Learning Features of Brain Network for Anomaly Detection

Jiaxin Liu†, Wei Zhao†, Ye Hong†, Sheng Gao‡, Xi Huang‡ and Yingjie Zhou†*
for the Alzheimer's Disease Neuroimaging Initiative§

†College of Computer Science, Sichuan University, Chengdu, Sichuan, China, 610065
‡Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, Sichuan, China, 610000

***Abstract*—Brain network is a kind of biological networks, which could express complex connectivity among brain functional components. The node in brain network denotes region of interest, which executes specific function in the brain. The edge in brain network represents the connection relationship between nodes. Neuropsychiatric disorders can cause changes in the brain's nerves, which will further change the related characteristics of brain network. The detection for neuropsychiatric disorders with brain network could be treated as an anomaly detection problem. Recent researches have explored using complex networks or graph mining to deal with the detection problem. However, they neither ignore local structural features in critical regions nor fail to comprehensively extract the structural features for brain network. In this paper, we propose a feature learning method to build effective representations for brain network. By treating the closed frequent graph as a node, these representations contain both connection relationships in critical regions and local/global structural features for critical regions, which could benefit the detection with brain network. Experiments using real world data indicate that the proposed method could improve the detection ability of existing machine learning methods in the literatures.**

***Index Terms*—brain network, anomaly detection, feature learning**

I. Introduction

The rapid improvement of the information technology brings remarkable convenience to our daily life and promotes the communication in the world. Taking the inspiration from the information technology, the knowledge of complex networks has been applied to model complex systems, Internet, economic networks and even biological networks. For instance, in [1], the international finance is abstracted into a weighted directed network, which reflects strong interdependence in the international finance. The analysis based on the complex networks brings a new perspective to understand the complexity of internal structure and interactions in complex systems.

In this paper, we focus on analyzing a kind of biological networks, i.e., brain network, which could be an efficient expression of the connection relationships in the brain. The node in brain network denotes Region Of Interest (ROI), which executes specific function in the brain. The edge in brain network represents the connection relationship between two ROIs. Several works related with brain network have suggested that many neuropsychiatric disorders, e.g., mild cognitive impairment[2] and Alzheimer[3], are related to the brain networks. Thus, using brain network to accurately and effectively identify various neuropsychiatric disorders can help to take appropriate measures in time for targeted treatment, which is of great significance.

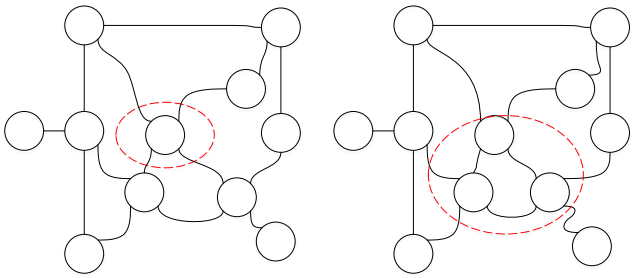
Recent works[2, 4] have explored using complex networks or graph mining to deal with the identification of neuropsychiatric disorders. The analysis based on complex networks usually uses some basic characteristics of networks, e.g., the degree distribution, to express the connection relationships in the brain network. This kind of expression treats each node (as shown in Figure 1(a)) in the brain network as analysis object and considers all nodes to be of equal importance, which ignores local structural features in critical regions of brain network. Different from the analysis based on the complex networks, the analysis based on the graph mining treats the subgraph (as shown in Figure 1(b)) as analysis object. It could reveal important local characteristics in the brain network. However, it fails to comprehensively extract the structural features for brain network, which ignores the local or global structural relationships between the graph and other parts of brain network. Due to the complexity of human brain, these two kinds of expressions for brain network are limited.

To tackle this, we propose a feature learning method to build effective representations for brain network. The feature learning method consists of three parts. Firstly, through a preprocessing process, the brain network is constructed by using the real world fMRI data. Secondly, the closed frequent graphs of brain network are mined by $CloseGraph$[5]. The closed frequent graphs, which are the critical regions in the brain network, reveal the most significant local structures in the brain network. Thirdly, inspired from complex network analysis[6],

This work is partly supported by National Natural Science Foundation of China (NSFC) with grant number 61801315.

*All corresponding to Yingjie Zhou (Email: yjzhou@scu.edu.cn).

§Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.



(a) The analysis based on the complex networks

(b) The analysis based on the graph mining

Figure 1. Two types of analysis for brain networks. The analysis based on the complex networks treats each node as the analysis object, which might ignore the local structural features in critical regions. The analysis based on the graph mining analyzes each subgraph separately, which might fail to express the structural features for brain networks.

three kinds of graph-based parameters are calculated to build the feature representations for brain network, while treating the closed frequent graph as a node. These representations not only describe the internal connection relationships of critical regions, but also express the local/global structural features between critical regions and other nodes in the brain network. The major contributions of this paper could be summarized as follows:

- We propose a feature learning method to build effective representations for the brain network. By treating the closed frequent graph as a node, the representations express three kinds of relationships in brain network, i.e., the internal connection relationships of critical regions, the local structural features of the relationships between critical regions and their neighbors, and the global structural features of the relationships between critical regions and the whole brain network. The proposed feature learning method could be easily integrated with popular machine learning methods for anomaly detection.
- We conducted experiments that employ learned features from the proposed method for anomaly detection by using real world data.

The rest of the paper is organized as follows. Section II introduces the preliminaries, including the background knowledge of parameters in complex networks and closed frequent graph. Section III presents the proposed method. In Section IV, we evaluate the proposed method by using real world data. We conclude the paper in Section V.

II. Preliminaries

A. Complex network parameters

Due to the complexity of the network, it is challenge to comprehensively express the connectivity and structural characteristics of the network. The weighted clustering coefficient[6] and weighted average efficiency[7] are two common parameters that express the local and global characteristics in the network, respectively.

The weighted clustering coefficient of a node indicates how close is it to its neighbors. The formula of weighted clustering coefficient is as follows:

$$C_i = \frac{2}{N_i\left(N_i-1\right)} \sum_{j,k} \left(\tilde{w}_{i,j}\tilde{w}_{j,k}\tilde{w}_{k,i}\right)^{1/3} \tag{1}$$

Where j and k represent the two different neighbors of node i. N_i is the number of neighbors of node i. $\tilde{w}_{i,j}$, $\tilde{w}_{j,k}$ and $\tilde{w}_{k,i}$ represent the edge weights between nodes (i,j), (j,k) and (k,i), respectively.

The weighted average efficiency represents the average efficiency of information transmission between nodes in the network, and it is inversely proportional to the average path length between nodes in the network. The formula of weighted average efficiency is as follows:

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{i,j}} \tag{2}$$

Where i and j represent the two nodes in graph G. N is the number of nodes in G. $d_{i,j}$ represents minimum path length between node i and node j.

B. Closed frequent graph

The closed frequent graphs are several special subgraphs of a graph dataset, which indicate the important information in a graph dataset. The definition of the closed frequent graphs of a graph dataset is shown in [5], which is briefly described as follows. Graph g is a closed frequent graph of the graph dataset D, if g is a frequent subgraph of D and there exists no proper supergraph[5] of g, which has the same support as g. The frequent subgraph g has the property that $support(g) \geqslant minSup$ (the minimum support threshold). The $support(g)$ represents that the number of graphs (contained in D), of which g is a subgraph.

With the varying value of $minSup$, different closed frequent graphs in the brain network are mined. When $minSup$ is set in a small value, the number of closed frequent graphs is large while the degree of frequency of each graph is low. Some redundant information might be contained in the closed frequent graphs. When $minSup$ is set in a large value, the number of closed frequent graphs is small while the degree of frequency of each graph is high. Some significant information might be removed from the closed frequent graphs. Thus, a appropriate value for $minSup$ needs to be set.

III. Proposed Method

The overall method we proposed is shown in Figure 2. There are three parts in this method. In the first part, through a preprocessing process, the fMRI data of each sample is transformed into the brain network. Secondly, according to the labels, the samples in the training set are divided into two groups, i.e., the Healthy Controls(HC) and Alzheimer's Disease(AD), and the closed frequent graphs of each group are mined by $CloseGraph$[5]. Thirdly, based on the closed frequent graphs, the graph-based features are constructed by the proposed method. These features not only describe the internal connection relationships of the critical regions in the brain network, but also express the local/global structural

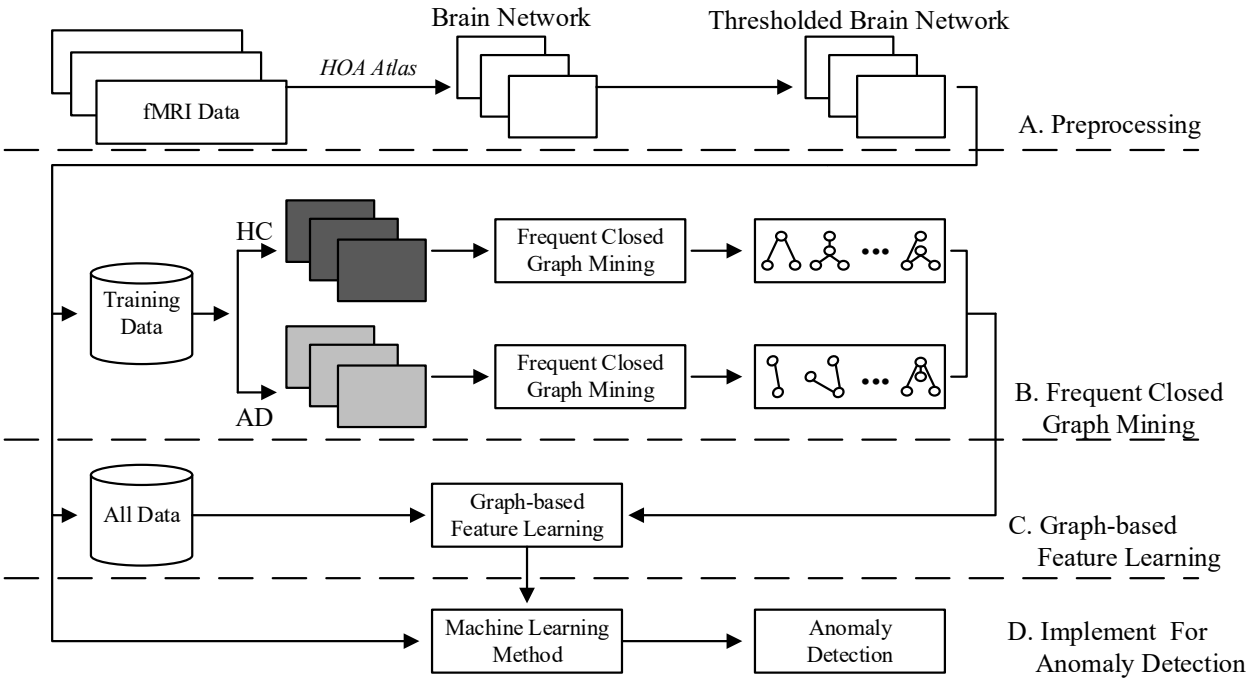


Figure 2. Overview of the processed method. The proposed method contains three parts. First, the fMRI data is preprocessed to make up the brain network. Then, the closed frequent graphs of HC group and AD group are mined. Third, based on the mined closed frequent graphs, the graph-based features of each sample are constructed by three kinds of graph-based parameters. By using these features, the proposed method could be integrated with machine learning methods to implement anomaly detection for the brain network.

information between the critical regions and other nodes of the brain network. Various machine learning methods could integrate the proposed feature learning method to detect anomalies for the brain network. In order to understand the proposed method, the details are described as follows.

A. Preprocessing

In this preprocessing process, the thresholded brain network of each sample is constructed by converting the fMRI data into a network. The fMRI data used in the experiment is collected by Alzheimer's Disease Neuroimaging Initiative(ADNI). All 297 subjects have whole-brain MRI scans on 3-Tesla GE Medical Systems. The details of the parameters set during the scan are described in detail in [2].

In order to construct the brain network for each sample, an advanced brain network construction method proposed in [2] is implemented. Some efficient tools, i.e., FSL[8] and PANDA[9], are used in our preprocessing process. FSL is a comprehensive image analysis library of fMRI data and PANDA is a toolbox for pipeline processing of fMRI data. With the Harvard–Oxford atlas(HOA atlas)[10], we divided the cerebral cortex and subcortical structures into multiple regions of interest(ROIs). A parameter is calculated to represent the connectivity between ROIs. The details of nodes and edges in the brain network are described as follows.

- The nodes in the brain network represent the ROIs that are predefined in the brain atlas.
- The edges in the brain network represent the connectivity between ROIs. In this work, we use the reciprocal of the number of brain fibers(FN) between two ROIs as the weight of the edge that connects the two ROIs. The smaller the weight of the edge is, the closer the relationship between the two ROIs connected by this edge is. In order to analyze the network efficiently, we filter out some edges in the brain network, when the weight of the edge is more than $\frac{1}{4}$.

B. Closed frequent graph mining

Closed frequent graph mining aims to obtain the most representative subgraphs to reveal important characteristics of the brain network. Neuropsychiatric disorders usually cause specific changes in the brain's nerves and these changes will make the characteristics of the brain network different from those of normal people. Therefore, these changes will reveal significant information that is useful to distinguish neuropsychiatric patients from normal people.

In order to express the changes mentioned above, the samples in the training set are divided into two groups, i.e., the Healthy Controls(HC) and Alzheimer's Disease(AD). By utilizing the $CloseGraph$[5], the closed frequent graphs of each group are mined. The closed frequent graphs indicate the important characteristics of the brain networks in a group. In this way, the closed frequent graphs will reveal the differences in the characteristics of the two groups. The closed frequent graphs will be used for the graph-based feature learning.

C. Graph-based feature learning

With the closed frequent graphs, the graph-based features for each sample are constructed by three kinds of graph-based

parameters. These features describe the internal connection relationships of critical regions, and express the local/global structural features in the brain network. The feature learning process is described as follows. First, all the closed frequent graphs are combined into a set, which contains N closed frequent graphs. Then, based on the brain network and closed frequent graphs, three kinds of graph-based parameters for each sample are calculated by specially designed formulas. Note that when a closed frequent graph is not included by a sample, all three calculated parameters for this sample are set to 0. Finally, the graph-based features for each sample are represented as a vector of length $3N$.

The formulas of three kinds of graph-based parameters used in the proposed method are described as follows. Firstly, the graph-based local clustering coefficient is a variation of clustering coefficient, which aims to express the internal connection relationships between a critical region and its neighbors. Secondly, the graph-based local importance is the global efficiency (described in Section II-A) for the local network of graph g divided the global efficiency for the local residual network of g. The local network of graph g is a network that contains the nodes of graph g and the neighbors of g, as well as the edges between these nodes. The local residual network of graph g is a network that removes the nodes and edges included by g from the local network of graph g. The graph-based local importance indicates the effect on the local connectivity of a brain network when a critical region is removed. Thirdly, the graph-based global importance is the global efficiency for the brain network divided the global efficiency for the global residual network of graph g. The global residual network of graph g is a network that removes the nodes and edges included by g from the brain network. The graph-based global importance indicates the effect on the global connectivity of a brain network when a critical region is removed.

The formula of graph-based local clustering coefficient is as follows:

$$C(BN,g) = \frac{1}{P(P-1)} \sum_{\substack{i,j \in S(g) \\ i \neq j}} (\hat{w}(g,i)\hat{w}(g,j)\hat{w}(i,j))^{1/3} \tag{3}$$

Where BN represents a brain network, and g represents a closed frequent graph. P is the number of neighbors of g, and $S(g)$ is the neighbors of the g in the brain network BN. $\hat{w}(g,i)$ and $\hat{w}(g,j)$ represent the minimum weight of the edge between graph g and the node i, graph g and the node j, respectively. $\hat{w}(i,j)$ represents the minimum weight of the edge between the node i and node j.

The formula of graph-based local importance is as follows:

$$LI(BN,g) = \frac{LR(LR-1)\sum_{\substack{i,j \in R(g) \\ i \neq j}} \frac{1}{d(i,j)}}{LN(LN-1)\sum_{\substack{i',j' \in \tilde{R}(g) \\ i' \neq j'}} \frac{1}{d(i',j')}} \tag{4}$$

Where BN and g have the same definition in Eq. (3). LN is the number of nodes of the local network of graph g. LR is the number of nodes of the local residual network of graph g. $R(g)$ is the nodes of the local network of graph g, and $\tilde{R}(g)$ is the nodes of the local residual network of graph g. $d(i,j)$ represents the shortest path length between node i and j and $d(i',j')$ represents the shortest path length between node i' and j'.

The formula of graph-based global importance is as follows:

$$GI(BN,g) = \frac{GR(GR-1)\sum_{\substack{i,j \in T(g) \\ i \neq j}} \frac{1}{d(i,j)}}{N(N-1)\sum_{\substack{i',j' \in \tilde{T}(g) \\ i' \neq j'}} \frac{1}{d(i',j')}} \tag{5}$$

Where BN and g have the same definition in Eq. (3). N is the number of nodes of the brain network BN. GR is the number of nodes of the global residual network of graph g. $T(g)$ is the nodes of the brain network BN and $\tilde{T}(g)$ is the nodes of the global residual network of graph g. $d(i,j)$ represents the shortest path length between node i and j, and $d(i',j')$ represents the shortest path length between node i' and j'.

The graph-based features are representations of the internal connection relationships and the local/global structural features of critical regions in the brain network. By using these features as complementary representations of the brain network, the proposed method could be easily integrated with the machine learning methods.

D. Implement for anomaly detection

Based on the features learned in the proposed method, the machine learning methods, including Random Forest(RF)[11] and Deep Neural Network(DNN)[12], are employed for anomaly detection in the brain network. The scikit-learn library[13] and tensorflow are used to efficiently implement RF and DNN. The detection model could be trained using the graph-based features or a combination of these features with other features. The parameters of the detection model are described in section IV-C.

IV. Evaluation

In this section, we conduct experiments to evaluate the performance of some basic parameters and the features learned from the proposed method by using different machine learning methods. First, we describe the dataset and metrics used for the evaluation. Then, experiments are conducted to evaluate the performance of the proposed method on different aspects.

A. Dataset

We conduct experiments on a real world fMRI dataset, which contains 297 samples, including 158 Healthy Controls(HC) and 139 Alzheimer's Disease(AD). The overview of demographic and clinical information of the samples are shown in Table I.

TABLE I
A BRIEF DESCRIPTION OF THE DATASET.

	AD	*HC*
Total	139	158
The number of male subjects	82	69
The number of female subjects	57	89
The average age of subjects	75.89	75.13

B. Performance metric

We use the detection rate(DR), false positive rate(FPR) and accuracy(ACC) to evaluate the performance of our model. The detection rate measures the proportion of actual positives that are correctly identified and the false positive rate measures the proportion of actual negatives that are incorrectly identified. The accuracy measures the proportion of samples that are correctly identified. For convenient expression for above metrics, we represent the number of true positives, false positives, true negatives and false negatives as TP, FP, TN, FN, respectively. The formulas of detection rate, false positive rate and accuracy are presented below:

$$Detection\ Rate = \frac{TP}{TP + FN} \tag{6}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \tag{7}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{8}$$

C. Results and discussion

Due to the limitation of the number of samples in the dataset, we used 5-fold cross validation in the experiments. The results presented are the average of the results of each fold.

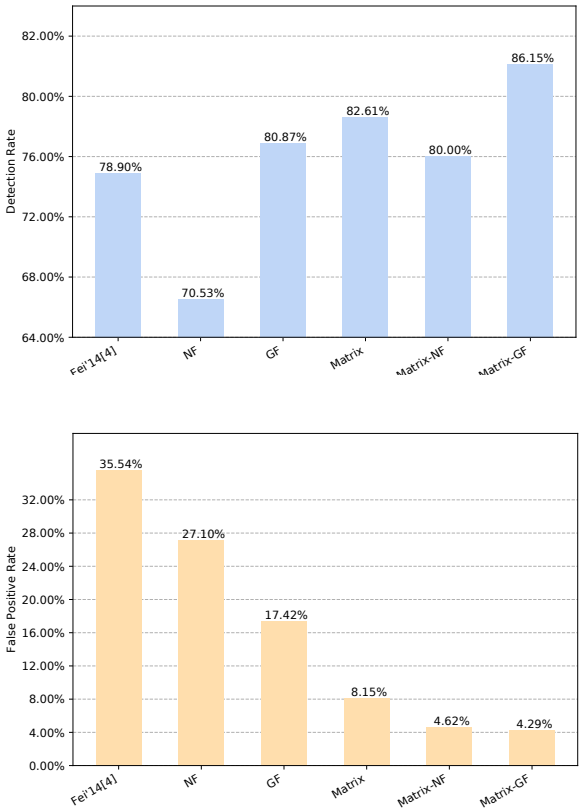
In the mining process of closed frequent graphs, the $minSup$ is set to 85%. The number of closed frequent graphs for group HC and group AD in each fold is shown in Table II. The number of closed frequent graphs in group HC is more than that in group AD, which indicates the connectivity relationships of the brain networks of group HC are closer than that of group AD.

TABLE II
THE RESULTS OF CLOSED FREQUENT GRAPH MINING FOR GROUP HC AND GROUP AD IN EACH FOLD.

Fold	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
group AD	286	156	227	275	138
group HC	626	348	489	606	265

In the experiments, we use different types of features to train models and compare the detection performance of these models to evaluate the effectiveness of these features. Three kinds of features are used in the experiments, including the original connectivity matrix, the node-based features and the features learned by the proposed method. The notations "Matrix", "NF", and "GF" indicate that the model is trained by using the original connectivity matrix, node-based features, and the proposed graph-based features, respectively. We also use their combinations, which are represented as "Matrix-NF" and "Matrix-GF". The node-based features are calculated by using each node in the brain network as the input of the formulas presented in section III-C. By comparing the proposed graph-based features with the node-based features mentioned above, it can be seen whether the critical regions better express the characteristics of the brain network than the nodes, thereby improving the detection performance of general machine learning models.

1) Experiment a: We implemented a DNN to test the performance of the models using different features during the training process. In the experiment, the DNN contains 5 fully connected layers. A dropout layer with the drop probability of 0.1 is connected to the last fully connected layer to avoid overfitting. The cross entropy is used as the loss function for optimization. The detection results of each method are shown in Figure 3(a)&(b).



(b) False Positive Rates

Figure 3. Comparison for the Detection Rate and False Positive Rate of six models.

As is shown in Figure 3(a)&(b), we compare the detection performance of six models, including the model proposed in [4] and five DNN models using different features, for training. Among these six models, the model using "Matrix-GF" features achieves the best detection performance. Compared with the model proposed in [4], the DR of the model using "Matrix-GF" features increases by 7.25%, while its FPR reduces by 31.25%. In addition, the models using "GF" features achieve better detection performance than those using "NF" features. The experiment results indicate that the DNN models using the proposed graph-based features could have a good detection performance compared with other models.

2) Experiment b: In the experiment, we compare the influence of graph-based features and node-based features on the detection performance of the RF model and the DNN

model. We implemented a RF model and a DNN model for comparison. The number of decision trees in the RF model is set to 200. The parameters setting for the DNN model has been introduced in experiment a.

TABLE III
THE COMPARISON OF THE PROPOSED GRAPH-BASED FEATURES AND THE NODE-BASED PARAMETERS.

	RF			*DNN*		
	DR	*FPR*	*ACC*	*DR*	*FPR*	*ACC*
NF	**77.04%**	36.13%	70.00%	70.53%	27.10%	72.00%
GF	68.89%	20.65%	74.48%	80.87%	17.42%	81.85%
Matrix-NF	67.27%	**8.00%**	77.93%	80.00%	4.62%	88.00%
Matrix-GF	71.03%	14.48%	**78.28%**	**86.15%**	**4.29%**	**91.11%**

As is shown in Table III, compared with the RF model using "NF" features, the ACC of the RF model using "GF" features increases by 4.48% and its FPR reduces by 15.48%; the DR of the RF model using "GF" features decreases by 8.15%, however, this decrease is less than the corresponding decrease of the FPR. The DNN model using "GF" features performs better than the DNN model using "NF" features. When a model is trained by using the combined features, i.e., "Matrix-NF" or "Matrix-GF", the detection performance of the model is improved. The model using "Matrix-GF" features has better performance than the model using "Matrix-NF" features. The experiment results indicate that the graph-based features could improve the detection performance of the RF model and the DNN model, compared with the node-based features.

V. CONCLUSION

This paper proposed a feature learning method to build effective representations for brain network, which could improve the detection ability of existing machine learning methods for the neuropsychiatric disorders. Based on closed frequent graph mining and graph-based feature learning, the features learned by this method expressed the internal connection relationships in the critical regions of brain network, as well as the global and local structural features related to the critical regions. The experiment results showed that the method can improve the ability of existing machine learning methods to detect neuropsychiatric disorders. In the future work, we will explore deep reinforcement learning[14, 15] and behavior analysis[16] to further improve the detection ability of the model.

REFERENCES

[1] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, "Economic networks: The new challenges," *science*, vol. 325, no. 5939, pp. 422–425, 2009.

[2] L. Zhan, Y. Liu, Y. Wang, J. Zhou, N. Jahanshad, J. Ye, and P. M. Thompson, "Boosting brain connectome classification accuracy in alzheimer's disease using higher-order singular value decomposition," *Frontiers in neuroscience*, vol. 9, p. 257, 2015.

[3] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of alzheimer's disease," *Alzheimer's & dementia*, vol. 3, no. 3, pp. 186–191, 2007.

[4] F. Fei, B. Jie, and D. Zhang, "Frequent and discriminative subnetwork mining for mild cognitive impairment classification," *Brain connectivity*, vol. 4, no. 5, pp. 347–360, 2014.

[5] X. Yan and J. Han, "Closegraph: mining closed frequent graph patterns," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 286–295.

[6] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, p. 440, 1998.

[7] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical review letters*, vol. 87, no. 19, p. 198701, 2001.

[8] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[9] Z. Cui, S. Zhong, P. Xu, G. Gong, and Y. He, "Panda: a pipeline toolbox for analyzing brain diffusion images," *Frontiers in human neuroscience*, vol. 7, p. 42, 2013.

[10] R. S. Desikan, F. Segonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman *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, 2006.

[11] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[14] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4005–4018, 2018.

[15] X. Chen, Z. Zhao, C. Wu, M. Bennis, H. Liu, Y. Ji, and H. Zhang, "Multi-tenant cross-slice resource orchestration: A deep reinforcement learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2377–2392, 2019.

[16] Z. Qin, Y. Wang, H. Cheng, Y. Zhou, Z. Sheng, and V. C. Leung, "Demographic information prediction: A portrait of smartphone application users," *IEEE Transactions on Emerging Topics in Computing*, vol. 6, no. 3, pp. 432–444, 2016.