

Classification of Mild Cognitive Impairment Based on a Combined High-Order Network and Graph Convolutional Network

XUEGANG SONG^{1,2}, AHMED ELAZAB^{1,2,3}, AND YUEXIN ZHANG^{4,5}

¹National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Shenzhen University, Shenzhen 518060, China

²Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Health Science Center, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China

³Computer Science Department, Misr Higher Institute of Commerce and Computers, Mansoura 35516, Egypt

⁴School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

⁵Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576

Corresponding author: Yuexin Zhang (smileyuexin@163.com)

This work was supported by the China Postdoctoral Science Foundation under Grant 2019M653014.

ABSTRACT Detection of early stages of Alzheimer's disease (AD) (i.e., mild cognitive impairment (MCI)) is important because it can delay or prevent progression to AD. The current researches of MCI classification are mainly based on static low-order functional connectivity network (FCN) and image information. However, static FCN cannot reflect time-varying dynamic behavior, low-order FCN overlooks inter-region interactions, and ignoring non-image information is not suitable especially when the size of dataset is small. In this paper, a method based on a combined high-order network and graph convolutional network (GCN) is proposed. The combined high-order network combines static, dynamic and high-level information to construct FCN while GCN is used to include non-image information to improve classifier's performance. Firstly, dynamic FCNs and static FCN are constructed by using a sliding window approach. Secondly, dynamic high-order FCNs and static high-order FCN based on the topographical similarity are then constructed. Thirdly, a novel combination method is proposed to utilize dynamic high-order FCNs and static high-order FCN to form a combined high-order FCN. Fourthly, features of the combined high-order FCNs are extracted by using a recursive feature elimination method. Lastly, after inputting extracted features into the GCN, in which MCI-graph establishes interactions between individuals and populations by using non-image information, the GCN outputs the binary classification results. Experimental results on Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (adni.loni.ucla.edu) show that our framework has good performance.

INDEX TERMS Mild cognitive impairment, binary classification, combined high-order network, graph convolutional network.

I. INTRODUCTION

Alzheimer's disease (AD), the most common type of dementia (accounting for 60%-80%), is a very serious and irreversible brain disorder [1], [2]. As the early stage of AD, mild cognitive impairment (MCI), has annual 10%-15% conversion rate and more than 50% conversion rate within 5 years turn to AD [3]. In MCI stages, with certain cognitive training and pharmacological treatment, the deterioration process can be delayed or stopped [4], [5]. Nevertheless, once

The associate editor coordinating the review of this manuscript and approving it for publication was Marcin Woźniak .

MCI has developed to AD, there are no effective therapies so far. Due to high conversion rate and severity of deterioration, the diagnosis and then classification of MCI are very important and therefore have drawn much attention during recent years [6]–[8].

The accurate diagnosis of MCI stages is a very challenging task due to its mild and subtle clinical symptoms [9]. It becomes even more challenging to differentiate between the early MCI (EMCI) and late MCI (LMCI). Current researches mainly focus on aided diagnosis or classification of MCI by using neuroimaging data. An effective neuroimaging modality is resting-state functional magnetic resonance

imaging (rs-fMRI). It is based on the blood-oxygenation-level-dependent (BOLD) signal that measures changes of blood flow velocity and oxygen levels caused by brain activities. The rs-fMRI has been widely used to classify MCI/AD in recent years [10]–[14]. Based on rs-fMRI, functional connectivity network (FCN), which describes the correlation of BOLD signals between different brain regions, can be measured and then used as main biomarker to classify MCI.

Modelling FCN is the first key step to guarantee classification performance. In literature, there are two popular FCN modeling methods, including Pearson's correlation coefficient (PCC) [15]–[17] and sparse representation [18]–[20]. PCC method is the first and most widely used algorithm due to its simplicity and computation efficiency. In contrast, sparse representation method is more complicated and has more computational cost, however, it can construct networks with better discriminability. The two modeling methods are both for the whole time series to reveal stationary nature. Recently it is showed in a typical resting-state fMRI experiment that the connectivity between different brain regions exhibits meaningful variations on top of correlational patterns of spontaneous fMRI signal fluctuations [21], [22]. As a consequence, many studies have been devoted to assess and characterize dynamic FCN (dFCN) in many related fields [23]–[27]. Static FCN (sFCN) aims to reveal global stationary nature and dFCN aims to real local dynamic nature. For MCI classification in existing research, the combinations of sFCN and dFCN is ignored. In this view, we propose to combine dynamic and static FCNs based on PCC to classify MCI stages.

Apart from FCN modeling methods, the classifier is also a key step to determine classification performance. The widely used traditional classifiers include random forest (RF), gradient boosting decision tree (GBDT), multi-layer perceptron classifier (MLP), and support vector machine (SVM). These classifiers treat each feature individually and ignore the relationship between individuals and populations. In recent years, the graph networks have received extensive attention as they can provide a powerful and intuitive way of establishing relationship between individuals and populations. For a graph, vertices represent features of every individual's image and edges are used to capture the similarities between each pair of vertices. The earliest work on graph neural networks (GNN) was proposed by Scarselli in 2009 [28], which extends existing neural network methods and maps graphs into Euclidean space for processing data. This work is considered the foundation for the development of graph networks. Afterwards, graph networks get extensive studies [29]–[33], particularly the graph convolutional network (GCN) was proposed by Kipf in 2017 [34]. The GCN mainly applies the convolution of Fourier transform and Taylor's expansion formula into graph networks to achieve good filtering performance. As its excellent performance, GCN has received popular recognition and has been successfully used to perform classification [35]–[37]. In general, the advantage of graph-based models is that, they can contain pairwise interactions and

integrate both image and non-image information. Based on the fact that the image information is acquired at different equipment, diverse imaging protocols and scanners are used. Therefore, the image information shows a better comparable relationship between features acquired from same type of equipment. At the same time, gender also provides a valuable information as different gender can result in differences in images. In this paper, we propose to use GCN to include gender and equipment type information to improve the performance of classifier.

In view of above status, a method based on a combined high-order network and GCN is proposed. In the proposed method, a combined high-order network is used to include static and dynamic high-level information. Non-image information is included in the GCN to establish interactions between individuals and populations to improve the performance of classifier. Overall, the main contributions of this paper are as follows:

- 1) Propose a novel combined high-order FCN that combines static, dynamic and high-level characteristic to construct FCN for MCI classification.
- 2) A MCI-graph construction strategy is used to include gender and equipment type information to better establish the relationship between individuals and populations.
- 3) Perform extensive experiments on relatively big dataset from ADNI dataset (rs-fMRI) including 184 subjects. Our experiments show that the proposed method based on combined high-order network and GCN leads to significant improvement of the classification performance superior to the-state-of-the-art methods.

The reminder of this paper is organized as follows: The proposed methodology is introduced in details in Section II. In Section III, experiments and results are given with comparisons to other competing methods. Discussion and performance analysis are given in Section IV. Finally, conclusions are summarized in Section V.

II. METHODOLOGY

The method proposed in this paper is mainly based on two points. First, by constructing a combined high-order FCN, we can include global static nature and local dynamic nature. Second, by establishing the MCI-graph using non-image information, the extracted features can be corrected and the performance of classifier can be improved. In addition, data preprocessing and feature extraction method are also included in whole process. The overview of the proposed method is shown in Fig. 1. In this figure, image information of every subject is represented by the combined high-order FCN which is described in subsection B.

A. DATA PREPROCESSING

The GREYNA toolbox [38] is used in this work to preprocess the rs-fMRI data. Our preprocessing follows the following steps. 1) The first 10 acquired rs-fMRI volumes of each subject are discarded, then the remaining 170 volumes are corrected by matching all time points to intermediate

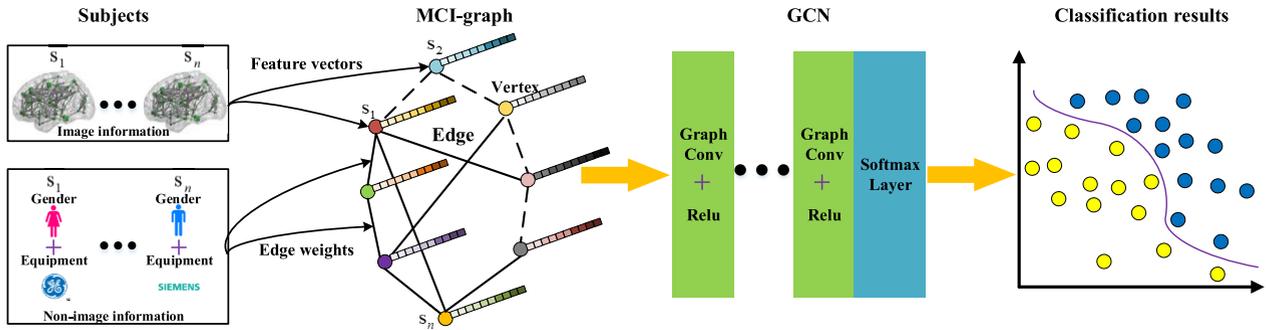


FIGURE 1. Overview of the proposed method for MCI classification. Image information of every subject is represented by the combined high-order FCN.

time points. 2) Apply head movement correction to remove the head-motion artifacts in rs-fMRI time-series. 3) Perform spatial normalization with DARTEL [39] to normalize volumes to the MNI atlas space, then resample the voxel size to $3 \times 3 \times 3$ mm. 4) Perform smooth filtering by employing the Gaussian kernel with 4 mm full width half maximum to remove low-frequency drift and high-frequency noise. 5) Regress regional mean time series to eliminate the influence of white matter signal, cerebrospinal fluid signal, and head movement signal. 6) Use the automated anatomical labelling (AAL) [40] to segment brain space into 90 region of interests (ROIs). Finally, we obtain the time series of 90 ROIs for each individual.

B. COMBINED HIGH-ORDER NETWORK CONSTRUCTION

Traditional low-order FCN overlooks inter-region interactions and dynamic characteristics, which limits its application for brain disease classification. In this paper, we combine dynamic and static high-order networks. As traditional static network is an extreme case where window length is maximized to the entire time scale, only dynamic high-order network is described in this subsection. A sliding window correlation method [41], [42] is employed, which uses a certain size window to intercept signal along the time direction and finally a series of functional connections results varying in time are obtained. These functional connections construct dynamic high-order FCNs. The detailed description is described as follows:

Firstly, the entire rs-fMRI time series are segmented into multiple sub-series by using a sliding window approach [43]. Given a matrix $X = \{x_1, x_2, \dots, x_N\} \in \mathbb{R}^{1 \times N}$, where N denotes the number of subjects, $x = [x(1), x(2), \dots, x(R)] \in \mathbb{R}^{M \times R}$, R denotes the total number of ROIs, and M denotes the length of rs-fMRI time series. Then we can get sub-series with a number $D = [(M - L)/s] + 1$, where M denotes the length of time series, L denotes the length of the sliding window, s denotes the size of moving step.

Secondly, PCC is used to construct the functional connectivity between each ROI pairs. For every sub-series, we can calculate PCC by using the formula: $c_{ij}^{(d)} = corr \{x^{(d)}(i), x^{(d)}(j)\}$, where $corr\{\cdot\}$ is the pairwise

correlation coefficient between region $x^{(d)}(i)$ and region $x^{(d)}(j)$ of the d -th sub-series. By calculating PCC, we finally get symmetric matrix $c^{(d)} = [c_{ij}^{(d)}] \in \mathbb{R}^{R \times R}$, and a series of temporal FCNs matrix $C^{(d)} = \{c_1^{(d)}, c_2^{(d)}, \dots, c_N^{(d)}\} \in \mathbb{R}^{1 \times N}$.

Thirdly, in order to reveal high-level information, the dynamic high-order FCNs are constructed, which are calculated as: $h_n^{(d)} = (c_n^{(d)})^T c_n^{(d)}$, where $h_n^{(d)}$ denotes high-order dFCNs of the d -th sub-networks in the n -th subjects. The traditional high-level sFCN is an extreme case where window length is maximized to the entire time scale.

Finally, by combining these high-order dFCNs $\{h_1^{(d)}, h_2^{(d)}, \dots, h_n^{(d)}\}$ and a high-order sFCN, we can get one final combined high-order FCN. In this paper, the combined high-order FCN is computed by weighted averaging high-order dFCNs and high-order sFCN, which is calculated as below:

$$FCN = a \times \sum_{i=1}^{i=N} dFCN_i \div N + (1 - a) \times sFCN \quad (1)$$

where a is a weight coefficient, and its value is set between 0 and 1. N is the number of dFCNs. Fig. 2 illustrates the process of dynamic high-order networks construction based on the PCC of rs-fMRI data.

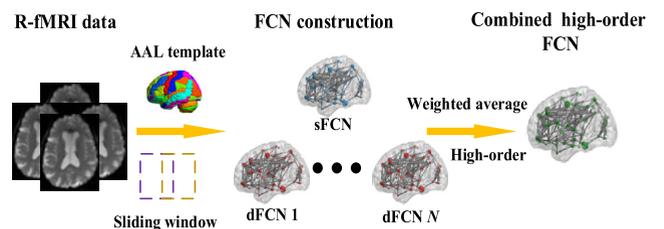


FIGURE 2. Overview of combined high-order FCN construction.

C. GRAPH CONVOLUTIONAL NETWORK CLASSIFIER

1) FEATURE EXTRACTION AND SELECTION

After constructing combined high-order FCN, we employ a recursive feature elimination (RFE) method for feature extraction. The RFE method can recursively remove attributes and build the model using the remaining

attributes [44], [45]. In details, given an external estimator (ridge classifier in this paper) that assigns weights to features, the goal of RFE is to select features by recursively considering smaller and smaller sets of features. First, the estimator is trained on the initial set of features and the importance of each feature is obtained. Then, the least important features are pruned from current set of features. That procedure is recursively repeated on the pruned set until the desired number of features to be selected is eventually reached. In this paper, the RFE function in (scikit-learn, python library) is used.

As demonstrated in [36], RFE is the best method to extract features in the classification of MCI compared to the other feature extraction algorithm (including principle component analysis (PCA), MLP and autoencoder). Therefore, RFE is adopted to extract features in the paper. For RFE, it is not known in advance how many features are sufficient and valid to distinguish diseases. In this work we investigate the effect of different number of features on the performance in Discussion Section.

2) MCI-GRAPH CONSTRUCTION

The MCI-graph is the key process for GCN, which establishes the interactions between individuals and populations. Reasonable construction of graph structure can correct features in process of features input and features dissemination. If the construction of MCI-graph is inaccurate, the performance can be worse compared to a simple classifier. The graph includes graph edges which include non-image information and graph vertices which represent features of every individual's image information. The definition of graph edges determines the performance of MCI-graph as graph edges play the role to filter features by using its neighbors' features instead of treating individuals' features individually. The image information is acquired at different equipment, diverse imaging protocols and scanners are used. As a result, the image information will have a better comparable relationship between features acquired from same type equipment. Beside equipment, gender is also a valuable information as different genders will result in some differences in images. In this subsection, the non-image phenotypic information, gender and acquisition equipment, is considered in MCI-graph construction.

Thus, the population graph's adjacency matrix W is defined as follows:

$$W(v, w) = \text{Sim}(A_v, A_w) \times (r_G(M_h(v), M_h(w)) + r_E(M_h(v), M_h(w))) \quad (2)$$

where $\text{Sim}(A_v, A_w)$ represents image-features (extracted features from RFE) similarity between individual v and individual w , which means high similarity will be assigned a big weight. M_h represents non-image phenotypic information. r represents distance between phenotypic features, r_G represents distance of gender, and r_E represents distance of acquisition equipment. r is the key to construct MCI-graph that directly controls the performance. In this work, we define

r as a unit-step function as follows:

$$r(M_h(v), M_h(w)) = \begin{cases} 1, & \text{if same} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

In the Eq. 3, if the individual v and individual w have same phenotypic information, r will be set 1, otherwise r will be set 0. For example, if individual v and individual w have same gender information, $r_G(M_h(v), M_h(w)) = 1$. And if they have same acquisition equipment type, $r_E(M_h(v), M_h(w)) = 1$. If they have different gender and equipment type information, $r_G(M_h(v), M_h(w)) = 0$ and $r_E(M_h(v), M_h(w)) = 0$.

The similarity measure is defined as:

$$\text{Sim}(A_v, A_w) = \exp\left(-\frac{[\rho(x(v), x(w))]^2}{2\sigma^2}\right) \quad (4)$$

where ρ represents the correlation distance, σ represents the width of the kernel. This formula is based on the fact that subjects belonging to same classes will show much more similarity in features than subjects belonging to different classes [36].

3) GCN STRUCTURE

The model architecture of GCN part is illustrated in Fig. 1. The proposed model consists of two graph convolution layers activated by rectified linear unit (ReLU) function and one SoftMax output layer. The strategy of GCN construction is studied in Discussion Section. In training process, whole MCI-graph is inputted and a subset of graph nodes are labelled in training set. In testing process, training set with labelled graph nodes and validation set with unlabeled graph nodes are both inputted into the trained model. Labelled training set and unlabeled validation set interact with each other in the process of features input and features propagation of GCN, and this makes the process become a semi-supervised classification scheme. A cross entropy loss function is used to evaluate the performance of recognition, and a nested leave-one-out cross-validation method is used in the testing process.

III. EXPERIMENTS AND RESULTS

Our model is evaluated on the ADNI database (adni.loni.ucla.edu) by using a nested leave-one-out cross-validation method. The total number of subjects getting from ADNI database is 184, which includes 40 LMCI patients, 77 EMCI patients, and 67 normal control (NC) as shown in Table 1. Scanning equipment includes GE, SIEMENS and PHILIPS. GCN parameters that are used in experiments are as follows: dropout rate is 0.1, l_2 regularization is 5×10^{-4} , learning rate is 0.002, number of epochs is 1000, 2 GCN layers, number of neurons per layer is 128 and number of extracted features is 100. Gender and acquisition equipment are used to build edges, and the default polynomial order is set to $K = 3$. Classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC) are used as evaluation metrics.

TABLE 1. Demographic information of the subjects used in our study.

Group	LMCI (40)	EMCI (77)	NC (67)
Male/Female	15M/25F	38M/39F	39M/28F
Age (mean±SD)	76.0±7.6	74.8±6.1	76.5±4.5
GE/SIEMENS/PHILIPS	7/7/26	24/19/34	20/18/29

We divide this section into 4 parts to describe the performance of the proposed algorithm. Firstly, we validate the performance of static and dynamic combination strategy. In this subsection, we study the effect of weighted coefficient a in Eq. 1. Secondly, we study the influence of sliding window’s parameters. Thirdly, we study the impact of non-image information, including gender and acquisition equipment. Lastly, we compare the performance of the proposed method with other different well-established classifiers to validate its outstanding performance.

A. INFLUENCE OF STATIC AND DYNAMIC COMBINATION STRATEGY

Generally, sFCN reveals global stationary nature whereas dynamic FCN describes local dynamic information. In this paper, we propose to combine global stationary nature and local dynamic information by a weighted averaging algorithm to better construct FCN for MCI classification. In this view, we test the performance of the proposed FCN construction strategy, and ACC, SEN, SPE, and AUC are shown in Table 2. The performance comparison of the proposed method with different weighted coefficient a is shown in Fig. 3, and the ROC curves are shown in Fig. 4. In this subsection, GCN which includes acquisition equipment and gender information is used to output the binary classification results. Weighted coefficient a in Eq. 1 is set from 0 to 1 and it represents the proportion of high-order dFCNs.

As shown in Table 2, compared with sFCN construction strategy ($a = 0$), combining dFCN and sFCN can improve

TABLE 2. Influence of static and dynamic combination strategy on the performance of the proposed method.

Task	weight (a)	ACC (%)	SEN (%)	SPE (%)	AUC (%)
EMCI vs. NC	0.0	65.5	49.2	85.4	85.0
	0.1	80.3	73.1	89.0	88.5
	0.2	74.5	62.6	89.0	88.7
	0.3	76.2	65.6	89.0	88.8
	0.4	77.8	68.6	89.0	88.3
	0.5	76.2	68.6	85.4	84.5
	0.6	81.1	74.6	89.0	87.1
	0.7	71.3	56.7	89.0	88.4
	0.8	81.1	74.6	89.0	88.9
	0.9	82.7	77.6	89.0	88.5
1.0	77.8	68.6	89.0	88.9	
LMCI vs. NC	0.0	87.8	85.0	92.5	93.9
	0.1	88.7	83.5	97.5	97.5
	0.2	83.1	76.1	95.0	96.1
	0.3	87.8	85.0	92.5	93.8
	0.4	81.3	74.6	92.5	94.5
	0.5	70.0	62.6	82.5	86.1
	0.6	87.8	85.0	92.5	92.8
	0.7	81.3	76.1	90.0	91.8
	0.8	82.2	76.1	92.5	93.1
	0.9	79.4	73.1	90.0	91.4
1.0	78.5	71.6	90.0	91.0	
LMCI vs. EMCI	0.0	83.7	81.8	87.5	89.5
	0.1	82.9	79.2	90.0	89.7
	0.2	76.9	72.7	85.0	87.4
	0.3	80.3	72.7	95.0	94.0
	0.4	79.4	76.6	85.0	84.9
	0.5	79.4	74.0	90.0	90.0
	0.6	80.3	71.4	97.5	95.5
	0.7	80.3	72.7	95.0	93.4
	0.8	82.9	79.2	90.0	88.6
	0.9	86.3	83.1	92.5	93.0
1.0	78.6	75.3	85.0	84.4	

performance for all three classification tasks. For EMCI vs. NC, the performance reaches the best with a set as 0.9. For LMCI vs. NC, the performance reaches the best with a set

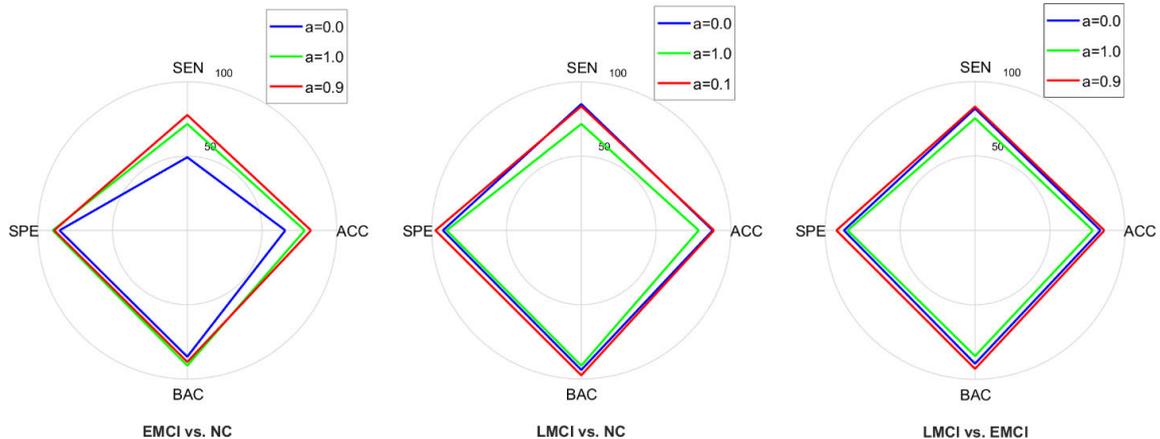


FIGURE 3. In three classification problems, the performance comparison of the proposed method with different weighted coefficient a . In every classification task, the red line represents the performance of the method with optimal weighted coefficient.

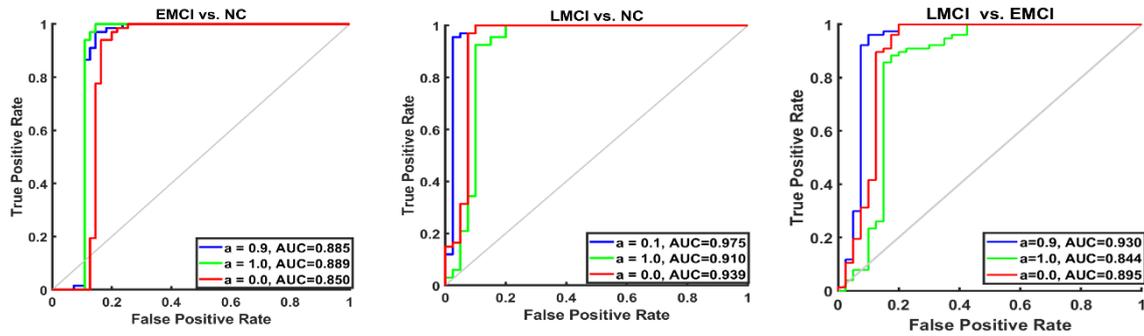


FIGURE 4. ROC curves of the proposed method with different weighted coefficient a for different classification tasks.

as 0.1. For LMCI vs. EMCI, the performance reaches the best with a set as 0.9.

In general, by using the proposed method, ACC can get improved 17.2% for EMCI vs. NC, 0.9% for EMCI vs. NC, and 2.6% for EMCI vs. LMCI compared with that of sFCN construction strategy. Besides ACC, SEN, SPE and AUC also get improved. Therefore, we will use the best combination strategy ($a = 0.9$ for EMCI vs. NC and EMCI vs. LMCI, $a = 0.1$ for LMCI vs. NC) in the following sections.

In the process of experiment, we change the sample amount distribution among different categories, and the results show that the optimal combination strategy has a little fluctuation. For our dataset, the combination strategy ($a = 0.9$ for EMCI vs. NC and EMCI vs. LMCI, $a = 0.1$ for LMCI vs. NC) reaches the optimal performance. We get the optimal combination strategy by testing all weighted coefficient a , but the dynamic information how to affect performance is unknown. In future works, we will study which and how dynamic information affect efficiency. In addition, we will collect more datasets for better training and validation.

B. INFLUENCE OF THE SLIDING WINDOW'S PARAMETERS

The parameters of sliding window including step size s and window size L are very important, which can affect the performance of the proposed method. In this subsection, we test the performance with setting $s = 5, 10, 15$ and $L = 10, 20, 30, 40, 50, 60, 70, 80$. The experimental results are summarized in Table 3, Table 4 and Table 5.

As shown in Table 3, Table 4 and Table 5, the proposed method reaches the best performance with setting window size $L = 30$ and step size $s = 5$ for all three classification tasks. Thus in other experiments in this paper, the window size L and step size s are set to 30 and 5, respectively.

C. INFLUENCE OF THE NON-IMAGE INFORMATION

The MCI-graph is the main component of graph networks, which establishes the interactions between individuals and populations. Graph edges include non-image information and graph vertices represent features of every individual's image information. As different types of acquisition equipment and different gender cause differences in image quality,

TABLE 3. Influence of window size L with step size $s = 5$.

Task	L	ACC (%)	SEN (%)	SPE (%)	AUC (%)
EMCI vs. NC	10	78.6	70.1	89.0	89.6
	20	85.2	82.0	89.0	89.6
	30	82.7	77.6	89.0	88.5
	40	74.5	62.6	89.0	87.4
	50	79.5	73.1	87.2	88.2
	60	67.2	49.2	89.0	86.7
	70	72.9	61.1	87.2	85.5
	80	77.0	71.6	83.6	82.6
LMCI vs. NC	10	84.1	79.1	92.5	94.9
	20	87.8	85.0	92.5	94.1
	30	88.7	83.5	97.5	97.5
	40	81.3	74.6	92.5	93.8
	50	90.6	88.0	95.0	96.0
	60	82.2	76.1	92.5	93.2
	70	87.8	85.0	92.5	93.9
	80	81.3	74.6	92.5	93.9
LMCI vs. EMCI	10	71.7	64.9	85.0	75.1
	20	72.6	63.6	90.0	87.5
	30	86.3	83.1	92.5	93.0
	40	85.4	81.8	92.5	94.8
	50	72.6	59.7	97.5	95.1
	60	76.9	68.8	92.5	93.5
	70	70.9	64.9	82.5	83.6
	80	78.6	72.7	90.0	90.3

the non-image information is considered as given in Eq. 2. In this subsection, the influence of acquisition equipment and gender information is evaluated.

The results of the test are shown in Table 6. The performance comparison of the proposed method with different non-image information is shown in Fig. 5, and the corresponding ROC curves are shown in Fig. 6. In Table 6, 'S' represents $Sim(A_v, A_w)$, 'G' represents $r_G(M_h(v), M_h(w))$, and 'E' represents $r_E(M_h(v), M_h(w))$, respectively in Eq. 2. Also, 'SG' means that $Sim(A_v, A_w)$ and $r_G(M_h(v), M_h(w))$ are included in Eq. 2. 'SGE' means that $Sim(A_v, A_w)$, $r_G(M_h(v), M_h(w))$ and $r_E(M_h(v), M_h(w))$ are included in Eq. 2.

As shown in Table 6, for all three classification tasks without applying gender and acquisition equipment information

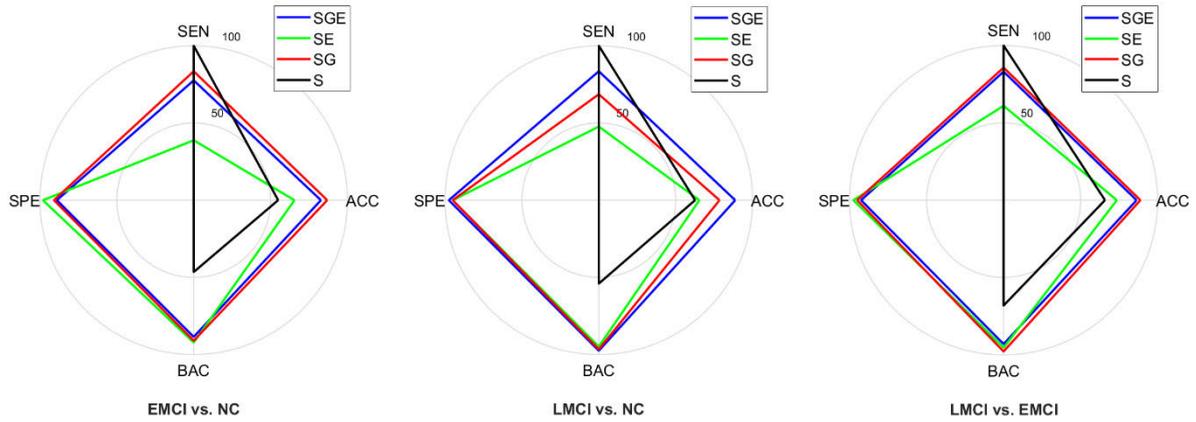


FIGURE 5. In three classification problems, the performance comparison of the proposed method with different non-image information.

TABLE 4. Influence of window size L with step size s = 10.

Task	L	ACC (%)	SEN (%)	SPE (%)	AUC (%)
EMCI vs. NC	10	84.4	80.5	89.0	88.9
	20	81.9	74.6	90.9	90.2
	30	75.4	64.1	89.0	88.3
	40	75.4	64.1	89.0	88.4
	50	75.4	64.1	89.0	88.5
	60	79.5	74.6	85.4	85.1
	70	79.5	71.6	89.0	87.3
	80	77.0	67.1	89.0	87.5
LMCI vs. NC	10	81.3	77.6	87.5	88.8
	20	80.3	74.6	90.0	92.9
	30	82.2	76.1	92.5	94.0
	40	76.6	65.6	95.0	96.6
	50	88.7	83.5	97.5	97.2
	60	80.3	74.6	90.0	91.4
	70	82.2	77.6	90.0	92.4
	80	85.0	79.1	95.0	95.9
LMCI vs. EMCI	10	88.0	88.3	87.5	89.1
	20	76.0	70.1	87.5	86.7
	30	72.6	64.9	87.5	87.0
	40	76.9	70.1	90.0	90.0
	50	74.3	63.6	95.0	95.6
	60	80.3	75.3	90.0	90.3
	70	90.5	88.3	95.0	95.1
	80	78.6	72.7	90.0	89.9

TABLE 5. Influence of window size L with step size s = 15.

Task	L	ACC (%)	SEN (%)	SPE (%)	AUC (%)
EMCI vs. NC	10	82.7	77.6	89.0	88.7
	20	80.3	71.6	90.9	89.9
	30	81.9	76.1	89.0	88.9
	40	75.4	64.1	89.0	87.3
	50	78.6	74.6	83.6	84.1
	60	81.1	77.6	85.4	84.7
	70	81.9	77.6	87.2	86.2
	80	70.4	55.2	89.0	88.7
LMCI vs. NC	10	86.9	85.0	90.0	92.5
	20	88.7	88.0	90.0	92.7
	30	85.0	80.5	92.5	93.1
	40	87.8	83.5	95.0	95.2
	50	85.0	79.1	95.0	96.0
	60	86.9	83.5	92.5	93.1
	70	82.2	74.6	95.0	96.0
	80	77.5	68.6	92.5	92.1
LMCI vs. EMCI	10	77.7	74.0	85.0	85.9
	20	78.6	70.1	95.0	91.0
	30	79.4	71.4	95.0	92.6
	40	82.0	79.2	87.5	88.8
	50	84.6	80.5	92.5	92.3
	60	72.6	64.9	87.5	85.5
	70	66.6	57.1	85.0	85.4
	80	76.0	68.8	90.0	91.4

in MCI-graph construction, the performance is very low, especially SPE is 0. The performance of applying only gender information are better than that applying only equipment information, and this shows that gender has a greater impact than equipment on performance.

In general, gender and acquisition equipment are significant non-image information for MCI classification. Including the non-image information, we can construct better relationship between individuals and populations, and eventually improve performance of MCI classification.

D. CLASSIFIERS COMPARISON

We further compare results of the proposed method with other different well-established classifiers. As GCN has been demonstrated to have excellent performance for MCI classification, methods in papers [36], [37] did not consider combination strategy of sFCN and dFCN which is investigated by comparisons in this subsection. Beside GCN, SVM, RF, GBDT and MLP classifiers are also included. For fair comparisons, the cross-entropy loss function is employed as the loss function for all classifiers in this subsection. For the MLP

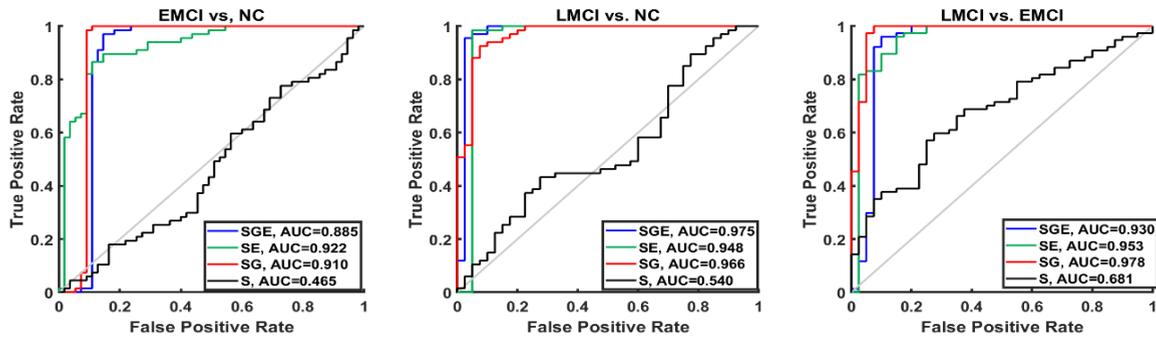


FIGURE 6. ROC curves of the proposed method with different non-image information for different classification tasks.

TABLE 6. Influence of gender and acquisition equipment information on the performance of the proposed method.

Task	Strategy	ACC (%)	SEN (%)	SPE (%)	AUC (%)
EMCI vs. NC	SGE	82.7	77.6	89.0	88.5
	SE	65.5	38.8	98.1	92.2
	SG	86.8	83.5	90.9	91.0
	S	54.9	100	0	46.5
LMCI vs. NC	SGE	88.7	83.5	97.5	97.5
	SE	65.4	47.7	95.0	94.8
	SG	78.5	68.6	95.0	96.6
	S	62.6	100	0	54.0
LMCI vs. EMCI	SGE	86.3	83.1	92.5	93.0
	SE	73.5	61.0	97.5	95.3
	SG	88.8	85.7	95.0	97.8
	S	65.8	100	0	68.1

method, the parameters are set to the same settings with the GCN implementation, including number of epochs, number of features, number of hidden layers, dropout, learning rate, seed and regularization. For SVM, penalty parameter is 10, kernel is 'linear', and gamma is 0.1. For RF, number of trees in the forest is 1000, maximum depth of the tree is 3, and random seed is 0. For GBDT, learning rate is 0.01, number of epochs is 200, and maximum depth of the tree is 5. The comparison results of different classifiers (Ours, GCN, SVM, RF, GBDT and MLP) are shown in Table 7. Fig. 7 provides the histograms and Fig. 8 provides the ROC curves of these methods. The comparison results of the proposed method with other methods are shown in Table 8.

Apart from the proposed method, GCN has better performance than the other methods which is shown in Table 7. The excellent performance of GCN demonstrates the effectiveness of non-image information whereas other methods do not include non-image information. Establishing the interactions between individual elements to correct features, which use neighbors' features based on weighted average method instead of treating features individually, can improve classification performance significantly. Based on the fact that the connectivity between different brain regions exhibits meaningful variations on top of correlational patterns of spontaneous fMRI signal fluctuations, we propose our combined

TABLE 7. Experimental results of different classifiers and their effect on the performance.

Task	Method	ACC (%)	SEN (%)	SPE (%)	AUC (%)
EMCI vs. NC	Ours	82.7	77.6	89.0	88.5
	GCN	65.5	49.2	85.4	85.0
	RF	51.7	44.3	74.8	55.5
	GBDT	55.7	42.0	67.4	53.8
	SVM	58.1	49.7	71.0	64.6
	MLP	57.4	46.4	68.5	54.8
LMCI vs. NC	Ours	88.7	83.5	97.5	97.5
	GCN	87.8	85.0	92.5	93.9
	RF	66.2	49.2	81.7	60.3
	GBDT	61.1	42.8	80.2	50.3
	SVM	56.0	38.5	78.2	61.0
	MLP	59.0	40.2	80.2	54.5
LMCI vs. EMCI	Ours	86.3	83.1	92.5	93.0
	GCN	83.7	81.8	87.5	89.5
	RF	66.3	54.5	77.7	56.0
	GBDT	52.7	42.5	62.0	53.5
	SVM	60.3	54.1	78.0	59.0
	MLP	63.2	55.4	73.4	56.8

high-order network that includes static and dynamic information to construct FCN. Compared with GCN, the proposed method can achieve better performance with ACC and AUC improved by 17.2% and 3.5% for EMCI vs. NC, 0.9%, and 3.6% for LMCI vs. NC, 2.6% and 3.5% for LMCI vs. EMCI. These results prove that the combined high-order network, which includes static and dynamic high-order information, can better construct FCN and eventually result in significant performance improvement. Table 8 shows the comparison of the proposed method in this paper with other methods in corresponding papers. We can observe that our proposed method achieves quite promising performance for MCI classification. By using core-i7-8700K and PyCharm, the average running time of our method for one classification task is about 12.5 minutes, average running time of GCN method and MLP method is about 11.2 minutes and average running time of other methods is about 3 minutes. The main running time of our method, GCN method and MLP method is used to train system as the optimal number of epochs is 1000 whereas optimal number of epochs of other methods are about 200.

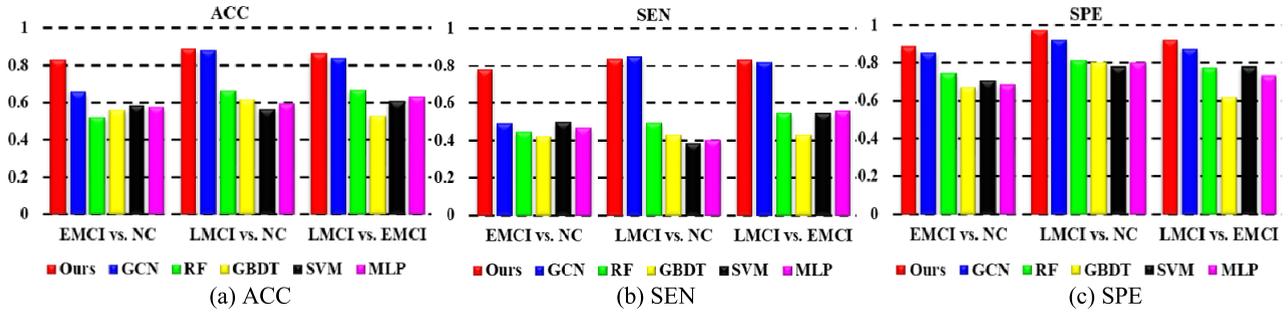


FIGURE 7. The comparison of different classifiers on the performances for different classification tasks.

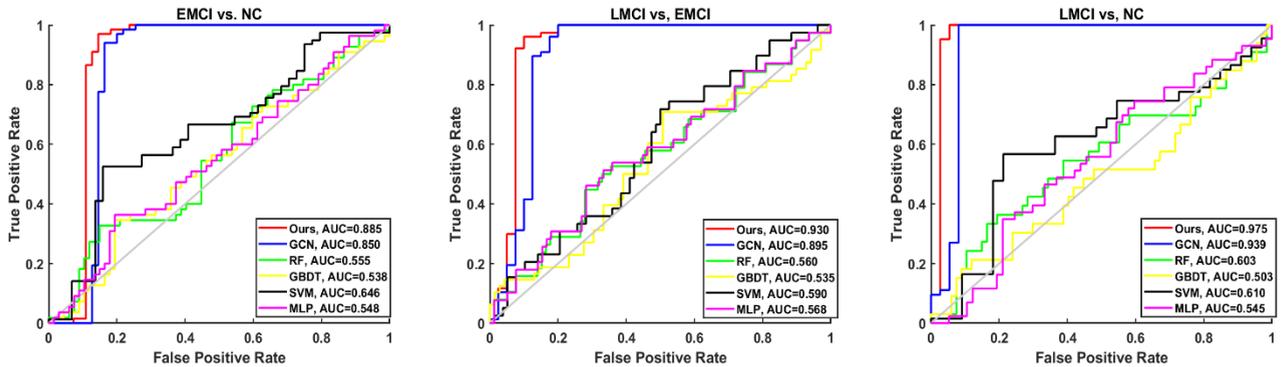


FIGURE 8. ROC curves of different classifiers for different classification tasks.

TABLE 8. Performance evaluation of proposed method against other competing methods.

Study	Subjects	Methods, # ROIs	Tasks	ACC (%)	SEN (%)	SPE (%)
Ours	40 LMCI, 77 EMCI, 67 NC	Combined high-order Network and GCN, 90	EMCI vs. NC	82.7	77.6	89.0
			LMCI vs. NC	88.7	83.5	97.5
			LMCI vs. EMCI	86.3	83.1	92.5
Wee et al. [46]	29 EMCI, 30 NC	Fused multiple graphical lasso, 90	EMCI vs. NC	79.6	75.8	70.0
Yu et al. [47]	50 MCI, 49 NC	Weighted Sparse Group Representation, 90	MCI vs. NC	84.8	91.2	78.5
Guo et al. [18]	33 EMCI, 32 LMCI, 28 NC	Multiple Features of Hyper-Network, 90	EMCI vs. NC	72.8	78.2	67.1
			LMCI vs. NC	78.6	82.5	72.1
Li et al. [48]	36 MCI, 37 NC	Ultra-Group Constrained Orthogonal Forward Regression, 90	MCI vs. NC	80.8	80.5	81.0

Compared with GCN method and MLP method, the running time of our method is acceptable.

In this section, we studied the performance of static and dynamic combination strategy, studied the influence of sliding window’s parameters, studied the impact of non-image information and compared the performance of the proposed method with other different well-established classifiers. The results show the following conclusions: Combing dFCN and sFCN can improve performance for all three classification tasks. Optimal window size L is 30 and optimal step size s is 5 for all classification tasks. Including gender and acquisition equipment information can improve the performance of classification. Our proposed method achieves quite promising performance for MCI classification. In next section, we further analysis and discuss other parameters.

IV. DISCUSSION

In this paper, a novel method based on a combined high-order network (combining static and dynamic high-order networks) and GCN (including non-image information) is proposed. For FCN construction, the whole rs-fMRI series are segmented by a sliding window approach to generate a certain number of rs-fMRI sub-series. Through a number of experiments, we find the best performance is achieved when window size and sliding size are set to 30 and 5, respectively. These rs-fMRI sub-series are then used to generate a series of dynamic high-order FCNs, which aims to explore dynamic characteristics and reveal high-level information. To take the advantages of both static and dynamic information, dynamic high-order FCNs and static high-order FCN are combined by weighted average method. In this paper, weighting coefficient

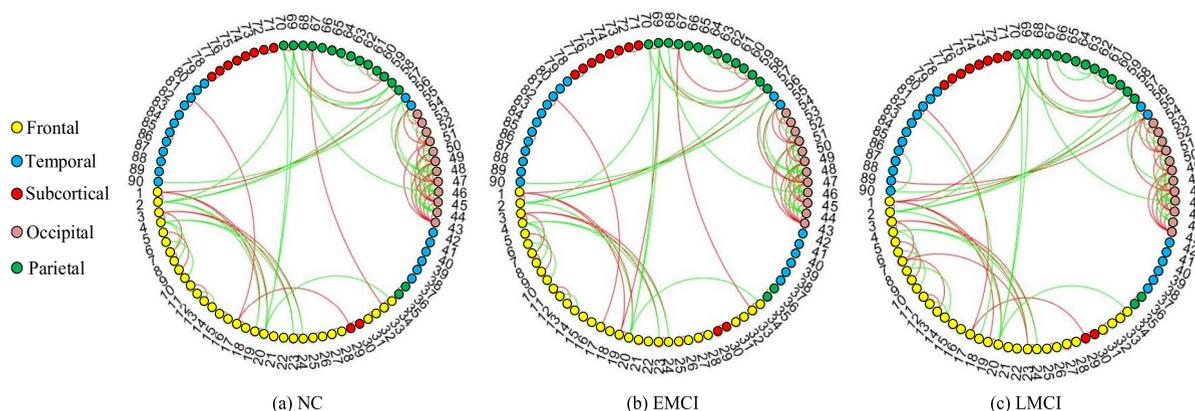


FIGURE 9. Brain FCNs of the proposed method.

is set as a constant value according to experimental results. For the three classification tasks, the performance of these four evaluation metrics (ACC, SEN, SPE, and AUC) can get improvement by combining static and dynamic information.

Extensive experiments show that there are differences among individuals' static brain FCNs whereas almost no difference between individuals' dynamic brain FCNs, and combined brain FCNs have better stability. In Fig. 5, combined brain FCNs with 100 top features, which are constructed according to all subjects' top features, are shown. And Fig. 9 shows that the top 100 features of different stages of MCI are consistent.

Representative features of constructed FCN are then extracted by using RFE method. Since sample size is limited, number of features need to be set to a reasonable value as excessive number will increase the burden on system and insufficient number is not abundant to represent the individual's information. We test the influence of number of extracted features through experiments. The performance of less than 100 features for classification of EMCI vs. NC is not satisfactory, and performance increases along with the number of extracted features increases. Once reaching about 100 features, the performance becomes stable with the number of extracted features varies from 100 to 1600. More than 1600 features also yield low performance. For the classification of LMCI vs. NC, the performance of less than 50 features increases along with the number of extracted features increases, and the performance maintains stable with the number of extracted features varies from 50 to 600. Eventually more than 600 features will result in performance deterioration. Finally, for the classification of LMCI vs. EMCI, the performance is stable with the number of extracted features varies from 100 to 1600, and more than 1600 features will result in performance deterioration. For different classification tasks, required minimum number of features is a little different with the range from 50 to 100. Meanwhile, after the classification performance remains stable within a reasonable range, performance deteriorates when the number of extracted feature excel one big value. In view of experimental results of different classification tasks, we set

the number of extracted features to 100 for all classification tasks in the paper. This number can satisfy the demand of classification accuracy and ensure the operational efficiency.

For classifier, we design the GCN to complete classification. The information of acquisition equipment and gender is considered in MCI-graph construction. We test the influence of the non-image information, and experimental results show that, by including the non-image information, the performance of MCI classification can make great improvement. The results demonstrate that there are much differences between image information getting from different gender and acquisition equipment. By contrasting the performance of different graph construction strategies, it shows single-layer GCN and three-layer GCN are performing worse than two-layer GCN network. Single-layer GCN network cannot fully express classification system and three-layer GCN network is redundant and inefficient. To validate the performance of the proposed method, we compare the results of the proposed method with that of other different well-established classifiers. Comparing experimental results show that GCN has better performance than RF, GBDT, SVM, and MLP classifiers. This demonstrates that non-image information is good supplement to image information for MCI classification. By comparing the results of our proposed method with the results of GCN, it shows that ACC, SEN, SPE, and AUC can be improved. The performance improvement shows that combining time-varying dynamic information can better describe changes in brain FCN and eventually improve the MCI classification performance.

For the loss function in the proposed method, we computed SoftMax cross-entropy and L_2 loss. The convergence curve of loss function are shown in Fig. 10. As shown, the proposed method has good convergence performance. And the number of epochs in all experiments is set 1000.

Although combining dynamic information in FCN can improve performance of MCI classification, the dynamic information which takes effect is unknown, and this results in insufficient study on combination strategies. In the future work, we will study which and how dynamic information affect efficiency. In addition, we will collect more

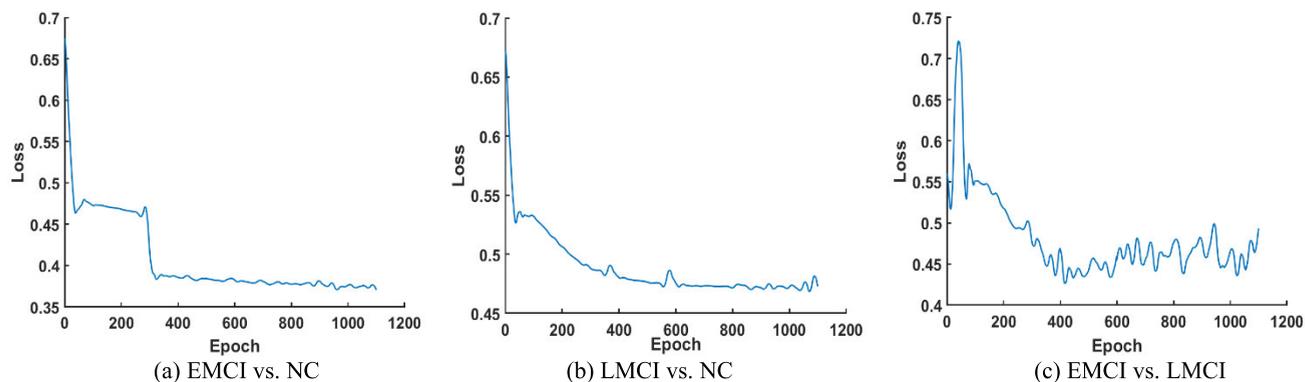


FIGURE 10. Convergence curve of the loss function.

datasets for better training and validation. The accurate diagnosis of MCI stages is a very challenging task and it becomes even more challenging to differentiate between the EMCI and LMCI. So far as we known, all works of MCI classification based on neuroimaging data are limited to two-classification task. As shown in TABLE 8, the performance of MCI two-classification is not good enough to study multi-classification task. In future work, we will try to study multi-classification task.

V. CONCLUSION

In this paper, we introduce a novel method based on a combined high-order network and GCN for MCI classification. The proposed method utilizes a combined high-order network to contain time-varying dynamic information and utilize GCN to contain image information (rs-fMRI image) and non-image information (gender and acquisition equipment). Experimental results show that, two-layer GCN network including acquisition equipment and gender information with 100 extracted features, has the best performance. In general, combining dynamic FCNs and sFCN can better describe dynamic and static characteristics in FCN. In addition, establishing interactions between individuals and populations by using non-image information can significantly improve the performance of classifier.

REFERENCES

- [1] A. Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 14, no. 3, pp. 367–429, Mar. 2018.
- [2] O. Zanetti, S. B. Solerte, and F. Cantoni, "Life expectancy in Alzheimer's disease (AD)," *Archives Gerontol. Geriatrics*, vol. 49, pp. 237–243, Oct. 2009.
- [3] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, S. Belleville, P. H. Brodaty, D. Bennett, P. H. Chertkow, P. J. L. Cummings, P. M. de Leon, P. H. Feldman, P. M. Ganguli, P. H. Hampel, P. P. Scheltens, M. C. Tierney, P. P. Whitehouse, and P. B. Winblad, "Mild cognitive impairment," *Lancet*, vol. 367, no. 9518, pp. 1262–1270, 2006.
- [4] Y. Li, H. Yang, K. Li, P.-T. Yap, M. Kim, C.-Y. Wee, and D. Shen, "Novel effective connectivity network inference for MCI identification," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, Sep. 2017, pp. 316–324.
- [5] J. Wang, X. Zuo, Z. Dai, M. Xia, Z. Zhao, X. Zhao, J. Jia, Y. Han, and Y. He, "Disrupted functional brain connectome in individuals at risk for Alzheimer's disease," *Biol. Psychiatry*, vol. 73, no. 5, pp. 472–481, Mar. 2013.
- [6] B. Lei, P. Yang, Y. Zhuo, F. Zhou, D. Ni, S. Chen, X. Xiao, and T. Wang, "Neuroimaging retrieval via adaptive ensemble manifold learning for brain disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1661–1673, Jul. 2019.
- [7] R. Yu, L. Qiao, M. Chen, S.-W. Lee, X. Fei, and D. Shen, "Weighted graph regularized sparse brain network construction for MCI identification," *Pattern Recognit.*, vol. 90, pp. 220–231, Jun. 2019.
- [8] Y. Li, J. Liu, X. Gao, B. Jie, M. Kim, P.-T. Yap, C.-Y. Wee, and D. Shen, "Multimodal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification," *Med. Image Anal.*, vol. 52, pp. 80–96, Feb. 2019.
- [9] Y. Zhang, H. Zhang, X. Chen, M. Liu, X. Zhu, S.-W. Lee, and D. Shen, "Strength and similarity guided group-level brain functional network construction for MCI diagnosis," *Pattern Recognit.*, vol. 88, pp. 421–430, Apr. 2019.
- [10] M. R. Brier, J. B. Thomas, A. M. Fagan, J. Hassenstab, D. M. Holtzman, T. L. Benzinger, J. C. Morris, and B. M. Ances, "Functional connectivity and graph theory in preclinical Alzheimer's disease," *Neurobiol. Aging*, vol. 35, no. 4, pp. 757–768, Apr. 2014.
- [11] H.-J. Park and K. Friston, "Structural and functional brain networks: From connections to cognition," *Science*, vol. 342, no. 6158, Oct. 2013, Art. no. 1238411.
- [12] A. M. Anter, Y. Wei, J. Su, Y. Yuan, B. Lei, G. Duan, W. Mai, X. Nong, B. Yu, C. Li, Z. Fu, L. Zhao, D. Deng, and Z. Zhang, "A robust swarm intelligence-based feature selection model for neuro-fuzzy recognition of mild cognitive impairment from resting-state fMRI," *Inf. Sci.*, vol. 503, pp. 670–687, Nov. 2019.
- [13] S. J. Teipel, A. Wohler, C. Metzger, T. Grimmer, C. Sorg, M. Ewers, E. Meisenzahl, S. Klöppel, V. Borchardt, M. J. Grothe, M. Walter, and M. Dyrba, "Multicenter stability of resting state fMRI in the detection of Alzheimer's disease and amnesic MCI," *NeuroImage, Clin.*, vol. 14, pp. 183–194, Jan. 2017.
- [14] B. Chen, S. Wang, W. Sun, X. Shang, H. Liu, G. Liu, J. Gao, and G. Fan, "Functional and structural changes in gray matter of Parkinson's disease patients with mild cognitive impairment," *Eur. J. Radiol.*, vol. 93, pp. 16–23, Aug. 2017.
- [15] L. Q. Uddin, A. M. C. Kelly, B. B. Biswal, F. X. Castellanos, and M. P. Milham, "Functional connectivity of default mode network components: Correlation, anticorrelation, and causality," *Hum. Brain Mapping*, vol. 30, no. 2, pp. 625–637, Jan. 2008.
- [16] M. R. Arbabshirani, E. Damaraju, R. Phlypo, S. Plis, E. Allen, S. Ma, D. Mathalon, A. Preda, J. G. Vaidya, T. Adali, and V. D. Calhoun, "Impact of autocorrelation on functional connectivity," *NeuroImage*, vol. 102, pp. 294–308, Nov. 2014.
- [17] Y. Mu, X. Liu, and L. Wang, "A Pearson's correlation coefficient based decision tree and its parallel implementation," *Inf. Sci.*, vol. 435, pp. 40–58, Apr. 2018.
- [18] H. Guo, F. Zhang, J. Chen, Y. Xu, and J. Xiang, "Machine learning classification combining multiple features of a hyper-network of fMRI data in Alzheimer's disease," *Frontiers Neurosci.*, vol. 11, p. 615, Nov. 2017.
- [19] P. Unnikrishnan, V. K. Govindan, and S. D. M. Kumar, "Enhanced sparse representation classifier for text classification," *Expert Syst. Appl.*, vol. 129, pp. 260–272, Sep. 2019.

- [20] Z. Chen, C. Huang, and S. Lin, "A new sparse representation framework for compressed sensing MRI," *Knowl.-Based Syst.*, vol. 188, Jan. 2020, Art. no. 104969.
- [21] C. Chang and G. H. Glover, "Time-frequency dynamics of resting-state brain connectivity measured with fMRI," *NeuroImage*, vol. 50, no. 1, pp. 81–98, Mar. 2010.
- [22] M. G. Preti, T. A. Bolton, and D. Van De Ville, "The dynamic functional connectome: State-of-the-art and perspectives," *NeuroImage*, vol. 160, pp. 41–54, Oct. 2017.
- [23] A. Kucyi, M. J. Hove, M. Esterman, R. M. Hutchison, and E. M. Valera, "Dynamic brain network correlates of spontaneous fluctuations in attention," *Cerebral Cortex*, vol. 27, no. 3, pp. 1831–1840, Mar. 2017.
- [24] E. Simony, C. J. Honey, J. Chen, O. Lositsky, Y. Yeshurun, A. Wiesel, and U. Hasson, "Dynamic reconfiguration of the default mode network during narrative comprehension," *Nature Commun.*, vol. 7, no. 1, p. 12141, Jul. 2016.
- [25] H.-J. Park, K. J. Friston, C. Pae, B. Park, and A. Razi, "Dynamic effective connectivity in resting state fMRI," *NeuroImage*, vol. 180, pp. 594–608, Oct. 2018.
- [26] D. Vidaurre, R. Abeysuriya, R. Becker, A. J. Quinn, F. Alfaro-Almagro, S. M. Smith, and M. W. Woolrich, "Discovering dynamic brain networks from big data in rest and task," *NeuroImage*, vol. 180, pp. 646–656, Oct. 2018.
- [27] B. Jie, M. Liu, and D. Shen, "Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease," *Med. Image Anal.*, vol. 47, pp. 81–94, Jul. 2018.
- [28] F. Scarselli, M. Gori, A. Chung Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.
- [29] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.
- [30] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," Dec. 2013, *arXiv:1312.6203*. [Online]. Available: <http://arxiv.org/abs/1312.6203>
- [31] M. Zhao, R. H. M. Chan, T. W. S. Chow, and P. Tang, "Compact graph based semi-supervised learning for medical diagnosis in Alzheimer's disease," *IEEE Signal Process. Lett.*, vol. 21, no. 10, pp. 1192–1196, Oct. 2014.
- [32] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
- [33] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2224–2232.
- [34] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*. [Online]. Available: <http://arxiv.org/abs/1609.02907>
- [35] R. Levie, F. Monti, X. Bresson, and M. M. Bronstein, "CayleyNets: Graph convolutional neural networks with complex rational spectral filters," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 97–109, Jan. 2019.
- [36] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert, "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.
- [37] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, Apr. 2018.
- [38] J. Wang, X. Wang, M. Xia, X. Liao, A. Evans, and Y. He, "GRETNA: A graph theoretical network analysis toolbox for imaging connectomics," *Frontiers Hum. Neurosci.*, vol. 9, p. 386, Jun. 2015.
- [39] M. Goto, J. Alzheimer's Disease Neuroimaging Initiative, O. Abe, S. Aoki, N. Hayashi, T. Miyati, H. Takao, T. Iwatsubo, F. Yamashita, H. Matsuda, H. Mori, A. Kunitatsu, K. Ino, K. Yano, and K. Ohtomo, "Diffeomorphic anatomical registration through exponentiated lie algebra provides reduced effect of scanner for cortex volumetry with atlas-based method in healthy subjects," *Neuroradiology*, vol. 55, no. 7, pp. 869–875, Apr. 2013.
- [40] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "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.
- [41] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, Nov. 2012.
- [42] E. Damaraju, E. A. Allen, A. Belger, J. M. Ford, S. McEwen, D. H. Mathalon, B. A. Mueller, G. D. Pearlson, S. G. Potkin, A. Preda, J. A. Turner, J. G. Vaidya, T. G. van Erp, and V. D. Calhoun, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage, Clin.*, vol. 5, pp. 298–308, Jul. 2014.
- [43] J. E. Chen, M. Rubinov, and C. Chang, "Methods and considerations for dynamic analysis of functional MR imaging data," *Neuroimaging Clinics North Amer.*, vol. 27, no. 4, pp. 547–560, Nov. 2017.
- [44] Z. Shao, S. Yang, F. Gao, K. Zhou, and P. Lin, "A new electricity price prediction strategy using mutual information-based SVM-RFE classification," *Renew. Sustain. Energy Rev.*, vol. 70, pp. 330–341, Apr. 2017.
- [45] S. Sahran, D. Albashish, A. Abdullah, N. A. Shukor, and S. H. M. Pauzi, "Absolute cosine-based SVM-RFE feature selection method for prostate histopathological grading," *Artif. Intell. Med.*, vol. 87, pp. 78–90, May 2018.
- [46] C.-Y. Wee, S. Yang, P.-T. Yap, and D. Shen, "Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification," *Brain Imag. Behav.*, vol. 10, no. 2, pp. 342–356, Jun. 2015.
- [47] R. Yu, H. Zhang, L. An, X. Chen, Z. Wei, and D. Shen, "Connectivity strength-weighted sparse group representation-based brain network construction for MCI classification," *Hum. Brain Mapping*, vol. 38, no. 5, pp. 2370–2383, Feb. 2017.
- [48] Y. Li, H. Yang, B. Lei, J. Liu, and C.-Y. Wee, "Novel effective connectivity inference using ultra-group constrained orthogonal forward regression and elastic multilayer perceptron classifier for MCI identification," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1227–1239, May 2019.



XUEGANG SONG received the Ph.D. degree in instrument science and technology from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2018. He is currently a Postdoctoral Researcher with Shenzhen University. His research interests include deep learning, mechanical vibration, and signal processing.



AHMED ELAZAB received the Ph.D. degree in pattern recognition and intelligent system from the Shenzhen Institutes of Advanced Technology, University of Chinese Academy of Sciences, China, in 2017. He is currently an Assistant Professor with the Computer Science Department, Misr Higher Institute for Commerce and Computers, Mansoura, Egypt. He is also a Postdoctoral Fellow with the School of Biomedical Engineering, Shenzhen University, Shenzhen, China. His current research interests include machine and deep learning, pattern recognition, medical image analysis, and computer-aided diagnosis.



YUESHIN ZHANG received the M.S. degree in optical engineering from Beihang University, China, in 2015. She is currently pursuing the Ph.D. degree with Southeast University and an exchange Ph.D. Student supported by the China Scholarship Council with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. Her current research interests include deep learning, load forecasting, and signal processing.

...