NF-GAT: A Node Feature-Based Graph Attention Network for ASD Classification

Shuaiqi Liu, Beibei Liang, Siqi Wang, Bing Li, *Member, IEEE*, Lidong Pan, and Shui-Hua Wang, *Member, IEEE*

***Abstract*—*Goal:* The purpose of this paper is to recognize autism spectrum disorders (ASD) using graph attention network. *Methods:* we propose a node features graph attention network (NF-GAT) for learning functional connectivity (FC) features to achieve ASD diagnosis. Firstly, node features are modelled based on functional magnetic resonance imaging (fMRI) data, with each subject modelled as a graph. Next, we use the graph attention layer to learn the node features and gets the node information of different nodes for ASD classification. *Results:* Compared with other models, the NF-GAT has significant advantages in terms of classification results. *Conclusions:* NF-GAT can be effectively used for ASD classification.**

***Index Terms*—Autism spectrum disorder, classification, functional connectivity, graphical attention network.**

Manuscript received 7 February 2023; revised 21 March 2023 and 12 April 2023; accepted 12 April 2023. Date of publication 26 April 2023; date of current version 11 June 2024. This work was supported in part by the Research Project of Hebei University Intelligent Financial Application Technology R&D Center under Grant XGZJ2022022, in part by the National Natural Science Foundation of China under Grant 62172139, in part by the Natural Science Foundation of Hebei Province under Grant F2022201055, in part by the Science Research Project of Hebei Province under Grant BJ2020030, in part by China Postdoctoral under Grant 2022M713361, in part by the Natural Science Interdisciplinary Research Program of Hebei University under Grant DXK202102, in part by the Open Project Program of the National Laboratory of Pattern Recognition under Grant 202200007, in part by the Open Foundation of Guangdong Key Laboratory of Digital Signal and Image Processing Technology under Grant 2020GDDSIPL-04, and in part by the High-Performance Computing Center of Hebei University. The review of this paper was arranged by Editor Y.-D. Zhang. *(Corresponding author: Lidong Pan.)*

Shuaiqi Liu is with the College of Electronic and Information Engineering, Machine Vision Engineering Research Center of Hebei Province, Hebei University, Baoding 071002, China, and also with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: shdkj-1918@163.com).

Beibei Liang, Siqi Wang, and Lidong Pan are with the Key Laboratory of Digital Medical Engineering of Hebei Province, College of Electronic and Information Engineering, Hebei University, Baoding 071002, China (e-mail: beibeiliang_hbu@163.com; sqwang_hbu@163.com; ldpan_hbu@163.com).

Bing Li is with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: bli@nlpr.ia.ac.cn).

Shui-Hua Wang is with the School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China (e-mail: shuihuawang@ieee.org).

This article has supplementary downloadable material available at https://doi.org/10.1109/OJEMB.2023.3267612, provided by the authors.

Digital Object Identifier 10.1109/OJEMB.2023.3267612

***Impact Statement*—A deep learning model, NF-GAT, is constructed, which uses rs-fMRI data to define node features and graph attention network to recognize ASD.**

I. Introduction

AUTISM spectrum disorder (ASD) [1] is a serious mental illness that affects the social behavior and communication skills of individuals. Due to its non-invasive, high temporal and spatial resolution [2], fMRI has been widely used in clinical and basic research in a variety of disciplines, including brain neuroscience, cognitive science and psychology [3], [4], [5]. Resting-state functional magnetic resonance imaging (rs-fMRI) data have been used to study ASD [6].

In computer-aided diagnosis, machine learning [7] and deep learning [8] have made good achievements in the classification and recognition of ASDs. However, graph structures such as knowledge graphs, brain connections, etc. are often irregular data structures which belong to non-Euclidean space [9]. Irregular data are difficult to be processed by convolutional neural networks (CNN) [10]. At the same time, many researchers have found that graph neural networks (GNNs) are powerful in modelling non-Euclidean data such as brain connectivity networks [11], [12], [13]. Graph convolutional networks (GCNs) are one of GNNs, which nicely extend convolutional operations to the graph domain. As brain connectivity graphs are irregular graph structures, GCNs are well suited to handle such data structures. For example, Parisot et al. [14] first proposed an application of GCNs to ASD classification. The classification performance of this method is significant improved to traditional machine learning methods. Jiang et al. [12] considered both the topological information of fMRI data and the correlation between subjects and used hierarchical GCN learn features for ASD classification. Wen et al. [13] used different thresholds to obtain different views, and then used GCN to learn common features between different views for ASD classification.

The above algorithm is well applied in emotion recognition. However, the performance suffers from several limitations due to a variety of problems. Firstly, the GCN-based classification algorithm requires the entire graph structure as the input of the model. Secondly, the GCN model is largely constrained by the graph structure, and the trained model cannot be applied to graphs of different structures. Thirdly, in the GCN model, all edges in the graph have the same weight. To solve these problems, Velickovié et al. [15] proposed a graph attention network (GAT). GAT

TABLE 1
SUBJECT DEMOGRAPHIC-INFORMATION (MEAN ± STD)

Site	ASD		TD	
	Age(year)	Sex(M/F)	Age(year)	Sex(M/F)
Pitt	18.3 ± 7.0	21/3	18.7 ± 6.7	22/4
Olin	17.1 ± 3.3	11/3	16.9 ± 3.6	13/0
OHSU	11.4 ± 2.2	12/0	10.2 ± 1.0	13/0
SDSU	15.3 ± 1.8	8/0	14.0 ± 1.9	13/6
Trinity	17.0 ± 3.2	19/0	17.1 ± 3.8	25/0
UM	12.9 ± 2.5	38/9	15.4 ± 3.4	55/18
USM	23.6 ± 8.4	43/0	20.9 ± 8.3	24/0
Yale	13.1 ± 3.0	14/8	13.6 ± 2.1	11/8
CMU	26.0 ± 5.4	4/2	27.8 ± 4.4	3/2
Leuven	17.0 ± 4.1	23/3	18.4 ± 5.0	26/4
KKI	10.7 ± 1.3	9/3	10.1 ± 1.2	15/6
NYU	14.8 ± 7.1	64/10	15.8 ± 6.2	72/26
Stanford	10.2 ± 1.6	9/3	9.8 ± 1.7	9/4
UCLA	13.1 ± 2.4	42/6	12.7± 2.1	32/5
MaxMun	28.4 ± 13.2	16/3	25.2 ± 8.4	26/1
Caltech	24.0 ± 7.6	4/1	28.2 ± 12.2	6/4
SBL	34.0 ± 6.6	12/0	33.6 ± 6.8	14/0
Total	17.1 ± 8.0	349/54	16.8 ± 7.2	378/90

aggregates only the features of a node's neighboring nodes as the new features of that node by introducing an attention mechanism when aggregating feature information. At the same time, GAT can learn different weights between nodes, so that each input of GAT is not the whole graph structure but a part of the graph, and which can be applied to different graph structures to capture the correlation between nodes greatly. On the basis of GAT, we construct a GAT based on node features (NF-GAT), which mainly uses rs-fMRI data to construct brain graph structures through brain atlas and define nodal features, and then learn node features through the graph attention layer to finally achieve the ASD classification. The main contributions are summarized as follows:

1) We propose NF-GAT model that can apply the learned node features to the ASD classification, and the experimental results verify the effectiveness of the proposed algorithm.
2) We give a new way to construct nodal features based on rs-fMRI data, which can facilitate the use of functional brain information for ASD classification.
3) The node features are mapped by a function and then weighted sum, and experiments show that the NF-GAT can better improve the classification performance.

II. MATERIALS AND METHODS

We construct NF-GAT to extract features from brain functional networks and to classify ASD. The proposed method is described below.

A. Data Sets

The datasets used in this paper are drawn from the public database which is Autism Brain Imaging Data Exchange I (ABIDE I) database [16]. Table 1 summarizes the demographic information for subjects. There are a total of 1112 subjects in the ABIDE dataset. And after pre-processing operation and the high-quality visual examination, 871 subjects are selected for the experiment.

B. Construction of Functional Connectivity Features

FC features are constructed by calculating similarity on pre-processed fMRI data. The process of constructing FC features is as follows. Firstly, the brain is divided into N ROIs by using the Havard Oxford (HO) atlas [17]. Secondly, the Pearson correlation coefficient is used to calculate the correlation, which can be defined as R_{ij} between brain region r_i and r_j. Finally, a correlation coefficient matrix of $N \times N$ is constructed for each subject, where N is the number of ROIs. This matrix is the functional connectivity matrix (FCM), which can be expressed as:

$$M = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ R_{21} & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ R_{N1} & R_{N2} & \cdots & R_{NN} \end{bmatrix} \tag{1}$$

Each row vector $\vec{h}_i$ of the FCM is chosen for the FC features. $\vec{h}_i$ can be expressed as:

$$\vec{h}_i = \{R_{i1}, R_{i2}, \cdots, R_{iN}\} \tag{2}$$

C. NF-GAT Model

The NF-GAT model structure is shown in Fig. 1. The NF-GAT model is mainly divided into two parts: the construction of the node feature graph and GAT. In the first part, each ROI is used as a node. Combining FC features extracted from fMRI data, we construct a graph containing node features and connecting edges between nodes for each subject. The graph of each subject is described as $\boldsymbol{G} = \{\boldsymbol{h}, \boldsymbol{R}\}$, where $\boldsymbol{h}$ is the node feature set and $\boldsymbol{R}$ is the edge connected between any two nodes. The second part is the GAT, which can learn the node features and achieve ASD recognition.

1) Construction of the Node Feature Graph: Each ROI is defined as a node. The weight value of the edge is the absolute value of R_{ij}, i.e., $|R_{ij}|$. The set of edges is defined by $\boldsymbol{R} = \{|R_{i1}|, |R_{i2}|, \cdots, |R_{ij}|\}$, $i, j \in [1, 110]$. Each row of constructed FC features is used as a node feature, and all the nodes are combined together to obtain the set of node features $\boldsymbol{h} = \{\vec{h}_1, \vec{h}_2, \cdots, \vec{h}_N\}, \vec{h}_i \in \mathbb{R}^F$, where F is the node feature dimension. Let the nodal feature graph constructed for each subject be $\boldsymbol{G} = \{\boldsymbol{h}, \boldsymbol{R}\}$.

2) Graph Attention Layers and Attention Mechanisms: The graph attention layer (GAL) proposed by Veličković et al. [15] is used for learning node representations. We set the number Z of GAL to 2. The input to the GAL is the node features $\boldsymbol{h} = \{\vec{h}_1, \vec{h}_2, \cdots, \vec{h}_N\}, \vec{h}_i \in \mathbb{R}^F$ and the output is a new set of node features $\boldsymbol{h'} = \{\vec{h}'_1, \vec{h}'_2, \cdots, \vec{h}'_N\}, \vec{h}'_i \in \mathbb{R}^{F'}$. A weight matrix $\boldsymbol{W} \in \mathbb{R}^{F' \times F}$ is trained for all nodes. To calculate the node representation, the one-hop neighborhood nodes of the node are aggregated for each node by using self-attentive mechanism.

The attention coefficient is defined as:

$$c_{ij} = a\left(\boldsymbol{W}\vec{h}_i, \boldsymbol{W}\vec{h}_j\right) \tag{3}$$

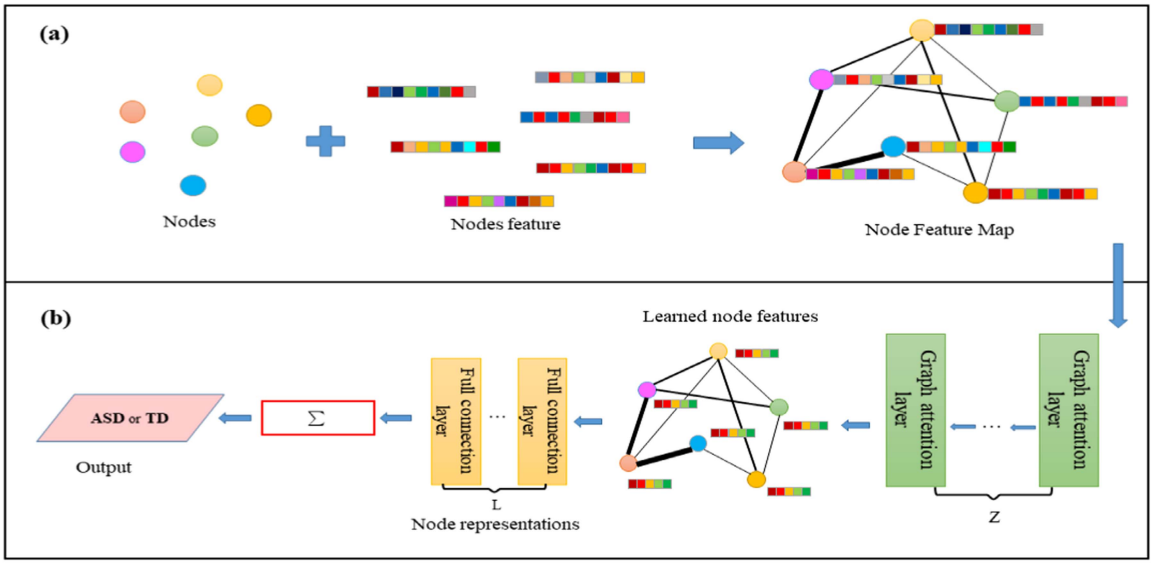

Fig. 1. The structure of NF-GAT.

where $a : \mathbb{R}^{F'} \times \mathbb{R}^{F'} \rightarrow \mathbb{R}$ is a shared attention mechanism that includes both self-attention and neighborhood attention, and the c_{ij} indicates the importance of node j to node i.

Softmax is introduced to regularize all neighboring nodes $j \in N_i$ of i, that is:

$$m_{ij} = softmax\left(c_{ij}\right) = \frac{\exp\left(c_{ij}\right)}{\sum_{k \in N_i} \exp\left(c_{ij}\right)} \tag{4}$$

The attention mechanism a is a single-layer feedforward neural network, $\vec{a} \in \mathbb{R}^{2F'}$ is the weight matrix which connected the layers in the neural network to each other, and a LeakyReLU is also added to the output layer. To sum up, the attention cross correlation coefficient can be obtained by:

$$\begin{aligned} m_{ij} &= softmax\left(LeakyReLU\left(c_{ij}\right)\right) \\ &= \frac{\exp\left(LeakyReLU\left(\vec{a}^T\left[\boldsymbol{W}\vec{h}_i \parallel \boldsymbol{W}\vec{h}_j\right]\right)\right)}{\sum_{k \in N_i} \exp\left(LeakyReLU\left(\vec{a}^T\left[\boldsymbol{W}\vec{h}_i \parallel \boldsymbol{W}\vec{h}_k\right]\right)\right)} \end{aligned} \tag{5}$$

where $\cdot^T$ represents matrix transposition and $\parallel$ represents joining the left and right matrices together.

The attention coefficients between the different nodes after regularization are obtained by the above operation and could be used to predict the output features of each node as shown in (6).

$$\vec{h}'_i = \sigma\left(\sum_{j \in N_i} m_{ij} \boldsymbol{W}\vec{h}_j\right) \tag{6}$$

where σ is the non-linear activation function, and j traversed in $j \in N_i$represents all nodes adjacent to i.

3) Multi-Head Attention: We use an attention mechanism constructed by a multi-head attention extension. The features learned by multiple attention heads can describe signals from different sides. Specifically, K independent self-attention mechanisms can be defined as (7), and then we connect the node feature representations obtained by each attention to establish the final node representation, which can be calculated as follows:

$$\vec{h}'_i = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{j \in N_i} m_{ij}^k \mathbf{W}^k \vec{h}_j\right) \tag{7}$$

where K denotes the number of the attention heads, K=3. m_{ij}^k is the normalized attention coefficient calculated from the k-th attention head. $\boldsymbol{W}^k$ denotes the weight matrix of the input linear transformation under the k-th attention head with the size of $\frac{F'}{K} \times F$.

Since the intermediate layer is the node feature representation obtained by each attention connected with an output dimension of KF'. For the final GAL performing multiple attentions, K-averaging is used in the final layer instead of the connection operation to make the final output is the same as the initial input dimension and delay the application of the final non-linear function, the node representation in the final layer is:

$$\vec{h}'_i = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in N_i} m_{ij}^k \boldsymbol{W}^k \vec{h}_j\right) \tag{8}$$

The multi-head attention of the node $\vec{h}_1$ on its neighborhood is updated with the target node based on the calculated weights.

4) Classification: The node features are transformed into node information through a two-layer full connection layer. In first full connection layer, the $\vec{h}'_i$ is mapped to the node information of each node I_i by the sigmoid function, which is:

$$I_i = sigmoid(\boldsymbol{W}^I \vec{h}'_i) \tag{9}$$

where $\boldsymbol{W}^I \in 1 \times F'$ is a shared weighted vector.

TABLE 2
MODEL TRAINING DETAILS

Name	Parameter
Operating system	Ubuntu 18.4
RAM	32G
CPU	Intel(R) Xeon(R) CPU I7
GPU	NVIDIA 2080Ti
Optimizer	Adam
Learning Rate	0.00005
Dropout	0.1
Batch size	10
Maximum training epoch	1000
Early stop mechanism	30

In second full connection layer, a weight is assigned to each node's information to be used in calculating each node's contribution to the final predicted outcome, and the weight is calculated as:

$$\boldsymbol{E} = softmax\left(\boldsymbol{W}^{E}\boldsymbol{I}\right) \tag{10}$$

where $\boldsymbol{W}^{E} \in N \times N$, $\boldsymbol{I} = \{I_1, I_2, \cdots, I_N\}$, and finally the weights $\boldsymbol{E}$ are linearly weighted sum with the nodes $\boldsymbol{I}$ to obtain the predictions of the NF-GAT model, that is:

$$p = \sum_{i=1}^{N} E_i I_i \tag{11}$$

III. RESULTS

A. Model Training and Algorithm Evaluation Indicators

The NF-GAT model is based on the Python language and implemented by using the TensorFlow framework. A 10-fold cross validation is used to evaluate the model's performance. The model training details are shown in Table 2. The loss function is a cross-entropy loss function, which is defined as

$$L = -\left[y \log \hat{y} + (1-y) \log (1-\hat{y})\right] \tag{12}$$

We use *accuracy*, *sensitivity*, *specificity*, *Fl value*, area under curve (*AUC*) and matthews correlation coefficient (*MCC*) as objective evaluation metrics. The *MCC* is defined as

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \tag{13}$$

where *TP* is the true positive rate, *FP* is the false positive rate, *TN* is the true negative rate, *FN* is the false negative rate. The *MCC* is essentially a correlation coefficient between the actual category and the predicted binary classification, which returns a value between -1 and +1. +1 indicates a perfect prediction, 0 indicates a random prediction, and -1 indicates that the predicted outcome is the opposite of the actual.

B. Comparison With Other Algorithms

To verify the effectiveness of the proposed model, we compare the NF-GAT with several other state-of-the-art ASD classification models.

The input of the traditional method is to remove the lower triangular and diagonal parts of the FCM. The remain part of the FCM is flatten into a one-dimensional vector as the classification features, whose dimension size is $[N \times (N-1)]/2$.

The classification results of the different algorithms are shown in Table 3. The best results are shown in bold. As can be seen, the NF-GAT outperforms other state-of-the-art classification methods. Firstly, compared with the traditional classification, our method has improved significantly. Secondly, compared with FCNN, NF-GAT achieves 4.47%, 8.24%, 1.42%, 6.94% and 8.79% improvement in five evaluation metrics. NF-GAT uses the GAL to learn node features, focus on the features of neighboring nodes and can assign different weights to different nodes in a neighborhood. So NF-GAT is better than GCN methods. At the same time, our method is roughly about 4% higher than FC-HAT in terms of accuracy, which may be due to the fact that the brain graph features learned by the hypergraph method cannot distinguish well between ASD and TD. In summary, the NF-GAT outperforms other classification methods in ASD classification.

C. Ablation Study

To demonstrate the validity of the proposed model framework, an ablation study is conducted. 1) We splice the learned node feature vectors together and spread them into a one dimensional vector as the input of subsequent network for classification, and name the model Flatten-GAT. 2) We average the node information to instead of weighted summation. (11) changes as (14). The model is named Average-GAT. The comparison results are shown in Table 4.

$$p = \frac{\sum_{i=1}^{N} I_i}{N} \tag{14}$$

In Table 4, our model achieves the best overall performance. The poor performance of the Flatten-GAT may be due to the loss of node information in the process of spreading the learned node feature vectors into a one-dimensional vector. The objective metrics of the Average-GAT are not as good as the NF-GAT due to averaging the information of each node, ignoring the different contributions of different nodes to the classification.

IV. DISCUSSION

It can be seen from Table 3 that different classification methods use different brain atlases. In order to test whether the results observed above depend on the choice of brain atlas, the NF-GAT is evaluated on six brain atlases other than the HO atlas, which are the Anatomical Automatic Labeling (AAL) atlas [22], the Cambion Craddock 200 (CC200) atlas [23], the Cameron Craddock 400 (CC400) atlas [24], Talariach and Tournoux (TT) atlas [25], Eikhoff-Zilles (EZ) atlas [26], and Dosenbach 160 atlas [27]. The information of these seven brain atlas is shown in Table 5.

The classification performance on the above atlases is shown in Fig. 2. The NF-GAT shows the best performance on all the

TABLE 3
CLASSIFICATION RESULTS OF DIFFERENT ALGORITHMS ON THE ABIDE I DATASET

Categories	Methods	Atlas	*Accuracy*	*Sensitivity*	*Specificity*	*F*1	*AUC*
Traditional methods	SVM	HO	65.44%	58.12%	72.15%	61.75%	65.13%
	RF	HO	66.02%	57.89%	73.47%	62.05%	65.68%
	GB	HO	62.68%	58.86%	66.20%	60.23%	62.53%
Non-graph deep learning methods	FCNN [17]	HO	69.81%	63.05%	75.63%	65.82%	72.62%
	PCCE+CNN [7]	HO	70.31%	71.67%	73.33%	-	73.00%
	IRHL-FC [8]	AAL	69.19%	64.79%	73.46%	-	76.00%
GCN methods	sGCN [11]	CC200	67.54%	64.73%	60.12%	-	64.11%
	DeepGCN [19]	HO	73.71%	-	-	69.68%	75.00%
	Hi-GCN [12]	AAL	67.23%	65.93%	68.43%	-	74.00%
	MVS-GCN [13]	CC200	69.89%	70.18%	63.05%	-	69.11%
GAT methods	GAT-LI [20]	HO	68.02%	**74.06%**	62.26%	69.31%	73.58%
	FC-HAT [21]	AAL	70.90%	70.00%	72.30%	-	-
Our method	NF-GAT	HO	**74.28%**	71.29%	**77.05%**	**72.76%**	**81.41%**

The bold entities represents the best value for objective evaluation index.

TABLE 4
RESULTS OF ABLATION EXPERIMENTS

Methods	*Accuracy*	*Sensitivity*	*Specificity*	*F*1	*MCC*
Flatten-GAT	71.41%	69.14%	73.52%	69.85%	42.82%
Average-GAT	71.63%	66.93%	75.91%	68.59%	44.08%
NF-GAT	**74.28%**	**71.29%**	**77.05%**	**72.76%**	**48.60%**

The bold entities represents the best value for objective evaluation index.

TABLE 5
DIFFERENT BRAIN ATLASES AND THEIR NUMBER OF ROIS

Atlas	Number of ROIs
HO	110
AAL	116
CC200	200
CC400	392
TT	97
EZ	116
Dosenbach160	161

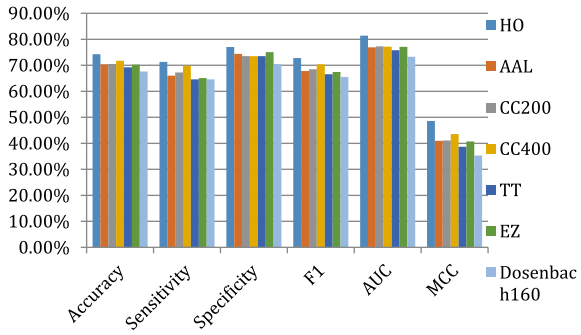

Fig. 2. Results of NF-GAT on seven brain atlases.

metrics of the HO atlas, especially compared with the Dosenbach160 atlas, the accuracy is improved by about 13.3%. This suggests that the HO atlas contains more features for classification than the other brain atlases. Therefore, all other experiments in this paper are conducted on the HO atlas.

Furthermore, we compared with Flatten-GAT and Average-GAT on several brain atlases. The results are shown in Supplementary Materials section. The classification result is related to each node, so we use BrainNet Viewer [28] shown the contribution of each node to the classification, the detailed analysis is included in the Supplementary Materials section.

V. CONCLUSION

A deep learning model, NF-GAT, is constructed, which uses rs-fMRI data to define node features and graph attention network to recognize ASD. A graph containing node features and connected edges between nodes is constructed for each subject, the node features are learned through the GAL. The node features are subjected to a function graph and the weighted sum of the node features is ultimately used to predict the probability of ASD. The results show that our model outperforms other models. Because the phenotypic information of the subjects also affects the classification, the work of this method is not enough. In addition, ASD classification studies should consider the effects of different atlases in future.

REFERENCES

[1] G. Arbanas, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Arlington, VA, USA: DSM-IV-TR, 2015.
[2] N. K. Logothetis et al., "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, vol. 412, no. 6843, pp. 150–157, 2001.
[3] K. Vakamudi et al., "Real-time presurgical resting-state fMRI in patients with brain tumors: Quality control and comparison with task-fMRI and intraoperative mapping," *Hum. Brain Mapping*, vol. 41, no. 3, pp. 797–814, Nov. 2019, doi: 10.1002/hbm.24840.
[4] S. Liu, L. Yin, S. Miao, J. Ma, and S. Hu, "Multimodal medical image fusion using rolling guidance filter with CNN and nuclear norm minimization," *Curr. Med. Imag. Rev.*, vol. 16, no. 10, pp. 1243–1258, Aug. 2020, doi: 10.2174/1573405616999200817103920.
[5] S. Liu et al., "Diffusion tensor imaging denoising based on Riemannian geometric framework and sparse Bayesian learning," *J. Med. Imag. Health Inform.*, vol. 9, no. 9, pp. 1993–2003, Dec. 2019, doi: 10.1166/jmihi.2019.2832.

[6] Z. - A. Huang, Z. Zhu, C. H. Yau, and K. C. Tan, "Identifying autism spectrum disorder from resting-state fMRI using deep belief network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 2847–2861, Jul. 2021, doi: 10.1109/TNNLS.2020.3007943.

[7] J. Ronicko et al., "Diagnostic classification of autism using resting-state fMRI data improves with full correlation functional brain connectivity compared to partial correlation," *J. Neurosci. Methods*, vol. 345, Nov. 2020, Art. no. 108884, doi: 10.1016/j.jneumeth.2020.108884.

[8] W. Jung et al., "Inter-regional high-level relation learning from functional connectivity via self-supervision," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2021, pp. 284–293, doi: 10.1007/978-3-030-87196-3.

[9] A. P. Norden, "The structure of the connection on a manifold of lines in a non-Euclidean space," *Izv.vys.uebn.zaved.matematika*, vol. 12, pp. 84–94, 1912.

[10] J. Ma, W. Wang, and L. Wang, "Irregular convolutional neural networks," in *Proc. IEEE 4th IAPR Asian Conf. Pattern Recognit.*, 2017, pp. 268–273.

[11] S. I. Ktena et al., "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, Apr. 2018, doi: 10.1016/j.neuroimage.2017.12.052.

[12] J. Hao J et al., "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.compbiomed.2020.104096.

[13] G. Wen et al., "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.compbiomed.2022.105239.

[14] S. Parisot et al., "Spectral graph convolutions for population-based disease prediction," in *Proc. 20th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2017, pp. 177–187, doi: 10.1007/978-3-319-66179-7_21.

[15] P. Velikovi et al., "Graph attention networks," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–9.

[16] A. D. 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, pp. 659–667, Jun. 2013, doi: 10.1038/mp.2013.78.

[17] R. S. Desikan et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, Jul. 2006, doi: 10.1016/j.neuroimage.2006.01.021.

[18] J. Hu et al., "Interpretable learning approaches in resting-state functional connectivity analysis: The case of autism spectrum disorder," *Comput. Math. Methods Med.*, vol. 2020, May 2020, Art. no. 1394830, doi: 10.1155/2020/1394830.

[19] 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.

[20] J. Hu et al., "GAT-LI: A graph attention network based learning and interpreting method for functional brain network classification," *BMC Bioinf.*, vol. 22, Jul. 2021, Art. no. 379, doi: 10.1186/s12859-021-04295-1.

[21] J. Ji, Y. Ren, and M. Lei, "FC–HAT: Hypergraph attention network for functional brain network classification," *Inf. Sci.*, vol. 608, pp. 1301–1316, Aug. 2022, doi: 10.1016/j.ins.2022.07.041.

[22] 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, doi: 10.1006/nimg.2001.0978.

[23] R. C. Craddock, G. A. James, P. E. Holtzheimer III, 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, doi: 10.1002/hbm.21333.

[24] M. Kunda, S. Zhou, G. Gong, and H. Lu, "Improving multi-site autism classification based on site-dependence minimisation and second-order functional connectivity," *IEEE Trans. Med. Imag.*, vol. 42, no. 1, pp. 55–65, Jan. 2023, doi: 10.1109/TMI.2022.3203899.

[25] L. Laitinen, "Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging," *Clin. Neurol. Neurosurgery*, vol. 91, no. 3, pp. 277–278, 1989, doi: 10.1016/0303-8467(89)90128-5.

[26] S. B. Eickhoff et al., "A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data," *NeuroImage*, vol. 25, no. 4, pp. 1325–1335, May 2005, doi: 10.1016/j.neuroimage.2004.12.034.

[27] N. U. Dosenbach et al., "Prediction of individual brain maturity using fMRI," *Science*, vol. 329, no. 5997, pp. 1358–1361, Sep. 2010, doi: 10.1126/science.1194144.

[28] M. Xia, J. Wang, and Y. He, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS One*, vol. 8, no. 7, Jul. 2013, Art. no. e68910, doi: 10.1371/journal.pone.0068910.

[29] J. R. Andrews-Hanna et al., "Functional-anatomic fractionation of the brain's default network," *Neuron*, vol. 65, no. 4, pp. 550–562, Feb. 2010, doi: 10.1016/j.neuron.2010.02.005.

[30] T. Suddendorf and M. C. Corballis, "The evolution of foresight: What is mental time travel, and is it unique to humans?," *Behav. Brain Sci.*, vol. 30, no. 3, pp. 299–313, Jun. 2007, doi: 10.1017/S0140525X07001975.

[31] N. Kim et al., "Aberrant neural activation underlying idiom comprehension in Korean children with high functioning autism spectrum disorder," *Yonsei Med. J.*, vol. 59, no. 7, pp. 897–903, Sep. 2018, doi: 10.3349/ymj.2018.59.7.897.

[32] Y. Chen, A. Liu, X. Fu, J. Wen, and X. Chen, "An invertible dynamic graph convolutional network for multi-center ASD classification," *Front. Neurosci.*, vol. 15, pp. 1662–453X, Feb. 2022, doi: 10.3389/fnins.2021.828512.