Multi-Scale Dynamic Graph Learning for Brain Disorder Detection With Functional MRI

Yunling Ma, Qianqian Wang, Liang Cao, Long Li, Chaojun Zhang, Lishan Qiao, and Mingxia Liu, *Senior Member, IEEE*

***Abstract*— Resting-state functional magnetic resonance imaging (rs-fMRI) has been widely used in the detection of brain disorders such as autism spectrum disorder based on various machine/deep learning techniques. Learning-based methods typically rely on functional connectivity networks (FCNs) derived from blood-oxygen-level-dependent time series of rs-fMRI data to capture interactions between brain regions-of-interest (ROIs). Graph neural networks have been recently used to extract fMRI features from graph-structured FCNs, but cannot effectively characterize spatiotemporal dynamics of FCNs, *e.g.*, the functional connectivity of brain ROIs is dynamically changing in a short period of time. Also, many studies usually focus on single-scale topology of FCN, thereby ignoring the potential complementary topological information of FCN at different spatial resolutions. To this end, in this paper, we propose a multi-scale dynamic graph learning (MDGL) framework to capture multi-scale spatiotemporal dynamic representations of rs-fMRI data for automated brain disorder diagnosis. The MDGL framework consists of three major components: 1) multi-scale dynamic FCN construction using multiple brain atlases to model multi-scale topological information, 2) multi-scale dynamic graph representation learning to capture spatiotemporal information conveyed in fMRI data, and 3) multi-scale feature fusion and classification. Experimental results on two datasets show that MDGL outperforms several state-of-the-art methods.**

***Index Terms*— Resting-state functional MRI, multi-scale dynamic graph, brain disorder, detection.**

I. Introduction

RESTING-STATE functional magnetic resonance imaging (rs-fMRI), as a noninvasive and nonradioactive technique, has been widely used to estimate neural activity of the whole brain over time [1], [2]. Numerous studies have applied rs-fMRI to automated detection of brain disorders such as autism spectrum disorders using various machine/deep learning techniques [3], [4], [5], [6], [7]. Conventional learning methods typically rely on brain functional connectivity networks (FCNs) derived from blood-oxygen-level-dependent (BOLD) time series to capture interactions between brain regions-of-interest (ROIs) [8], [9], [10]. Some earlier studies use human-engineered FCN features (*e.g.*, closeness centrality) for fMRI-based disease diagnosis, requiring specific expertise and prior knowledge of brain functional connectivity [11], [12], [13].

Graph neural network (GNN) is a powerful deep learning method that can automatically learn graph representation in a data-driven manner [14], [15]. Due to the graph-structured nature of the brain FCNs, GNNs have been widely used to analyze abnormal or dysfunctional patterns of brain FCNs for brain disease detection [16], [17]. However, most existing studies directly feed static FCNs into GNNs to learn network representation for downstream tasks, without considering temporal dynamics within fMRI time series (that is, the functional connectivity between brain ROIs is dynamically changing over time) [18], [19]. It is essential to model spatiotemporal dynamic information conveyed in fMRI data for brain disease analysis [20], [21], [22]. A few GNN-based studies propose to incorporate spatiotemporal information of FCNs into fMRI analysis, but they generally focus on single-scale topology of brain FCNs using one single atlas for brain ROI partition [16]. Intuitively, multi-scale FCNs derived from different brain atlases with multi-resolution ROI partitions can provide complementary topological information about the brain and help improve the learning performance of GNN models.

To this end, we propose a multi-scale dynamic graph learning (MDGL) framework to capture multi-resolution spatiotemporal dynamic representations of rs-fMRI data for automated ASD diagnosis. As shown in Fig. 1, we first partition the brain into multi-scale ROIs with Automated Anatomical Labeling (AAL) atlas [23] and Craddock200 (CC200) [24] atlas, respectively, to capture multi-scale topological information. Considering the dynamic nature of FCNs, we use sliding windows to split fMRI time series in the temporal dimension to construct multi-scale dynamic FCNs. Then, a multi-scale dynamic graph representation learning strategy is designed to capture multi-scale spatiotemporal information conveyed in fMRI data. Finally, we fuse multi-

Manuscript received 13 February 2023; revised 1 June 2023; accepted 22 August 2023. Date of publication 29 August 2023; date of current version 4 September 2023. The work of Yunling Ma was supported by the Taishan Scholar Program of Shandong Province. The work of Lishan Qiao was supported in part by the National Natural Science Foundation of China under Grant 61976110 and Grant 11931008 and in part by the Natural Science Foundation of Shandong Province under Grant ZR202102270451. *(Corresponding authors: Mingxia Liu; Lishan Qiao.)*

Yunling Ma and Chaojun Zhang are with the School of Mathematics Science, Liaocheng University, Liaocheng 252000, China.

Qianqian Wang and Mingxia Liu are with the Department of Radiology and Biomedical Research Imaging Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (e-mail: mingxia_liu@med.unc.edu).

Liang Cao and Long Li are with the Taian Tumor Prevention and Treatment Hospital, Tai'an, Shandong 271000, China.

Lishan Qiao is with the School of Mathematics Science, Liaocheng University, Liaocheng 252000, China, and also with the School of Computer Science and Technology, Shandong Jianzhu University, Jinan, Shandong 250101, China (e-mail: qlishan@163.com).

Digital Object Identifier 10.1109/TNSRE.2023.3309847

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

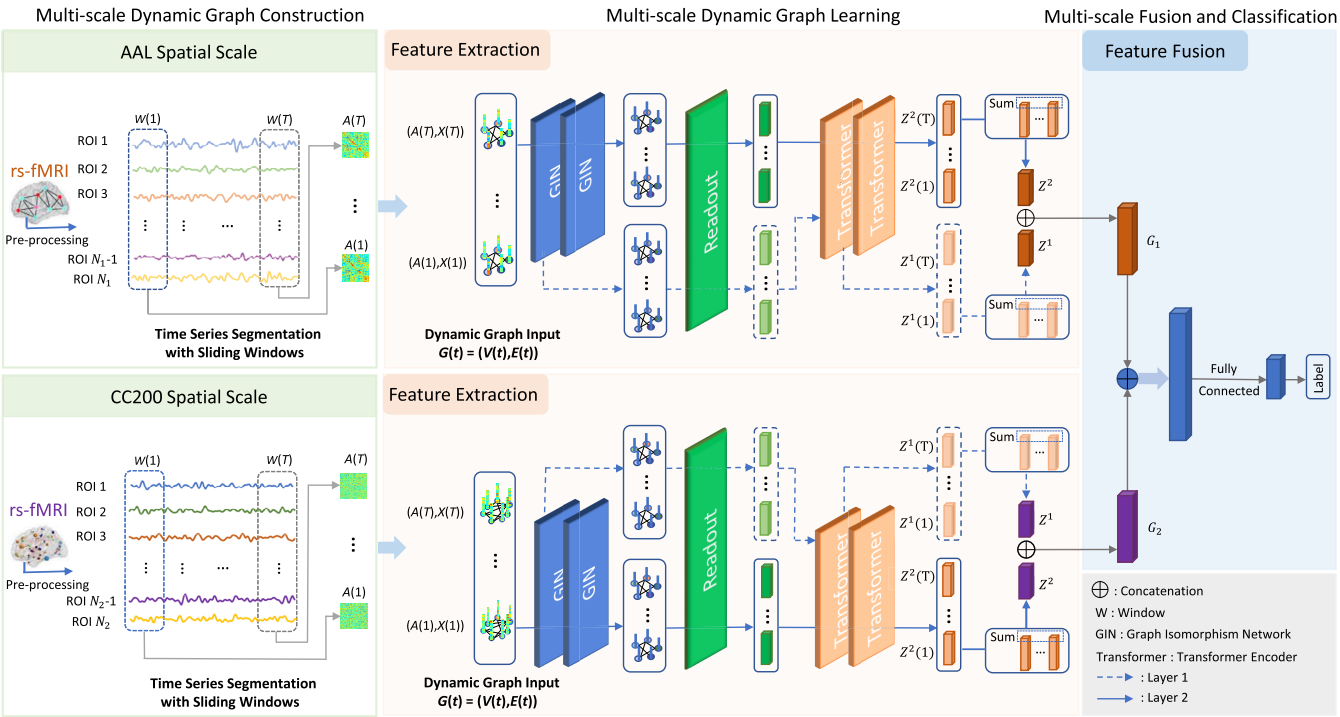

Fig. 1. Overview of multi-scale dynamic graph learning (MDGL), consisting of three major components: (1) multi-scale dynamic graph construction using different brain atlases (*e.g.*, AAL [23] and CC200 [24]) with resting-state fMRI (rs-fMRI), (2) multi-scale dynamic graph learning via graph isomorphism network (GIN), Readout operation, and Transformer encoder, (3) multi-scale feature fusion for disease detection.

scale dynamic graph representations and feed the final FCN representation into a fully connected layer and a Softmax layer for automated brain disorder diagnosis. Therefore, the proposed framework not only focuses on the topological information at different spatial scales, but also considers the dynamic properties of FCNs. Experimental results on the Autism Brain Imaging Data Exchange (ABIDE) initiative [25] and the REST-meta-MDD consortium [26] demonstrate that our proposed MDGL outperforms several state-of-the-art methods in automated identification of two brain disorders with rs-fMRI data. The code is available at https://github.com/myl1030/MDGL.git.

The major contributions of this work are summarized below.

- A multi-scale dynamic graph learning (MDGL) framework is designed to leverage complementary information present in different scales and spatiotemporal dynamics of fMRI data to accurately detect brain disorders.
- The MDGL can capture spatiotemporal dynamic features from fMRI at various spatial resolutions. Experiments have shown it to be superior to several state-of-the-art methods in distinguishing autistic patients from healthy controls, with strong discriminatory power.
- The MDGL is a general framework that can be applied to the analysis of major depressive disorder (MDD) and gender classification. We conducted experiments on two datasets for MDD detection and gender classification, with results suggesting the effectiveness of our proposed method in classification and its strong generalization ability when applied to different datasets.

The rest of this paper is organized as follows. We briefly review related studies in Section II. In Section III, we introduce the materials and the proposed method. In Section IV, we present experimental setup, competing methods, and experimental results. In Section V, we discuss major parts of our proposed method and point out several limitations of the current work and potential future research directions. Finally, we summarize this paper in Section VI.

II. Related Work

A. Traditional Learning Methods for Brain FCN Analysis

Brain FCNs derived from fMRI data can capture abnormal or dysfunctional neural activities in the brain to aid in disease diagnosis. Traditional fMRI-based studies typically extract features (*e.g.*, local clustering coefficients and connectivity strength) from FCNs for disease classification [27]. For example, Liu et al. [28] extracted connectivity strength features from FCNs to aid in the diagnosis of diseases. Wee et al. [29] extracted local clustering coefficients from structural and functional brain networks as feature representations of each subject respectively. Although the traditional machine learning methods achieved good results in the analysis of FCNs, the definition of human-engineered FCN features usually requires specific expertise and prior knowledge of brain functional connectivity. Moreover, the extraction of hand-crafted FCN features is separated from the downstream task, and thus may be suboptimal representations for specific tasks.

B. Graph Isomorphism Network for Brain FCN Analysis

Graph neural networks (GNNs) that can automatically learn graph representation have been increasingly used for graph-structured FCN analysis. Whereas most GNNs often rely on empirical intuition, heuristics, and trial-and-error design. Xu et al. [30] proposed a theoretical framework for analyzing the representational power of GNN and developed a novel variant of GNN, the graph isomorphic network (GIN), with discriminative power comparable to that of the Weisfeiler-Lehman (WL) test. Given advancements in deep learning techniques, GIN has increasingly been adopted for modeling the topological features of brain FCN. For example, Kim et al. [31] developed a graph representation learning method based on the GIN for FCN analysis, achieving good performance in the task of graph classification. Patel et al. [32] proposed a multimodal GIN that uses fMRI task data to classify adolescent gender at different stages. Yang et al. [33], on the other hand, improved the GIN to learn the features of FCNs while taking into account the importance of each node in the classification and in the identification of brain diseases. Meng et al. [34] developed IsoNN, a novel isomorphic neural network that matches the input graph to a template and extracts its isomorphic features for representation learning. Furthermore, Zhao et al. [35] proposed a new classification framework based on rs-fMRI data and developed a graph encoder with convolution by a GIN for supervised training. However, these studies only consider the static FCN for fMRI representation learning and ignore the temporal dynamics in fMRI time series [18], [19].

C. Dynamic Functional Connectivity Based on fMRI

Dynamic functional connectivity is an extension of traditional FC that captures changes in whole-brain connectivity over time [36]. It has been shown that the interaction patterns of the intrinsic network can be captured in the fluctuations of FC [36], [37]. Moreover, dynamic functional connectivity features have demonstrated superior performance in detecting diseases (such as schizophrenia and depression) compared to static analysis of FCNs [38], [39], [40]. Thus, the assumption that connections between brain regions remain static throughout numerous fMRI-based studies has some limitations. A few GNN models have been recently developed for analyzing temporal dynamics of brain FCNs. For example, Wang et al. [10] designed a spatial-temporal dependency model to capture temporal features of fMRI for brain disease diagnosis, where long short-term memory (LSTM) modules were used to capture temporal dependency of the whole brain based on rs-fMRI data. Gadgil et al. [16] designed a spatiotemporal graph convolutional network (ST-GCN) to capture dynamic features of FCNs by adding convolutional layers in the temporal dimension. Kim et al. [41] designed a novel method for learning dynamic graph learning with spatial-temporal attention. However, these methods only pay attention to single-scale topological information of brain FCNs using one single brain template for brain ROI division [16]. Intuitively, multi-scale FCNs derived from different brain atlases with multi-resolution ROI partitions can provide complementary topological information of the brain and help improve the learning performance of GNN models.

TABLE I

PHENOTYPIC INFORMATION ON 184 SUBJECTS FROM NEW YORK UNIVERSITY (NYU) IN THE AUTISM BRAIN IMAGING DATA EXCHANGE (ABIDE) DATASET [25]. ASD: AUTISM SPECTRUM DISORDER; NC: NORMAL CONTROL; M/F: MALE/FEMALE; FIQ: FULL-SCALE INTELLIGENCE QUOTIENT. VALUES OF AGE AND FIQ ARE REPORTED IN TERMS OF MEAN±STANDARD DEVIATION

Category	Subject	Gender (M/F)	Age (Years)	FIQ
ASD	79	68/11	14.52 ± 6.97	107.91 ± 16.62
NC	105	79/26	15.81 ± 6.25	113.15 ± 13.12

III. MATERIALS AND METHOD

A. Materials and Data Preprocessing

In the main experiments, we used rs-fMRI data from the New York University (NYU) site in the ABIDE dataset [25]. The NYU site is the largest imaging site in ABIDE with 184 subjects, including 79 autism spectrum disorder (ASD) patients and 105 normal controls (NCs). Table I reports the demographic and clinical information of the study population. The rs-fMRI data from NYU were obtained by 3.0 Tesla Allegra with the following parameters: echo time (TE) $= 15\,ms$, repetition time (TR) $= 2,000\,ms$, inversion angle $= 90$, 33 slices, and 180 time points.

Data preprocessing of rs-fMRI data is performed through the Data Processing Assistant for Resting-State fMRI (DPARSF) [42] pipeline, which includes the following six major steps: (1) discarding the first five volumes, (2) head motion correction, (3) spatial normalization and smoothing, (4) interference signal removal, (5) bandpass filtering $(0.01 - 0.10\ Hz)$ of BOLD time series, (6) dividing the brain into ROIs based on a certain atlas to obtain ROI-level BOLD time series. We choose AAL [23] and CC200 [24] to parcellate the brain into 116 and 200 ROIs, respectively.

B. Proposed Method

In this work, we propose a multi-scale dynamic graph learning (MDGL) framework to capture multi-scale spatiotemporal dynamic representations of rs-fMRI data for automated brain disease diagnosis. As shown in Fig. 1, the MDGL consists of three major components: (1) multi-scale dynamic FCN construction using multiple brain atlases to model multi-scale topological information, (2) multi-scale dynamic graph representation learning to capture spatiotemporal information of FCNs, and (3) multi-scale feature fusion and classification.

A brain FCN of each subject can be modeled as an undirected graph, with each node denoting a specific ROI and each edge denoting functional connection between paired ROIs [43], [44]. The FCN of each subject is expressed as $G = (V, E)$, where V and E denote the sets of nodes and edges, respectively. Considering the dynamic nature of brain FCNs, the dynamic FC graph of each subject is described as $G(t) = \{(V(t), E(t))\}_{t=1}^{T}$, where $V(t)$ and $E(t)$ are the sets of

nodes and edges at the t-th time window/segment, respectively, and T is the total number of time windows.

1) Multi-Scale Dynamic Graph Construction: The brain of each subject can be partitioned into different regions-of-interest (ROIs) based on a specific brain atlas. Brain FCNs have been widely used to help capture the interactions between ROIs for brain disease analysis. Considering the dynamic nature of functional connectivity between ROIs over time, constructing dynamic FCNs via sliding windows can model spatiotemporal dynamics in fMRI [45], [46]. On the other hand, since FCNs constructed at different spatial scales can provide underlying complementary topological information, we will construct multi-scale dynamic FCNs based on multiple brain atlases for each subject in this work.

Specifically, we first use the sliding window technique with window step η to segment fMRI time series to model temporal information in fMRI. Denote the BOLD signal of the i-th ROI at the t-th time window as $x_i(t) \in \mathbb{R}^B, i = (1, \cdots, N)$, where N is the number of ROIs, B is the length of each time window, and the number of overlapping time points for every two adjacent windows is $B - \eta$. Then, we construct dynamic FCNs for each subject based on Pearson correlation (PC), a simple and classical method for estimating FCNs. Since two brain atlases (*i.e.*, AAL with 116 ROIs and CC200 with 200 ROIs) are used for multi-scale ROI partition, we can generate multi-scale dynamic FCNs for each subject, represented as $P(t) = \{P^{116}(t), P^{200}(t)\}_{t=1}^{T}$.

After constructing multi-scale dynamic FCNs, we need to generate a node/ROI-level feature matrix and an adjacency matrix for each subject for subsequent analysis. Here, we use PC coefficients that describe connectivity strength connected to each node as node-level features. Thus, the multi-scale node-level feature matrix can be defined as $X(t) \in \{X^{116}(t), X^{200}(t)\}$, where $X(t) = P(t)$. Besides, each FCN constructed by PC is a dense graph, which may contain noisy or uninformative edges. Following previous studies [47], we sparsify the constructed FCNs with a certain proportion. In this work, we keep the first 30% strongest edges (that is, the first 30% largest elements in $X(t)$ are set to 1, and others are set to 0). The multi-scale adjacency matrix can be represented as $A(t) = \{A^{116}(t), A^{200}(t)\}$, where $A^N(t) \in \{0, 1\}$.

2) Multi-Scale Dynamic Graph Learning: With multi-scale FCNs as input, we propose to capture spatiotemporal dynamic information of fMRI via multi-scale dynamic graph representation learning with a graph isomorphism network (GIN), a Readout operation, and a Transformer encoder.

With the development of deep learning, GNN is widely used to capture topological information in brain graphs/FCNs and update node features. Generally, GNN consists of the AGGREGATE and COMBINE functions, formulated as follows:

$$\begin{aligned} a_v^{(k+1)} &= \Psi^{(k+1)}(\{h_u^{(k)} : u \in N(v)\}) \\ h_v^{(k+1)} &= \Phi^{(k+1)}(h_v^{(k)}, a_v^{(k+1)}) \end{aligned} \tag{1}$$

where $N(v)$ is a set of neighboring nodes of the center node v, $h_v^{(k)}$ is the feature vector of the node v at the k-th layer, $\Psi(\cdot)$ denotes the AGGREGATE function for aggregating the features of neighboring nodes, and $a_v^{(k+1)}$ is the neighborhood information after aggregation at the layer $k+1$. Also, $\Phi(\cdot)$ denotes the COMBINE function that updates the features of node v at the layer $k+1$ via nonlinear transformation based on aggregated neighborhood feature $a_v^{(k+1)}$ and the feature $h_v^{(k)}$ of the node v itself at the layer k.

Based on different AGGREGATE and COMBINE functions, researchers have developed many variants of GNNs, and GIN is a typical example that has been widely used for graph representation learning [31]. The GIN replaces the traditional mean aggregator on GNNs and uses summation operations to aggregate neighborhood node features. And GIN introduces a multilayer perceptron (MLP) to transform the aggregated neighborhood node features and the original node feature in a nonlinear way, where the importance weights of the original node information can be adaptively learned by the supervision of category labels. Mathematically, a GIN layer can be formulated as follows:

$$h_v^{(k+1)} = \theta^{(k+1)}((1 + \varepsilon^{(k+1)})h_v^{(k)} + \sum\nolimits_{u \in N(v)} h_u^{(k)}) \tag{2}$$

where $h_v^{(k)}$ represents the feature vector of the node v at the k-th layer of GIN, $\theta(\cdot)$ is defined as an MLP operation, and ε is the learnable weights.

As shown in Fig. 2, in our MDGL framework, we use the adjacency matrix $A \in A(t)$ and the node feature matrix $X \in X(t)$ as input. The node features are updated by aggregating the neighboring nodes through the two GIN layers. Since we consider the dynamic nature of nodes and FCNs, the node feature for each time window will be updated by K ($K = 2$ in this work) GIN layers. Taking a specific time window as an example, the output node feature matrix F after two GIN layers can be formulated as follows:

$$F = f(X, A) = \delta(\varepsilon^{(1)} I + A)[\delta(\varepsilon^{(0)} I + A) X W^{(0)}] W^{(1)} \tag{3}$$

where $\delta(\cdot)$ is the ReLU activation function after each GIN layer, ε is a learnable parameter initialized to zero for the purpose of knowing the weight of node information retained in that layer by the previous layer, $W^{(0)}$ and $W^{(1)}$ denote the node information weight matrices learned by MLP for the first and second GIN layers, respectively.

To perform graph classification, we input the learned node feature F to a readout layer to obtain a graph-level feature vector. We use Sero [41] as an attention-based readout method that follows the MLP based on squeezed excitation networks [48], and it has been shown that such spatial dimension squeezed excitation modules can effectively boost the performance of GNN models [49]. The Sero learns the degree of contribution of different nodes/ROIs to the classification performance by MLP attention, defined as follows:

$$S = sigmoid(Q^{(2)} \sigma(Q^{(1)} F \phi_{mean})) \tag{4}$$

where S is the attention weights of nodes/ROIs, $\phi_{mean} = [\frac{1}{N}, \frac{1}{N}, \cdots, \frac{1}{N}]^T$, and $Q^{(1)}$, $Q^{(2)}$ are the learnable weight matrices in MLP. The right part of Fig. 2 illustrates the

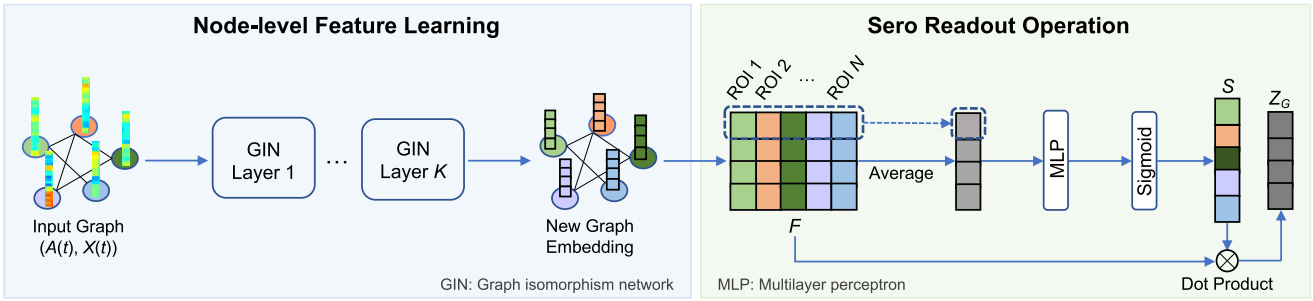

Fig. 2. Node-level feature learning via K graph isomorphism network (GIN) layers and a Sero readout operation [41] for spatial attention learning in the proposed MDGL framework, where $A(t)$ is the adjacency matrix and $X(t)$ is the node feature matrix. Here, F is the aggregated node feature matrix after the aggregation operation via GIN layers, S is the spatial attention weight of N nodes/ROIs, and Z_G is the graph-level representation of the input subject with spatial attention.

Sero readout operation. The graph representation with spatial attention output from the Sero readout layer is defined as:

$$Z_G = FS \tag{5}$$

Another purpose of constructing dynamic graphs is to capture temporal dynamic information in fMRI series and explore the contribution degree of dynamic FCNs in the temporal dimension. To obtain temporal attention, we use a single-head Transformer encoder [50] based on the sequence of graph representations. The self-attention weights learned by the Transformer encoder can measure the importance of temporal dimension. Thus, a new sequence of dynamic graph representations $Z(t)$ with temporal and spatial attention is generated after Transformer layers. Assuming that the new sequence of graph representations obtained by the k-th Transformer layer is defined as $Z^k(t)$, we sum the sequence of graph representations $Z^k(t)$, as defined as follows:

$$Z^{(k)} = sum(Z^{(k)}(t)), \quad t \in \{1, \cdots, T\} \tag{6}$$

Then, we obtain the final graph representation G by concatenating the learned feature vectors $\{Z^{(k)}\}_{k=1}^{K}$, where K is the number of Transformer encoder layers. As shown in Fig. 1, the multi-scale dynamic graph constructed in Section III-B.1 is learned parallelly at two spatial scales via the GIN layers, the readout layers, and two Transformer encoder layers. The final dynamic graph representations $G_1 \in \mathbb{R}^D$ and $G_2 \in \mathbb{R}^D$ with spatiotemporal attention at different scales are obtained, where D is the dimension of learned graph features.

3) Multi-Scale Feature Fusion and Classification: After multi-scale dynamic graph learning, we obtain the dynamic graph features (*i.e.*, G_1 and G_2) at two different scales for each subject. To fully utilize the potential complementary information provided by different spatial scales, we design a feature fusion module to integrate multi-scale features and use the fused final graph features for the disease classification/prediction task. The feature fusion module is denoted as:

$$G_1 \,||\, G_2 \in \mathbb{R}^{2D} \tag{7}$$

where $||$ indicates concatenation operation. Each subject's multi-scale dynamic graph representation is a $2D$-dimensional representation vector, which is then fed into a fully connected layer and a Softmax layer for prediction.

To train the MDGL model, the cross-entropy loss function L_c is used to optimize the model parameters. To avoid the issue of gradient disappearance or gradient explosion during optimization, we add an orthogonal constraint to the node feature matrix, defined as follows:

$$L_{otho} = || \frac{1}{m} F^T F - I \,||_2 \tag{8}$$

where $m = max(F^T F)$. Then, the total loss function L of the MDGL model can be expressed as:

$$L = L_c + \lambda L_{otho} \tag{9}$$

where L_c is the cross-entropy loss function, L_{otho} is the orthogonal constraint at two scales, and λ is the hyperparameter to balance the contributions from the two terms in Eq. 9.

C. Implementation Details

The proposed MDGL model is implemented on Pytorch with NVIDIA Quadro P2200, trained by minimizing the total loss function L in Eq. 9. We set the number of network layers (both GIN layer and transformer encoder layer) $K = 2$, the embedding dimension of each scale $D = 64$, the window length $B = 40$, the window step size $\eta = 3$, and the regularization factor $\lambda = 1.0 \times 10^{-6}$. The MDGL model is optimized using the Adam algorithm, with a learning rate of 0.001, training epochs of 40, a batch size of 8, and a dropout rate of 0.5. In the Discussion section, we will explore the effect of hyperparameters on the performance of MDGL.

IV. Experiments

A. Experimental Settings

The MDGL model is trained in an end-to-end supervised training mode. A five-fold cross-validation strategy is used in the experiments. Specifically, we first randomly divide all subjects into five subsets. Then, we use four subsets as training set and take the remaining one subset as testing set. Concurrently, internal cross-validation is applied for the selection of hyper-parameters (*i.e.*, the parameters discussed in Section V). Specifically, for each fold of the divided training set, a secondary division is implemented. Here, 80% of the data is used for model training, and the remaining 20% is used

as a validation set for testing. The parameters of the model are fine-tuned based on the classification results of the validation set, enabling us to choose the best hyper-parameters. The classification results in terms of mean and standard deviation among these five folds are recorded. Five metrics are used for performance evaluation, including (1) classification accuracy, (2) precision, (3) recall, (4) area under the receiver operating characteristic curve (AUC), and (5) F1-score.

B. Methods for Comparison

We compare our MDGL with three traditional machine learning methods and eight deep learning methods. For fair comparison, these competing methods employ very similar learning schemes to MDGL, as detailed below.

1) MCC+SVM: Based on PC coefficients, we first construct two FCNs using multi-scale atlases (*e.g.*, AAL and CC200 atlases) for each subject. Then, we extract closeness centrality (CC) of two-scale FCNs as features, obtaining 116-dimensional and 200-dimensional vectors respectively. Finally, the two feature vectors are concatenated and the obtained 316-dimensional vector is fed into a linear support vector machine (SVM) with default parameters for classification [51].

2) MLCC+SVM: Similar to MCC+SVM, we first construct multi-scale FCNs using different atlases for each subject. Then, we extract local clustering coefficient (LCC) [29] of each node from two FCNs as feature vectors (116-dimensional and 200-dimensional, respectively). Finally, we concatenate the two feature vectors to obtain a 316-dimensional representation, followed by an SVM for classification.

3) MEC+SVM: This method first extracts eigenvector centrality (EC) [52], which measures the transitive influence of nodes in a graph, as network representation for each single-scale FCN. The final 316-dimensional representation is fed into an SVM for disease classification.

4) CNN: Considering that many convolutional neural networks (CNNs) [53], [54] have been used for neuroimage analysis, we also compare our method with a CNN model. In CNN, we divide ROIs for each subject based on the AAL atlas and compute the FC matrix as input to the CNN. Here, the CNN contains two convolutional layers and a fully connected layer. The channel numbers of the two convolutional layers are 4 and 8, respectively, and the corresponding sizes of the convolution kernels are 3×3 and 5×5, respectively.

5) GIN: The method first constructs a single-scale FCN for each subject based using the AAL atlas, and then uses them as input of GIN. The GIN consists of two convolutional layers, a readout layer and a fully connected layer, and the output dimension of each convolutional layer is 64.

6) GCN: Similar to GIN, we use the AAL atlas to divide the ROIs and calculate the PC coefficients between the ROIs to construct the FCN, which is then fed into a GCN model. The GCN contains two convolutional layers, a readout layer and a fully connected layer, and each convolutional layer has an output dimension of 64.

7) ST-GCN: The spatiotemporal graph convolutional network (ST-GCN) [16] can help model dynamic features of functional connectivities. In this method, we first take the BOLD signal of ROIs divided by AAL atlas as inputs. Then, two spatial-temporal graph convolution units are used for learning spatiotemporal features in fMRI series. Finally, we obtain the graph-level representation via global average pooling and feed it into a fully connected layer for disease classification. Each convolution unit has an output dimension of 64.

TABLE II
RESULTS (MEAN±STANDARD DEVIATION) ACHIEVED BY DIFFERENT METHODS IN THE TASK OF ASD VS. NC CLASSIFICATION, AND THE BEST RESULTS ARE SHOWN IN BOLD

Method	Accuracy	Precision	Recall	AUC	F1-score
MCC+SVM	0.60±0.09	0.63±0.24	0.23±0.12	0.62±0.11	0.32±0.14
MLCC+SVM	0.61±0.09	0.69±0.21	0.21±0.16	0.65±0.08	0.29±0.19
MEC+SVM	0.61±0.06	0.68±0.21	0.27±0.13	0.63±0.11	0.35±0.12
CNN	0.65±0.04	0.66±0.11	0.45±0.12	0.69±0.06	0.52±0.07
GIN	0.64±0.08	0.61±0.15	0.57±0.09	0.65±0.05	0.57±0.08
GCN	0.64±0.04	0.59±0.06	0.56±0.18	0.67±0.06	0.55±0.11
ST-GCN	0.57±0.07	0.58±0.22	0.42±0.12	0.57±0.08	0.44±0.04
MCNN	0.67±0.06	0.62±0.09	0.53±0.24	0.73±0.06	0.54±0.22
MGIN	0.65±0.08	0.60±0.11	0.56±0.16	0.70±0.06	0.57±0.13
MGCN	0.70±0.09	0.67±0.14	0.62±0.15	0.70±0.10	0.63±0.11
MST-GCN	0.60±0.05	0.54±0.05	0.64±0.28	0.63±0.08	0.55±0.14
MDGL (Ours)	**0.72±0.07**	**0.71±0.14**	**0.66±0.15**	**0.74±0.10**	**0.66±0.09**

8) MCNN: This is a multi-scale variant of CNN, where we use CNN to extract features from FCNs at two different scales (based on AAL and CC200). Then, we concatenate the features at the two scales learned from CNNs and input the concatenated vector to a fully connected layer for classification. The MCNN model consists of two convolutional layers and a fully connected layer.

9) MGIN: This method first uses AAL and CC200 atlases for ROI partition respectively, and calculates the PC coefficient of each ROIs time series to obtain two FC matrices (size: 116×116 and 200×200). Similar to our MDGL, only the top 30% strongest edges are kept in the FC matrices at two scales, followed by two parallel branches for FCN feature learning with each one using a GIN as the backbone. The learned features are further concatenated, followed by a fully-connected layer and a Softmax layer for classification. Each GIN backbone consists of two convolutional layers and a readout layer.

10) MGCN: This is a multi-scale variant of GCN, where the input contains FCNs constructed using two atlases. Similar to MDGL, the learned features at two scales will be concatenated and fed into a fully connected layer for classification.

11) MST-GCN: In this method, we first construct FCNs at different scales (size: 116×116 and 200×200) based on the AAL atlas and the CC200 atlas, respectively. At each scale, ROI-based BOLD signals of each subject with are fed into two spatiotemporal convolution (ST-GC) modules [16] for feature learning. Each ST-GC module consists of a spatial convolution layer and a temporal convolution layer. The output of each ST-GC module will be fed to the average readout layer. It finally concatenates the learned features at two scales, followed by a fully connected layer and a Softmax layer for classification.

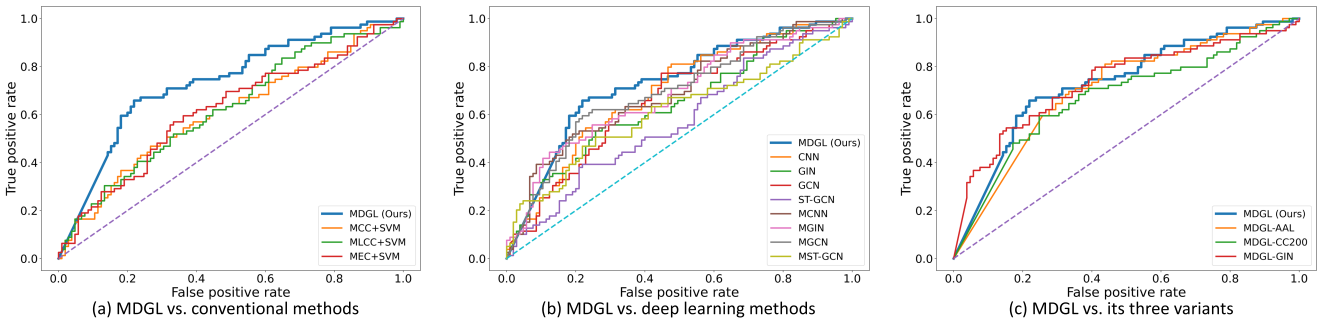

Fig. 3. Receiver operating characteristic (ROC) curves achieved by different methods in the task of ASD vs. NC classification.

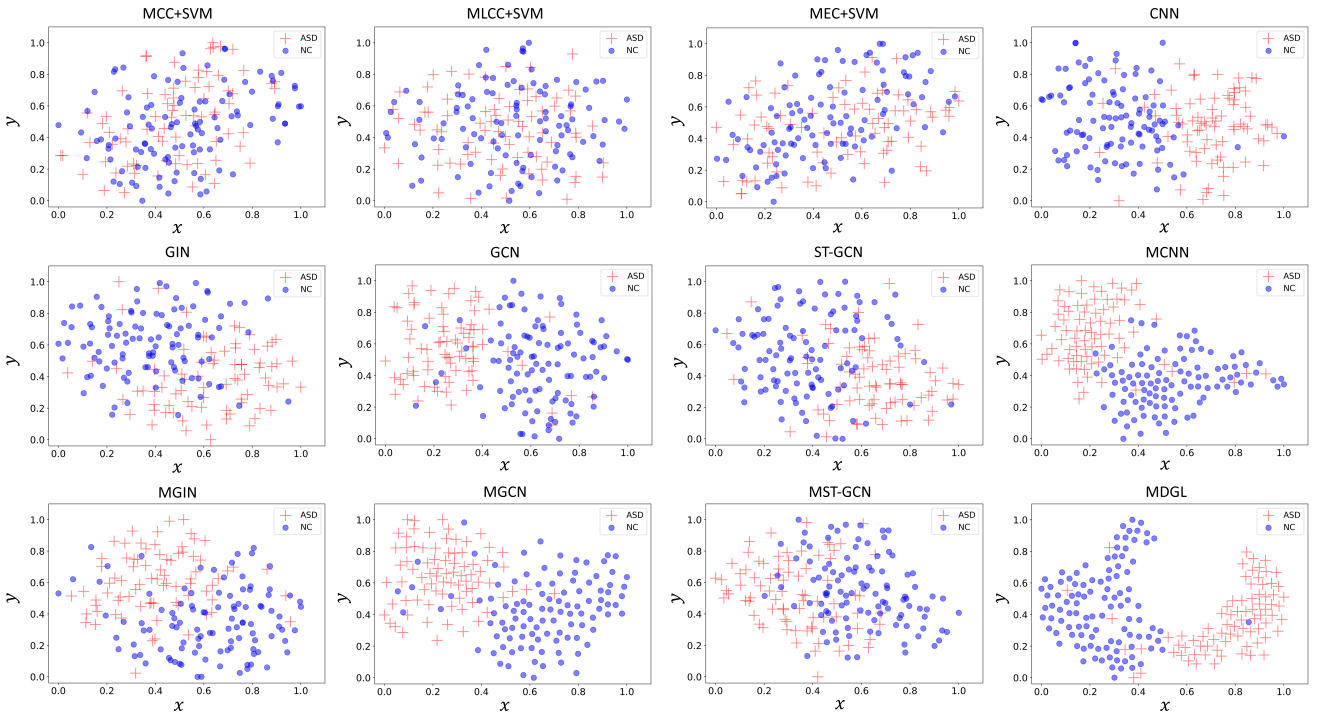

Fig. 4. Data distribution visualization of rs-fMRI features of eleven competing methods and the proposed MDGL via t-SNE [55]. The horizontal and vertical axes represent two dimensions after t-SNE projection.

C. Experimental Results

We report the experimental results of the proposed MDGL and eleven competing methods in ASD vs. NC classification in Table II. In Fig. 3 (a)-(b), we show the ROC curves achieved by our MDGL and the competing methods. From Table II and Fig. 3, we have the following interesting findings.

First, our MDGL and eight deep learning methods generally outperform the three conventional methods (*i.e.*, MCC+SVM, MLCC+SVM and MEC+SVM) that rely on handcrafted features. For example, the MDGL achieves the improvement of 11% and 9% in terms of ACC and AUC values, respectively, compared with the best traditional method (*i.e.*, MLCC+SVM). The possible reason is that the representation automatically learned by deep learning models is more beneficial to downstream classification tasks. *Second*, multi-scale deep learning methods (*i.e.*, MCNN, MGIN, MGCN and MST-GCN) are generally superior to their single-scale counterparts (*e.g.*, CNN, GIN, GCN and ST-GCN) that only use singe-scale FCNs. This result implies that brain FCNs constructed at multiple spatial scales may provide complementary topological information for disease classification, thus helping improve classification performance. *Third*, compared to other multi-scale deep models (*e.g.*, MCNN, MGIN and MGCN) without considering spatiotemporal dynamics of brain FCN, the proposed MDGL achieves better performance, which implies that mining underlying dynamic information in fMRI series is very important to brain disease detection. *Besides*, when comparing the two multi-scale methods (*i.e.*, MST-GCN and our MDGL) that consider spatiotemporal dynamics of FCNs, the MST-GCN is inferior to our MDGL. This may be due to that the MST-GCN can only capture local temporal information via convolution operations that are limited by kernel sizes [56], while our method with transformer layers can learn more discriminative features in the temporal dimension.

TABLE III
STATISTICAL SIGNIFICANCE ANALYSIS RESULTS BY COMPARING THE PROPOSED MDGL AND ELEVEN COMPETING METHODS IN ASD DETECTION

Pairwise Comparison	p-value	$p<0.05$
MDGL vs. MCC+SVM	2.74×10^{-2}	Yes
MDGL vs. MLCC+SVM	2.65×10^{-2}	Yes
MDGL vs. MEC+SVM	2.60×10^{-2}	Yes
MDGL vs. CNN	2.97×10^{-2}	Yes
MDGL vs. GIN	2.55×10^{-2}	Yes
MDGL vs. GCN	3.09×10^{-2}	Yes
MDGL vs. ST-GCN	3.99×10^{-2}	Yes
MDGL vs. MCNN	1.41×10^{-2}	Yes
MDGL vs. MGIN	1.96×10^{-2}	Yes
MDGL vs. MGCN	7.07×10^{-3}	Yes
MDGL vs. MST-GCN	1.44×10^{-2}	Yes

TABLE IV
CLASSIFICATION RESULTS OF OUR MDGL AND ITS VARIANTS IN ASD VS. NC CLASSIFICATION, AND BEST RESULTS ARE MARKED IN BOLD

Method	Accuracy	Precision	Recall	AUC	F1-score
MDGL-AAL	0.65±0.07	0.60±0.11	**0.76±0.11**	0.71±0.05	0.65±0.04
MDGL-CC200	0.67±0.04	0.62±0.04	0.60±0.18	0.69±0.09	0.59±0.09
MDGL-GIN	0.70±0.09	0.69±0.11	0.55±0.14	0.73±0.05	0.61±0.12
MDGL (Ours)	**0.72±0.07**	**0.71±0.14**	0.66±0.15	**0.74±0.10**	**0.66±0.09**

We further compare the probability scores (predicted by MDGL and a competing method) of subjects belonging to each category via t-test.

In Table III, we report the obtained p-values in t-test, from which one can see that p-values are lower than the significance level (0.05). This suggests that the MDGL is significantly different from the competing methods.

D. Visualization of Learned fMRI Features

We visualize rs-fMRI features produced by different methods using t-Distributed Stochastic Neighbor Embedding (t-SNE) [55], as shown in Fig. 4. From this figure, we can see that the feature distributions of two categories (including ASD and NC) generated from three traditional methods (*e.g.*, MCC+SVM, MLCC+SVM and MEC+SVM) have no distinct difference, compared with that generated from MDGL and deep learning methods. Furthermore, compared with eight deep learning models, our MDGL can yield features with clearer distribution differences for the two categories. This further validates the strong discriminative power of our MDGL in distinguishing ASD patients from normal controls.

V. DISCUSSION

A. Ablation Study

To investigate the role of our multi-scale learning and Transformer-based temporal feature learning strategies, we compare MDGL and its three variants (called **MDGL-AAL**, **MDGL-CC200** and **MDGL-GIN**, respectively). The MDGL-AAL and MDGL-CC200 only consider single-scale information based on the AAL and CC-200 atlases, respectively. Similar to MDGL, these two methods consist of two GIN layers, a Sero readout layer, two Transformer encoder layers, a fully connected layer and a Softmax layer. The MDGL-GIN employs the same multi-scale FCN input as MDGL and a very similar network architecture (except for two Transformer encoders). That is, no Transformer is used in MDGL-GIN.

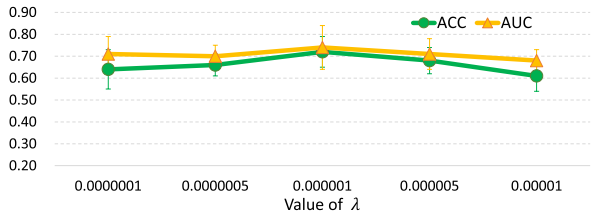

Fig. 5. Accuracy and AUC values achieved by MDGL with different parameter values of λ in ASD and NC classification.

For a fair comparison, we set the model parameters of the MDGL methods and its three variants to be consistent. The experimental results of four methods in ASD vs. NC classification on ABIDE are reported in Table IV, with ROC curves shown in Fig. 3 (c). It can be seen from Table IV and Fig. 3 (c) that our MDGL achieves better performance in most cases, when compared with its two single-scale variants (*e.g.*, MDGL-AAL and MDGL-CC200). This result demonstrates that constructing FCNs at different spatial scales can provide complementary topological information to improve classification performance. Besides, our MDGL method is superior to the MDGL-GIN that discards Transformer encoders, which validates the necessity of modeling temporal dynamics conveyed in resting-state fMRI time series.

B. Influence of Hyperparameter

In Eq. (9), the hyperparameter λ is used to adjust the contribution of the orthogonal regularization. We report the classification accuracy and AUC values of MDGL with different λ values in Fig. 5. As is shown in Fig. 5, the ACC and AUC values generally first become larger and then become smaller with the increase of hyperparameter λ. With $\lambda = 10^{-6}$, the MDGL achieves the best performance in terms of ACC and AUC. Also, it can be seen that when the value of λ is too small (*e.g.*, $\lambda = 10^{-7}$), the performance of MDGL is not good. The possible reason is that when the orthogonal regularization constraint is too weak (with a small λ), the learned node feature matrix is far away from an orthogonal matrix, which may lead to gradient explosion or gradient dispersion and reduce the classification performance. But too strong orthogonal regularization constraint (*e.g.*, $\lambda = 10^{-5}$) can also degrade model performance, which may be due to that too strong orthogonal regularization constraint weakens the impact of cross-entropy loss that is oriented to the downstream task.

C. Influence of FCN Sparsity

In this work, we first construct preliminary multi-scale FC networks based on Pearson's correlation (PC). Then, we set the top $q\%$ largest values of each FC matrix to 1 and set the remaining values to 0, resulting in an adjacency matrix for subsequent analysis. In the main experiments, the FCN

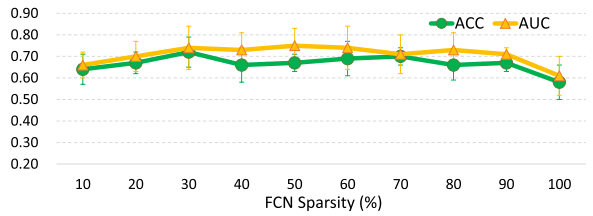

Fig. 6. Accuracy and AUC values achieved by MDGL with different FCN sparsity thresholds in ASD vs. NC classification.

sparsity threshold q is set as 30. To explore the impact of FCN sparsity on the model, we perform experiments under different FCN sparsity values within the range $\{10, 20, \cdots, 100\}$, and report the experimental results of MDGL in Fig. 6. As shown from Fig. 6, the MDGL achieves the best classification performance when the top 30% strongest edges are retained. The performance of MDGL is relatively worse when the FCN sparsity threshold q is set as 100 or 10. The possible reason is that too dense brain FCNs contain some noisy or redundant information, but too sparse FCNs may lose useful topological information. Therefore, constructing preliminary FCNs with proper sparsity is important for subsequent analysis.

D. Influence of Multi-Scale Feature Fusion Strategy

In the main experiments, we adopt the feature-level fusion strategy (via simple concatenation followed by a fully connected layer) to integrate multi-scale dynamic graph representations. To further explore the effect of different fusion strategies on our model, we compare the MDGL with its variant **MDGL-*i*** with a decision-level fusion strategy. Specifically, at each spatial scale, we first feed the constructed FCNs into GIN layers, readout layers and the Transformer encoder layers for learning dynamic graph representation. Then, we separately feed each single-scale representation into a fully connected layer to generate estimated probability scores. These two-scale probability scores will be used to obtain final probability scores for classification through a weighted summation strategy. We change the weights between the AAL and CC200 spatial scales within the range $i = \{\frac{0.2}{0.8}, \frac{0.5}{0.5}, \frac{0.8}{0.2}\}$, and report the experimental results of MDGL-i in Fig. 7.

From Fig. 7, we can see that the MDGL-3 with a higher weight of the AAL spatial space achieves relatively better performance in most cases, compared with the other decision-level fusion methods (*e.g.*, MDGL-1 and MDGL-2). In addition, our proposed MDGL with feature-level fusion is generally superior to those methods that use feature-level fusion (*e.g.*, MDGL-1, MDGL-2 and MDGL-3). This further demonstrates the effectiveness of fusing multi-scale dynamic representation at the intermediate feature level.

E. Influence of Network Architecture

In the main experiments, we set the number of GIN and Transformer encoder layers as $K = 2$ and the embedding dimension of hidden layers as $D = 64$. To investigate the influence of different network architectures, we vary the network depth K and the embedding dimension D used in MDGL, and present the experimental results of MDGL in Table V. From Table V, we can see that our MDGL can yield better performance when the network depth K is set to 2 and 3 (with $D = 64$). With $K = 1$ and $K = 4$, the MDGL cannot generate good results. The reason may be that shallow networks can not effectively learn dynamic graph representations, while too deep network may degrade the model generalization caused by possible overfitting problems. *Besides*, when the network depth is fixed (*i.e.*, $K = 2$), the MDGL using different embedding dimensions achieves comparable AUC results, which implies MDGL is robust to the network width.

TABLE V
RESULTS OF OUR MDGL IN ASD VS. NC CLASSIFICATION BASED ON DIFFERENT NETWORK ARCHITECTURES, WITH K DENOTING THE NUMBER OF GIN AND TRANSFORMER ENCODER LAYERS AND D DENOTING THE EMBEDDING DIMENSION

$[K,\ D]$	Accuracy	Precision	Recall	AUC	F1-score
[1, 64]	0.65±0.06	0.60±0.08	0.53±0.12	0.70±0.06	0.56±0.09
[2, 64]	**0.72±0.07**	**0.71±0.14**	**0.66±0.15**	**0.74±0.10**	**0.66±0.09**
[3, 64]	0.70±0.07	0.65±0.08	0.60±0.18	0.70±0.08	0.61±0.12
[4, 64]	0.66±0.09	0.62±0.11	0.58±0.12	0.70±0.09	0.59±0.09
[2, 32]	0.63±0.06	0.56±0.10	0.55±0.26	0.70±0.04	0.53±0.16
[2, 64]	**0.72±0.07**	**0.71±0.14**	0.66±0.15	**0.74±0.10**	**0.66±0.09**
[2, 96]	0.68±0.06	0.65±0.10	0.61±0.12	**0.74±0.06**	0.62±0.08
[2, 128]	0.66±0.06	0.61±0.09	**0.67±0.10**	0.72±0.05	0.63±0.05
[2*, 64]	0.67±0.05	0.64±0.12	0.59±0.24	0.71±0.07	0.58±0.12

TABLE VI
CLASSIFICATION RESULTS ACHIEVED BY MDGL USING DIFFERENT READOUT METHODS IN ASD VS. NC CLASSIFICATION, WITH BEST RESULTS SHOWN IN BOLD

Method	Accuracy	Precision	Recall	AUC	F1-score
Mean	0.66±0.07	0.53±0.26	0.51±0.26	0.71±0.07	0.51±0.26
Max	0.63±0.04	0.59±0.09	0.53±0.11	0.66±0.05	0.54±0.04
Garo	0.70±0.06	0.66±0.09	0.65±0.04	0.73±0.07	0.65±0.05
Sero (Ours)	**0.72±0.07**	**0.71±0.14**	**0.66±0.15**	**0.74±0.10**	**0.66±0.09**

It is worth noting that in the main experiments, we concatenate the features learned by the first and second layers as the final features for classification. To compare the features of different layers, we directly use the output features from the last layer (*i.e.*, Layer 2 in Fig. 1) for classification, and the results are also reported in Table V, denoted as [2*, 64]. As can be seen, the operation of retaining the output of each layer in our proposed method achieves better results. This implies that the knowledge learned by the shallow network may contain features that are beneficial for the classification task.

F. Effectiveness of Attention-Based Readout Operation

In the main experiments, we use Sero, an attention-based readout method, to transform the learned node features into a graph-level feature vector. To explore the effects of different readout operations, we use Garo, Mean and Max as readout methods for our proposed model, respectively. Specifically, Garo follows the attention of transformer based on key query embedding, while Mean is the average readout operation and Max is the maximum readout operation. Results achieved by MDGL with different readout operations are shown in Table VI. This table suggests that MDGL achieves the best

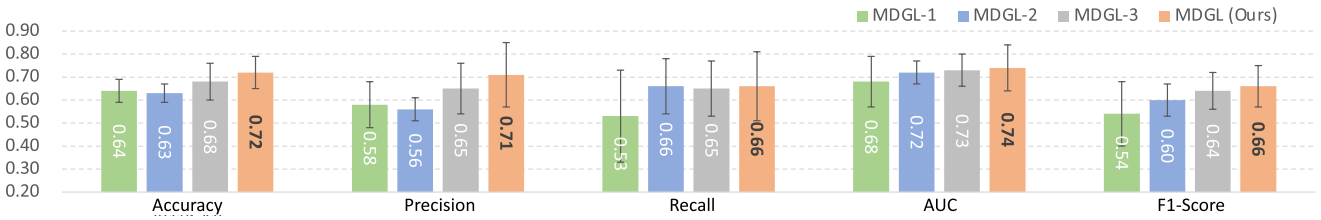

Fig. 7. Results of our MDGL and its variant using decision-level fusion (called MDGL-*i*) in ASD vs. NC classification. MDGL-1, MDGL-2 and MDGL-3 denote the weight ratios of $\frac{0.2}{0.8}$, $\frac{0.5}{0.5}$ and $\frac{0.8}{0.2}$ between the two spatial scales based on AAL and CC200, respectively.

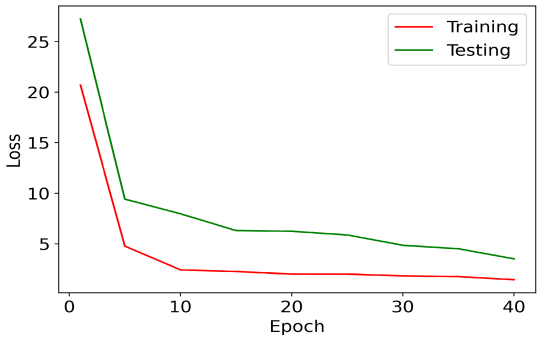

Fig. 8. Convergence of MDGL on the training and testing sets in ASD vs. NC classification.

TABLE VII
RESULTS (MEAN±STANDARD DEVIATION) OF DIFFERENT METHODS IN MDD VS. NC CLASSIFICATION, WITH BEST RESULTS SHOWN IN BOLD

Method	Accuracy	Precision	Recall	AUC	F1-score
MCC+SVM	0.55±0.06	0.57±0.05	0.59±0.07	0.54±0.07	0.58±0.05
MLCC+SVM	0.53±0.05	0.56±0.06	0.57±0.07	0.53±0.04	0.56±0.04
MEC+SVM	0.52±0.03	0.55±0.04	0.55±0.08	0.56±0.03	0.55±0.04
CNN	0.56±0.03	0.58±0.03	0.62±0.12	0.58±0.04	0.59±0.07
GIN	0.56±0.02	0.59±0.06	0.64±0.20	0.59±0.03	0.59±0.07
GCN	0.52±0.04	0.54±0.03	0.64±0.07	0.55±0.03	0.59±0.04
ST-GCN	0.53±0.03	0.55±0.02	0.64±0.16	0.51±0.02	0.58±0.07
MCNN	0.55±0.02	0.58±0.03	0.59±0.13	0.57±0.04	0.57±0.07
MGIN	0.57±0.03	**0.60±0.04**	0.62±0.15	0.59±0.05	0.60±0.06
MGCN	0.56±0.02	0.57±0.03	**0.65±0.05**	0.56±0.04	0.61±0.01
MST-GCN	0.53±0.03	0.54±0.02	0.62±0.17	0.53±0.05	0.57±0.09
MDGL (Ours)	**0.58±0.04**	**0.60±0.06**	**0.65±0.05**	**0.61±0.12**	**0.62±0.05**

TABLE VIII
RESULTS OF STATISTICAL SIGNIFICANCE ANALYSIS BY COMPARING THE PROPOSED MDGL AND EACH OF ELEVEN COMPETING METHODS IN MDD DETECTION

Pairwise Comparison	p-value	$p<0.05$
MDGL vs. MCC+SVM	3.26×10^{-5}	Yes
MDGL vs. MLCC+SVM	1.43×10^{-6}	Yes
MDGL vs. MEC+SVM	1.10×10^{-6}	Yes
MDGL vs. CNN	1.79×10^{-6}	Yes
MDGL vs. GIN	1.66×10^{-5}	Yes
MDGL vs. GCN	3.99×10^{-6}	Yes
MDGL vs. ST-GCN	1.05×10^{-4}	Yes
MDGL vs. MCNN	2.28×10^{-5}	Yes
MDGL vs. MGIN	5.01×10^{-6}	Yes
MDGL vs. MGCN	3.88×10^{-5}	Yes
MDGL vs. MST-GCN	2.68×10^{-7}	Yes

TABLE IX
RESULTS (MEAN±STANDARD DEVIATION) OF DIFFERENT METHODS IN GENDER CLASSIFICATION, WITH BEST RESULTS SHOWN IN BOLD

Method	Accuracy	Precision	Recall	AUC	F1-score
MCC+SVM	0.69±0.03	0.69±0.03	0.67±0.05	0.77±0.03	0.68±0.03
MLCC+SVM	0.63±0.04	0.64±0.03	0.60±0.08	0.71±0.03	0.62±0.03
MEC+SVM	0.69±0.04	0.69±0.05	0.68±0.06	0.75±0.04	0.68±0.03
CNN	0.77±0.02	0.77±0.04	0.77±0.07	0.86±0.02	0.77±0.03
GIN	0.75±0.04	0.76±0.06	0.73±0.06	0.82±0.04	0.74±0.03
GCN	0.67±0.05	0.77±0.06	0.47±0.10	0.75±0.06	0.58±0.09
ST-GCN	0.74±0.08	0.86±0.14	0.65±0.23	0.91±0.02	0.69±0.12
MCNN	0.80±0.04	0.79±0.06	0.80±0.04	0.89±0.03	0.80±0.04
MGIN	0.79±0.04	0.79±0.05	0.80±0.04	0.86±0.03	0.79±0.04
MGCN	0.72±0.05	0.74±0.06	0.65±0.04	0.78±0.04	0.70±0.05
MST-GCN	0.75±0.12	0.91±0.10	0.60±0.32	**0.93±0.02**	0.64±0.30
MDGL (Ours)	**0.86±0.01**	**0.90±0.05**	**0.82±0.08**	0.92±0.01	**0.85±0.02**

classification performance. Moreover, the two attention-based readout operations (*i.e.*, Sero, Garo) outperform the other two methods, demonstrating the effectiveness of using the attention-based readout approach in MDGL.

G. Generalizability to Another Brain Disease

To validate the generalizability of the proposed MDGL, we apply it to fMRI-based diagnosis of another brain disease, major depressive disorder (MDD). We perform experiments on the largest site (*i.e.*, Site 20) of the REST-meta-MDD consortium [26] with rs-fMRI. We discard the fMRI data of the two subjects containing all-zero columns, and use the remaining 531 subjects (with 281 MDD patients and 250 NCs) for MDD detection. Consistent with the main experiments, we first construct multi-scale dynamic FCNs based on AAL and CC200 atlases, and then learn multi-scale dynamic graph representation, followed by multi-scale feature fusion for MDD detection. Table VII reports the experimental results of our MDGL and eleven competing methods in MDD vs. NC classification. It can be seen from Table VII that our proposed MDGL also shows an overall better performance than the competing methods for fMRI-based MDD detection. This validates the generalizability of our MDGL in handling rs-fMRI data for automated brain disease detection. In addition, we report in Table VIII the p-values obtained by t-test, from which it can be seen that the p-values are all below the significance level (0.05). This indicates that MDGL is significantly different from other competing methods in detecting MDD.

H. Reproducibility on Large-Scale Dataset

To further explore the generalization ability of our MDGL, we also evaluate its effectiveness on a larger Human Connectome Project (HCP) dataset. The publicly available HCP1200 processed fMRI data from 812 subjects are used for gender classification, including 410 female subjects and 402 male subjects. The fMRI data were minimally

preprocessed according to [57], [58], where the artifacts were removed using ICA+FIX [16], [59]. Finally, the fMRI images are segmented by group-ICA, and the corresponding time series of each brain region is obtained. Here, we use the fMRI time series divided into 100 and 200 brain regions as two different scales, which were then fed into the MDGL framework for gender classification. Table IX shows the experimental results of our method and 11 competing methods on the HCP dataset, and it can be seen that our method can still achieve good classification performance. These results further validate the generalization ability of our method on different datasets.

I. Convergence Analysis

We visualize the objective function loss changes of our MDGL to gain an intuitive understanding of model convergence during training for ASD vs. NC classification. Since the loss is unstable during the training process, in order to better observe the change in the loss, we record the loss of MDGL on the training set and testing set every 5 epochs in Fig. 8. It can be observed from Fig. 8 that at the beginning of the training period, the loss values for both the training and testing sets are relatively large. As the training progresses, the loss values gradually decrease and then level off, indicating that our proposed method has converged.

J. Limitations and Future Work

There are several limitations in the current work. *First*, we capture multi-scale spatiotemporal dynamic fMRI features at the whole-graph-level via a single-head Transformer encoder and assign different weights to the graph-level features, without explicitly modeling the node-level temporal dynamics. It is interesting to incorporate both node-level and whole-graph-level temporal dynamics into a unified framework, which will be our future work. *Second*, this work focuses on only resting-state fMRI data, without considering other data modalities (*e.g.*, structural MRI and demographic information). Multimodal data could provide rich and complementary information about the brain from different perspectives and thus can be used to further improve learning performance, which will be our further work. *Third*, there exhibits significant data heterogeneity across sites in multisite studies such as ABIDE and REST-meta-MDD. Alleviating inter-site heterogeneity would help improve data quality and greatly increase sample size [60], which is an interesting direction for future research. *Additionally*, our model was trained on small-sized fMRI datasets (*i.e.*, 184 subjects from ABIDE and 531 subjects from REST-meta-MDD), while there exist large-scale unlabeled fMRIs in other multisite studies [57], [61]. We will use unsupervised and self-supervised learning strategies [62], [63] to pretrain the model for fMRI feature extraction and fine-tune it on downstream tasks in the future.

VI. Conclusion

In this paper, we propose a multi-scale dynamic graph learning (MDGL) framework for brain disorder detection with resting-state functional MRI. We consider the temporal information in fMRI while also focusing on the topological information at different spatial resolutions. Specifically, we first construct multi-scale dynamic FCNs by using multiple brain atlases to delineate the brain regions of each subject and consider dynamic properties of brain FCNs derived from fMRI. To simulate the spatiotemporal dynamic features of functional MRI at different spatial resolutions, a multi-scale dynamic graph representation learning framework is designed to capture spatiotemporal information of FCNs. Finally, multi-scale dynamic graph representations are fused and fed into fully connected layers for disease diagnosis. Experimental results on three datasets demonstrated the effectiveness of MDGL in brain disorder detection and gender classification.

References

[1] P. M. Matthews and P. Jezzard, "Functional magnetic resonance imaging," *J. Neurol., Neurosurg. Psychiatry*, vol. 75, no. 1, pp. 6–12, 2004.

[2] X. Liu et al., "Subcortical evidence for a contribution of arousal to fMRI studies of brain activity," *Nature Commun.*, vol. 9, no. 1, pp. 1–10, Jan. 2018.

[3] M. Jiang et al., "Anatomy-guided spatio-temporal graph convolutional networks (AG-STGCNs) for modeling functional connectivity between gyri and sulci across multiple task domains," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 5, 2022, doi: 10.1109/TNNLS.2022.3194733.

[4] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101765.

[5] M. Khosla, K. Jamison, G. H. Ngo, A. Kuceyeski, and M. R. Sabuncu, "Machine learning in resting-state fMRI analysis," *Magn. Reson. Imag.*, vol. 64, pp. 101–121, Dec. 2019.

[6] Y. Chen et al., "Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 14, 2022, doi: 10.1109/TNNLS.2022.3154755.

[7] Y. Fang, M. Wang, G. G. Potter, and M. Liu, "Unsupervised cross-domain functional MRI adaptation for automated major depressive disorder identification," *Med. Image Anal.*, vol. 84, Feb. 2023, Art. no. 102707.

[8] N. C. Dvornek, P. Ventola, K. A. Pelphrey, and J. S. Duncan, "Identifying autism from resting-state fMRI using long short-term memory networks," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 362–370.

[9] D. Zhang, J. Huang, B. Jie, J. Du, L. Tu, and M. Liu, "Ordinal pattern: A new descriptor for brain connectivity networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1711–1722, Jul. 2018.

[10] 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.

[11] L. Cerliani, M. Mennes, R. M. Thomas, A. Di Martino, M. Thioux, and C. Keysers, "Increased functional connectivity between subcortical and cortical resting-state networks in autism spectrum disorder," *JAMA Psychiatry*, vol. 72, no. 8, pp. 767–777, 2015.

[12] K. Dadi et al., "Benchmarking functional connectome-based predictive models for resting-state fMRI," *NeuroImage*, vol. 192, pp. 115–134, May 2019.

[13] 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. Imag.*, vol. 39, no. 3, pp. 644–655, Mar. 2020.

[14] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[15] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[16] S. Gadgil, Q. Zhao, A. Pfefferbaum, E. V. Sullivan, E. Adeli, and K. M. Pohl, "Spatio-temporal graph convolution for resting-state fMRI analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 528–538.

[17] H. Cui, W. Dai, Y. Zhu, X. Li, L. He, and C. Yang, "Interpretable graph neural networks for connectome-based brain disorder analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2022, pp. 375–385.

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

[19] H. Tang et al., "Hierarchical brain embedding using explainable graph learning," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–5.

[20] F. I. Karahanoğlu and D. Van De Ville, "Transient brain activity disentangles fMRI resting-state dynamics in terms of spatially and temporally overlapping networks," *Nature Commun.*, vol. 6, no. 1, pp. 1–10, Jul. 2015.

[21] E. Al-Sharoa, M. Al-Khassaweneh, and S. Aviyente, "Tensor based temporal and multilayer community detection for studying brain dynamics during resting state fMRI," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 3, pp. 695–709, Mar. 2019.

[22] B. Jie, M. Liu, and D. Shen, "Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease," *Med. Image Anal.*, vol. 47, pp. 81–94, Jul. 2018.

[23] N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, Jan. 2002.

[24] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "A whole brain fMRI atlas generated via spatially constrained spectral clustering," *Hum. Brain Mapping*, vol. 33, no. 8, pp. 1914–1928, Aug. 2012.

[25] 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.

[26] C.-G. Yan et al., "Reduced default mode network functional connectivity in patients with recurrent major depressive disorder," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 18, pp. 9078–9083, 2019.

[27] W. L. Hamilton, "Graph representation learning," *Synth. Lectures Artif. Intell. Mach. Learn.*, vol. 14, no. 3, pp. 1–159, 2020.

[28] F. Liu et al., "Multivariate classification of social anxiety disorder using whole brain functional connectivity," *Brain Struct. Function*, vol. 220, no. 1, pp. 101–115, Jan. 2015.

[29] C.-Y. Wee et al., "Identification of MCI individuals using structural and functional connectivity networks," *NeuroImage*, vol. 59, no. 3, pp. 2045–2056, Feb. 2012.

[30] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" 2018, *arXiv:1810.00826*.

[31] B.-H. Kim and J. C. Ye, "Understanding graph isomorphism network for rs-fMRI functional connectivity analysis," *Frontiers Neurosci.*, vol. 14, p. 630, Jun. 2020.

[32] B. Patel, A. Orlichenko, and Y.-P. Wang, "Multimodal graph isomorphism network to study fMRI connectivity in developmental stages of adolescence," *Proc. SPIE*, vol. 12036, pp. 520–526, Apr. 2022.

[33] S. Yang, D. Jin, J. Liu, and Y. He, "Identification of young high-functioning autism individuals based on functional connectome using graph isomorphism network: A pilot study," *Brain Sci.*, vol. 12, no. 7, p. 883, Jul. 2022.

[34] L. Meng and J. Zhang, "IsoNN: Isomorphic neural network for graph representation learning and classification," 2019, *arXiv:1907.09495*.

[35] T. Zhao and G. Zhang, "Detecting major depressive disorder by graph neural network exploiting resting-state functional MRI," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2023, pp. 255–266.

[36] V. D. Calhoun, R. Miller, G. Pearlson, and T. Adalı, "The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery," *Neuron*, vol. 84, no. 2, pp. 262–274, Oct. 2014.

[37] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, Mar. 2014.

[38] J. Chen et al., "Alterations of static functional connectivity and dynamic functional connectivity in motor execution regions after stroke," *Neurosci. Lett.*, vol. 686, pp. 112–121, Nov. 2018.

[39] X. Duan et al., "Effect of risperidone monotherapy on dynamic functional connectivity of insular subdivisions in treatment-naive, first-episode schizophrenia," *Schizophrenia Bull.*, vol. 46, no. 3, pp. 650–660, Apr. 2020.

[40] W. Liao, J. Li, X. Duan, Q. Cui, H. Chen, and H. Chen, "Static and dynamic connectomics differentiate between depressed patients with and without suicidal ideation," *Hum. Brain Mapping*, vol. 39, no. 10, pp. 4105–4118, Oct. 2018.

[41] B.-H. Kim, J. C. Ye, and J.-J. Kim, "Learning dynamic graph representation of brain connectome with spatio-temporal attention," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 4314–4327.

[42] C. Yan and Y. Zang, "DPARSF: A MATLAB toolbox for 'pipeline' data analysis of resting-state fMRI," *Frontiers Syst. Neurosci.*, vol. 4, p. 1377, Jan. 2010.

[43] L. Qiao, L. Zhang, S. Chen, and D. Shen, "Data-driven graph construction and graph learning: A review," *Neurocomputing*, vol. 312, pp. 336–351, Oct. 2018.

[44] S. I. Ktena et al., "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, Apr. 2018.

[45] R. Hindriks et al., "Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI?" *NeuroImage*, vol. 127, pp. 242–256, Feb. 2016.

[46] Y. Feng et al., "The relationships between dynamic resting-state networks and social behavior in autism spectrum disorder revealed by fuzzy entropy–based temporal variability analysis of large-scale network," *Cerebral Cortex*, vol. 33, no. 3, pp. 764–776, Jan. 2023.

[47] A.-K. Seghouane and A. Iqbal, "Basis expansion approaches for regularized sequential dictionary learning algorithms with enforced sparsity for fMRI data analysis," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1796–1807, Sep. 2017.

[48] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[49] A. G. Roy, N. Navab, and C. Wachinger, "Recalibrating fully convolutional networks with spatial and channel 'squeeze and excitation' blocks," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 540–549, Aug. 2018.

[50] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[51] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Jan. 1995.

[52] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Netw.*, vol. 29, no. 4, pp. 555–564, Oct. 2007.

[53] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

[54] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, Jan. 2018.

[55] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[56] K. Han et al., "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.

[57] S. M. Smith et al., "Resting-state fMRI in the human connectome project," *NeuroImage*, vol. 80, pp. 144–168, Oct. 2013.

[58] M. F. Glasser et al., "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, Oct. 2013.

[59] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, "Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers," *NeuroImage*, vol. 90, pp. 449–468, Apr. 2014.

[60] H. Guan and M. Liu, "Domain adaptation for medical image analysis: A survey," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022.

[61] C. R. Jack et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.

[62] Y. Sun et al., "Self-supervised learning with application for infant cerebellum segmentation and analysis," *Nature Commun.*, vol. 14, no. 1, p. 4717, Aug. 2023.

[63] G.-J. Qi and J. Luo, "Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2168–2187, Apr. 2022.