

Received May 1, 2019, accepted May 23, 2019, date of publication June 5, 2019, date of current version June 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920978

# Multi-Scale Feature Combination of Brain Functional Network for eMCI Classification

ZHUQING JIAO<sup>ID</sup><sup>1</sup>, ZHENGWANG XIA<sup>1</sup>, XUELIAN MING<sup>1</sup>, CHUN CHENG<sup>1</sup>,  
AND SHUI-HUA WANG<sup>ID</sup><sup>2,3</sup>

<sup>1</sup>School of Information Science and Engineering, Changzhou University, Changzhou 213164, China

<sup>2</sup>School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough LE11 3TU, U.K.

<sup>3</sup>Department of Informatics, University of Leicester, Leicester LE1 7RH, U.K.

Corresponding authors: Zhuqing Jiao (jzq@cczu.edu.cn), Chun Cheng (chengchun@cczu.edu.cn), and Shui-Hua Wang (shuihuawang@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grant 51877013 and Grant 51307010, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20181463, in part by the Science and Technology Program of Changzhou City under Grant CE20185038, in part by the University Natural Science Research Program of Jiangsu Province under Grant 17KJB510003 and Grant 13KJB510002, and in part by the Qing Lan Project of Jiangsu Province.

**ABSTRACT** People suffering from mild cognitive impairment (MCI) are at an increased risk of developing Alzheimer's disease (AD) or another dementia. High prevalence will possibly be reduced if early interventions could be applied to the stage of early MCI (eMCI). In network-based classification, brain functional networks are often constructed, relying on the entire time series. It can lead to the neglect of the complex and dynamic interaction relationships among brain regions. As a result, the features derived from this type of functional network may fail to serve as an effective disease biomarker. To address this problem, we proposed a multi-scale feature combination framework for the eMCI classification. In this framework, global static features, time-varying features, and more refined features could be able to flexibly extract from static functional networks, dynamic functional networks, and high-order functional networks, respectively. Then, they are utilized to train and test the classification model in the form of feature combination. The experimental results have verified that the proposed method achieves superior classification accuracy than other competed methods in the eMCI classification, indicating a great potential in understanding the dysfunction of the brain regions.

**INDEX TERMS** Brain functional network, multi-scale feature combination, Alzheimer's disease (AD), early mild cognitive impairment (eMCI), classification accuracy.

## I. INTRODUCTION

Alzheimer's disease (AD), a progressive, irreversible neurodegenerative disease, is the most common type of dementia which accounts for 50% to 80% of dementia cases [1]–[3]. It is serious enough to affect patients' daily life without any effective clinical treatment so far. As reported, the incidence of AD doubles every five years after age 65 and one of every 85 persons will be affected by this disease by 2050 [4], [5]. Mild cognitive impairment (MCI) is an intermediate stage between the expected cognitive decline in normal aging and the more pronounced decline in dementia. It may progress to probable AD with an average conversion rate of 10% to 15% per year, and more than 50% within

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

five years [6]. Due to high conversion rate and increasing life expectancy, early interventions of MCI are important to reduce the risk of developing AD by providing medications as well as non-medication approaches. Therefore, accurate identification of early MCI (eMCI) has drawn much attention to researchers during the last decades [7], [8]. However, an increasing number of studies have suggested there are very subtle changes in the brains of eMCI sufferers, and thus early detections of MCI in neuroimaging data can be very challenging [9], [10].

Vast literatures show that the pathological manifestations of AD can be observed in a few years or even a decade before clinical symptoms appear [11]–[13]. Nowadays, neuroimaging techniques have been widely applied for AD/MCI detection, including structural magnetic resonance imaging (MRI) [14], [15], diffusion tensor imaging (DTI) [16], [17],

positron emission tomography (PET) [18], [19] and functional MRI (fMRI) [20]–[22]. Resting-state fMRI (RS-fMRI) allows doctors and researchers to view activity or problems within the human brain without invasive neurosurgery. It has been successfully applied to identify MCI or AD for its ability to detect biomarkers before clinical symptoms [23]–[26]. Furthermore, many researchers generally assume that the interaction patterns among brain regions keep stationary. They tend to compute the functional connectivity of brain regions based on the entire time series of RS-fMRI data [27]–[29]. This method may overlook the complex and dynamic interaction patterns among brain regions, which are essentially time-varying, frequency-varying or multi-task [30]–[32]. In fact, some recent research studies have demonstrated that functional connectivity contains rich dynamic temporal information. Damaraju *et al.* [33] used the entire time series and sliding time windows respectively to analyze schizophrenia disease and advocated the use of dynamic analysis to better understand functional connectivity. Leonardi *et al.* [34] assumed non-stationary functional connectivity can reflect additional and rich information about brain organization. Therefore, the study of dynamic functional network can help to further explore the brain's working mode, and can better understand the functional organization of the brain, which will be of great help for the diagnosis of brain diseases.

At present, the research direction of brain functional network analysis can be roughly divided into two categories. One is group differences study based on group contrast analysis, such as default mode network (DMN) [35], small-world network [36], and hippocampal network [37], etc. The other one is individual classification and prediction depending on machine learning [38], [39]. In the first type of research work, researchers mainly look for some abnormal functional connectivity through complex network theory and some group comparison analysis methods, which reveal the difference in functional connectivity between brain disease patients and normal human brains. However, such studies generally only seek evidence supporting a certain hypothesis, and cannot automatically complete the classification of individuals, which has great limitations in clinical applications. In the second type of research work, individuals can be automatically classified by training the classification model through machine learning methods. Yu *et al.* [40] proposed a sparse brain functional network modeling framework for the diagnosis of eMCI and AD, and its classification accuracy is up to 84.8%. Chen *et al.* [41] constructed a high-order functional network model to eliminate redundant features, and used Support vector machine (SVM) classifiers for MCI classification. Finally, the classification accuracy of eMCI subjects reached 88.14%. Thus, many studies confirm that the functional network models with more features tend to achieve better classification results.

Currently, the network-based MCI classification research method is roughly divided into three steps: (i) constructing brain functional networks or brain structural networks;

(ii) extracting various features based on the constructed functional networks for training the classification model; (iii) classifying and predicting the unknown samples based on the trained model. Nevertheless, the above procedure is typically performed for extracting single-scale features from a single network model. It may be difficult to make an effective biomarker for disease diagnosis, which can mislead to the clinical application of the above research results. In view of this, extracting multi-scale features from different network models may help to improve the classification accuracy of eMCI to some extent. On this issue, we use innovative multi-scale feature combination method to extract more refined features from brain functional network for eMCI classification. The global static features, time-varying features, and more refined features are respectively derived from static functional networks, dynamic functional networks, and high-order functional networks to train and test the classification model. The rest of the article is organized as follows. In Section II, we introduced the materials and methods, including the proposed framework, data acquisition and processing, and feature extraction. In Section III, we described the experimental results on classification performance, and discuss the influence of variable parameters. In Section IV, we concluded the entire article.

## II. MATERIALS AND METHODS

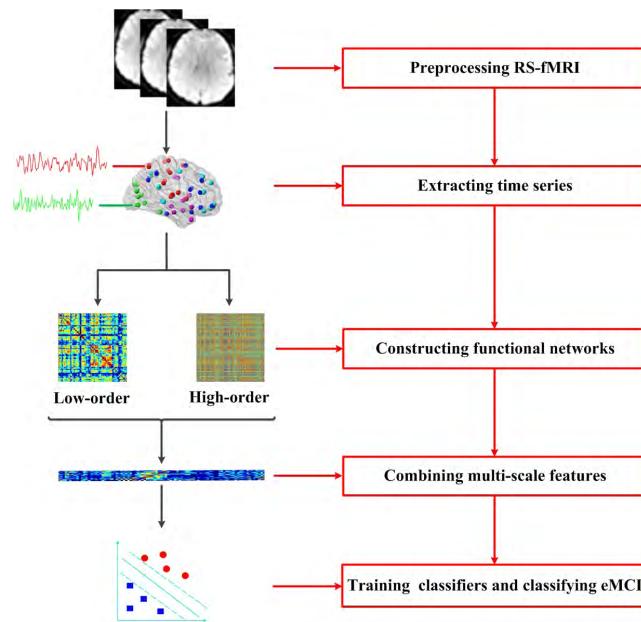
### A. PROPOSED FRAMEWORK

Figure 1 lists the proposed framework of eMCI classification based on multi-scale feature combination. It mainly consists of five steps: (i) preprocessing the original RS-fMRI data of all subjects; (ii) extracting the time series of all brain regions from RS-fMRI data; (iii) constructing low-order (static and dynamic) functional networks and high-order functional networks, respectively; (iv) combining multi-scale features from different network models and performing feature selection to remove redundant features; (v) training SVM classifiers via the combined features and classifying the subjects.

### B. DATA ACQUISITION AND PROCESSING

The RS-fMRI data are obtained from Alzheimer's disease neuroimaging initiative (ADNI) project. Subjects in the current study were selected from ADNIGo and ADNI2 dataset since they include eMCI data (For more details about imaging parameters, please see the ADNI protocols at <http://adni.loni.ucla.edu>). There are totally 60 participants, including 30 eMCI subjects (14F/16M, aged  $73.7 \pm 4.8$  years) and 30 normal subjects (13F/17M, aged  $72.6 \pm 4.5$  years). It should be noted that eMCI subjects did not differ much from normal subjects in terms of brain function, and it is not easy to distinguish them. For this the reason we chose eMCI subjects rather than MCI subjects as the research object.

All subjects were scanned with the same protocol using 3.0 T Philips Achieva scanners. The following parameters were used: repetition time (TR) = 3000 ms, echo time (TE) = 30 ms, flip angle =  $80^\circ$ , imaging matrix =  $64 \times 64$ ,



**FIGURE 1.** The framework for eMCI classification based on multi-scale feature combination.

48 slices, 140 volumes, and slice thickness = 3.3 mm. The RS-fMRI data are preprocessed using the SPM8 toolbox (<http://www.fil.ion.ucl.ac.uk>) with further correction and normalization. The scanning time for each subject is 7 min (i.e., 140 volumes), and the subjects with more than 2.5 min of large frame-wise displacement ( $FD > 0.5$ ) were not included in this study (i.e., excluded before data inclusion). For magnetization equilibrium, the first 3 volumes of each subject were discarded before preprocessing, and then the remaining 137 volumes were involved in subsequent analysis. A rigid-body transformation was adopted to correct head motion during the scan; and the subjects of head motion larger than 2 mm or  $2^\circ$  were not included in this study. The RS-fMRI images were parcellated into 90 regions according to the automated anatomical labeling (AAL) template [42]. The mean time series of each brain region was band-pass filtered (0.015~0.15 Hz). Head motion parameters (Friston-24), mean blood oxygenation level dependent (BOLD) signal of white matter, and mean BOLD signal of cerebrospinal fluid were all regressed out of the RS-fMRI data to further reduce artifacts.

We utilized the regional fMRI time series to construct correlation matrices by means of *Pearson* correlation coefficients. These matrices described frequency-dependent correlations, a measure of functional connectivity, between spatially-distinct brain regions which represent the stability of brain activity in a relatively long period of time. An increasing number of studies have suggested that brain functional network is not stationary but spontaneously changes over time [33], [43], [44]. Specifically, sliding windows are performed over the entire fMRI time series to generate multiple short overlapping segments [45]. For each segment, a static

functional network is constructed for assessing the short-term correlation between brain regions. The dynamic functional networks (obtained from all segments) describe the dynamics of short-term functional connectivity along the time, and thus also forming the correlation matrix for every pair of brain regions.

Suppose the length of the sliding window is  $N$  and the step size between two successive windows is  $s$ . Let  $\mathbf{x}_i^l(k) \in R^N$  denote the  $k$ -th segment of subseries extracted from the entire time series  $\mathbf{x}_i^l$  of the  $l$ -th subject, which comprises  $N$  image volumes. The total number of segments generated by this approach is given by  $K = [(M - N)/s] + 1$ , thus we have  $1 \leq k \leq K$ . For the  $l$ -th subject, the  $k$ -th segment of subseries in all  $R$  regions can be expressed in a matrix form as

$$\mathbf{X}^l(k) = [\mathbf{x}_1^l(k), \mathbf{x}_2^l(k), \dots, \mathbf{x}_R^l(k)] \in R^{N \times R} \quad (1)$$

where  $R$  is the total number of regions.

The *Pearson* correlation coefficient for the  $l$ -th subject is computed as  $C_{ij}^l(k)$ , which represents the pairwise *Pearson* correlation coefficients between the  $i$ -th and the  $j$ -th regions of the  $l$ -th subject using the  $k$ -th segment of subseries. Then, taking  $\{\mathbf{x}_i^l(k)\}$  as the vertices and  $\{C_{ij}^l(k)\}$  as the weights of edges connecting each pair of vertices, we can establish  $K$  dynamic functional networks  $\mathbf{D}_L^l(k)$ , ( $k = 1, 2, \dots, K$ ) for the  $l$ -th subject, which describes the temporal variation of the connection strength for all region pairs. For each region pair  $(i, j)$  in the  $l$ -th subject, we can concatenate all  $C_{ij}^l(k)$  for  $1 \leq k \leq K$  to obtain a correlation time series as

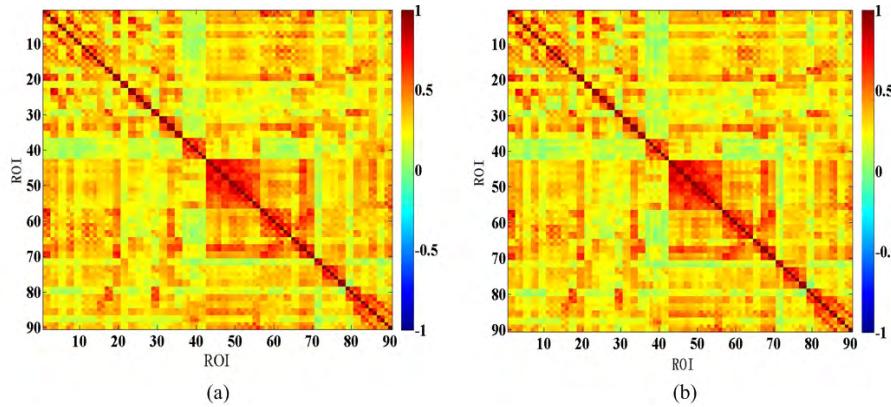
$$\mathbf{y}_{ij}^l = [C_{ij}^l(1), C_{ij}^l(2), \dots, C_{ij}^l(K)] \in R^K \quad (2)$$

The high-order functional network is constructed by measuring the similarity between the topographical profiles of two regions' functional connectivity (i.e., correlation's correlation). Accordingly, we calculate *Pearson* correlation coefficients again for the  $l$ -th subject as  $H_{ij,pq}^l$ , which indicates how the correlation between the  $i$ -th and the  $j$ -th regions influence the correlation between the  $p$ -th and the  $q$ -th regions. In other words, the correlation coefficient  $H_{ij,pq}^l$  is able to characterize more complex and abstract interaction patterns among brain regions. Taking  $\{\mathbf{y}_{ij}^l\}$  as new vertices and  $\{H_{ij,pq}^l\}$  as the weights of new edges, each connecting vertices  $\mathbf{y}_{ij}^l$  and  $\mathbf{y}_{pq}^l$ , we can finally obtain a dynamic high-order functional network  $\mathbf{D}_H^l$ .

### C. FEATURE EXTRACTION

The weighted-graph local clustering coefficient, which describes the local connectivity of a network, is computed to quantify the probability that the neighbors of the node are also connected to each other [46]. For the sake of brevity, suppose there is a network with  $N$  nodes. The weight of the edge connecting node  $i$  and node  $j$  is denoted by  $w_{ij}$  ( $1 \leq i, j \leq N$ ). Then, the weighted-graph local clustering coefficient for node  $i$  is defined as:

$$C_i = \frac{2 \sum_{i,j \in \Delta_i} (w_{ij})^{\frac{1}{3}}}{|\Delta_i|(|\Delta_i| - 1)} \quad (3)$$



**FIGURE 2.** Static functional networks: (a) eMCI subjects, (b) normal subjects.

where  $\Delta_i$  denotes the set of nodes directly connected to node  $i$ , and  $J(\theta) = \frac{1}{2}(X\theta - Y)^T(X\theta - Y) + \lambda \|\theta\|_1 + \beta\theta^T X^T L X \theta$  is the number of elements in  $\Delta_i$ .

It is a challenging task to solve pattern recognition or classification problems with data of high dimensionality [47]. With the presence of irrelevant or redundant features, machine learning methods tend to overfit and become less generalizable. Therefore, feature selection is a useful and important means to identify relevant features for dimensionality reduction and improving generalization performance. In machine learning field, least absolute shrinkage and selection operator (LASSO) is an effective feature selection algorithm when dealing with a small number of subjects with high-dimensional features due to its simplicity and efficiency [48]. Different from other feature selection methods, LASSO can achieve feature selection by minimizing a penalized objective function, which tends to assign zero weight to the most irrelevant and redundant feature. The loss function of LASSO is defined as:

$$J(\theta) = \frac{1}{2} \|Y - X^T \theta\|^2 + \lambda \|\theta\|_1 \quad (4)$$

where  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{N \times d}$  is a feature matrix, and  $N$  is the number of subjects, each  $\mathbf{x}_i$  occupies one row and it represents the combined multi-scale features.  $Y = \{y_i | y_i = +1 \text{ or } y_i = -1\}_{i=1}^N$  is a set of corresponding category labels of subjects ( $y_i = +1$  for normal subjects,  $y_i = -1$  for eMCI subjects),  $\theta$  denotes the regression coefficient of feature, and  $\lambda$  is the regularization parameter, whose role is to balance the complexity of the model.

SVM is a set of related supervised learning algorithms for data analysis and patterns recognition that is applied to classification and regression analysis [49]. It constructs a hyperplane or a set of hyperplanes in a high or infinite dimensional space, which can be utilized for classification, regression, or other tasks [50], [51]. Its kernel function maps the original feature in a high-dimensional space where original chaotic feature distribution becomes linearly separable. A satisfactory separation achieved by the hyperplane has the

largest distance to the nearest training data points of any class, since in general the larger the margin the lower the generalization error of the classifier.

The output form of SVM can be presented as follows:

$$f(\mathbf{X}, \mathbf{W}) = \text{sgn}(\sum_{i=1}^N w_i x_i + b) \quad (5)$$

where  $N$  is the number of samples,  $\mathbf{X}$  is the input vector of the subject, and  $x_i$  is the  $i$ -th element of the vector  $\mathbf{X}$ .  $\mathbf{W}$  is the output vector of SVM,  $w_i$  is the  $i$ -th element of the vector  $\mathbf{W}$ , and  $b$  is the residual term.

The optimization objective function is defined as follows:

$$\mathbf{J}(\mathbf{W}) = \mathbf{W}^T \mathbf{W} = \|\mathbf{W}\|^2 \quad (6)$$

The constraint is demonstrated as follows:

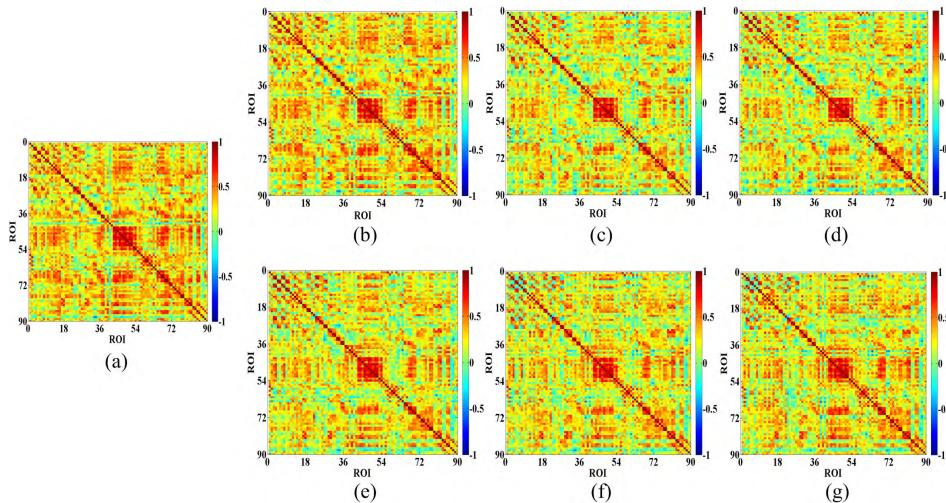
$$y_i(w_i x_i + b) \geq 1, \quad i = 1, 2, \dots, l \quad (7)$$

where  $y_i$  is the sample category labels. Accordingly, the objective function  $\mathbf{J}(\mathbf{W})$  is to make sure the optimality of the classification, and Eq.(7) is to guarantee the classification accuracy.

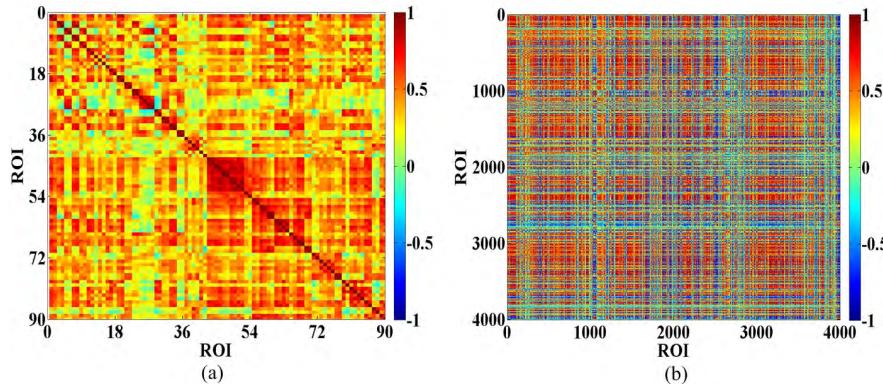
### III. RESULTS AND DISCUSSION

#### A. LOW-ORDER AND HIGH-ORDER FUNCTIONAL NETWORKS

Figure 2(a) and Figure 2(b) list the static functional network for all normal subjects and eMCI subjects, respectively. It is apparent that there is very subtle difference between normal subjects and eMCI subjects. This simple comparison makes more intuitive that there are very few effective features as the distinction, and it is very difficult to classify eMCI through static functional networks. Figure 3(a) shows a static functional network of normal subjects, which provides a global view of the interactions between different brain regions. The positive and negative values of the correlation matrix reflect a competitive relationship between brain regions, namely functional differentiation and functional integration of the brain. However, the static functional network does not



**FIGURE 3.** Low-order functional networks of normal subjects: (a) static functional network, (b)-(g) segments of dynamic functional network.



**FIGURE 4.** Functional network of eMCI subjects (a) static functional network, (b) high-order functional network.

reflect the time-varying characteristics of brain functional connectivity. Consequently, a group of dynamic functional networks is generated by the sliding window method so as to better understand the time-varying information contained in RS-fMRI time series. As shown in Figure 3(b) - Figure 3(g), it is clear that there are only a few subtle differences between the segments of the dynamic functional network, which characterize the time-varying process of functional connectivity.

The high-order functional network is constructed on the base of dynamic functional networks. It contains time-varying information of the brain functional network and reflects more complex brain region interactions, and is therefore able to reveal higher-level and more complex relationships than low-order functional networks. Figure 4 shows the static functional network and the high-order functional network of eMCI subjects. It is apparent that the high-order functional network has higher dimensions and contains more functional connectivity information between the brain regions of eMCI subjects.

Furthermore, high-order correlation matrix can model interactive relationships among brain regions without introducing too many parameters, and thus avoiding over-fitting to a certain extent in the case of small samples. Considerably, the high-order functional network can extract interaction information from four different brain regions at most, while low-order functional network can only reflect the interaction information between two brain regions.

## B. CLASSIFICATION PERFORMANCE

The features of static functional networks, dynamic functional networks, and high-order functional networks are combined in different ways to assist the diagnosis of eMCI. It is helpful to validate the effectiveness of the proposed method in eMCI diagnosis. We evaluate the performances of all competing methods quantitatively based on accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC) [52]. AUC measures the probability that a classifier will rank a randomly chosen positive sample higher than a randomly chosen

**TABLE 1.** Classification performance of different methods for eMCI subjects and normal subjects.

Method	ACC (%)	SEN (%)	SPE (%)	AUC
SFN	79.65 %	82.68 %	76.94 %	0.8011
DFN	82.54 %	85.79 %	77.04 %	0.8024
SFN+DFN	86.69 %	87.57 %	81.29 %	0.8548
SFN+HFN	88.93 %	92.12 %	86.77 %	0.8873
<b>SFN+DFN+HFN</b>	<b>91.13%</b>	<b>93.17 %</b>	<b>87.92 %</b>	<b>0.9211</b>

negative one. Letting TP, TN, FP, and FN denote true positive, true negative, false positive and false negative, respectively, then ACC, SEN, and SPE can be defined as:

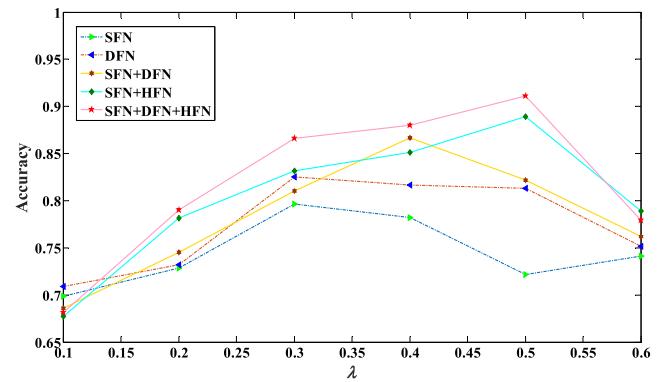
$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (8)$$

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (10)$$

In this section, after constructing three different brain network models, five different feature combination methods are adopted to compare and analyze the effects of these different methods in eMCI classification. Wherein, the weighted clustering coefficients of the network models are extracted as features for training classifier. These network models include static functional network (SFN), dynamic functional network (DFN) and high-order functional network (HFN). The five competing methods are: (i) SFN; (ii) DFN; (iii) SFN+DFN; (iv) SFN+HFN; (v) SFN+DFN+HFN. Table 1 summarizes the classification performances of different methods where the best scores are highlighted in bold.

As shown in Table 1, the classification performance of multi-scale feature combination (SFN+DFN+HFN) is superior to that of single-scale features (SFN, DFN). However, the DFN method is better than the SFN method, which verifies that the time-varying features of DFN contribute greatly to improve the accuracy of eMCI diagnosis. It is noteworthy that the feature combination of three network models achieves the best classification performance, and the classification accuracy reaches ACC=91.13%, which is obviously superior to other methods. Compared with other methods, the best classification accuracy obtained by this method can be explained from three aspects. On the one hand, static low-order functional networks contain a global view of the interaction between different brain regions. On the other hand, **time-varying features in dynamic functional networks help to improve the classification accuracy of eMCI**. In addition, a large number of complex interactive features contained in high-order functional networks are conducive to improve classification accuracy. The classification accuracy of method (iv) is 2.24% higher than that of method (iii). This demonstrates that the complex interactive information of brain regions in high-order networks is more effective than

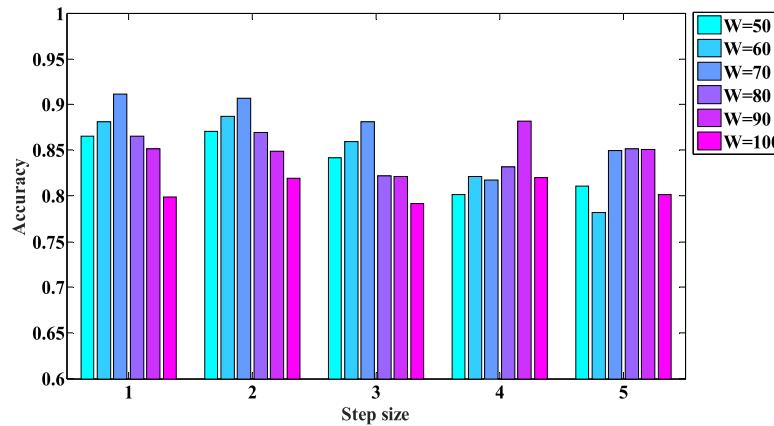
**FIGURE 5.** Accuracy of eMCI classification under different regularization parameters.

the time-varying information in dynamic networks to improve the accuracy of eMCI classification.

### C. INFLUENCE OF VARIABLE PARAMETERS

Feature selection is the subsequent step of multi-scale feature combination in eMCI classification framework. It contributes to remove redundant features and improve the generalization performance of machine learning model. As aforementioned, LASSO contains a regularization parameter  $\lambda$  to adjust the complexity of the model for feature selection. Figure 5 shows the specific results of classification accuracy between eMCI subjects and normal subjects under different regularization parameters. The range of the regularization parameter is  $\lambda \in [0.1, 0.2, 0.3, 0.4, 0.5, 0.6]$ . The reason for the maximum value of  $\lambda$  to 0.6 is that when  $\lambda$  is greater than 0.65, the result of feature selection is that all features have zero weight, that is no features are selected which is meaningless, so the range of values after 0.6 is abandoned.

In Figure 5, the dotted line is a change line graph of the classification accuracy of single-scale features, and the solid line is a change line graph of the classification accuracy of the combined feature. As can be seen from the figure, in most cases, the classification accuracy of the combined features is higher than that of the single-scale features. Specifically, the combined features achieve the best classification accuracy as  $\lambda = 0.4$  or  $\lambda = 0.5$ , while the single-scale features obtain the best classification accuracy as  $\lambda = 0.3$ . The reason for



**FIGURE 6.** Accuracy of eMCI classification under variable parameters.

this difference may be that the total number of features in single model is very small comparing with combined features. As the value of  $\lambda$  increases, the complexity of the model decreases and the selected features become less and less. It ultimately leads to a worsening of the fitting ability of the classification model. Accordingly, regularization parameters have great influences on the classification results, thus it is necessary to be more careful in adjusting them.

The above results suggest that multi-scale feature combination can achieve the best classification accuracy when the regularization parameter is set as  $\lambda = 0.5$ . However, both sliding window width and step size have important influences on the classification results for dynamic functional networks as well as high-order functional networks. Figure 6 shows the classification accuracy of multi-scale feature combination for the eMCI subjects and the normal subjects under different step sizes and window widths. The horizontal axis indicates the range of the values of step size and the vertical axis indicates the ACC value of classification. The step size ranges from S=1 to S=5, and the window width ranges from W=50 to W=100 with an interval of 10. As seen from Figure 6, in most cases, the classification accuracy reaches highest when the window width W=70 and the classification accuracy gets lowest when the window width W=100. Specially, it can be observed that with the increase of step size, the optimal classification accuracy shows a downward trend. This may be due to the fact that large step size ignores part of the dynamic information of functional connections between brain regions over time, which leads to a gradual decline in the optimal classification accuracy.

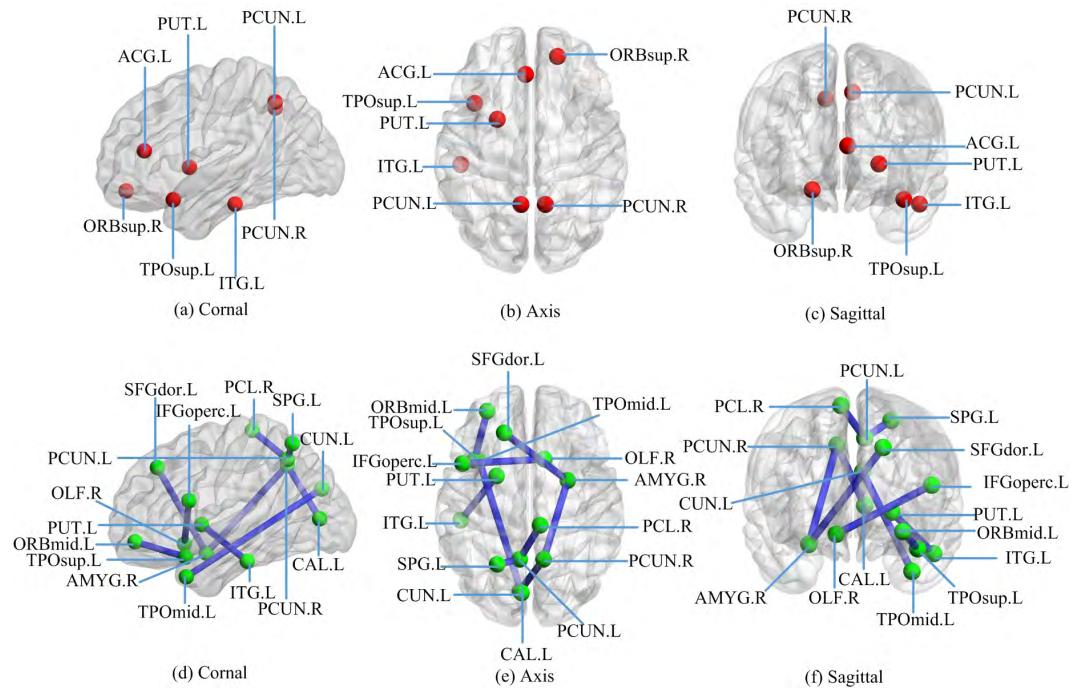
The weighted clustering coefficients are extracted as the features for classification. In low-order functional networks they locate in some specific brain regions, while in high-order functional networks they reflect the functional connection strengths between brain regions. Therefore, we visualize all the feature selection results using the BrainNet Viewer toolbox (<https://www.nitrc.org>) [53]. As shown in Figure 7, these red and green nodes represent brain regions, and the blue links represent functional connections. Figure 7(a) - Figure 7(c) are

the selected brain regions from the low-order brain network, and Figure 7(d) - Figure 7(f) are the selected functional connections from the high-order functional network.

Evidently, some features associated with the brain regions in DMN are selected, including precuneus (PCUN.L, PCUN.R) and inferior temporal gyrus (ITG.L). It is widely believed that DMN plays an important role in cognitive function, and abnormalities in DMN are often observed in a range of neurological diseases. Besides DMN, other brain regions or functional connections may also have important implications for the diagnosis of MCI, such as orbital part of middle frontal gyrus (ORBmid.L, ORBmid.L-TPOsup.L), lenticular nucleus (PUT.L, PUT.L-ITG.L), superior temporal gyrus (TPOsup.L) and middle temporal gyrus (TPOmide.R, TPOmid.R-CUN.L). These results are basically consistent with the previous research results [41], [45].

#### D. FEATURE SELECTION RESULTS

According to the above experimental results, the multi-scale feature combination (SFN+DFN+HFN) achieves the best classification effect (ACC=91.13%) when the regularization parameter  $\lambda = 0.5$ , the step size S=1 and the sliding window width W = 70. At the same time, when the step size S=1 or S=2, the optimal classification accuracy does not change much. Therefore, in the analysis process of this section, based on the super-parameters of the optimal classification result, the step size is set to S=1, the sliding window width is W = 70, and the regularization parameter is set from  $\lambda = 0.4$  to  $\lambda = 0.6$  with an interval of 0.05. Table 2 shows the results of feature selection under different regularization parameters. The results listed in the table are related brain regions and functional connections with feature weights higher than selected mean values of all selected features. It is worth noting that LASSO sets the weights of some features to negative values, but the experimental results show that these features are beneficial to classification. LASSO only judges that they are negatively correlated with the category labels when selecting features. Thus, we first calculate the absolute



**FIGURE 7. Feature selection results: (a)-(c) selected brain regions from low-order brain network, (d)-(f) selected functional connections from high-order functional network.**

**TABLE 2. High-weight selected features under different regularization parameters.**

Regularization parameters	Features number	Selected high-weight features	
		Brain regions	Functional connections
0.4	13	ORBsup.R ACG.L CAL.L PCUN.L PCUN.R TPOsup.L	ORBmid.L-TPOsup.L IFGoperc.L-OLF.R AMYG.R-PCUN.R SPG.L-PCUN.L PCUN.L-PCL.R PCG.L-ACG.L ITG.L-PUT.L
0.45	10	ORBsup.R PCUN.L PCUN.R TPOsup.L	ORBmid.L-TPOsup.L IFGoperc.L-OLF.R AMYG.R-PCUN.R PCUN.L-PCL.R PCG.L-ACG.L ITG.L-PUT.L
0.5	8	ORBsup.R PCUN.L PCUN.R TPOsup.L	ORBmid.L-TPOsup.L AMYG.R-PCUN.R PCUN.L-PCL.R ITG.L-PUT.L
0.55	5	ORBsup.R PCUN.R	AMYG.R-PCUN.R PCUN.L-PCL.R ITG.L-PUT.L
0.6	3	PCUN.R	ORBmid.L-TPOsup.L AMYG.R-PCUN.R

value of all feature weights before calculating the mean of them.

Table 2 intuitively shows that, fewer and fewer high-weight features are selected with the increasing regularization parameters. Accordingly, the regularization parameters have very good regulation effects in the process of

feature selection. After comparing the high-weight selected features of brain regions, we find that the selected brain regions show serious overlap under different regularization parameters, such as ORBsup.R and PCUN.R. And for the high-weight selected features of functional connections, the selected functional connections also show serious overlap,

such as ORBmid.L-TPOsup.L and AMYG.R-PCUN.R. It can be concluded that these brain regions and functional connections can play important roles in the classification and diagnosis of eMCI. By comprehensively comparing the selected brain regions and functional connections, we further find that almost all the same brain regions are involved under same regularization parameters. For example, ACG.L, PCUN.L, PCUN.R and TPOsup.L, as well as the abnormal functional connections associated with them, all have high weights in the process of feature selection when  $\lambda = 0.4$ .

#### IV. CONCLUSION

In summary, we develop a novel eMCI classification framework based on the combination of multi-scale features. The static functional network, the dynamic functional network and the high-order functional network are constructed by Pearson correlation coefficients and sliding time windows. And the weighted clustering coefficients are extracted in the form of feature combination to train and test the classification model, respectively. The experimental results show that multi-scale feature combination has greater advantages in eMCI classification than single network model. In addition, the regularization parameters of feature selection algorithm also have a great influence on the classification results. Simultaneously, the width and the step size of sliding window have certain influences on the accuracy of the eMCI classification. Some dynamic information of functional connectivity between brain regions may be ignored with the increase of the step size. It is worth noting that the methods in this study do not verify the generalization performance of different feature selection algorithms and classifiers. Therefore, our future work should focus on the improvement of feature selection algorithms and the parameter optimization of classifier models.

#### CONFLICT OF INTEREST

We have no conflicts of interest to disclose with regard to the subject matter of this paper.

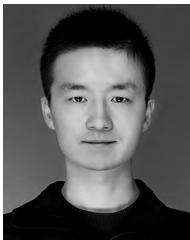
#### REFERENCES

- [1] W. Thies and L. Bleiler, "Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 8, no. 2, pp. 131–168, 2012.
- [2] J. Guillou, Y. Attal, O. Colliot, V. La Corte, B. Dubois, D. Schwartz, M. Chavez, and F. De Vico Fallani, "Loss of brain inter-frequency hubs in Alzheimer's disease," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 10879.
- [3] Y. Zhang, S. Wang, Y. Sui, M. Yang, B. Liu, H. Cheng, J. Sun, W. Jia, P. Phillips, and J. M. Gorriz, "Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization," *J. Alzheimer's Disease*, vol. 65, no. 3, pp. 855–869, 2018.
- [4] L. J. Bain, K. Jedrzweski, M. Morrison-Bogorad, M. Albert, C. Cotman, H. Hendrie, and J. Q. Trojanowski, "Healthy brain aging: A meeting report from the Sylvan M. Cohen annual retreat of the University of Pennsylvania Institute on aging," *Alzheimer's dementia*, vol. 4, no. 6, pp. 443–446, 2008.
- [5] 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, Jul. 2007.
- [6] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow, J. L. Cummings, M. de Leon, H. Feldman, M. Ganguli, H. Hampel, P. Scheltens, M. C. Tierney, P. Whitehouse, and B. Winblad, "Mild cognitive impairment," *Lancet*, vol. 367, no. 9518, pp. 1262–1270, 2006.
- [7] Z. Dong, A. Liu, S. Wang, G. Ji, Z. Zhang, and J. Yang, "Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine," *J. Med. Imag. Health Inf.*, vol. 5, no. 7, pp. 1395–1403, Dec. 2015.
- [8] Y. Zhang, S. Wang, P. Phillips, Z. Dong, G. Ji, and J. Yang, "Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOT-VAC," *Biomed. Signal Process. Control*, vol. 21, pp. 58–73, Aug. 2015.
- [9] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *NeuroImage*, vol. 104, no. 1, pp. 398–412, Jan. 2015.
- [10] Y. Zhang, H. Zhang, X. Chen, and D. Shen, "Constructing multi-frequency high-order functional connectivity network for diagnosis of mild cognitive impairment," in *Connectomics in NeuroImaging* (Lecture Notes in Computer Science), vol. 10511. Basel, Switzerland: Springer, 2017, pp. 9–16.
- [11] S. C. Johnson, T. W. Schmitz, C. H. Moritz, M. E. Meyerand, H. A. Rowley, A. L. Alexander, K. W. Hansen, C. E. Gleason, C. M. Carlsson, M. L. Ries, S. Asthana, K. Chen, E. M. Reiman, and G. E. Alexander, "Activation of brain regions vulnerable to Alzheimer's disease: The effect of mild cognitive impairment," *Neurobiol. Aging*, vol. 27, no. 11, pp. 1604–1612, 2006.
- [12] J. L. Whitwell, S. A. Przybelski, S. D. Weigand, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack, Jr., "3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease," *Brain*, vol. 130, no. 7, pp. 1777–1786, 2007.
- [13] Y. Zhang, S. Wang, P. Phillips, J. Yang, and T.-F. Yuan, "Three-dimensional eigenbrain for the detection of subjects and brain regions related with Alzheimer's disease," *J. Alzheimer's Disease*, vol. 50, no. 4, pp. 1163–1179, 2016.
- [14] L. K. McEvoy, C. Fennema-Notestine, J. C. Roddey, D. J. Hagler, Jr., D. Holland, and D. S. Karow, "Alzheimer disease: Quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment," *Radiology*, vol. 251, no. 1, pp. 195–205, 2009.
- [15] Y.-D. Zhang, X.-Q. Chen, T.-M. Zhan, Z.-Q. Jiao, Y. Sun, Z.-M. Chen, Y. Yao, L.-T. Fang, Y.-D. Lv, and S.-H. Wang, "Fractal dimension estimation for developing pathological brain detection system based on Minkowski–Bouligand method," *IEEE Access*, vol. 4, pp. 5937–5947, 2016.
- [16] S. Haller, D. Nguyen, C. Rodriguez, J. Emch, G. Gold, A. Bartsch, K. O. Lovblad, and P. Giannakopoulos, "Individual prediction of cognitive decline in mild cognitive impairment using support vector machine-based analysis of diffusion tensor imaging data," *J. Alzheimer's Disease*, vol. 22, no. 1, pp. 315–327, 2010.
- [17] C.-Y. Wee, P.-T. Yap, W. Li, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen, "Enriched white matter connectivity networks for accurate identification of MCI patients," *NeuroImage*, vol. 54, no. 3, pp. 1812–1822, 2011.
- [18] I. Garali, M. Adel, S. Bourennane, and E. Guedj, "Brain region ranking for 18FDG-PET computer-aided diagnosis of Alzheimer's disease," *Biomed. Signal Process. Control*, vol. 27, pp. 15–23, May 2016.
- [19] S. H. Nozadi and S. Kadoury, "Classification of Alzheimer's and MCI patients from semantically parcelled PET images: A comparison between AV45 and FDG-PET," *Int. J. Biomed. Imag.*, vol. 2018, pp. 1247430:1–1247430:13, 2018.
- [20] C.-Y. Wee, P.-T. Yap, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen, "Resting-state multi-spectrum functional connectivity networks for identification of MCI patients," *PLoS ONE*, vol. 7, no. 5, 2012, Art. no. e37828.
- [21] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, "Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI," *NeuroImage*, vol. 112, pp. 232–243, May 2015.
- [22] B. Hart, I. Cribben, and M. Fiecas, "A longitudinal model for functional connectivity networks using resting-state fMRI," *NeuroImage*, vol. 178, pp. 687–701, Sep. 2018.
- [23] E. L. Dennis and P. M. Thompson, "Functional brain connectivity using fMRI in aging and Alzheimer's disease," *Neuropsychol. Rev.*, vol. 24, no. 1, pp. 49–62, Mar. 2014.
- [24] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "Network modelling methods for fMRI," *NeuroImage*, vol. 54, no. 2, pp. 875–891, 2011.

- [25] O. Sporns, "The human connectome: A complex network," *Schizophrenia Res.*, vol. 136, no. Suppl. 1, p. S28, Apr. 2012.
- [26] Y. I. Sheline and M. E. Raichle, "Resting state functional connectivity in preclinical Alzheimer's disease," *Biol. Psychiatry*, vol. 74, no. 5, pp. 340–347, 2013.
- [27] M. Kaiser, "A tutorial in connectome analysis: Topological and spatial features of brain networks," *NeuroImage*, vol. 57, no. 3, pp. 892–907, 2011.
- [28] K. C. Wang, G. B. Wu, X. Hou, D. T. Wei, H. S. Liu, and J. Qiu, "Segmentation and application of functional network from group to individual," *Sci. Bull.*, vol. 61, no. 27, pp. 3022–3035, 2016.
- [29] Z. Jiao, K. Ma, H. Wang, L. Zou, and Y. Zhang, "Research on node properties of resting-state brain functional networks by using node activity and ALFF," *Multimedia Tools Appl.*, vol. 77, no. 17, pp. 22689–22704, 2018.
- [30] M. J. Tobia, K. Hayashi, G. Ballard, I. H. Gotlib, and C. E. Waugh, "Dynamic functional connectivity and individual differences in emotions during social stress," *Human Brain Mapping*, vol. 38, no. 12, pp. 6185–6205, 2017.
- [31] W. H. Thompson and P. Fransson, "The frequency dimension of fMRI dynamic connectivity: Network connectivity, functional hubs and integration in the resting brain," *NeuroImage*, vol. 121, pp. 227–242, Nov. 2015.
- [32] X. Wang, Y. S. Ren, and W. S. Zhang, "Multi-task fused lasso method for constructing dynamic functional brain network of resting-state fMRI," *J. Image Graph.*, vol. 22, no. 7, pp. 978–987, 2017.
- [33] 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, 2014.
- [34] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J.-M. Annoni, M. Schlueter, P. Vuilleumier, and D. Van De Ville, "Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest," *NeuroImage*, vol. 83, pp. 937–950, Dec. 2013.
- [35] M. D. Greicius, K. Supekar, V. Menon, and R. F. Dougherty, "Resting-state functional connectivity reflects structural connectivity in the default mode network," *Cerebral Cortex*, vol. 19, no. 1, pp. 72–78, 2009.
- [36] Z. Q. Jiao, K. Ma, Y. L. Rong, H. Wang, and L. Zou, "Adaptive synchronisation of small-world networks with Lorenz chaotic oscillators," *Int. J. Sensor Netw.*, vol. 24, no. 2, pp. 90–97, 2017.
- [37] G. Buzsaki, D. L. Buhl, K. D. Harris, J. Csicsvari, B. Czéh, and A. Morozov, "Hippocampal network patterns of activity in the mouse," *Neuroscience*, vol. 116, no. 1, pp. 201–211, 2003.
- [38] X. Chen, H. Zhang, S. W. Lee, and D. Shen, "Hierarchical high-order functional connectivity networks and selective feature fusion for MCI classification," *Neuroinformatics*, vol. 15, no. 3, pp. 271–284, 2017.
- [39] Y. Zhang, H. Zhang, X. Chen, S.-W. Lee, and D. Shen, "Hybrid high-order functional connectivity networks using resting-state functional MRI for mild cognitive impairment diagnosis," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 6530.
- [40] 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, 2017.
- [41] X. Chen, H. Zhang, Y. Gao, C.-Y. Wee, G. Li, and D. Shen, "High-order resting-state functional connectivity network for MCI classification," *Hum. Brain Mapping*, vol. 37, no. 9, pp. 3282–3296, 2016.
- [42] N. Tzourio-Mazoyer, 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, 2002.
- [43] W. H. Thompson and P. Fransson, "Bursty properties revealed in large-scale brain networks with a point-based method for dynamic functional connectivity," *Sci. Rep.*, vol. 6, Dec. 2016, Art. no. 39156.
- [44] X. Chen, H. Zhang, L. Zhang, C. Shen, S.-W. Lee, and D. Shen, "Extraction of dynamic functional connectivity from brain grey matter and white matter for MCI classification," *Human Brain Mapping*, vol. 38, no. 10, pp. 5019–5034, 2017.
- [45] 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, 2016.
- [46] Y.-D. Zhang, S.-H. Wang, X.-J. Yang, Z.-C. Dong, G. Liu, P. Phillips, and T.-F. Yuan, "Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine," *SpringerPlus*, vol. 4, p. 716, Nov. 2015.
- [47] Y.-D. Zhang, S. Chen, S.-H. Wang, J.-F. Yang, and P. Phillips, "Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine," *Int. J. Imag. Syst. Technol.*, vol. 25, no. 4, pp. 317–327, 2015.
- [48] Y. Zhang and S. Wang, "Detection of Alzheimer's disease by displacement field and machine learning," *PeerJ*, vol. 3, Sep. 2015, Art. no. e1251.
- [49] 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.
- [50] C.-Y. Wee, P.-T. Yap, and D. Shen, "Diagnosis of autism spectrum disorders using temporally distinct resting-state functional connectivity networks," *CNS Neurosci. Therapeutics*, vol. 22, no. 3, pp. 212–219, 2016.
- [51] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 27, pp. 9673–9678, 2005.
- [52] C. Echávarri, P. Aalton, H. B. M. Uylings, H. I. L. Jacobs, P. J. Visser, E. H. B. M. Gronenschild, F. R. J. Verhey, and S. Burgmans, "Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease," *Brain Struct. Function*, vol. 215, nos. 3–4, pp. 265–271, 2011.
- [53] M. Xia, J. Wang, and Y. He, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS ONE*, vol. 8, no. 7, 2013, Art. no. e68910.



**ZHUQING JIAO** received the B.S. degree from the Qilu University of Technology, Jinan, China, in 2005, and the M.S. and Ph.D. degrees from Jiangnan University, Wuxi, China, in 2008 and 2011, respectively. He is currently an Associate Professor with the School of Information Science and Engineering, Changzhou University, Changzhou, China. His current research interests include intelligent computing, computer-aided diagnosis, and brain networks.



**ZHENGWANG XIA** received the B.S. degree from Yuncheng University, Yuncheng, China, in 2016. He is currently pursuing the M.S. degree with the School of Information Science and Engineering, Changzhou University, Changzhou, China. His research interest includes intelligent computing of medical image.



**XUELIAN MING** received the bachelor's degree from the Changzhou University Huaide College, Jingjiang, China, in 2017. She is currently pursuing the M.S. degree with the School of Information Science and Engineering, Changzhou University, Changzhou, China. Her research interests include complex networks and cognitive computing.



**CHUN CHENG** received the B.S. degree from Nantong Medical University, Nantong, China, in 1985. He is currently a Professor with Changzhou University, Changzhou, China, and the Nanjing University of Chinese Medicine, Nanjing, China. His research interest includes neuroscience.



**SHUI-HUA WANG** received the B.S. degree from Southeast University, Nanjing, China, in 2008, the M.S. degree from The City University of New York, New York, NY, USA, in 2012, and the Ph.D. degree from Nanjing University, Nanjing, in 2017. She is currently a Postdoctoral Researcher with the School of Architecture, Building and Civil Engineering, Loughborough University, U.K. Her research interests include biomedical image processing and computer-aided diagnosis.

• • •