

# Multimodal Autism Spectrum Disorder Diagnosis Method Based on DeepGCN

Mingzhi Wang<sup>ID</sup>, Jifeng Guo<sup>ID</sup>, Yongjie Wang<sup>ID</sup>, Ming Yu<sup>ID</sup>, and Jingtian Guo<sup>ID</sup>

**Abstract**— Multimodal data play an important role in the diagnosis of brain diseases. This study constructs a whole-brain functional connectivity network based on functional MRI data, uses non-imaging data with demographic information to complement the classification task for diagnosing subjects, and proposes a multimodal and across-site WL-DeepGCN-based method for classification to diagnose autism spectrum disorder (ASD). This method is used to resolve the existing problem that deep learning ASD identification cannot efficiently utilize multimodal data. In the WL-DeepGCN, a weight-learning network is used to represent the similarity of non-imaging data in the latent space, introducing a new approach for constructing population graph edge weights, and we find that it is beneficial and robust to define pairwise associations in the latent space rather than the input space. We propose a graph convolutional neural network residual connectivity approach to reduce the information loss due to convolution operations by introducing residual units to avoid gradient disappearance and gradient explosion. Furthermore, an EdgeDrop strategy makes the node connections sparser by randomly dropping edges in the raw graph, and its introduction can alleviate the overfitting and oversmoothing problems in the DeepGCN training process. We compare the WL-DeepGCN model with competitive models based on the same topics and nested 10-fold cross-validation show that our method achieves 77.27% accuracy and 0.83 AUC for ASD identification, bringing substantial performance gains.

**Index Terms**— Medical imaging, autism spectrum disorders, brain networks, graph convolutional neural network, multimodal.

## I. INTRODUCTION

AUTISM spectrum disorder (ASD) is a neurodevelopmental disorder characterized by social impairment and

Manuscript received 4 July 2023; revised 22 August 2023 and 5 September 2023; accepted 8 September 2023. Date of publication 12 September 2023; date of current version 20 September 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61300098, in part by the Natural Science Foundation of Heilongjiang Province of China under Grant LH2019C003, and in part by the Fundamental Research Funds for the Central Universities of China under Grant 2572019BH03. (*Corresponding author: Ming Yu.*)

This work involved human subjects or animals in its research. The authors confirms that all human/animal subject research procedures and protocols are exempt from review board approval.

Mingzhi Wang, Yongjie Wang, Ming Yu, and Jingtian Guo are with the College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China (e-mail: wzm@nefu.edu.cn; dlzm@nefu.edu.cn).

Jifeng Guo is with the College of Computer Science and Engineering, Guilin University of Aerospace Technology, Guilin 540004, China (e-mail: gjf710621@163.com).

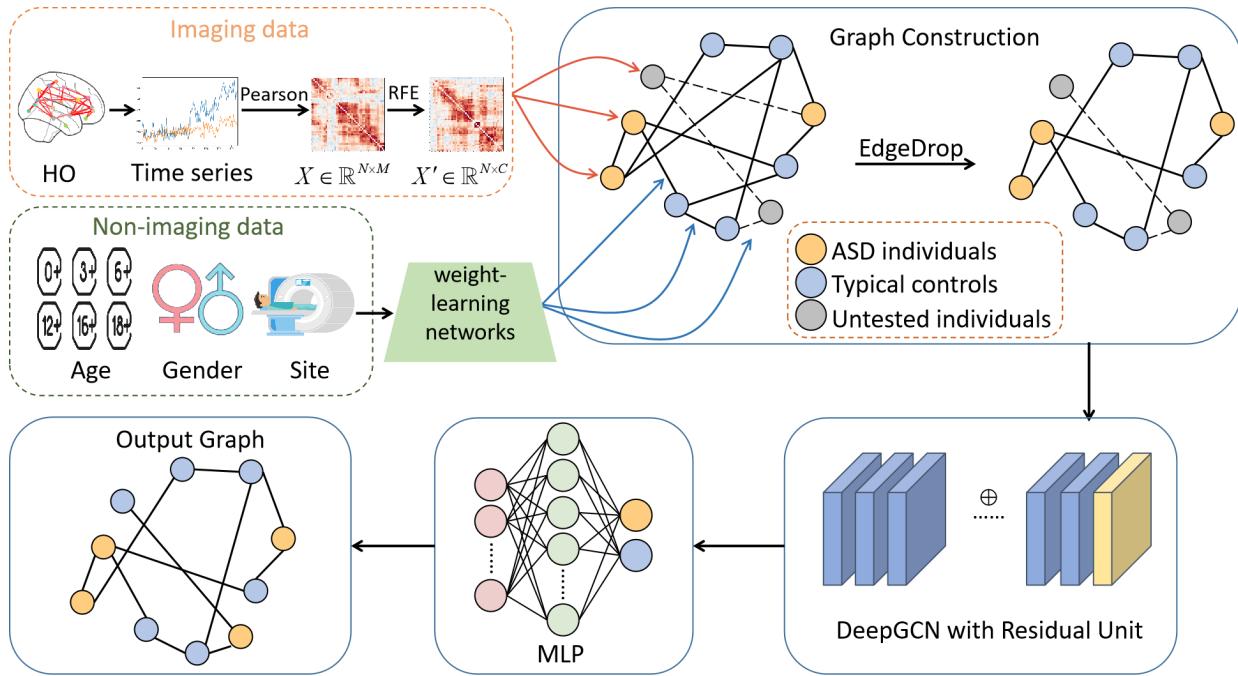
Digital Object Identifier 10.1109/TNSRE.2023.3314516

repetitive stereotypic behaviors [1]. Early diagnosis and intervention can improve the prognosis of ASD, and the earlier the age of treatment, the more significant the improvement [2]. Doctors determine whether a patient has an ASD based on their behavior and daily performance, which is highly subjective [3]. Therefore, there is an urgent need for an objective diagnostic method to assist physicians in diagnosing patients.

With the development of modern medical technology and artificial intelligence [4], machine learning-based methods for analyzing and studying brain magnetic resonance imaging (MRI) of patients with ASD have achieved excellent results [5]. A study by Dekhil et al. [6] converted the time series to power spectral density for 34 independent sets of components to analyze spatial graphs and used sparse autoencoders to reduce the input dimensionality for input into a support vector machine (SVM). Li et al. [7] developed a novel deep neural network framework that first trained a Stacked sparse autoencoder to learn functional connectivity patterns from an existing database of subjects. Heinsfeld et al. [8] used an autoencoder to learn whole-brain functional connectivity features. They constructed a deep neural network and used a transfer learning strategy to successfully classify 1035 subjects in the autism dataset. In recent years, non-imaging datasets have been collected in large numbers. Factors such as patient genetic sequences, gender, and intelligence quotient (IQ) play an important role in disease diagnosis [9]. Multimodal complementation of imaging and non-imaging data will help improve the performance of classification algorithms. However, non-imaging data often has high dimensionality, which limits the representational capacity of traditional machine learning methods [10].

Deep learning methods offer the possibility of fusing multimodal data to diagnose mental disorders [11]. Peng et al. [12] showed that deep learning can improve the classification of brain age and sex prediction. Shi et al. [13] used multimodal stacked denoised sparse autoencoders to fuse cross-sectional and longitudinal features estimated from brain MRIs. Khosla et al. [14] used the largest and most heterogeneous dataset ( $n = 774$ , site = 17) using a 3D convolutional neural network (CNN). However, non-graph-based deep learning methods are not conducive to multimodal modeling and typically only support the use of a single-modal, which limits the performance of the model.

In recent years, graph neural networks have been proposed to solve this problem [15], [16], [17]. A graph convolutional



**Fig. 1.** Overview of the proposed WL-DeepGCN.

network (GCN) [18] extends the convolution operation on Euclidean data to non-Euclidean graph data. Parisot et al. [19] conducted research on ASD diagnosis using a GCN, using non-imaging data such as age and gender to construct graph edges, obtaining 69.5% accuracy, which is the first successful attempt to apply a GCN to an ASD classification task. Jiang et al. [20] proposed a hierarchical GCN framework (Hi-GCN) to learn graph feature embeddings while considering network topological information and subject associations and obtained 73.1% accuracy. Huang and Chung [21] proposed a graph convolutional neural network called EV-GCN, which uses pairwise association encoders to construct graphs. During training, jump connections were used to avoid the overfitting problem. The above methods study shallow GCNs. In the field of image classification, the CNN depth is increasing [22] to extract better and express the raw features and improve model performance. However, there is a shortage of research on deep-layer GCNs applied to ASD diagnosis.

The DeepGCN can more effectively learn the association between neighboring nodes. Cao et al. [23] proposed a 16-layer GCN for ASD classification, randomly dropped edges at each layer, and obtained a model accuracy of 73.7%. Due to the limited data sample size [24], a DeepGCN is prone to gradient explosion and gradient disappearance problems, while the associations between non-imaging data are not well utilized. For the above issues, in this study, we propose combining the DeepGCN with a weight-learning network (WL-DeepGCN) for ASD diagnosis. The main contributions of our work are as follows.

- (1) We propose a new DeepGCN framework for ASD diagnosis using brain functional networks for classification. A weight-learning network automatically exploits the pairwise associations of non-imaging data in the latent

space for constructing graph edge weights, building an adaptive population graph model. Our method provides the potential for extending to the diagnosis of other psychiatric disorders.

- (2) We propose the residual connections of a graph convolutional neural network to avoid the problems of the DeepGCN gradient explosion and gradient disappearance. The GCN residual unit can reduce the feature information overfitting problem caused by the convolution operation.
- (3) We introduce an EdgeDrop strategy. Random edge dropping in the graph during model training can make the node connections sparser. This reduces the over-smoothing aggregation speed and reduces subsequent information loss.
- (4) Our proposed WL-DeepGCN takes sufficient advantage of the DeepGCN. Experiments are conducted on a public dataset to evaluate the proposed method compared with existing excellent methods. The nested 10-fold cross-validation results show that the proposed method facilitates feature learning and improves classification performance. The source code for the proposed architecture is publicly available at <https://github.com/Charles-wmz/WL-DeepGCN>.

## II. METHODOLOGY

An overview of the proposed method is shown in Fig. 1. We use a population graph [25] structure to represent the subject data and model the diagnostic task as a node classification task. First, a functional connectivity network [26] is constructed in a whole-brain functional connectivity manner [27], thus defining the population graph node features. Second, the edges in the population graph are described as

similarities in site, gender, and age for different subjects. As shown in Fig. 1, the non-imaging data are used to optimize the connectivity of the population graph in the potential space using a weight-learning network (WL). The completed graphs are processed using the EdgeDrop strategy to drop edges randomly. Finally, we use the DeepGCN learn features representations and perform the final classification using a multilayer perceptron (MLP) [28]. The residual connectivity of GCN is introduced to acquire a more profound feature representation of each node, aggregating the untested nodes and finally generating a fully labeled output graph. The WL-DeepGCN model realizes end-to-end training [29] by semi-supervised learning [30].

### A. Data Acquisition and Preprocessing

Our experiments were implemented on the **Autism Brain Imaging Data Exchange (ABIDE)** [31], an open-access multimodal data repository. ABIDE-I collected data on 1112 subjects from 17 sites worldwide. ABIDE-I contains MRI for each subject, and sufficient phenotypic information. In this work, we use resting-state functional MRI (rs-fMRI) [32], [33] and non-imaging data, using a publicly available preprocessed version of the dataset provided by the Preprocessed Connectomes Project initiative, with preprocessing performed by the Configurable Pipeline for the Analysis of Connectomes (C-PAC). To ensure **data quality and methodological comparability**, we excluded some data that lacked time series, incomplete brain coverage, severe head movement, and other scanning artifacts. Finally, we used the 871 subjects data, which contained 403 individuals with ASD and 468 typical controls (TC). The study used the Harvard - Oxford (HO) Cortical Structural Atlas [34] to define network nodes. The HO atlas combines processed cortical and subcortical regions of interest (ROIs) using nearest neighbor interpolation to segment the functional resolution, yielding 111 ROIs. Each participant's average time series of ROIs was extracted and used for subsequent functional connectivity network construction.

### B. Graph Nodes

We represent participants as  $N$  subjects, and each node represents **a feature extracted by the subject from the imaging data**. A concept widely used to generate features from fMRI data is the strength of functional connectivity (FC) between brain regions. Since a functional connectivity matrix is a real symmetric matrix, to avoid feature redundancy, we remove the values of the upper triangle of the matrix, and the main diagonal of the matrix [35]. Finally, only the strictly lower triangular part is retained to represent the degree of functional connectivity of any two ROIs, and the final retained matrix vectorization is expanded into a one-dimensional vector to obtain an  $M$  dimensional feature vector required for classification. The above processing is implemented for each subject to get the input matrix  $X \in \mathbb{R}^{N \times M}$ , where each vector in the matrix represents the features of a node. Since the HO atlas divides the fMRI data of each subject into 111 ROIs, each node has 6105 dimensional feature vectors.

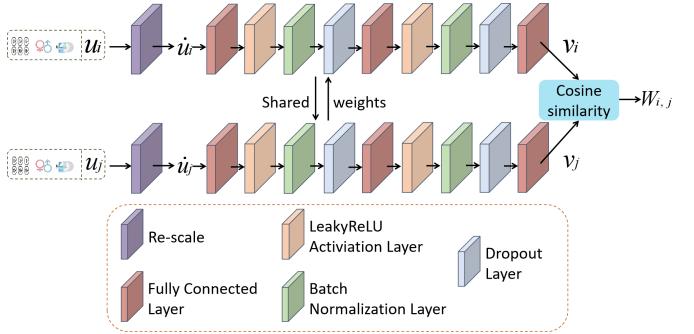


Fig. 2. Framework for the weight-learning network.

We obtained feature vectors with high dimensionality, especially concerning a graph size of 871 nodes, which will have a negative impact on the performance of the algorithm. We use recursive feature elimination (RFE) [36] to reduce the dimensionality of the feature vectors. For the whole population graph, we obtain the input matrix  $X' \in \mathbb{R}^{N \times C}$  after feature elimination.

### C. Graph Edges

In constructing the graph, we use edges to represent the description of the association between nodes, and the weights of the edges encode the similarity between subjects and influence the aggregation function in the graph convolution. A weight-learning network establishes intersubject weights based on the information provided by non-imaging phenotypic data.

Since multimodal data have different statistical properties and are prone to vanishing gradient problems during training, we normalize and standardize the non-imaging input data  $u$  to rescaled vector  $\hat{u}$ . Specifically, we constructed dictionaries from the phenotypic data  $u$  consisting of site, gender, and age for each sample. Gender is denoted by (0, 1), the 17 sites are coded as (0, 1, 2, ..., 16), and the age takes values in the range (6, 64). We normalized the sites and ages to between (0, 1). The feature matrix is then normalized to obtain the rescaled  $\hat{u}$ . The rescaled vector passes through three fully connected layers whose output dimensions are  $h_1$ ,  $h_2$ , and  $h_3$ . Except for the last layer, a LeakyReLU activation layer, a batch normalization layer, and a dropout layer are added after each fully connected layer. The activation function is used to introduce nonlinear factors to improve model expressiveness. The batch normalization layer is used to solve the problem of numerical instability in the neural network so individual features of the same batch are not similar, and the network can be trained more efficiently. To avoid overfitting, the dropout layer in the WL network randomly drops features with a WL-dropout probability. The framework of the weight-learning network is shown in Fig. 2. To calculate the edge weights  $w$  between two nodes, we perform the above operation with two non-imaging inputs,  $u_i$  and  $u_j$ , while sharing the weights, and we calculate the cosine similarity [37] of the two potential feature vectors by Formula (1).

$$w(i, j) = \frac{v_i^T v_j}{2 \|v_i\| \|v_j\|} + 0.5 \quad (1)$$

where  $v_i$  and  $v_j$  denote the feature representations of  $ui$  and  $uj$  respectively after weight sharing network as inputs for cosine similarity. As shown in Fig. 2.

According to the validation of the study by Huang and Chung [21], it is beneficial and robust to define pairwise associations on the latent space rather than the input space. Our experiments also show that WL performs better in the ASD classification task.

#### D. EdgeDrop Strategy

There are two main reasons why graph neural networks fail to deepen: overfitting and oversmoothing. To solve these two problems, this paper proposes an EdgeDrop strategy inspired by the population graph improvement in the GCN training process introduced by Yu et al. [38]. The main idea of EdgeDrop is to randomly remove some edges from the constructed population graph according to a certain proportion before the training starts. Specifically,  $Eq$  nonzero elements of the adjacency matrix  $A$  are randomly selected and set to zero in the training.  $E$  is the total number of edges in the raw population graph, and  $q$  is the drop rate. After dropping, a new adjacency matrix is obtained as the final input to the DeepGCN.

EdgeDrop can be seen as a data enhancement technique. In the training phase, different random edge drops in the raw graph also enhance the randomness and diversity of the input data, which can alleviate the overfitting problem. EdgeDrop can also be seen as a message-passing reducer, where the message passing between neighboring nodes in GCNs is achieved by connecting edges. The random deletion of some edges makes the nodes more sparsely connected, which to a certain extent, alleviates the oversmoothing problem caused by the deepening of GCN layers.

#### E. Residual Connections for the GCN

We use Chebyshev polynomials [19], [39] proposed by Defferrard et al. [40] to approximate convolutional kernels in graph convolutional neural networks. CNNs allow the design and training of deep models, from which we derive the following relationship between layers of the DeepGCN.

$$y^{(l+1)} = G \left( D^{-\frac{1}{2}} A D^{-\frac{1}{2}} y^{(l)} W^{(l)} \right) \quad (2)$$

where  $y^{(l)}$  is the feature representation of layer  $l$ .  $y^{(0)} = x$ .  $W^{(l)}$  is the learnable parameter of layer  $l$ ,  $G$  is the activation function, and  $D$  is degree matrix of adjacency matrix  $A$ .

In a DeepGCN, the problem of gradient explosion and gradient disappearance [41] will cause the feature vectors to converge to the same value. To solve this problem, we propose residual connectivity for graph convolutional neural networks to enhance the feature representation. We are inspired by ResNet [42], which converts a solved network mapping function  $H(x)$  to a residual network mapping function  $H(x) = F(x) + x$ . Updating the weight values of  $F(x)$  can get infinitely close to the optimal solution.

We use residual units to implement DeepGCN with residual connections, and the model is shown in Fig. 3. The constructed graph  $X' \in \mathbb{R}^{N \times C}$  is input to  $G_h$  convolutional layers.

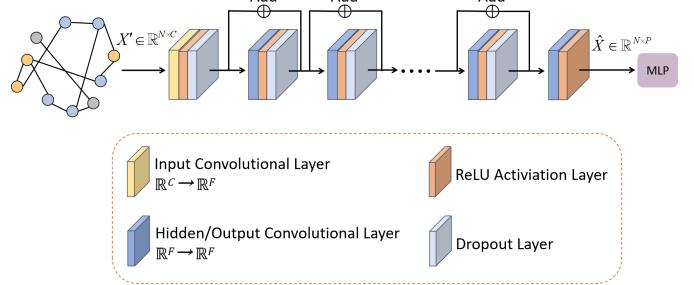


Fig. 3. Residual connections for the deep graph convolutional neural networks.

Except for the last convolutional layer (output layer), each convolutional layer contains a ReLU activation layer followed by a dropout layer. Finally, the new feature representation  $\hat{X} \in \mathbb{R}^{N \times P}$  is input to an MLP, and  $P$  is the number of units in the hidden and output layers. The residual unit merges the output of this layer and the output of the previous layer as the input to the next layer, which results in the following relationship between the DeepGCN layers:

$$y^{(l+1)} = S \left( y^{(l)}, W^{(l)} \right) + y^{(l)} \quad (3)$$

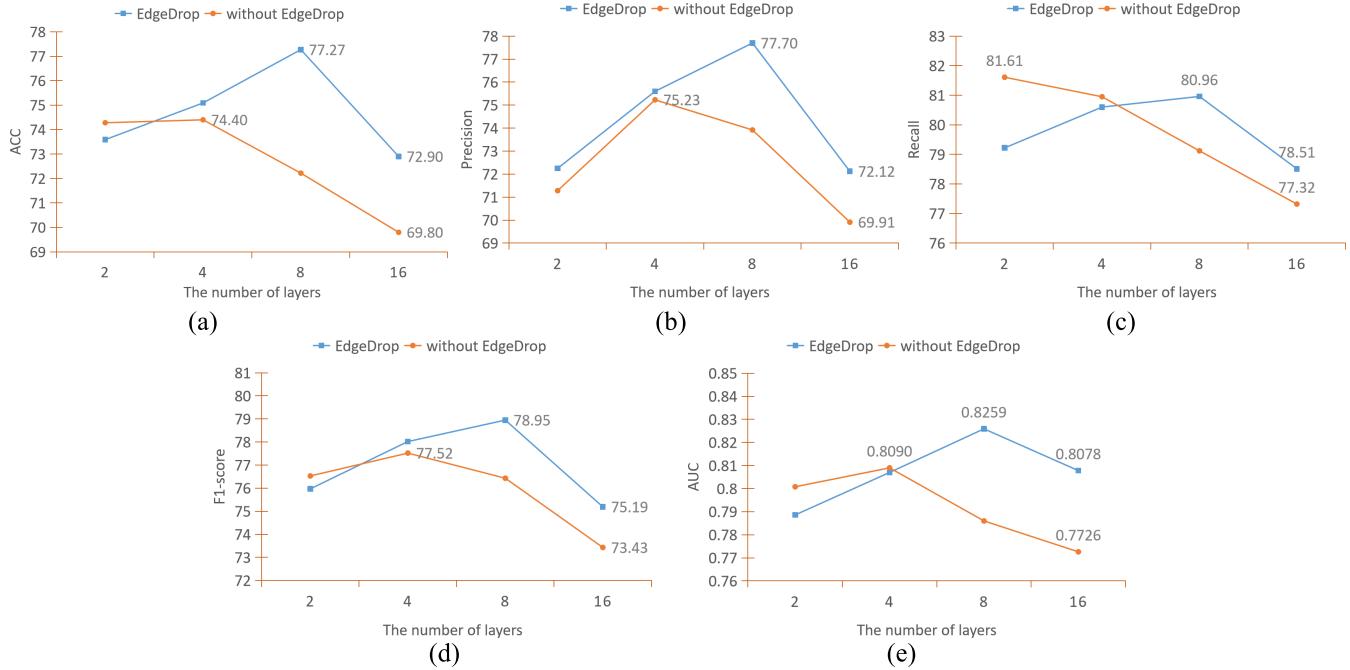
where  $S$  represents the mapping function, which can be a graph convolution operation in GCN, which transforms the feature  $y^{(l)}$  and weight  $W^{(l)}$  of the layer.

It is worth noting that the output layer does not use residual units, which means that the MLP input is only the output of the last DeepGCN convolutional layer and does not contain the output before the output layer. The proposed residual connectivity of the graph convolutional neural network can reduce the problem of feature information overfitting caused by the spectral graph convolution operation and improve the performance of the model. The dropout layer is applied to the new feature representation in the hidden layer, which can further avoid the gradient problem in the DeepGCN.

The pseudocode is shown in Algorithm 1 to facilitate understanding of the WL-DeepGCN.

### III. EXPERIMENTS AND RESULTS

We evaluate the performance of the model by executing a nested 10-fold cross-validation [43] on the ABIDE-I dataset. The outer loop of the nested cross-validation is set to 10, as well as the inner loop. In the outer loop, the dataset is divided into training set and test set. In the inner loop, the training set is further divided into new training set and validation set. The test set of each fold is only used to evaluate the model's performance and is not involved in adjusting model parameters or feature selection. This effectively avoids the problem of feature peeking during the training process, which can lead to artificially high model performance but weak generalization ability in conventional 10-fold cross-validation. We performed preprocessing and executed RFE to extract 2000-dimensional feature vectors. The non-imaging data age, gender, and site are selected to learn the edge weights of the graph. We set  $h_1 = 256$  and  $h_2 = 128$  in the WL network, and the dropout rate of the network to 0.3. In the GCN, we use



**Fig. 4.** Classification results of different network depths and EdgeDrop strategy under 10-fold cross-validation.

#### Algorithm 1 WL-DeepGCN Pseudocode

```

Input: Raw features  $X$ , phenotypic features  $u$ , label  $Y$ , convolution layers  $G_h$ . GCN residual connections mapping functions  $S$ 
Output: Predicted probability of each subject  $Y_{pre}$ .
1:  $X$  are extracted to obtain  $X'$ .  $u$  are processed through the weight-learning network.
2:  $X'$  and  $u$  are used to construct the initial graph  $G$ .
3:  $G$  is randomly dropped using the EdgeDrop strategy to obtain  $G'$ .
4: for  $gh \leftarrow 1, \dots, G_h$  do
5:   if  $gh = 1$  do
6:      $G'$  is used as input to the GCN to learn information about neighboring nodes and clustering is conducted to obtain a new graph  $G'_{gh}$ .
7:   else do
8:      $G'_{gh}$  is used as input to the GCN to learn information about neighboring nodes, perform clustering and use GCN residual connections to obtain a new graph  $G'_{gh+1}$ .
9:    $G'_{gh+1} = S(G'_{gh}) + G'_{gh}$ ,  $G'_{gh} = G'_{gh+1}$ 
10:  Gradient and cross-entropy losses are updated using the Adam optimization algorithm  $l$ .
11: end if
12:end for
```

the 3 order Chebyshev polynomial approximation convolution kernel with  $p = 16$  units in the hidden layers and output layers. We train a 200-epoch graph-based model using the Adam optimizer. For the WL-DeepGCN, the learning rate is set to 0.001, the dropout rate to 0.2, and  $q = 0.3$  in the EdgeDrop strategy. We employ an early stopping mechanism

with an early stopping patience of 20 epochs to avoid overfitting. All models are developed using the open-source machine learning library PyTorch, and experiments are performed on a GeForce GTX 3060 GPU.

#### A. Performance Evaluation

For performance evaluation, we use the accuracy (ACC), precision, recall, F1 score and area under the curve (AUC) as evaluation metrics. A true positive (TP) is the correct classification for positive classes. A true negative (TN) is the correct classification for negative classes. A false positive (FP) is the incorrect prediction of a positive. A false negative (FN) is the incorrect prediction of a negative. These evaluation indicators are defined as follows:

$$\text{ACC} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \quad (4)$$

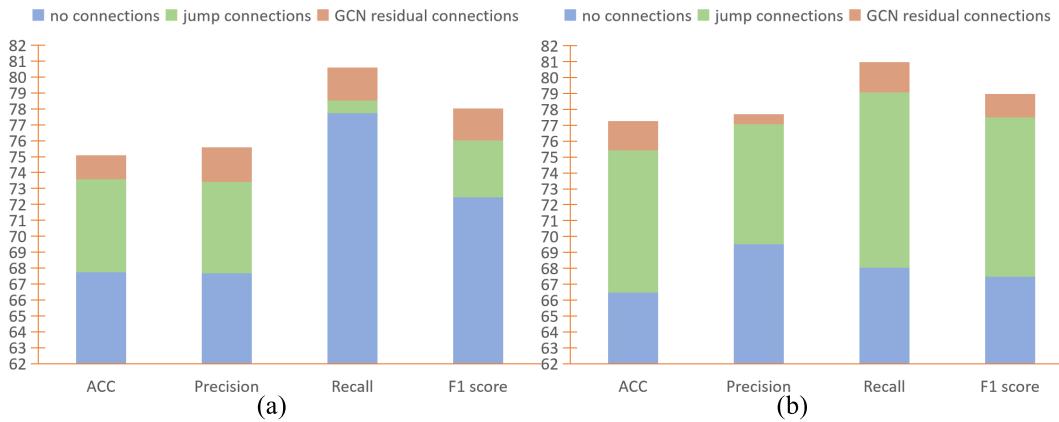
$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (6)$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \quad (7)$$

#### B. Effect of Network Structure on Results

To evaluate the effect of GCN depth on the classification, we conducted experiments setting the number of convolutional layers  $G_h = 2, 4, 8, 16$  in the training settings. Additionally, to verify the effectiveness of EdgeDrop for ablation experiments, we conducted experiments without EdgeDrop strategy in the model under the same experimental conditions to observe the effect of EdgeDrop on the DeepGCN. Fig. 4 shows how the number of GCN layers and the EdgeDrop strategy affect the results under nested 10-fold cross-validation.



**Fig. 5.** Classification results of our proposed model with nested 10-fold cross-validation in three connection modes: no connections, jump connections and GCN residual connections. (a) performance of the model under 4-layer GCN; (b) performance of the model under 8-layer GCN.

As depicted in Fig. 4, the performance of the model without EdgeDrop starts to decline significantly after a slight improvement in performance at the 4-layer GCN as the network architecture deepens. The accuracy of the model with EdgeDrop continues to increase but peaks at a GCN depth of 8 layers. Combining the performance metrics of different network structures, we conclude that a DeepGCN with eight convolutional layers and the EdgeDrop strategy has the best classification performance.

In the experiments, the classification accuracy increases with the deepening of the model after adding the EdgeDrop strategy. This indicates that the EdgeDrop strategy increases the diversity and randomness of the data by randomly dropping some edges, which can alleviate the overfitting and oversmoothing problems that occur in the DeepGCN training process. The EdgeDrop strategy has good generalization and can be applied to any network model with a graph structure without additional parameters. As seen in Fig. 4, after the number of network layers reaches 8, the performance of the model begins to decrease, and the underlying reason is the limitation of the network depth on the model expression. However, after 2 layers, the performance of the model with the EdgeDrop strategy is always higher than that of the model without the EdgeDrop strategy. This also verifies that the EdgeDrop strategy can improve the DeepGCN performance.

### C. Effect of Connections Pattern on Results

In addition to our proposed residual connectivity for GCN, researchers have investigated DeepGCN gradient explosion and gradient disappearance. Huang and Chung [21] proposed jump connectivity for a GCN to fuse the hidden features. We perform experiments with no connections, jump connections, and GCN residual connections under 4-layer and 8-layer DeepGCN using the EdgeDrop strategy to verify the effectiveness of GCN residual connections. The experimental results are shown in Fig. 5 and Table I.

As shown in Fig. 5, the orange color represents the GCN residual connections, which are at the top of the metrics and outperforms the other two connection approaches. The jump connections have some effect on improving the model performance compared to the no connections approach because the

**TABLE I**  
AUC OF OUR PROPOSED MODEL WITH THREE TYPES OF CONNECTIONS: NO CONNECTIONS, JUMP CONNECTIONS AND GCN RESIDUAL CONNECTIONS

Connections Pattern	AUC under 4-layer	AUC under 8-layer
No connections	78.91	72.50
Jump connections	79.27	81.16
GCN residual connections	80.70	82.59

jump connections fuse some of the features at different depths in the final output, reducing the problem of information loss due to convolution operations.

From Fig. 5 and Table I, we can observe that under the 8-layer DeepGCN, the jump connections pattern achieves 8.97% higher accuracy, 10.02% higher F1 score, and 8.66% higher AUC compared to the no connections pattern. Furthermore, the GCN residual connections pattern achieves 1.84% higher accuracy, 1.46% higher F1 score, and 1.43% higher AUC than the jump connections pattern. The DeepGCN with 4-layer has the same trend, which indicates that the results of the experiments are very robust. The experiments show that adding residual units to the DeepGCN allows each layer to fuse the hidden features of the upper layer before the input. Compared to the jump connections' last fused features, our proposed GCN residual connections approach is more accurate than the jump connections. GCN residual connections can better maintain the integrity of information and reduce information loss, thus avoiding the problems of gradient explosion and gradient disappearance. Our experiments validate the great potential of residual connections in the DeepGCN.

### D. Effect of Weight-Learning Network on Results

To prove that the weight-learning network we proposed is effective, we will experiment with the model without a weight-learning network and the model with a weight-learning network under the condition that the number of network layers is 8, using GCN residual connections and EdgeDrop strategy. Models that do not use a weight-learning network build an unweighted graph, with edges randomly connected between nodes according to a uniform distribution. In this

TABLE II

MODEL PERFORMANCE OF UNWEIGHTED GRAPH AND WEIGHT-LEARNING NETWORK

The Method of Generating Edge Weights	ACC	Precision	Recall	AUC
Unweighted Graph	73.51	74.12	76.45	78.05
Weight-Learning Network	77.27	77.70	80.96	82.59

case, the calculation of GCN mainly focuses on the connection relationship between nodes without considering the specific weight value. Table II shows the experimental results of the two methods.

From the results of this set of experiments, the weight-learning network we proposed is beneficial to constructing the edges of the population graph and significantly impacts model performance. The weight-learning network initializes edge weights by learning the pairwise correlations of non-imaging data in the latent space. Compared with the unweighted graph without initialized edge weights, the ACC and AUC of the model can be improved by 3.26% and 5%, respectively, which proves the effectiveness of our proposed weight-learning network.

#### E. Comparison of Overall Performance

To validate the WL-DeepGCN model advances while ensuring comparable results, we further compared the WL-DeepGCN with several baselines and state-of-the-art models on the same topic. We have selected four classic methods, HOFC [44], GCN [45], DNN [8], and ASD-DiagNet [46], as the baselines. Hi-GCN [20] adopts a hierarchical perspective for graph embedding learning, considering the structural information within individual networks and the topic relevance in the global population network. This enables the capturing of fundamental embedding features. MVS-GCN [24] combines graph structure and multi-view graph embedding to identify potential functional subnetworks. EV-GCN [21] incorporates an edge-variational autoencoder to learn potential representations among nodes in the graph, followed by further processing and prediction using a 4-layer GCN. Table III shows the results of our comparison with existing methods. It is worth noting that, to ensure consistent experimental conditions, we employed nested 10-fold cross-validation rather than traditional 10-fold cross-validation. This cross-validation method, especially during the replication of EV-GCN, involves re-partitioning the dataset, where the newly partitioned dataset keeps the test set separate from the model training and feature selection processes. We can obtain more reliable and unbiased performance estimates for the models using nested cross-validation.

Table III presents the classification results on the ABIDE dataset. We can see that the WL-DeepGCN method proposed in this paper achieves the highest performance with 77.27% ACC and 82.59% AUC. The deeper network structure and optimization strategies provide improved feature representations for the model classification task.

Firstly, it is observed that methods based on GCN generally outperform non-graph-based models. In particular, the DNN,

TABLE III

PERFORMANCE COMPARISON BETWEEN OUR PROPOSED MODEL AND SEVERAL STATE-OF-THE-ART MODELS ON THE SAME TOPIC USING 10-FOLD CROSS-VALIDATION ON THE ABIDE-I DATASET

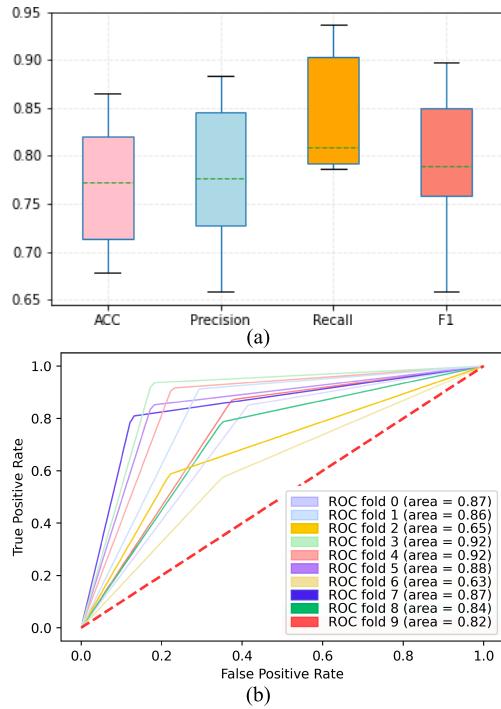
Method	ACC	Precision	Recall	AUC	Parameter (M)	p-value
HOFC	71.66	69.51	61.77	76.55	1.25	
GCN	70.45	68.09	74.47	73.22	0.10	
DNN	68.27	71.64	78.02	64.47	3.34	
ASD-DiagNet	70.04	69.05	70.73	71.39	4.67	< 0.05
Hi-GCN	73.10	65.37	71.40	81.60	0.69	
MVS-GCN	69.89	63.82	70.18	68.73	93.79	
EV-GCN	75.37	75.22	84.45	79.12	0.13	
Proposed	<b>77.27</b>	<b>77.70</b>	80.96	82.59	0.16	

as a single modality method using fMRI data, achieves only 68.27% accuracy. In addition, we can see that the accuracy and AUC of ASD-DiagNet are 70.04% and 71.39%, respectively, which are higher than the model proposed by Heinsfeld et al. [8]. This shows that the data augmentation method has a certain impact on model performance improvement. However, it is not easy to express a profound representation of features through an autoencoder, and using only a single-layer perceptron as a classifier does not achieve excellent classification results. This combined approach also uses only a single modality for research. This indicates that fusing multi-modal inputs provides more features information compared to using single-modal inputs. This aligns with the clinical practice, where clinical doctors consider not only radiological data but also additional patient information such as age, gender, etc., to improve the reliability of diagnoses. By integrating multiple modalities, the diagnostic process can be enhanced and made more accurate and robust. At the same time, it can be seen that the GCN-based method we proposed is superior to the classic methods such as DNN proposed by Heinsfeld et al. and ASD-DiagNet proposed by Eslami et al.

Secondly, for Hi-GCN, our proposed approach improved the accuracy by 4.17%. This improvement can be attributed to weight-learning network, which extracts more useful information by learning from the phenotypic data features compared to randomly initialized edge weights.

Furthermore, the accuracy of MVS-GCN is 69.89%, precision is 63.82%, recall is 70.18 and AUC is 68.73%. Compared with it, our method also significantly improved the performance of the model. The complementary information among multiple views provides richer feature information. However, MVS-GCN requires learning a more significant number of parameters, making the model difficult to converge and may result in weak generalization ability.

Finally, EV-GCN achieves better performance compared to the above two graph-based models. EV-GCN has an accuracy of 75.37% and an AUC of 79.12%. However, it still falls short compared to our proposed method. The most likely reason for this discrepancy lies in the depth limitation of GCN and the fixed hidden layer of the autoencoder. The limited depth of GCN restricts the model's ability to capture complex feature representations, and fixed hidden layer of the autoencoder hampers the learning of latent features across different topics.



**Fig. 6.** The average performance of our proposed method in nested 10-fold cross-validation. (a) Boxplots of the ACC, precision, recall and F1 score; (b) ROC curves.

In contrast, our proposed model addresses these limitations by introducing EdgeDrop and GCN residual connections. These techniques effectively alleviate the depth limitation and enable more flexible learning of latent features across different topics, resulting in superior performance compared to EV-GCN.

We can see from the number of parameters that MVS-GCN needs to learn more parameters and has the longest running time in the experiment. The GCN model proposed by Parisot et al. [45] has the least parameters. In addition, the model with the least number of parameters is EV-GCN, which has a running time of 38min 57s under the nested 10-fold cross-validation experiment. Further, we find that the number of parameters of the GCN-based models is generally lower than the other compared methods. A large number of parameters that need to be trained are simultaneously present in ASD-DiagNet. Its running time is longer and reaches 16h 54min 28s. Our proposed model only needs to train 0.16M parameters, and the running time is 32min 43s. Outperforms EV-GCN, which has more parameters, demonstrating that our method performs better in model structure and algorithm optimization. It demonstrates the efficiency of our proposed method. In addition, we use Student's t-tests to evaluate the difference between our method and other methods. We set the significance level  $\alpha = 0.05$  to determine the statistical significance. From Table III, we can observe that all  $p$ -values are less than 0.05. The statistical analysis further shows that our proposed method has significant advantages.

We present the detailed WL-DeepGCN performance metrics in Fig. 6. Fig. 6(a) shows a boxplot of the ACC, precision, recall, and F1 score of our proposed method in the nested 10-fold cross-validation case, and the green line represents

the mean values. Our method exhibits an average precision of 77.70%, indicating that the model possesses high confidence. In addition, the average recall of 80.96%, reflects that the model has a high breadth. The F1 score also yields an excellent result of 78.95%.

Fig. 6(b) shows the ROC curve changes for the test set in the nested 10-fold cross-validation case. From Fig. 6(b), it is clear that the area under the curve of the model is large, and the average AUC is 82.59% from Table III, which indicates that the model has very high robustness. In addition, the true positive rate of our WL-DeepGCN increases at the beginning of the ROC curve, which means that the model has a higher diagnostic rate and a lower misdiagnosis rate for ASD. This makes it possible to provide accurate and reliable results for clinical applications.

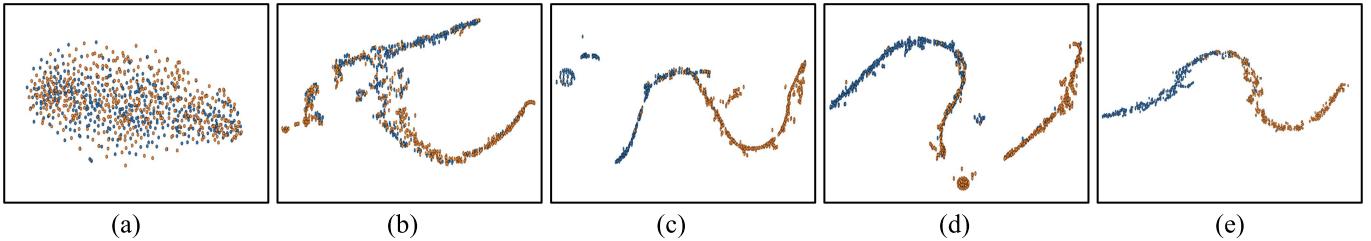
In our proposed WL-DeepGCN model, the weight-learning network for edge construction is one of the critical reasons why our method outperforms other methods. It is good at learning associations from non-imaging data in the latent space. The non-imaging data are an essential complement to subject information in the population graph in the form of edges and edge weights to represent the association between phenotypic information such as age, gender, and site among different subjects. In addition, GCN residual connections alleviate the gradient problem by integrating the output results of different layers. This is one of the ways we can solve the problem that GCN cannot be too deep. Through more profound feature expression, our method is superior to other models. Finally, EdgeDrop makes node connections more sparse by randomly dropping some edges. In summary, our method outperforms other methods. This method can be extended to other classification tasks for mental disorders, transforming the disease prediction problem into a GCN node classification task, thus providing an effective way to classify large-scale heterogeneous datasets.

#### F. Leave-One-Site-Out Classification

To evaluate model performance across sites, we performed Leave-one-site-one-out cross-validation. According to this evaluation method, data from one site is excluded from the training process and then used as a test set to evaluate the performance of the model. Table IV reports the accuracy versus other methods.

In the studies of Heinsfeld et al. [8], Eslami et al. [46], and Almuqhim and Saeed [47] they used the CC200 atlas. To ensure the comparability of the results, we also used CC200 in the experiment. From these results, our method shows better accuracy on 13 out of 17 sites. The model achieved an average accuracy of 74.2%, and outperforming other state-of-the-art methods. The accuracies of MAX\_MUN, SDSU, STANFORD, TRINITY, and USM are all lower than 70.0%. This suggests that the data from these sites have variability absent from other sites.

Overall, the WL-DeepGCN model exhibits better performance on more sites than the other three state-of-the-art methods, demonstrating our proposed model's robustness and generalizability.



**Fig. 7.** The blue nodes represent autism spectrum disorder (ASD), and the orange nodes represent typical controls (TC). **(a)** visualization of raw features in 2-dimensional space; **(b)** embedding representations learned by the MVS-GCN; **(c)** embedding representations learned by the Hi-GCN; **(d)** embedding representations learned by the EV-GCN; **(e)** embedding representations learned by the WL-DeepGCN.

**TABLE IV**  
LEAVE-ONE-SITE-OUT CROSS-VALIDATION RESULTS  
FOR DIFFERENT METHOD

Site	Size	Heinsfeld <i>et al.</i>	Eslami <i>et al.</i>	Almuqhim <i>et al.</i>	Our Proposed
CALTECH	15	52.3	52.8	56.7	<b>71.7</b>
CMU	11	45.3	68.5	70.6	<b>80.5</b>
KKI	33	58.2	69.5	72.6	<b>76.2</b>
LEUVEN	56	51.8	61.3	64.6	<b>74.3</b>
MAX_MUN	46	54.3	48.6	47.5	<b>69.7</b>
NYU	172	64.5	68.0	72.0	<b>79.1</b>
OHSU	25	74.0	<b>82.0</b>	72.0	78.6
OLIN	28	44.0	65.1	66.6	<b>80.0</b>
PITT	50	59.8	67.8	<b>73.1</b>	72.4
SBL	26	46.6	51.6	56.6	<b>79.7</b>
SDSU	27	63.6	63.0	64.2	<b>67.7</b>
STANFORD	25	48.5	64.2	53.2	<b>68.3</b>
TRINITY	44	61.0	54.1	57.5	<b>66.9</b>
UCLA	85	57.7	<b>73.2</b>	68.3	72.8
UM	120	57.6	63.8	67.8	<b>81.2</b>
USM	67	62.0	68.2	<b>70.0</b>	67.5
YALE	41	53.0	63.6	66.0	<b>75.6</b>
Average	51.2	56.1	63.8	64.7	<b>74.2</b>

### G. Visualization

To explicitly demonstrate the capability of our proposed WL-DeepGCN for feature learning and fusion on the ABIDE-I dataset, we use the t-SNE technique [48] for two-dimensional feature visualization. As shown in Fig. 7(a), the raw features of both the ASD and typical control (TC) types are randomly distributed in the binary mapping, and the nodes of both types are mixed unordered and without clear boundaries. This reflects the great difficulty in classifying directly using raw features. Fig. 7(b) shows that MVS-GCN exhibits some classification capability, but the data distribution remains complex and disordered. The aggregation results of Hi-GCN show significant improvement, with a more organized overall distribution. Although it already demonstrates good classification capability, Fig. 7(c) reveals that some of the subjects' data points are scattered far away. From Fig. 7(d), we can see the visualization results of EV-GCN, which has better inter-class discrimination ability than MVS-GCN and better intra-class aggregation ability than Hi-GCN. But there is still a particular gap compared with our proposed method. The visualization of node embedding after WL-DeepGCN convolution is shown in Fig. 7(e), where the TC nodes have overall similarity with the submap corresponding to the ASD class. The algorithm clearly distinguishes between

ASD patients and typical controls and reflects the discrepancy between the classes. This indicates that the fused multimodal features exhibit better intraclass clustering performance and interclass discriminability, demonstrating the effectiveness of the WL-DeepGCN.

### IV. CONCLUSION

We propose a multimodal, cross-site data-based DeepGCN model to identify autism spectrum disorders. The WL-DeepGCN is based on a GCN with whole-brain functional connectivity as the node feature representation. A weight-learning network is proposed to represent the similarity between non-imaging data. Deepening the network structure enables the GCN to learn the neighbor information of nodes better, thus aggregating nodes to complete the classification task of unlabeled nodes. To avoid the gradient disappearance and gradient explosion problems, we propose GCN residual connections to fuse the potential features and introduce the EdgeDrop to alleviate the overfitting and oversmoothing of DeepGCN, thus improving the robustness of the model. Experimental results on the ABIDE-I dataset show that the WL-DeepGCN method brings substantial performance improvement for predicting autism spectrum disorders compared with existing methods. There are still some limitations to our work. First, we have yet to fully address the problem of gradient disappearance in deeper GCNs, which means that backpropagation through deeper layers of the network is oversmoothed, ultimately leading to convergence of the features of the graph vectors to the same values. Although our research model has reached eight layers, continuing to deepen the number of layers will result in severe performance degradation. We can consider using low-rank approximation, local aggregation and improved gradient clipping methods to deepen the network and effectively reduce the computational cost of the model. They can alleviate the gradient disappearance problem in GCN from the aspects of weight matrix, graph structure and gradient itself. Second, the ASD-related dataset has a small number of samples, which also limits the performance of the model to a large extent. Although the data enhancement method proposed by ASD-DiagNet [46] alleviates this problem to some extent, the model performance is still far from the clinical requirements. In addition, in the ABIDE dataset, the male-to-female ratio of the samples is unbalanced, with more men than women. More female volunteers should be called in the follow-up work to solve

this problem. Finally, as with other approaches using deep learning for psychiatric disease diagnosis, our approach lacks interpretability for biomarkers. Although the structure of the network is interpretable of the network, the model is unable to address the effect of the feature set on the predicted outcome. Interpretability is essential in trusting such models, which is necessary to understand brain abnormalities and differences between controls and patients. In our future work, we will attempt to use techniques such as feature importance scores and saliency maps to identify the most critical nodes and connections in the GCN as candidates for biomarkers, which help to explain the regions and nodes of interest for the model.

### ACKNOWLEDGMENT

The authors would like to thank to the ABIDE for the public access and download of the data for further research.

### REFERENCES

- [1] C. Lord et al., "Autism spectrum disorder," *Nature Rev. Disease Primers*, vol. 6, no. 1, p. 5, Jan. 2020, doi: [10.1038/s41572-019-0138-4](https://doi.org/10.1038/s41572-019-0138-4).
- [2] S. Yazdani, A. Capuano, M. Ghaziuddin, and C. Colombi, "Exclusion criteria used in early behavioral intervention studies for young children with autism spectrum disorder," *Brain Sci.*, vol. 10, no. 2, p. 99, Feb. 2020, doi: [10.3390/brainsci10020099](https://doi.org/10.3390/brainsci10020099).
- [3] M. Liu, B. Li, and D. Hu, "Autism spectrum disorder studies using fMRI data and machine learning: A review," *Frontiers Neurosci.*, vol. 15, Sep. 2021, Art. no. 697870, doi: [10.3389/fnins.2021.697870](https://doi.org/10.3389/fnins.2021.697870).
- [4] Z. Pang, J. Guo, Z. Ma, W. Sun, and Y. Xiao, "Median stable clustering and global distance classification for cross-domain person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 3164–3177, May 2022, doi: [10.1109/TCSVT.2021.3103753](https://doi.org/10.1109/TCSVT.2021.3103753).
- [5] Y. Du, Z. Fu, and V. D. Calhoun, "Classification and prediction of brain disorders using functional connectivity: Promising but challenging," *Frontiers Neurosci.*, vol. 12, p. 525, Aug. 2018, doi: [10.3389/fnins.2018.00525](https://doi.org/10.3389/fnins.2018.00525).
- [6] O. Dekhil et al., "Using resting state functional MRI to build a personalized autism diagnosis system," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Washington, DC, USA, Apr. 2018, pp. 1381–1385.
- [7] H. Li, N. A. Parikh, and L. He, "A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes," *Frontiers Neurosci.*, vol. 12, p. 491, Jul. 2018, doi: [10.3389/fnins.2018.00491](https://doi.org/10.3389/fnins.2018.00491).
- [8] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," *NeuroImage, Clin.*, vol. 17, pp. 16–23, Aug. 2018, doi: [10.1016/j.nicl.2017.08.017](https://doi.org/10.1016/j.nicl.2017.08.017).
- [9] M. Quaak, L. van de Mortel, R. M. Thomas, and G. van Wingen, "Deep learning applications for the classification of psychiatric disorders using neuroimaging data: Systematic review and meta-analysis," *NeuroImage, Clin.*, vol. 30, Jan. 2021, Art. no. 102584, doi: [10.1016/j.nicl.2021.102584](https://doi.org/10.1016/j.nicl.2021.102584).
- [10] L. Zhang, M. Wang, M. Liu, and D. Zhang, "A survey on deep learning for neuroimaging-based brain disorder analysis," *Frontiers Neurosci.*, vol. 14, p. 779, Oct. 2020, doi: [10.3389/fnins.2020.00779](https://doi.org/10.3389/fnins.2020.00779).
- [11] T. M. Ghazal et al., "Supervised machine learning empowered multifactorial genetic inheritance disorder prediction," *Comput. Intell. Neurosci.*, vol. 2022, May 2022, Art. no. 1051388, doi: [10.1155/2022/1051388](https://doi.org/10.1155/2022/1051388).
- [12] H. Peng, W. Gong, C. F. Beckmann, A. Vedaldi, and S. M. Smith, "Accurate brain age prediction with lightweight deep neural networks," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101871, doi: [10.1016/j.media.2020.101871](https://doi.org/10.1016/j.media.2020.101871).
- [13] B. Shi, Y. Chen, P. Zhang, C. D. Smith, and J. Liu, "Nonlinear feature transformation and deep fusion for Alzheimer's disease staging analysis," *Pattern Recognit.*, vol. 63, pp. 487–498, Mar. 2017, doi: [10.1016/j.patcog.2016.09.032](https://doi.org/10.1016/j.patcog.2016.09.032).
- [14] M. Khosla, K. Jamison, A. Kuceyeski, and M. R. Sabuncu, "Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction," *NeuroImage*, vol. 199, pp. 651–662, Oct. 2019, doi: [10.1016/j.neuroimage.2019.06.012](https://doi.org/10.1016/j.neuroimage.2019.06.012).
- [15] T. Eslami, F. Almuqhim, J. S. Raiker, and F. Saeed, "Machine learning methods for diagnosing autism spectrum disorder and attention-deficit/hyperactivity disorder using functional and structural MRI: A survey," *Frontiers Neuroinform.*, vol. 14, Jan. 2020, Art. no. 575999, doi: [10.3389/fninf.2020.575999](https://doi.org/10.3389/fninf.2020.575999).
- [16] M. Khodatars et al., "Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: A review," *Comput. Biol. Med.*, vol. 139, Dec. 2021, Art. no. 104949, doi: [10.1016/j.combiomed.2021.104949](https://doi.org/10.1016/j.combiomed.2021.104949).
- [17] X. Li et al., "BrainGNN: Interpretable brain graph neural network for fMRI analysis," *Med. Image Anal.*, vol. 74, Dec. 2021, Art. no. 102233, doi: [10.1016/j.media.2021.102233](https://doi.org/10.1016/j.media.2021.102233).
- [18] F. M. Bianchi, D. Grattarola, L. Livi, and C. Alippi, "Graph neural networks with convolutional ARMA filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3496–3507, Jul. 2019.
- [19] S. Parisot et al., "Spectral graph convolutions for population-based disease prediction," in *Medical Image Computing and Computer Assisted Intervention*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, S. Duchesne, Eds. Cham, Switzerland: Springer, 2017, pp. 177–185.
- [20] H. Jiang, P. Cao, M. Xu, J. Yang, and O. Zaiane, "Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction," *Comput. Biol. Med.*, vol. 127, Dec. 2020, Art. no. 104096, doi: [10.1016/j.combiomed.2020.104096](https://doi.org/10.1016/j.combiomed.2020.104096).
- [21] Y. Huang and A. C. S. Chung, "Edge-variational graph convolutional networks for uncertainty-aware disease prediction," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, and S. K. Zhou, Eds. Cham, Switzerland: Springer, 2020, pp. 562–572.
- [22] Y. Wang, M. Bai, M. Wang, F. Zhao, and J. Guo, "Multiscale traffic sign detection method in complex environment based on YOLOv4," *Comput. Intell. Neurosci.*, vol. 2022, Oct. 2022, Art. no. 5297605.
- [23] M. Cao et al., "Using DeepGCN to identify the autism spectrum disorder from multi-site resting-state data," *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 103015, doi: [10.1016/j.bspc.2021.103015](https://doi.org/10.1016/j.bspc.2021.103015).
- [24] G. Wen, P. Cao, H. Bao, W. Yang, T. Zheng, and O. Zaiane, "MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis," *Comput. Biol. Med.*, vol. 142, Mar. 2022, Art. no. 105239, doi: [10.1016/j.combiomed.2022.105239](https://doi.org/10.1016/j.combiomed.2022.105239).
- [25] J. Pan, H. Lin, Y. Dong, Y. Wang, and Y. Ji, "MAMF-GCN: Multi-scale adaptive multi-channel fusion deep graph convolutional network for predicting mental disorder," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105823, doi: [10.1016/j.combiomed.2022.105823](https://doi.org/10.1016/j.combiomed.2022.105823).
- [26] Y.-M. Wang et al., "Altered brain structural and functional connectivity in schizophrenia," *Psychol. Med.*, vol. 52, no. 5, pp. 834–843, Apr. 2022, doi: [10.1017/s033291720002445](https://doi.org/10.1017/s033291720002445).
- [27] M. Liu et al., "Golgi apparatus-targeted aggregation-induced emission luminogens for effective cancer photodynamic therapy," *Nature Commun.*, vol. 13, no. 1, p. 2179, Apr. 2022, doi: [10.1038/s41467-022-29872-7](https://doi.org/10.1038/s41467-022-29872-7).
- [28] P. Bede et al., "Phenotypic categorisation of individual subjects with motor neuron disease based on radiological disease burden patterns: A machine-learning approach," *J. Neurological Sci.*, vol. 432, Jan. 2022, Art. no. 120079, doi: [10.1016/j.jns.2021.120079](https://doi.org/10.1016/j.jns.2021.120079).
- [29] S. Hu, B. Lei, S. Wang, Y. Wang, Z. Feng, and Y. Shen, "Bidirectional mapping generative adversarial networks for brain MR to PET synthesis," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 145–157, Jan. 2022, doi: [10.1109/tmi.2021.3107013](https://doi.org/10.1109/tmi.2021.3107013).
- [30] X. Chen et al., "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444, doi: [10.1016/j.media.2022.102444](https://doi.org/10.1016/j.media.2022.102444).
- [31] 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.

- [32] D. Zacà, J. Jovicich, F. Corsini, U. Rozzanigo, F. Chioffi, and S. Sarubbo, “ReStNeuMap: A tool for automatic extraction of resting-state functional MRI networks in neurosurgical practice,” *J. Neurosurgery*, vol. 131, no. 3, pp. 764–771, Sep. 2019, doi: [10.3171/2018.4.jns18474](https://doi.org/10.3171/2018.4.jns18474).
- [33] M. A. Reiter, A. Jahedi, A. R. J. Fredo, I. Fishman, B. Bailey, and R.-A. Müller, “Performance of machine learning classification models of autism using resting-state fMRI is contingent on sample heterogeneity,” *Neural Comput. Appl.*, vol. 33, no. 8, pp. 3299–3310, Apr. 2021.
- [34] 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,” *Human Brain Mapping*, vol. 33, no. 8, pp. 1914–1928, Aug. 2012, doi: [10.1002/hbm.21333](https://doi.org/10.1002/hbm.21333).
- [35] P. Lanka, D. Rangaprakash, M. N. Dretsch, J. S. Katz, T. S. Denney, and G. Deshpande, “Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets,” *Brain Imag. Behav.*, vol. 14, no. 6, pp. 2378–2416, Dec. 2020, doi: [10.1007/s11682-019-00191-8](https://doi.org/10.1007/s11682-019-00191-8).
- [36] X. Ding, F. Yang, and F. Ma, “An efficient model selection for linear discriminant function-based recursive feature elimination,” *J. Biomed. Informat.*, vol. 129, May 2022, Art. no. 104070, doi: [10.1016/j.jbi.2022.104070](https://doi.org/10.1016/j.jbi.2022.104070).
- [37] Y. Fan, M. Chen, and X. Pan, “GCRFLDA: Scoring lncRNA-disease associations using graph convolution matrix completion with conditional random field,” *Briefings Bioinf.*, vol. 23, no. 1, Jan. 2022, doi: [10.1093/bib/bbab361](https://doi.org/10.1093/bib/bbab361).
- [38] R. Yu, W.-B. Huang, T. Xu, and J. Huang, “DropEdge: Towards deep graph convolutional networks on node classification,” in *Proc. Int. Conf. Learn. Represent.*, Addis Ababa, 2019, pp. 1–184.
- [39] R. M. Ganji, H. Jafari, and D. Baleanu, “A new approach for solving multi variable orders differential equations with Mittag-Leffler kernel,” *Chaos, Solitons Fractals*, vol. 130, Jan. 2020, Art. no. 109405.
- [40] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, New York, NY, USA, 2016, pp. 3844–3852.
- [41] J. Wang, X. Li, J. Li, Q. Sun, and H. Wang, “NGCU: A new RNN model for time-series data prediction,” *Big Data Res.*, vol. 27, Feb. 2022, Art. no. 100296, doi: [10.1016/j.bdr.2021.100296](https://doi.org/10.1016/j.bdr.2021.100296).
- [42] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [43] Y. Zhang, S. C. Satapathy, L.-Y. Zhu, J. M. Górriz, and S. Wang, “A seven-layer convolutional neural network for chest CT-based COVID-19 diagnosis using stochastic pooling,” *IEEE Sensors J.*, vol. 22, no. 18, pp. 17573–17582, Sep. 2022, doi: [10.1109/jsen.2020.3025855](https://doi.org/10.1109/jsen.2020.3025855).
- [44] H. Zhang et al., “Topographical information-based high-order functional connectivity and its application in abnormality detection for mild cognitive impairment,” *J. Alzheimer’s Disease*, vol. 54, no. 3, pp. 1095–1112, Oct. 2016, doi: [10.3233/JAD-160092](https://doi.org/10.3233/JAD-160092).
- [45] 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.
- [46] T. Eslami, V. Mirjalili, A. Fong, A. R. Laird, and F. Saeed, “ASD-DiagNet: A hybrid learning approach for detection of autism spectrum disorder using fMRI data,” *Frontiers Neuroinform.*, vol. 13, p. 70, Nov. 2019, doi: [10.3389/fninf.2019.00070](https://doi.org/10.3389/fninf.2019.00070).
- [47] F. Almuqhim and F. Saeed, “ASD-SAENet: A sparse autoencoder, and deep-neural network model for detecting autism spectrum disorder (ASD) using fMRI data,” *Frontiers Comput. Neurosci.*, vol. 15, Apr. 2021, Art. no. 654315, doi: [10.3389/fncom.2021.654315](https://doi.org/10.3389/fncom.2021.654315).
- [48] L. van der Maaten, “Accelerating t-SNE using tree-based algorithms,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Oct. 2014.