The effect of node features on GCN-based brain network classification: an empirical study

Guangyu Wang[1], Limei Zhang[1,2] and Lishan Qiao[1,2]

[1] Liaocheng University, Liaocheng, China
[2] Shandong Jianzhu University, Jinan, China

ABSTRACT

Brain functional network (BFN) analysis has become a popular technique for identifying neurological/mental diseases. Due to the fact that BFN is a graph, a graph convolutional network (GCN) can be naturally used in the classification of BFN. Different from traditional methods that directly use the adjacency matrices of BFNs to train a classifier, GCN requires an additional input-node features. To our best knowledge, however, there is no systematic study to analyze their influence on the performance of GCN-based brain disorder classification. Therefore, in this study, we conduct an empirical study on various node feature measures, including (1) original fMRI signals, (2) one-hot encoding, (3) node statistics, (4) node correlation, and (5) their combination. Experimental results on two benchmark databases show that different node feature inputs to GCN significantly affect the brain disease classification performance, and node correlation usually contributes higher accuracy compared to original signals and manually extracted statistical features.

INTRODUCTION

Resting-state functional magnetic resonance imaging (rs-fMRI) is an effective non-invasive technique for recording spontaneous neural activity in the brain when the subjects are awake and relaxed in the absence of task-specific stimuli. Based on rs-fMRI, the interaction between different regions of interest (ROIs) can be described by the brain functional network (BFN) (*Van Den Heuvel & Pol, 2010*) that has become an increasingly important tool to diagnose neurological or mental diseases, such as mild cognitive impairment (MCI) and autism spectrum disorder (ASD) (*Lord et al., 2020*).

In consideration of the fact that BFN is a graph, a graph convolutional network (GCN) can naturally be used to extract features from BFN for brain disease classification. For example, *Parisot et al. (2018)* combined imaging and non-imaging data for population brain analysis using GCN. *Arya et al. (2020)* used GCN to fuse structural and functional MRI for ASD classification. Additionally, many improved versions based on GCN were also developed from multi-layer (*Yu et al., 2020*; *Song et al., 2022*) and multi-view (*Cao et al., 2021*; *Wen et al., 2022*) methods to detect brain diseases. Unlike traditional classifiers

Corresponding author
Limei Zhang, limeizhanglcu@163.com

that are trained directly by the adjacency matrices of BFNs, the GCN-based classification methods need the node feature matrix as an extra input. However, to the best of our knowledge, there is no systematic study to analyze the influence of different node features on the classification results.

In this article, we use/design four kinds of node feature matrices and evaluate their influence on GCN-based brain disorder classification. The first node feature is original signals (OS) directly extracted from the rs-fMRI. These signals are highly reproducible and provide data sets that can be easily compared across studies. The second node feature is one-hot encoding (OH) that can uniquely identify the location of each node. In addition, we compute several node statistics (NS), including local efficiency, node centrality, and local clustering coefficient, which are concatenated to generate the third type of node features. The fourth node feature is the correlation vector (CV) that reflects the relationship between the current ROI and other ROIs. For a more systematical evaluation, we also discuss the impact of different node feature combinations on the classification results.

Two classification tasks are conducted in this empirical study: (1) MCI identification (identifying subject with MCI from healthy controls) and (2) ASD identification (identifying subject with ASD from healthy controls). Experimental results suggest that different node feature inputs to GCN have significant effects on brain disease classification performance, and CV usually contributes more than other three features.

The rest of this article is organized as follows. In 'Methods', we introduce the relevant concepts and network architecture that will be used in this study. In 'Materials', we describe the data materials including data acquisition and data preprocessing. In 'Experiments', we report the experimental details and experimental setting. In 'Results' and 'Discussion', we report experimental results and discussion, respectively. Finally, we conclude this article in 'Conclusions'.

METHODS

In this section, we introduce the relevant notation/concepts and architecture of the network used in this study.

Problem formulation

We represent each subject as a graph $G = (V, E)$, where $V = v_1, v_2, \cdots, v_n$ denotes the node set, n is the number of nodes/ROIs, and $E = e_1, e_2, \cdots, e_{|\varepsilon|}$ stands for the edge set that is determined by the relationship between different nodes. Accordingly, the node feature matrix and graph adjacency matrix are denoted by $X \in R^{n \times d}$ and $A \in 0, 1^{n \times n}$, respectively, where $x_i \in R^d$ is the feature of node v_i, d is the number of feature dimensions and $A_{ij} = 1$ if $(v_i, v_j) \in E$, 0 otherwise. We split subjects/graphs into training and test sets. Given an adjacency matrix A and a feature matrix X in the training sets, we learn a GCN-based classification model and predict the labels of subjects in the test set.

Network architecture

Our analysis is based on a spectral GCN with two convolution layers (*Kipf & Welling, 2016*), which has been empirically verified as an effective architecture for brain disorder
identification in several recent works (*Ying et al., 2018*; *Abu-El-Haija et al., 2020*; *Chu et al., 2022*). Its mathematical model can be formulated as follows:

$$f(X,A) = ReLU(\hat{A} \cdot ReLU(\hat{A}XW^0) \cdot W^1), \tag{1}$$

where $\hat{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, $A = A + I^n$, I^n is an identity matrix, and D is a diagonal matrix whose diagonal element $D_{ii} = \sum_j A_{ij}$ represents the degree of the i-th node, W^0 and W^1 are two layers of model parameters that need to be learned from data, and *ReLU* is a nonlinear activation function.

Suppose that $F = f_i(X,A) \in R^{n \times d'}$ is the learned representations of nodes, where f_i is the embedding of node v_i, d' is the dimension of the feature after embedding. We can further obtain a graph-level representation $H^G = r(F) \in R^{d'}$ for G by aggregating the node-level embedding. The readout function $r(\cdot)$ is a concatenation of the maximum pooling and average pooling operation (*Lee, Lee & Kang, 2019*). It can be formulated as follows:

$$H^G = \frac{1}{n}\sum_{i=1}^{n} f_i \parallel \max_{i=1} f_i, \tag{2}$$

where $\parallel$ denotes concatenation.

MATERIALS

In this section, we describe the data preparation, including data acquisition and data preprocessing.

Data acquisition

In this article, two publicly available datasets are used to evaluate the effect of node feature on GCN-based method. One is from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and the other is from the Autism Brain Imaging Data Exchange (ABIDE) dataset. The details of subjects involved in the two datasets are shown in Table 1, including distribution of ASD/MCI, gender (M/F), age and imaging parameters echo time (TE) and repetition time (TR). Specifically, the ADNI dataset consists of 137 subjects, including 69 HCs and 68 MCIs. The fMRI data were scanned by a 3.0-T Philips MRI scanner and the scanning parameters included TR = 3,000 ms, TE = 30 ms, flip angle = 80°, number of slices = 48, slice thickness = 3.3 mm, total volume = 140. For ABIDE dataset, 184 subjects are involved, including 79 ASDs and 105 HCs. All rs-fMRI data are obtained by 3.0-T Siemens Allegra scanner with the following imaging parameters: TR/TE = 2,000/15 ms, slices number = 33, voxel thickness = 4.0 mm, flip angle = 90°, and the scanning time is 6 min, resulting in 180 time points.

Data preprocessing

All the rs-fMRI data are preprocessed through DPARSF (*Yan et al., 2016*) and statistical parametric mapping (SPM12) based on the following standard pipeline: we discard the first five volumes of each functional time series and slice timing to allow signal stabilization. Then, the remaining nuisance signals, including ventricle and white matter signals, are regressed, and the Friston 24-parameter model is used to regress out the high-order effect

Table 1 Demographic information of ADNI and ABIDE datasets.

Datasets	Class	Gender (M/F)	Age (Years)	TR/TE (ms)
ADNI	MCI	39/29	76.50 ± 13.50	3000/30
	HC	17/52	71.50 ± 14.50	3000/30
ABIDE	ASD	68/11	18.58 ± 11.45	2000/15
	HC	79/26	19.13 ± 11.85	2000/15

Notes.

M/F, Male/Female; TE/TR, Echo Time/Repetition Time.

of head motion. After that, the signals are spatially normalized to MNI space, and the fMRI data are spatially smoothed with the full-width-half-maximum of four mm. Finally, depending on the AAL atlas, the preprocessed rs-fMRI time series signals are partitioned into 116 ROIs and then averaged respectively to get a representative signal for each ROI.

EXPERIMENTS

Experimental details

As mentioned earlier, network architecture GCN with two convolution layers, due to its popularity and effectiveness, is used in our study. Furthermore, with the aim of conducting control experiments, we set the embedding dimension to 32, the number of epochs to 100, the learning rate to 0.001 and the weight decay to $1e^{-3}$. Four kinds of node features are involved to verify the classification performance as follows.

(1) *Original Signals (OS):* OS capture the spontaneous fluctuations of brain activity associated with different ROIs, which is a manifestation of functional connectivity of the brain.

(2) *One-hot Encoding (OH):* We associate each node with a one-hot indicator that uniquely identifies the spatial position of each ROI.

(3) *Node Statistics (NS):* We design eight node statistics and concatenate them into a feature vector (*Zhang et al., 2022*). In particular, these statistics include three definitions of local clustering coefficients (*Li, Shang & Yang, 2017*), four centralities (*Hamilton, 2020*) (*i.e.,* degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality), as well as local efficiency.

(4) *Correlation vector (CV):* CV reflect the relationship between the current ROIs and other ROIs. The most popular measure is Pearson's correlation (PC) as follows:

$$c_{ij} = \frac{(s_i - \bar{s}_i)^T (s_j - \bar{s}_j)}{\sqrt{(s_i - \bar{s}_i)^T (s_i - \bar{s}_i)} \sqrt{(s_j - \bar{s}_j)^T (s_j - \bar{s}_j)}}, \tag{3}$$

where $s_i \in R^m (i = 1, 2, \cdots, n)$ is the extracted rs-fMRI time series from the ith ROI, m is the number of time points, n is the number of ROIs, $\bar{s}_i \in R^m$ is the mean of s_i, and c_{ij} is the correlation coefficient between the i-th and j-th ROIs.

Table 2 Performance comparison of different node features on both ADNI datasets, with bold values indicating the best results.

Dataset	Features	Acc ± std	Sen ± std	Spe ± std	Pre ± std	F1 ± std	AUC ± std
ADNI	OS	0.621 ± 0.009	0.570 ± 0.047	0.763 ± 0.022	0.721 ± 0.018	0.569 ± 0.005	0.637 ± 0.016
	OH	0.553 ± 0.014	0.578 ± 0.011	0.515 ± 0.035	0.636 ± 0.016	0.606 ± 0.007	0.503 ± 0.002
	NS	0.605 ± 0.017	0.668 ± 0.099	0.538 ± 0.096	0.607 ± 0.022	0.631 ± 0.039	0.572 ± 0.085
	CV	0.767 ± 0.075	0.819 ± 0.086	0.736 ± 0.122	0.735 ± 0.133	0.766 ± 0.080	0.759 ± 0.073
	OS + OH	0.642 ± 0.031	0.833 ± 0.046	0.500 ± 0.030	0.555 ± 0.036	0.666 ± 0.024	0.687 ± 0.010
	OH + NS	0.636 ± 0.035	0.570 ± 0.004	**0.793 ± 0.055**	0.721 ± 0.026	0.569 ± 0.019	0.637 ± 0.027
	OH + CV	**0.785 ± 0.028**	**0.857 ± 0.092**	0.714 ± 0.064	**0.750 ± 0.044**	0.800 ± 0.027	**0.885 ± 0.020**
	NS + CV	0.721 ± 0.021	0.788 ± 0.049	0.420 ± 0.060	0.734 ± 0.015	**0.803 ± 0.019**	0.824 ± 0.015
	OH + OS + NS	0.642 ± 0.047	0.667 ± 0.110	0.545 ± 0.054	0.475 ± 0.078	0.545 ± 0.097	0.809 ± 0.095
	OH + OS + CV	0.585 ± 0.028	0.671 ± 0.065	0.500 ± 0.071	0.574 ± 0.026	0.617 ± 0.033	0.542 ± 0.013
	OS + NS + CV	0.657 ± 0.042	0.585 ± 0.171	0.728 ± 0.085	0.745 ± 0.015	0.573 ± 0.083	0.846 ± 0.020
	OH + OS + NS + CV	0.642 ± 0.028	0.600 ± 0.057	0.666 ± 0.010	0.500 ± 0.027	0.545 ± 0.040	0.682 ± 0.095

Experimental setting

In this study, we randomly select 80% of the subjects for training and the remaining 20% for testing. For all datasets, we report the mean of 100-run results to evaluate the performance of the involved methods. The performance metrics include

$$\text{Accuracy} = (TP + TN)/(TP + FP + TN + FN),$$
$$\text{Sensitivity} = TP/(TP + FN),$$
$$\text{Specificity} = TN/(TN + FP),$$
$$\text{Precision} = TP/(TP + FP),$$
$$F1 - \text{score} = 2 \times TP/(2 \times TP + FP + FN).$$

where TP, TN, FP, and FN indicate the true positive, true negative, false positive and false negative, respectively. Additionally, the area under curve (AUC) is adopted for measuring the classification performance.

RESULTS

In Tables 2 and 3, we report the classification performance corresponding to four kinds of node features and their combinations based on PC adjacency matrix. From the above results, we have the following observations.

First of all, for the single node feature, CV can generally result in a higher classification accuracy than the others, and the results are consistent on the two datasets and three differently estimated BFNs (adjacency matrices). This drives us to think that the relationship between ROIs may contain more comparable discriminative information for MCI and ASD identification.

Moreover, the OH gives the worst result among the four single features in most cases, especially on the ADNI dataset, which illustrates that only the spatial position of the ROIs is not enough to capture effective information for disease classification. Interestingly, however, the combination of OH and CV usually gets the best results. A possible reason is that spatial position encoded by OH can provide complementary information to CV for more discriminative feature representation/learning. For example, in Tables 2 and 3,

Table 3 Performance comparison of different node features on ABIDE datasets, with bold values indicating the best results.

Dataset	Features	Acc ± std	Sen±std	Spe± std	Pre± std	F1 ± std	AUC ± std
ABIDE	OS	0.584 ± 0.079	0.619 ± 0.110	0.551 ± 0.104	0.563 ± 0.101	0.589 ± 0.080	0.605 ± 0.072
	OH	0.538 ± 0.022	0.731 ± 0.023	0.474 ± 0.026	0.574 ± 0.023	0.607 ± 0.114	0.510 ± 0.095
	NS	0.552 ± 0.014	0.661 ± 0.087	0.538 ± 0.035	0.573 ± 0.005	0.676 ± 0.040	0.536 ± 0.023
	CV	0.626 ± 0.055	0.662 ± 0.092	0.573 ± 0.107	0.642 ± 0.012	0.653 ± 0.060	0.677 ± 0.071
	OS + OH	0.705 ± 0.034	0.550 ± 0.100	0.718 ± 0.081	**0.701 ± 0.075**	0.606 ± 0.063	0.683 ± 0.010
	OH + NS	0.600 ± 0.057	0.712 ± 0.148	0.450 ± 0.130	0.632 ± 0.037	0.662 ± 0.085	0.589 ± 0.010
	OH + CV	0.626 ± 0.015	0.500 ± 0.000	0.707 ± 0.021	0.528 ± 0.014	0.560 ± 0.009	0.565 ± 0.013
	NS + CV	**0.747 ± 0.031**	**0.757 ± 0.065**	**0.741 ± 0.069**	0.636 ± 0.048	**0.688 ± 0.028**	**0.850 ± 0.009**
	OH + OS + NS	0.552 ± 0.026	0.532 ± 0.037	0.690 ± 0.044	0.561 ± 0.039	0.505 ± 0.035	0.568 ± 0.016
	OH + OS + CV	0.578 ± 0.019	0.525 ± 0.084	0.609 ± 0.101	0.500 ± 0.018	0.500 ± 0.036	0.514 ± 0.015
	OS + NS + CV	0.600 ± 0.025	0.750 ± 0.081	0.490 ± 0.044	0.518 ± 0.022	0.612 ± 0.015	0.727 ± 0.075
	OH + OS + NS + CV	0.684 ± 0.040	0.528 ± 0.110	0.733 ± 0.074	0.590 ± 0.065	0.593 ± 0.097	0.782 ± 0.015

we note that such a combination achieve the best classification accuracy (78.5% on ANDI dataset and 74.7% on ABIDE dataset, respectively).

Finally, different feature matrices do have a significant impact on the classification performance. It can also be seen that the designed single features usually cannot achieve the best accuracy. In contrast, when we incrementally concatenate these node features, the accuracy is generally improved. However, concatenating all features cannot achieve the best accuracy, indicating that simply increasing the feature dimension does not guarantee a good classification effect. In practice, the medical data are generally limited, and more features means more parameters, especially for GCN-based deep methods, which may lead to the overfitting problem. In fact, even if the data amount is sufficient, more features do not necessarily work well, since some redundant features can easily degrade the performance (*Guyon & Elisseeff, 2003*).

DISCUSSION

In general, different BFN estimation methods have a significant influence on the classification performance (*Sun et al., 2021*; *Jiang et al., 2022*). Therefore, we discuss the effect of node features based on BFNs estimated by different methods, including PC, sparse representation (SR) (*Qiao et al., 2018*), and low-rank representation (LR) (*Qiao, Chen & Tan, 2010*). The visualizations of three BFNs are shown in Fig. 1.

We compare the classification results on different datasets under three adjacency matrices. We show the impact of different node features based on SR and LR adjacency matrices in Figs. 2 and 3, respectively. The results show that on ADNI dataset LR-based adjacency matrix generally leads to a good performance, while on ABIDE dataset SR-based adjacency matrix works better. This suggests that appropriate sparsity is beneficial for brain disease diagnosis compared to the fully connected matrix estimated by PC. This is consistent with many existing studies (*Dadi et al., 2019*; *Pervaiz et al., 2020*).

In this article, we mainly consider the commonly used features or design features based on classical node statistics. In fact, however, a graph generally contains complicated

(1) PC   (2) SR   (3) LR

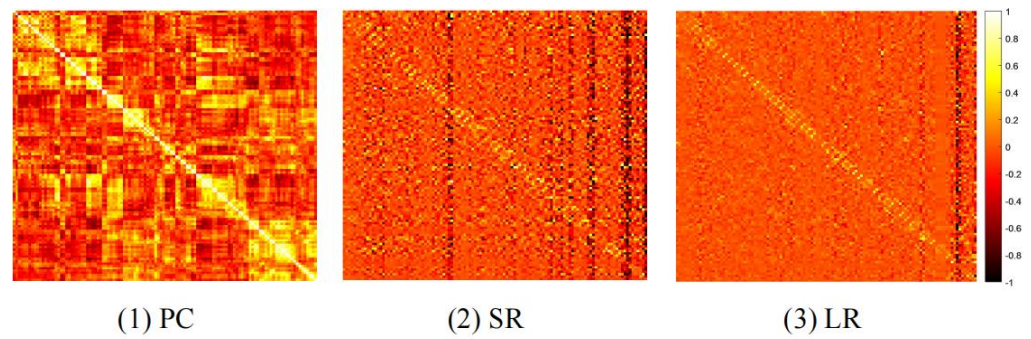



Figure 1 **The heat map of adjacency matrices estimated by PC, SR and LR.**

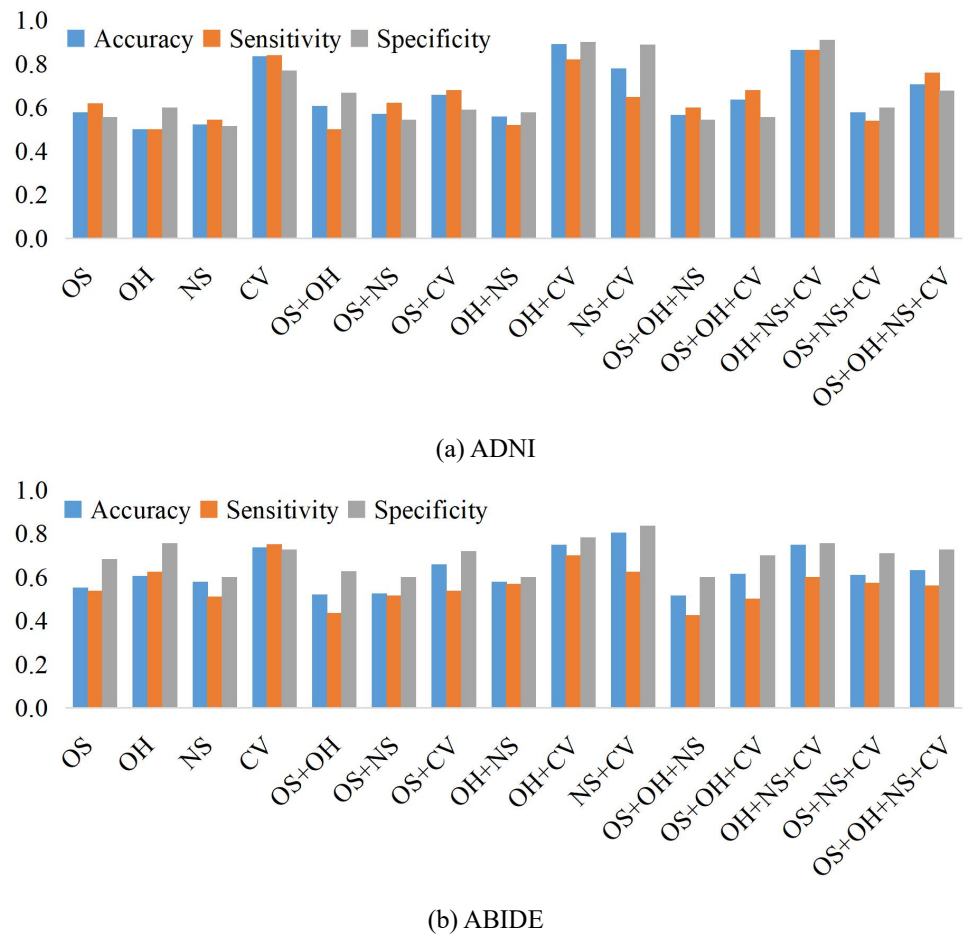
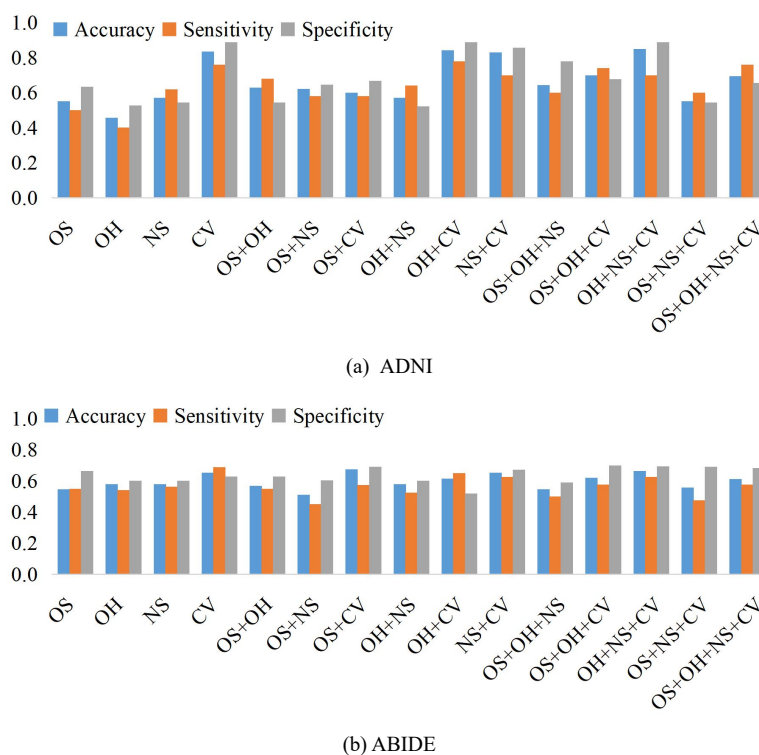



Figure 2 **The classification performance corresponding to four node features and their combinations on (A) ADNI and (B) ABIDE datasets, respectively, where the BFNs are estimated by SR.**

(a) ADNI

(b) ABIDE

Figure 3 **The classification performance corresponding to four node features and their combination on (A) ADNI and (B) ABIDE datasets, respectively, where the BFNs are estimated by LR.**

structures such as different motifs. Therefore, in theory, we can design various node statistics as the candidate of the node features in GCN. In addition, the brain network is generally expressed by a signed graph, and thus different possible triad configurations may be captured as the node features according to the structural balance theory (*Newman, 2018*). In the future, we will try more node features from different views or levels to improve the performance of brain disease classification.

CONCLUSIONS

In this article, we use/design four kinds of node feature matrices based on rs-fMRI data to empirically evaluate their influence on the GCN-based brain disorder classification tasks. In addition, we conduct control experiments of the feature matrix under three adjacency matrices that correspond to different BFN estimation methods. Experimental results demonstrate that different node features have significant effects on the classification performance. In general, CV, reflecting the relationship between ROIs, or features containing CV tends to result in higher classification performance. Additionally, we also note that combined features usually achieve higher accuracy than single features.

APPENDIX

In Table A1, we give the calculation formula and physical meaning of eight node statistics involved in the article. In particular, these statistics include three definitions of local clustering coefficients, four centralities (*i.e.,* degree centrality, betweenness centrality, closeness centrality and eigenvector centrality), as well as local efficiency. For consistence with the above description, we also represent each subject as a graph $G = (V, E)$, and $A = (a_{ij}) \in R^{n \times n} (i, j = 1, \ldots, n)$ is the adjacency matrix of graph G.

Table A1 The calculation formula and physical meaning of eight node statistics. Note that c is a constant, G_i refers to the subgraph formed by the neighbors of node i, and l_{jq} represents the shortest path length between nodes j and q.

Node statistics	Physical meaning	Calculation formula
Local clustering coefficient	The connectivity in node's neighborhoods	$NS_{i,O} = \frac{\sum_{j,q}(a_{ji}a_{iq}a_{jq})^{\frac{1}{3}}}{d_i(d_i-1)}$
		$NS_{i,Z} = \frac{\sum_{j,q}(a_{ji}a_{iq}a_{jq})}{(\sum_q a_{iq})^2 - \sum_q a_{iq}^2}$
		$NS_{i,C} = \frac{\sum_{j,q}(a_{ji}a_{iq}a_{jq})}{\sum_{j \neq q}|a_{ji}a_{iq}|}$
Degree centrality	The number of edges associated with a node	$NS_{i,D} = \sum_j a_{ij}$
Betweenness centrality	The number of shortest paths between any two pairs of other nodes for a node	$NS_{i,B} = \sum_{j \neq i \neq q} \frac{l_{jq}(i)}{l_{jq}}$
Closeness centrality	The closeness of a node to other nodes	$NS_{i,S} = \frac{n-1}{\sum_{i \neq j} l_{ij}}$
Eigenvector centrality	The importance of its neighbor nodes	$NS_{i,E} = c \sum_j a_{ij} x_j$
Local efficiency	The local information transmission ability of network	$NS_{i,L} = \frac{1}{N_{G_i}(N_{G_i}-1)} \sum_{j \neq q \in G_i} \frac{1}{l_{jq}}$

Funding

This work was supported by the National Natural Science Foundation of China (No. 61976110, 62176112, 11931008), and the Natural Science Foundation of Shandong Province (No. ZR202102270451). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 61976110, 62176112, 11931008.
Natural Science Foundation of Shandong Province: ZR202102270451.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Guangyu Wang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Limei Zhang conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Lishan Qiao conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The ABIDE data is available at: http://fcon_1000.projects.nitrc.org/indi/abide/; NYU_0050954–NYU_0051154.

The ADNI data is available at http://adni.loni.usc.edu/. The raw measurements are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.14835#supplemental-information.

REFERENCES

Abu-El-Haija S, Kapoor A, Perozzi B, Lee J. 2020. N-GCN: multi-scale graph convolution for semi-supervised node classification. In: *Uncertainty in Artificial Intelligence. PMLR. Available at* https://proceedings.mlr.press/v115/abu-el-haija20a.html.

Arya D, Olij R, Gupta DK, Gazzar AEl, Wingen G, Worring M, Thomas RM. 2020. Fusing structural and functional MRIs using graph convolutional networks for autism classification. In: *Medical Imaging with Deep Learning. PMLR.* 44–61.

Cao M, Yang M, Qin C, Zhu X, Chen Y, Wang J, Liu T. 2021. Using deep GCN to identify the autism spectrum disorder from multi-site resting-state data. *Biomedical Signal Processing and Control* **70**:103015 DOI 10.1016/j.bspc.2021.103015.

Chu Y, Wang G, Cao L, Qiao L, Liu M. 2022. Multi-scale graph representation learning for autism identification with functional MRI. *Frontiers in NeuroInformatics* **15**:802305 DOI 10.3389/fninf.2021.802305.

Dadi K, Rahim M, Abraham A, Chyzhyk D, Milham M, Thirion B, Varoquaux G, Initiative ADN. 2019. Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage* **192**:115–134 DOI 10.1016/j.neuroimage.2019.02.062.

Guyon I, Elisseeff A. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**:1157–1182.

Hamilton WL. 2020. Graph representation learning. *Synthesis Lectures on Artifical Intelligence and Machine Learning* **14(3)**:1–159.

Jiang X, Qiao L, De Leone R, Shen D. 2022. Joint selection of brain network nodes and edges for mci identification. *Computer Methods and Programs in Biomedicine* **225**:107082 DOI 10.1016/j.cmpb.2022.107082.

Kipf TN, Welling M. 2016. Semi-supervised classification with graph convolutional networks. ArXiv preprint. arXiv:1609.02907.

Lee J, Lee I, Kang J. 2019. Self-attention graph pooling. In: *International Conference on Machine Learning. PMLR*. 3734–3743.

Li Y, Shang Y, Yang Y. 2017. Clustering coefficients of large networks. *Information Sciences* **382**:350–358.

Lord C, Brugha TS, Charman T, Cusack J, Dumas G, Frazier T, Jones EJ, Jones RM, Pickles A, State MW. 2020. Autism spectrum disorder. *Nature Reviews Disease Primers* **6(1)**:1–23 DOI 10.1038/s41572-019-0135-7.

Newman M. 2018. *Networks.* New York, Oxford: Oxford University Press.

Parisot S, Ktena SI, Ferrante E, Lee M, Guerrero R, Glocker B, Rueckert D. 2018. Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. *Medical Image Analysis* **48**:117–130 DOI 10.1016/j.media.2018.06.001.

Pervaiz U, Vidaurre D, Woolrich MW, Smith SM. 2020. Optimising network modelling methods for fMRI. *NeuroImage* **211**:116604 DOI 10.1016/j.neuroimage.2020.116604.

Qiao L, Chen S, Tan X. 2010. Sparsity preserving projections with applications to face recognition. *Pattern Recognition* **43(1)**:331–341 DOI 10.1016/j.patcog.2009.05.005.

Qiao L, Zhang L, Chen S, Shen D. 2018. Data-driven graph construction and graph learning: a review. *NeuroComputing* **312**:336–351 DOI 10.1016/j.neucom.2018.05.084.

Song X, Zhou F, Frangi AF, Cao J, Xiao X, Lei Y, Wang T, Lei B. 2022. Multi-center and multi-channel pooling GCN for early ad diagnosis based on dual-modality fused brain network. In: *IEEE Transactions on Medical Imaging*. Piscataway: IEEE.

Sun L, Xue Y, Zhang Y, Qiao L, Zhang L, Liu M. 2021. Estimating sparse functional connectivity networks *via* hyperparameter-free learning model. *Artificial Intelligence in Medicine* **111**:102004 DOI 10.1016/j.artmed.2020.102004.

Van Den Heuvel MP, Pol HEH. 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology* **20(8)**:519–534 DOI 10.1016/j.euroneuro.2010.03.008.

Wen G, Cao P, Bao H, Yang W, Zheng T, Zaiane O. 2022. Mvs-GCN: a prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. *Computers in Biology and Medicine* **142**:105239 DOI 10.1016/j.compbiomed.2022.105239.

Yan CG, Wang XD, Zuo XN, Zang YF. 2016. DPABI: data processing & analysis for (resting-state) brain imaging. *NeuroInformatics* **14(3)**:339–351 DOI 10.1007/s12021-016-9299-4.

Ying Z, You J, Morris C, Ren X, Hamilton W, Leskovec J. 2018. Hierarchical graph representation learning with differentiable pooling. In: *Advances in Neural Information Processing Systems. Available at https://proceedings.neurips.cc/paper/2018/file/e77dbaf6759253c7c6d0efc5690369c7-Paper.pdf* .

Yu S, Wang S, Xiao X, Cao J, Yue G, Liu D, Wang T, Xu Y, Lei B. 2020. Multi-scale enhanced graph convolutional network for early mild cognitive impairment detection. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 228–237.

Zhang Y, Xue Y, Wu X, Qiao L, Shen D. 2022. Selecting multiple node statistics jointly from functional connectivity networks for brain disorders identification. *Brain Topography* **35(5)**:559–571 DOI 10.1007/s10548-022-00914-z.