 0.784(0.075) | 0.720(0.087) | 0.708(0.090) | 0.752(0.084) | 0.828(0.069) | 0.659(0.034) | 0.676(0.016) | 0.683(0.009) | 0.677(0.007) | 0.711(0.030) | 0.681(0.032) | 0.679(0.027) | 0.727(0.021) | 0.703(0.058) | 0.709(0.067) | 0.689(0.057) | 0.690(0.073) | 0.723(0.033) | 0.681(0.050) | 0.712(0.065) | 0.698(0.029) | 0.615(0.044) | 0.643(0.047) | 0.677(0.052) | 0.670(0.008) | 0.697(0.055) | 0.678(0.038) | 0.650(0.100) | 0.754(0.030) | 0.659(0.032) | 0.676(0.015) | 0.683(0.009) | 0.680(0.007) | 0.710(0.030) | 0.680(0.030) | 0.681(0.025) | 0.726(0.021) | 0.615(0.044) | 0.643(0.047) | 0.677(0.052) | 0.670(0.008) | 0.697(0.055) | 0.678(0.038) | 0.650(0.100) | 0.754(0.030) | 0.659(0.032) | 0.676(0.015) | 0.683(0.009) | 0.680(0.007) | 0.710(0.030) | 0.680(0.030) | 0.681(0.025) | 0.726(0.021) |

of all comparison methods on HC-versus-eMCI, HC-versus-AD, and HC-versus-ASD classification tasks are depicted. In all classification tasks, the following consistent results can be observed. First, both hierarchical graph processing with data-driven pooling methods (DP, GP, and SAGP) and the multiscale-atlases-based methods outperform the single-scale-based GCN, suggesting the advantage of multiscale analyses. The multiscale analyses investigate more systematic characterizations of the brain system and thus can detect more sensitive information for brain disorder diagnosis. Second, all best performances are achieved by multiscale-atlases-based methods, rather than data-driven pooling methods, which suggests the advantage of biological-prior-based methods over data-driven methods in properly capturing multiscale brain features. This can attribute to the fact that a pure data-driven method estimates the hierarchy from scratch and its parameter learning could be biased by noise from data. On the other hand, the brain atlas can introduce prior of the brain hierarchy with no additional parameter estimation, which thus stabilizes and benefits the model learning. Third, among all the comparison methods, our proposed method achieves the best ACC and AUC. Most improvements are significant. In addition, when compared with MAGCN, i.e., MAHGNCN without AP, our MAHGNCN provides significantly higher AUCs in all four tasks. This demonstrates that our proposed MAHGNCN based on AP can best integrate information of multiscale-atlases-based FCNs for brain disorder diagnosis. Among all the multiscale-atlases-based methods, MV and WV are based on separately pretrained single-scale classifiers, and thus, their information fusion process is blind to the learning process and cannot achieve optimal performances. In the same line, MAGCN conducts the feature-level fusion after the features are independently extracted at each scale. The MAHGNCN enables a joint feature encoding for multiscale FCNs with the awareness of hierarchical relationships. This guarantees that the MAHGNCN can extract more accurate hierarchical feature representations for optimal diagnosis.

Besides, the performance on HC-versus-AD classification by MAHGNCN on the OASIS dataset is comparable to that yielded on the ADNI dataset, which could partially support the generalizability of our proposed model on AD identification across different datasets. In addition, our method outperforms
other methods in the HC-versus-ASD classification task using the ABIDE dataset in terms of ACC, SPE, and AUC, suggesting that our proposed method is generalizable to different brain disorders other than MCI and AD.

### B. Comparison With State-of-the-Art Methods

We further compare the performances of our method with performances from state-of-the-art studies on FCN-based diagnostic analyses. In Table III, we group the studies based on the classification tasks and report the number of subjects and the achieved performances. Due to the variation of sample selection, the performances may not be quantitatively comparable. In the HC-versus-eMCI and HC-versus-AD classification tasks, our method obtains better results than the advanced CNN-based methods. This could originate from the advantage of GCN over CNN, which can more properly process the graph data. Also, it may attribute to the advantage of utilizing prior knowledge in our method. Also, in the HC-versus-ASD classification task, our methods achieve slightly low but competitive results when compared to other GCN-based methods. A possible reason could be that the data heterogeneity of large datasets increases the difficulty of classification.

We further perform a comparison with the MMTGCN method [28] that also used the multiscale atlases. In particular, in the ADNI dataset, we select 148 MCI subjects (63 males, age = 74.2 ± 8.9), match them with 155 HCs (77 males, age = 74.9 ± 6.4), and apply our MAHGCN. Note that Yao et al. [28] also used the ADNI dataset and the ratio between HC and MCI in their study is not significantly different from ours (chi-square test, \( p = 0.473 \)). Our method obtains ACC = 0.790 ± 0.012 (CI = [0.773, 0.807]), SEN = 0.778 ± 0.056 (CI = [0.701, 0.856]), SPE = 0.804 ± 0.046 (CI = [0.740, 0.869]), and AUC = 0.791 ± 0.007 (CI = [0.781, 0.801]), which are lower than those reported in [28] in the HC-versus-MCI classification. However, we emphasize that their idea and our idea are different and could be integrated. Their idea relies more on the intersample relationship, which is different from our design. More specifically, MMTGCN introduces a "triplet sampling" strategy to consider the relationship among subjects at each scale and further aligns the distributions of the features among multiple scales by a "mutual learning strategy" to benefit the integration of multiscale information. Our method considers the interscale correlation in nodal features when hierarchically encoding the FCN from one subject. In addition, the discrepancy in the performance may also attribute to difference in the diagnostic task. Their strategy of using intersample relationships could be more beneficial for MCI detection, whereas our work could be more specific to the diagnosis for the eMCI state, which is earlier than MCI and thus is more clinically significant.

### C. Ablation Study

1) **Effects of the Range of Integrated Scales:** We have investigated the effect of changing the range of integrated scales using HC-versus-eMCI and HC-versus-AD (ADNI) classification tasks. In Fig. 2, it can also be observed that the ACCs and AUCs of MAHGCN are generally increasing with more scales included. We also try to integrate non-neighboring scales such as including 100 and 500, as well as 100, 300, and 500 in the HC-versus-eMCI classification, and observe similar increasing trends. The model with 100 and 500 obtained...
ACC = 0.714 ± 0.037 (CI = [0.678, 0.781]), SEN = 0.645 ± 0.104 (CI = [0.500, 0.789]), SPE = 0.755 ± 0.073 (CI = [0.654, 0.857]), AUC = 0.700 ± 0.028 (CI = [0.662, 0.738]), and the model with 100, 300, and 500 yielded ACC = 0.725 ± 0.078 (CI = [0.632, 0.848]), SEN = 0.674 ± 0.101 (CI = [0.534, 0.814]), SPE = 0.761 ± 0.130 (CI = [0.581, 0.942]), and AUC = 0.718 ± 0.025 (CI = [0.683, 0.753]). Both observations suggest that MAHGCN can effectively integrate information from multiscale FCNs for diagnosis. However, the performance may not grow continuously with further increases in the range of integrated scales. We expect a ceiling effect to appear, although currently not being observed in our experiments. The artifacts in parcellation at fine scales could affect a meaningful hierarchical encoding. Also, more integrated scales require increased parameters in MAHGCN, leading to the risk of overfitting on small-sample data.

2) Different Configurations of AP: We perform further ablation studies by changing the configurations of the MAHGCN. The HC-versus-eMCI classification, as the most difficult task, is performed to sensitively assess the changes due to modification of the model. We test the performance under different pooling schemes (i.e., using max, average, and sum pooling) in the AP. The corresponding models are called Max AP, Average AP, and Sum AP. Note that the “Sum AP” is a default configuration we use above.

Interestingly, unlike natural image processing where the average pooling and the max pooling are the most efficient pooling schemes, we find that the sum pooling significantly outperforms other types of pooling in terms of ACC, SPE, and AUC (Fig. 3). This may attribute to the essential difference between natural image data and brain network data. Average pooling normalizes the differences in the number of nodes under integration at the finer scales and flattens differences of nodes at the coarser scales, while max pooling only relies on the ROI with max features to represent other ROIs at the finer scale. These pooling schemes could ignore meaningful biological information for the brain system and thus produce decreased performance.

3) Effects of Skip Connections: In Fig. 4, we systematically test MAHGCNs with different schemes of AP and their corresponding versions without SCs. The SCs significantly improve AUC in MAHGCNs with all schemes of AP. This is reasonable since the final output from MAHGCN only provides an abstract feature shared by FCNs at all scales, while SCs allow the reservation of scale-specific information at all scales. In addition, the SC could overcome the gradient vanishing and benefit the optimization of MAHGCN.

4) Effects of Thresholds for Atlas Mapping Matrix: In this article, we simply use Th = 0, under which any sharing between the ROIs in atlases is considered overlapping ROIs. Other settings of Th are also explored, and the respective diagnosis performances are provided in Fig. 5. In terms of SEN and AUC, the performance increases with the reduction of Th. Blue: Th = 0.5; Red: Th = 0.25; Yellow: Th = 0. The asterisks (*) and circles (◦) have the same meaning as in Fig. 2.
Fig. 6. RSN-wise averaged class activation values for diagnosing eMCI and AD. The color codes and acronyms for different RSNs are the same as in Fig. 1(a). (a) and (b) Brain patterns of class activation values for eMCI and AD. (c) and (d) Distributions of RSN-wise averaged class activation values for eMCI and AD.

5) Effects of Preprocessing Pipeline: The ABIDE dataset provides preprocessed fMRI data using different pipelines, which are further quality-controlled (see https://preprocessed-connectomes-project.org/abide/). To test the effect of preprocessing on our method, we utilize those preprocessed data under two pipelines, i.e., DPARSF and connectome computation system (CCS, based on AFNI and FSL [46]), to investigate the effect of preprocessed pipeline on our MHAGCN. After checking data availability for both pipelines and matching the age and gender, we include 371 HC (311 males and age = 16.60 ± 6.70 years) and 325 ASD (284 males and age = 17.07 ± 8.32 years) for this analysis.

The MHAGCN achieves ACC = 0.694 ± 0.018 (CI = [0.670, 0.719]), SEN = 0.709 ± 0.060 (CI = [0.625, 0.793]), SPE = 0.684 ± 0.070 (CI = [0.587, 0.781]), AUC = 0.696 ± 0.017 (CI = [0.672, 0.721]) using DPARSF, and ACC = 0.706 ± 0.016 (CI = [0.684, 0.728]), SEN = 0.684 ± 0.093 (CI = [0.589, 0.846]), SPE = 0.702 ± 0.091 (CI = [0.557, 0.810]), and AUC = 0.702 ± 0.013 (CI = [0.684, 0.719]) using CCS. These two results show no significant difference using paired t-test ($p = 0.728$) and Wilcoxon signed-rank test ($p = 0.526$). The reduction of performance compared to results in Table I may attribute to the changes in sample size.

D. Explainability Analysis

To demonstrate the explainability of the prediction from MAHGCN, we apply Grad-CAM to explore important features during the decision-making by MAHGCN. In particular, the
group-wise CAMs of five models from cross validations are first obtained by averaging the CAMs from each individual within brain disorder groups. Then, the resulting group-wise CAMs are weight-averaged according to the AUC of each model, yielding the CAMs as brain patterns [Fig. 6(a) and (b)]. In Fig. 6(c) and (d), we further assign class activation values in ROIs to the corresponding RSNs and compute the RSN-wise averaged class activation values.

For AD diagnosis (ADNI dataset), a pronounced finding is that the limbic network (LIM) is detected as an important predictor shared by most of the scales (100, 300, 400, and 500). This is consistent with the widely accepted understanding that AD is associated with damage in the hippocampus related to the limbic network and thus typically exhibits amnestic cognitive impairment [47]. In addition, the default mode network (DMN) (related to executive functions) at scales of 200, 400, and 500 ROIs is found to be the second most contributive RSN, which is also consistent with the findings in the AD literature [47]. These consistencies could partially support the explainability of the features identified by our method. The feature distributions from the OASIS dataset are provided in Fig. S1. Consistently, the model identifies LIM (at scales of 100, 200, and 400) and DMN (500-ROI scale). However, the details (e.g., spatial locations and scales) are different, probably due to the heterogeneity of data between ADNI and OASIS.

For eMCI diagnoses, different scales rely on different RSNs as the most important features. Somatomotor (SM), LIM, visual (VIS), attention (ATT), and salience (SAL) networks can be identified as the predictive RSNs at different scales, and there is few RSN being consistently identified as a crucial predictor. This may also be reasonable due to complex symptoms of the eMCI involving impairments in multiple cognitive domains, including executive function, attention, and visuospatial ability or language, besides memory [48], [49]. This could deliver a new insight that eMCI damages the functioning of different cognitive domains at different scales.

For ASD diagnoses (Fig. S1), the SM, VIS, ATT, and SAL are indicated as important features for the model predictions, which implies that ASD symptoms may originate from deficits in sensory processing and attention. This is also in line with one of the prevalent understandings in ASD studies [50], [51].

V. CONCLUSION

We have presented a novel method for brain disorder diagnosis based on information from multiscale FCNs and predefined multiscale atlases. Using neuroimaging data from ADNI, OASIS, and ABIDE, we have demonstrated that our proposed method has a higher effectiveness in identifying brain functional deficits induced by MCI, AD, and ASD. The improvement of diagnosis by our proposed method comes from the usage of prior knowledge on brain parcellation as well as proper designs (based on brain hierarchy) to fuse multiscale spatial information in a deep-layer GCN structure. We have further investigated the features of atypical brain connectomes across different scales and provide novel clues about the pathology of MCI, AD, and ASD, paving a new path to explore multiscale-atlases-based FCN in brain disorder diagnostic studies. Though FCN is the focus in this study, we expect that our method or its variations could be applied to brain network data based on other modalities [52], [53], [54], [55], [56].

REFERENCES

[1] J. Sui, R. Jiang, J. Bustillo, and V. Calhoun, “Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: Methods and promises,” Biol. Psychiatry, vol. 88, no. 11, pp. 818–828, Dec. 2020.
[2] A. Khazaee, A. Ebrahimmzadeh, and A. Babajani-Feremi, “Identifying patients with Alzheimer’s disease using resting-state fMRI and graph theory,” Clin. Neurophysiol., vol. 126, no. 11, pp. 2132–2141, Nov. 2015.
[3] S. Haller et al., “Individual classification of mild cognitive impairment subtypes by support vector machine analysis of white matter DTI,” J. Neurol. Sci., vol. 34, no. 2, pp. 283–291, Feb. 2013.
[4] C.-Y. Lee et al., “Identification of MCI individuals using structural and functional connectivity networks,” NeuroImage, vol. 59, no. 3, pp. 2045–2056, Feb. 2012.
[5] 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.
[6] R. Anirudh and J. J. Thiagarajan, “Bootstrapping graph convolutional neural networks for autism spectrum disorder classification,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 3197–3201.
[7] C.-Y. Lee, C. Liu, A. Lee, J. S. Poh, H. Ji, and A. Qiu, “Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations,” NeuroImage, Clin., vol. 23, Jan. 2019, Art. no. 101929.
[8] Y. Qiu et al., “Loss of integrity of corpus callosum white matter hyperintensity penumbra predicts cognitive decline in patients with subcortical vascular mild cognitive impairment,” Frontiers Aging Neurosci., vol. 13, Feb. 2021, Art. no. 605900.
[9] T. A. Song et al., “Graph convolutional neural networks for Alzheimer’s disease classification,” in Proc. Int. Symp. Biomed. Imag., vol. 19, pp. 414–417.
[10] R. F. Betzel and D. S. Bassett, “Multi-scale brain networks,” NeuroImage, vol. 160, pp. 73–83, Oct. 2017.
[11] V. Bazinet, R. Vos de Wael, P. Hagmann, B. C. Bernhardt, and B. Misic, “Multiscale communication in cortico-cortical networks,” NeuroImage, vol. 243, Nov. 2021, Art. no. 118546.
[12] A. Ashourvan, Q. K. Telesford, T. Verstynen, J. M. Vettel, and D. S. Bassett, “Multi-scale detection of hierarchical community architecture in structural and functional brain networks,” PLoS ONE, vol. 14, no. 5, May 2019, Art. no. e0215520.
[13] D. Meinier, R. Lambiotte, A. Fornito, K. D. Ersche, and E. T. Bullmore, “Hierarchical modularity in human brain functional networks,” Frontiers Neuroinform., vol. 3, p. 37, Feb. 2009.
[14] J. Wang et al., “Parcellation-dependent small-world brain functional networks: A resting-state fMRI study,” Hum. Brain Mapping, vol. 30, no. 5, pp. 1511–1523, May 2009.
[15] A. Zalesky et al., “Whole-brain anatomical networks: Does the choice of nodes matter?” NeuroImage, vol. 50, no. 3, pp. 971–983, 2010.
[16] R. Wang, P. Lin, M. Liu, Y. Wu, T. Zhou, and C. Zhou, “Hierarchical connectome modes and critical state jointly maximize human brain functional diversity,” Phys. Rev. Lett., vol. 123, no. 3, Jul. 2019, Art. no. 038301.
[17] R. Wang, M. Liu, X. Cheng, Y. Wu, A. Hildebrandt, and C. Zhou, “Segregation, integration, and balance of large-scale resting brain networks configure different cognitive abilities,” Proc. Nat. Acad. Sci. USA, vol. 118, no. 23, Jun. 2021, Art. no. e202288118.
[18] J. Liang, S.-J. Wang, and C. Zhou, “Less is more: Wiring-economical modular networks support self-sustained firing-economical neural avalanches for efficient processing,” Nat. Sci. Rev., vol. 9, no. 3, 2021, Art. no. nwab102.
[19] R. F. Betzel, B. Mišić, Y. He, J. Rumshlag, X.-N. Zuo, and O. Sporns, “Functional brain modules reconfigure at multiple scales across the human lifespan,” 2015, arXiv:1510.08045.
[20] Y. Zhang, H. Zhang, X. Chen, S.-W. Lee, and D. Shen, “Hybrid high-order functional connectivity networks using resting-state functional MRI for mild cognitive impairment diagnosis,” Sci. Rep., vol. 7, no. 1, pp. 1–15, Jul. 2017.

[21] X. Xing et al., “Detection of discriminative neurological circuits using hierarchical graph convolutional networks in fMRI sequences,” in Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis (Lecture Notes in Computer Science). New York, NY, USA: Springer, 2020, pp. 121–130.

[22] X. Li et al., “BrainGNN: Interpretable brain graph neural network for fMRI analysis,” bioRxiv, vol. 16, May 2020, Art. no. 100057.

[23] R. Ying, C. Morris, W. L. Hamilton, J. You, X. Ren, and J. Leskovec, “Hierarchical graph representation learning with differentiable pooling,” in Proc. Adv. Neural Inf. Process. Syst., 2018, pp. 4805–4815.

[24] H. Gao and S. Ji, “Graph U-Nets,” in Proc. 36th Int. Conf. Mach. Learn., 2019, pp. 1–11.

[25] J. Lee, I. Lee, and J. Kang, “Self-attention graph pooling,” in Proc. 36th Int. Conf. Mach. Learn., Jun. 2019, pp. 6661–6670.

[26] H. Jia, P.-T. Yap, and D. Shen, “Iterative multi-atlas-based multi-image segmentation with tree-based registration,” NeuroImage, vol. 59, no. 1, pp. 422–430, Jan. 2012.

[27] H. Jia, G. Wu, Q. Wang, and D. Shen. “ABSORB: Atlas building by self-organized registration and bundling,” NeuroImage, vol. 51, no. 3, pp. 1057–1070, Jul. 2010.

[28] 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.

[29] A. Schaefer et al., “Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI,” Cerebral Cortex, vol. 28, no. 9, pp. 3095–3114, Sep. 2018.

[30] M. Liu, H. Zhang, F. Shi, and D. Shen, “Building dynamic hierarchical brain networks and capturing transient meta-states for early mild cognitive impairment diagnosis,” in Medical Image Computing and Computer Assisted Intervention—MICCAI (Lecture Notes in Computer Science), vol. 12907. Cham, Switzerland: Springer, 2021, pp. 574–583.

[31] P. S. Aisen, R. C. Petersen, M. Donohue, and M. W. Weiner, “Alzheimer’s disease neuroimaging initiative 2 clinical core: Progress and plans.” Alzheimer’s Dementia, vol. 11, no. 7, pp. 734–739, Jul. 2015.

[32] P. J. LaMontagne et al., “OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease,” medRxiv, vol. 2019, pp. 1–37, Dec. 2019.

[33] A. Di Martino et al., “The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism,” Mol. Psychiatry, vol. 19, no. 6, pp. 659–667, Jun. 2014.

[34] R. W. Cox, “AFNI: Software for analysis and visualization of functional magnetic resonance images,” Comput. Biomed. Res., vol. 29, no. 3, pp. 162–173, Jun. 1996.

[35] C.-G. Yan and Y.-F. Zang, “DPARSF: A MATLAB toolbox for ‘pipeline’ data analysis of resting-state fMRI,” Frontiers Syst. Neurosci., vol. 4, p. 13, Jan. 2010.

[36] S. Maknojia, N. W. Churchill, T. A. Schweizer, and S. J. Graham, “Resting state fMRI: Going through the motions,” Frontiers Neurosci., vol. 13, p. 825, Aug. 2019.

[37] B. T. Thomas Yeo et al., “The organization of the human cerebral cortex estimated by intrinsic functional connectivity,” J. Neurophysiol., vol. 106, no. 3, pp. 1125–1165, Sep. 2011.

[38] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in Proc. 5th Int. Conf. Learn. Represent., 2017, pp. 1–11.

[39] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention—MICCAI (Lecture Notes in Computer Science). New York, NY, USA: Springer, 2015, pp. 234–241.

[40] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 618–626.

[41] A. Paszke et al., “PyTorch: An imperative style, high-performance deep learning library,” in Proc. Adv. Neural Inf. Process. Syst., vol. 32, 2019, pp. 8026–8037.

[42] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in Proc. 3rd Int. Conf. Learn. Represent., 2015, pp. 1–11.

[43] N. C. Dvornek, P. Ventola, K. A. Pelphey, and J. S. Duncan, “Identifying autism from resting-state fMRI using long short-term memory networks,” Machine Learning in Medical Imaging, vol. 10541. New York, NY, USA: Springer, 2017, p. 362.

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

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

[46] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, “Review FSL,” NeuroImage, vol. 62, pp. 782–790, Jun. 2012.

[47] D. P. Veitch et al., “Understanding disease progression and improving Alzheimer’s disease clinical trials: Recent highlights from the Alzheimer’s disease neuroimaging initiative,” Alzheimer’s Dementia, vol. 15, no. 1, pp. 106–152, Jan. 2019.

[48] A. Busse, A. Hensel, U. Gruene, M. C. Angermeyer, and S. G. Riedel-Heller, “Mild cognitive impairment: Long-term course of four clinical subtypes,” Neurology, vol. 67, no. 12, pp. 2176–2185, Dec. 2006.

[49] M. Díaz-Mardomingo, S. García-Herranz, R. Rodríguez-Fernández, C. Venero, and H. Peraita, “Problems in classifying mild cognitive impairment (MCI): One or multiple syndromes?” Brain Sci., vol. 7, no. 12, p. 111, Sep. 2017.

[50] E. J. Marco, L. B. N. Hinkley, S. S. Hill, and S. S. Nagarajan, “Sensory processing in autism: A review of neurophysiologic findings,” Pediatr. Res., vol. 69, no. 8, pp. 48–54, May 2011.

[51] M. D. Thye, H. M. Bednarz, A. J. Herringshaw, E. B. Sartin, and R. K. Kana, “The impact of atypical sensory processing on social impairments in autism spectrum disorder,” Develop. Cognit. Neurosci., vol. 29, pp. 151–167, Jan. 2018.

[52] X. Chen, H. Zhang, L. Zhang, C. Shen, S. Lee, and D. Shen, “Extraction of dynamic functional connectivity from brain grey matter and white matter for MCI classification,” Hum. Brain Mapping, vol. 38, no. 10, pp. 5019–5034, Oct. 2017.

[53] Y. Fan et al., “Unaffected family members and schizophrenia patients share brain structure patterns: A high-dimensional pattern classification study,” Biol. Psychiatry, vol. 63, no. 1, pp. 118–124, Jan. 2008.

[54] F. Shi, P.-T. Yap, W. Gao, W. Lin, J. H. Gilmore, and D. Shen, “Altered structural connectivity in neonates at genetic risk for schizophrenia: A combined study using morphological and white matter networks,” NeuroImage, vol. 62, no. 3, pp. 1622–1633, Sep. 2012.

[55] T. Zhou, K. Thung, X. Zhu, and D. Shen, “Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis,” Hum. Brain Mapping, vol. 40, no. 3, pp. 1001–1016, Feb. 2019.

[56] Y. Fan et al., “Multivariate examination of brain abnormality using both structural and functional MRI,” NeuroImage, vol. 36, no. 4, pp. 1189–1199, Jul. 2007.
Han Zhang (Senior Member, IEEE) received the B.S. degree in electronic science and technology from Zhejiang University, Hangzhou, China, in 2005, and the Ph.D. degree in cognitive neuroscience from the National Key Laboratory for Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China, in 2011. He is currently an Associate Professor at ShanghaiTech University, Shanghai, China. His research interests focus on the brain functional network and its applications in neuroscience, brain development, and neurodegenerative disorders.

Feng Shi received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2008. He has been an Assistant Professor at The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, and Cedar Sinai Medical Center, Los Angeles, CA, USA. He is currently a Research Scientist at Shanghai United Imaging Intelligence, Shanghai, China. His research interests include medical image analysis, artificial intelligence, brain development, and neurodegenerative disorders.

Dinggang Shen (Fellow, IEEE) is currently a Professor and the Founding Dean of the School of Biomedical Engineering, ShanghaiTech University, Shanghai, China, and the Co-CEO of United Imaging Intelligence (UII), Shanghai. He was a Jeffrey Houpt Distinguished Investigator and a Full Professor (Tenured) with The University of North Carolina at Chapel Hill (UNC-CH), Chapel Hill, NC, USA, directing the Center of Image Analysis and Informatics, the Image Display, Enhancement, and Analysis (IDEA) Laboratory, and the Medical Image Analysis Core. He has published more than 1500 peer-reviewed articles in international journals and conference proceedings, with H-index of 130 and more than 70,000 citations. His research interests include medical image analysis, computer vision, and pattern recognition.

Dr. Shen is a fellow of the American Institute for Medical and Biological Engineering (AIMBE), the International Association for Pattern Recognition (IAPR), and the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society. He serves as an Editor-in-Chief for the Frontiers in Radiology and an editorial board member for eight international journals. He served on the Board of Directors at the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society, from 2012 to 2015, and was the General Chair of MICCAI 2019.