

Received February 6, 2019, accepted February 20, 2019, date of publication March 11, 2019, date of current version April 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903332

# Graph-Kernel Based Structured Feature Selection for Brain Disease Classification Using Functional Connectivity Networks

MI WANG<sup>1</sup>, BIAO JIE<sup>1</sup>, WEIXIN BIAN<sup>1</sup>, XINTAO DING<sup>1</sup>, WEN ZHOU<sup>1</sup>,  
ZHENGDONG WANG<sup>1</sup>, AND MINGXIA LIU<sup>2</sup>

<sup>1</sup>School of Computer and Information, Anhui Normal University, Wuhu 241003, China

<sup>2</sup>Department of Information Science and Technology, Taishan University, Taian 271000, China

Corresponding authors: Biao Jie (jbiao@nuaa.edu.cn) and Mingxia Liu (mxliu1226@gmail.com)

This work was supported in part by the National Natural Science Foundations of China under Grant 61573023, Grant 61703301, and Grant 61602009, in part by the Natural Science Foundation of Anhui Province under Grant 1708085MF145 and Grant 1808085MF171, in part by the Foundation for Outstanding Young in Higher Education of Anhui, China, under Grant gxyqZD2017010, in part by the AHNU Fundamental Research Funds under Grant 2016XJJ120, in part by the Taishan Scholar Program of Shandong Province, China, in part by the Scientific Research Foundation of Taishan University under Grant Y-01-2018019, and in part by the BZUU Fundamental Research Funds under Grant KJ2018A0821.

**ABSTRACT** Feature selection has been applied to the analysis of complex structured data, such as functional connectivity networks (FCNs) constructed on resting-state functional magnetic resonance imaging (rs-fMRI), for removing redundant/noisy information. Previous studies usually first extract topological measures (e.g., clustering coefficients) from FCNs as feature vectors, and then perform vector-based algorithms (e.g., *t*-test) for feature selection. However, due to the use of vector-based representations, these methods simply ignore important local-to-global structural information of connectivity networks, while such structural information could be used as prior knowledge of networks to improve the learning performance. To this end, we propose a graph kernel-based structured feature selection (gk-SFS) method for brain disease classification with connectivity networks. Different from previous studies, our proposed gk-SFS method uses the graph kernel technique to calculate the similarity of networks and thus can explicitly take advantage of the structural information of connectivity networks. Specifically, we first develop a new graph kernel-based Laplacian regularizer in our gk-SFS model to preserve the structural information of connectivity networks. We also employ an  $l_1$ -norm based sparsity regularizer to select a small number of discriminative features for brain disease analysis (classification). The experimental results on both ADNI and ADHD-200 datasets with rs-fMRI data demonstrate that the proposed gk-SFS method can further improve the classification performance compared with the state-of-the-art methods.

**INDEX TERMS** Functional connectivity network, graph kernel, feature selection, Laplacian regularizer, classification.

## I. INTRODUCTION

Advanced neuroimaging technologies, such as functional magnetic resonance imaging (fMRI), provide non-invasive ways to explore the function of the human brain, thus providing important insights into the basic cognitive processes of the brain [1]. These advanced technologies also provide the important way to achieve performances applicable in diagnosis of brain diseases [2], [3]. Functional connectivity networks (FCNs) constructed from fMRI data can characterize

the interactions of brain regions at the connectivity level, and have been widely applied to the analysis of various brain diseases, including Alzheimer's disease (AD) and its prodromal stage (*i.e.*, mild cognitive impairment, MCI) and Attention Deficit Hyperactivity Disorder (ADHD). Dysfunctions have been found in AD/MCI/ADHD patients in FCN. Recent, FCNs based on resting-state fMRI (rs-fMRI) have been applied to computer-aided diagnosis of brain diseases by using machine learning methods [4]–[8].

Conventional connectivity network based studies usually contains two essential steps, *i.e.*, feature extraction and feature selection. The aim of feature extraction is to extract

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

meaningful measures (*e.g.*, local clustering coefficients and connectivity strengths) from connectivity networks as feature representation for each subject. Feature selection focuses on selecting the most discriminative features to improve the performance of learned model. For example, Wee *et al.* [9] extract local clustering coefficient features from both structural and functional connectivity networks, followed by *t*-test based feature selection for MCI identification. Liu *et al.* [10] extract connectivity strength features from FCNs, and perform feature selection using F-scores to identify patients with social anxiety disorder. Wang *et al.* [11] use the Regional homogeneity (ReHo) measures extracted from fMRIs as feature representation of each subject, followed by the *t*-test algorithm to select the discriminative features for ADHD classification. Jie *et al.* [12] extract clustering coefficient features from hyper-connectivity networks, and perform a sparse-learning-based feature selection for MCI classification. These studies have shown that feature selection can not only improve the performances of disease diagnosis, but also help discover neuroimage-based biomarkers for better understanding the pathology of brain diseases.

Previous studies typically extract vector-wise measurements from FCNs and perform vector-based feature selection for subsequent classification. The major disadvantage of these studies is that they simply ignore the local-to-global structural information conveyed in the whole population of connectivity networks, while such information can be used as the prior knowledge to boost the classification performance. To solve this issue, we propose a graph kernel based structured feature selection (called gk-SFS) method for brain disease classification using rs-fMRI data. Different from previous feature selection algorithms, the proposed gk-SFS method uses graph kernels to calculate the similarity of connectivity networks, thus naturally preserving the structural information of all networks. Here, graph kernel is a kernel that measures the structural similarity of a pair of graphs/networks. Specifically, our gk-SFS model contains two regularization items: 1) a graph kernel based Laplacian regularizer that can preserve the local-to-global structural information of network data, and 2) a  $l_1$ -norm based sparsity regularizer [13], [14] that can yield the sparse solution of model, and thus ensure only a small number of important features to be selected for classification. We validate our proposed method on two public datasets with baseline rs-fMRI data, *i.e.*, 1) the ADNI dataset<sup>1</sup> that contains 43 Late MCI (IMCI) patients, 56 early MCI (eMCI) patients and 50 healthy controls (HCs), and 2) the ADHD-200 dataset<sup>2</sup> that includes 118 ADHD children and 98 HCs. The experimental results demonstrate the efficacy of our proposed gk-SFS method.

The rest of the paper is organized as following: in Section II, we briefly review the related studies. In Section III, we introduce the data used in the study and present our

proposed method. In Section IV, we give the experimental setting and results. In Section V, we discuss the experiments, results and limitations. Finally, we conclude our study in Section VI.

## II. RELATED WORKS

### A. FUNCTIONAL CONNECTIVITY NETWORK ANALYSIS

Studies on anatomy and physiology have suggested that cognitive processes of the human brain are related with the interactions between brain regions [15]. These interactions can be represented as connectivity networks with nodes corresponding to brain regions (*e.g.*, regions-of-interest, ROIs) [16] and edges denoting the interactions between brain regions. These connectivity networks provide an important way for helping us to better understand the functions of the brain and pathology of brain diseases, by exploring connectivity of networks using graph-based methods. Currently, FCNs have been widely applied to various kinds of tasks, including brain disease analysis [17]–[19] and video shots representation [20]. Existing brain disease analysis studies based on FCNs can be usually classified into two categories: 1) specific hypothesis test using group analysis algorithms, and 2) subject-level classification using machine learning algorithms. The first category focuses on testing specific hypothesis to identify abnormal connectivity and properties of networks in patients with brain diseases. For example, previous studies have demonstrated abnormal connectivity between the hippocampus and other regions in AD/MCI brains [21], [22]. And other studies also found disrupted small-world properties in connectivity networks of AD/MCI patients [23], [24]. One of disadvantages of these studies is that they usually investigate the connectivity in FCNs using group analysis methods, thus can not be used for automatic classification of brain diseases.

In the second category (*i.e.*, subject-level classification methods), various machine learning methods have been used to train learning models for identifying patients with brain diseases from HCs. For example, previous studies have proposed effective connectivity network based methods to accurately identify AD/MCI patients from HCs [4], [12], [25]–[27] and to classify ADHD patients from HCs [18]. These studies usually first extract meaningful network measures as feature representation for each subject, followed by vector-based feature selection for classification. However, they usually ignore the structural information of connectivity networks, thus leading to sub-optimal learning performance.

### B. FEATURE SELECTION

As a dimension reduction technique, feature selection helps eliminate redundant/noise features and is therefore widely used in a variety of applications for machine learning and pattern recognition. At present, there are a large number of feature selection methods that can be generally divided into two categories, *i.e.*, ranking-based methods and subset-based

<sup>1</sup><http://adni.loni.usc.edu>

<sup>2</sup>[http://fcon\\_1000.projects.nitrc.org/indi/adhd200/](http://fcon_1000.projects.nitrc.org/indi/adhd200/)

methods. In ranking-based methods, all features are first ranked and then selected based on a specific criterion. Typical feature ranking algorithms include Laplacian score [28], Fisher score [29], *t*-test based on *p*-values, and infinite latent feature selection (called inf-FS) method proposed in [30]. The major limitation of this kind of methods is that they only take into account the specific contribution of each single feature, without considering the possible complementary information conveyed in multiple features. In the second category, a subset of features is selected by optimizing a specific objective function. Typical methods in this category include recursive feature elimination (RFE) method [31], [32], LASSO method [13] and HSIC-LASSO method [33]. Subset-based methods can take into account the importance of combined features, thus usually achieve better performance compared with ranking-based methods.

In FCN analysis, feature selection is usually performed to choose important features and remove redundant/noisy features for improving the learning performance. Two simple feature selection methods have been widely applied currently, *i.e.*, *t*-test and LASSO methods. In the *t*-test method, the standard *t*-test algorithm is first used for measuring the discriminative power of each feature, followed by selecting important features based on their discriminative capability. Existing studies have shown that the *t*-test method typically achieves good performance in small sample problems [34], [35]. Different from the *t*-test method, the LASSO method selects features by minimizing an  $l_1$ -norm regularized objective function. Existing studies have shown that the LASSO method can achieve good performance when there are a large number of unrelated features but only a small amount of subjects [36].

### C. GRAPH KERNEL

Kernel methods, which implicitly map the original data to a high-dimensional space in which the samples are more likely linear separable, have been successfully applied to many tasks [37]. The kernel can also be regarded as a function that calculates the similarity of subjects. Besides on vector-wise features, kernels can also be constructed on complex structured data, such as graphs. Such kernels constructed on graphs are called graph kernels, which measure the topological similarity between graphs [38], [39] and have been used in image processing [40] and bioinformatics [41], [42]. Recently, graph kernels are also used for brain connectivity network analysis, including brain disease classification [43] and task-related state analysis [44].

Graph kernels are usually constructed by comparing small sub-structures (*i.e.*, sub-graphs), such as shortest-path [45] and walk [46]. For example, Shervashidze *et al.* [39] construct a subtree-based graph kernel (called Weisfeiler-Lehman subtree kernel) that can effectively capture the structural information of graphs. The key idea of Weisfeiler-Lehman subtree kernel is to construct kernel based on the Weisfeiler-Lehman graph isomorphism test. Specifically, given a pair of graphs/networks  $E_1$  and  $E_2$ , let  $P_i$  denote the set of labels for nodes in  $E_1$  and  $E_2$  at the  $i^{th}$  iteration

of Weisfeiler-Lehman test of isomorphism, and  $P_0$  denotes a set of initial labels occurred in  $E_1$  and  $E_2$ . If no initial label on nodes of graphs is available, we use the degree of each node as its label. In each iteration of Weisfeiler-Lehman graph isomorphism test, the label of each node is simultaneously updated using the labels of its neighboring nodes (*i.e.*, mapping the original label of node and labels of its neighboring nodes into a new label). The Weisfeiler-Lehman subtree kernel is defined as follows:

$$k(E_1, E_2) = \langle \emptyset(E_1), \emptyset(E_2) \rangle \quad (1)$$

where

$$\begin{aligned} \emptyset(E_1) = & (\beta_0(E_1, p_{0,1}), \dots, \beta_0(E_1, p_{0,|P_0|}), \dots, \\ & \beta_h(E_1, p_{l,1}), \dots, \beta_l(E_1, p_{l,|P_l|})), \end{aligned}$$

and

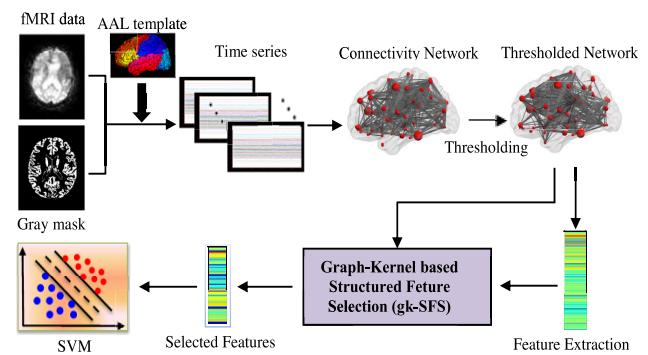
$$\begin{aligned} \emptyset(E_2) = & (\beta_0(E_2, p_{0,1}), \dots, \beta_0(E_2, p_{0,|P_0|}), \dots, \\ & \beta_l(E_2, p_{l,1}), \dots, \beta_l(E_2, p_{l,|P_l|})), \end{aligned}$$

where  $l$  represents the max number of iteration of Weisfeiler-Lehman graph isomorphism test,  $p_{i,j} \in P_i$  and  $|P_i|$  is the number of labels in  $P_i$ ,  $\beta_i(E_1, p_{i,j})$  and  $\beta_i(E_2, p_{i,j})$  denote the number of the label  $p_{i,j}$  occurred in networks  $E_1$  and  $E_2$ , respectively.

It's worth noting that each updated label in graph isomorphism test denotes a subtree pattern, and the number of nodes in each subtree becomes larger with the increment of the number of iterations. Therefore, the Weisfeiler-Lehman subtree kernel defined in Eq. 1 can capture both local and global structural information of the network for network similarity computation. Empirical studies have shown that the graph kernel defined in Eq. 1 outperforms conventional graph kernels [39].

### III. MATERIALS AND METHOD

Figure 1 presents the proposed gk-SFS based learning framework, including three main steps: 1) image pre-processing



**FIGURE 1.** Illustration of our proposed graph-kernel based learning framework for brain disease classification, including three major components: (1) rs-fMRI image pre-processing and functional connectivity network construction, (2) feature extraction and graph-kernel based structured feature selection (gk-SFS), and (3) support vector machine (SVM) based disease classification. rs-fMRI: resting-state functional MRI. BOLD: blood-oxygen level-dependent signal.

and FCN construction, 2) feature learning (*i.e.*, feature extraction and graph-kernel-based structured feature selection), and 3) brain disease classification. In the following subsections, we first introduce two datasets used in this study, as well as image pre-processing and FCN construction. Then, we present the proposed graph-kernel structured feature selection method, and the corresponding learning framework for brain disease classification.

#### A. MATERIALS

The first dataset used in this study is downloaded from ADNI database, which contains 149 subjects with rs-fMRI data, including 43 IMCI, 56 eMCI and 50 HCs. Data acquisition is performed as follows: the image resolution is  $2.29\text{--}3.31\text{ mm}$  for inplane, and slice thickness is  $3.31\text{ mm}$ , TE =  $30\text{ ms}$  and TR =  $2.2\text{--}3.1\text{ s}$ . Another dataset is ADHD-200 from New York University (NYU) site, which contains 216 subjects with rs-fMRI data, including 118 ADHD and 98 HCs. The acquisition of data in ADHD-200 is performed as follows: matrix size is  $49 \times 58$ , axial slices is 47, slice thickness is  $4\text{ mm}$ , FOV =  $240\text{ mm}$ , TR =  $2\text{ s}$ , TE =  $15\text{ ms}$ , flip angle =  $90^\circ$ , voxel size is  $3 \times 3 \times 4\text{ mm}$ . Demographic information of the subjects in two datasets is given in Table 1.

**TABLE 1. Demographic information of subjects in ADNI and ADHD-200 datasets (mean  $\pm$  standard deviation). MMSE: Mini-mental state examination, M=Male, F=Female.**

Dataset	Class	Subject #	Age	Sex(M/F)	MMSE
ADNI	IMCI	43	$72.1 \pm 8.2$	26/17	$27.2 \pm 2.0$
	eMCI	56	$71.2 \pm 6.8$	21/35	$28.1 \pm 1.5$
	HC	50	$75.0 \pm 6.9$	21/29	$28.9 \pm 1.6$
ADHD-200	ADHD	118	$11.2 \pm 2.7$	25/93	-
	HC	98	$12.2 \pm 2.1$	51/47	-

#### B. IMAGE PRE-PROCESSING AND NETWORK CONSTRUCTION

For images from the ADNI dataset, followed the work in [47], we use the standard pipeline for performing image pre-processing, including discarding the first 10 rs-fMRI volumes, slice timing correction, and head motion correction. The brain space of fMRI scans is partitioned into 90 ROIs using the Automated Anatomical Labeling (AAL) template [48] with a deformable registration method [49]. The band-pass filtering is performed within a frequency interval of  $[0.025\text{Hz}, 0.100\text{Hz}]$ . The BOLD signals from the gray matter tissue are extracted, and the mean time series of each ROI is calculated to construct FCNs. For the ADHD-200 dataset used in this study, we directly used the time series from the Athena preprocessed data. The details of image pre-processing for this dataset can be found on the Athena website.<sup>3</sup> Briefly, the data pre-processing steps includes: removing the first 4 image volumes, slice

<sup>3</sup><http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline>

timing correction, head motion correction, co-registering the image into template space, extracting the fMRI time series from gray matter regions, temporal band-pass filtering  $[0.009\text{Hz}, 0.08\text{Hz}]$ , partitioning the brain space into 90 ROIs using AAL template [48], and extracting regional mean time series of each ROIs.

Finally, for subjects in both datasets, we construct the corresponding FCNs using the pairwise Pearson correlation coefficient as the measures of functional connectivity between ROIs. And, we use the Fisher's r-to-z transformation on the elements of an FCN for normalizing the correlation coefficients.

#### C. GRAPH-KERNEL BASED STRUCTURED FEATURE SELECTION

Given a network set  $\mathcal{E} = \{\tilde{E}_1, \tilde{E}_2, \dots, \tilde{E}_N\}$ , where  $\tilde{E}_i$  denotes the thresholded network from the  $i^{\text{th}}$  subject,  $N$  is the number of subjects. Let  $X = [x_1, x_2, \dots, x_N]^T \in R^{N \times d}$  denotes a set of feature vectors for  $N$  subjects (with each vector corresponding to a specific subject). For example,  $x_i$  denotes the region-specific clustering coefficient features extracted from the  $i^{\text{th}}$  network/subject, and  $d$  is the feature dimension. Let  $Y = [y_1, y_2, \dots, y_N] \in R^N$  denote the response vector, where  $y_i$  represents the class label of the  $i^{\text{th}}$  subject.

Motivated by [50], we first introduce a graph-kernel-based Laplacian regularization term to preserve the local-to-global structural information of FCNs of all training subjects, *i.e.*,

$$\sum_{i,j}^N \|w^T x_i - w^T x_j\|^2 S_{i,j} = 2w^T X^T M X w, \quad (2)$$

where  $w$  is the weighted coefficients whose each element is associated with a specific feature.  $S = [S_{i,j}] \in R^{N \times N}$  is a similarity matrix that measures the similarity between subjects, and  $M = C - S$  is a Laplacian matrix. Note that  $C$  is a diagonal matrix whose diagonal elements is defined as  $C_{i,i} = \sum_{j=1}^N S_{i,j}$ . To preserve the structural information of FCNs, we use the graph kernel to measure the similarity between networks *i.e.*,

$$S_{i,j} = k(\tilde{E}_i, \tilde{E}_j) \quad (3)$$

where  $k(\tilde{E}_i, \tilde{E}_j)$  is a graph kernel defined in Eq. 1, which calculates the similarity between networks  $\tilde{E}_i$  and  $\tilde{E}_j$ .

From Eqs. 2-3, we can see that if two subjects have similar network architectures, they will be encouraged to be as close as possible after mapping. It is easy to see that Eq. 2 can is expected to well preserve the structural information of networks, by embedding the local-to-global structural information in the mapping process.

Based on the formulation in Eq. 2, the objective function of our proposed graph-kernel based structured feature selection (gk-SFS) model is defined as following:

$$\min_w \|Y - Xw\|_2^2 + \beta w^T X^T M X w + \lambda \|w\|_1, \quad (4)$$

where  $M$  is a Laplacian matrix defined in Eq. 2. Also,  $\beta$  and  $\lambda$  are two positive constants that balance the importances of three items. In practice, we usually use inner cross validation on the training data to determine their optimal values.

According to definition in Eq. 4, the objective function of our proposed gk-SFS method contains three items. The first item is a quadratic loss function that measures the difference between estimated and true values for training subjects. The second item is a graph-kernel-based Laplacian regularizer that preserves the structural information of network data. Here, we use the graph kernel to compute the similarity of networks, which can also capture the local and global structural information of networks, thus helping to learn more discriminative features. The last item is a sparsity regularizer with  $l_1$ -norm [51] that lead to a sparse solution of model. The features corresponding to non-zero factors in  $w$  will be selected for classification.

It is worth noting that, when  $\beta = 0$ , the proposed gk-SFS method will degenerate to the LASSO method [13]. And, the objective function defined in Eq. 4 can be effectively solved via using accelerated proximal gradient algorithm [52].

#### D. GK-SFS BASED DISEASE CLASSIFICATION

##### 1) NETWORK THRESHOLDING

Since weights of edges correspond to the Pearson correlation coefficient between ROIs, the constructed FCN of each subject is a full-connected weighted network. To characterize the topology of networks, we threshold FCNs of all subjects using a given value  $T$ . Specifically, for  $i$ -th FCN (*i.e.*,  $E_i$ ), we threshold it via the following formulation:

$$\tilde{A}_{p,q} = \begin{cases} 0, & \text{if } A_{p,q} < T \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

where  $A$  is the adjacency matrix of FCN  $E_i$ , and  $A_{p,q}$  is the element of matrix  $A$ , corresponding to the weight of edge between ROIs  $p$  and  $q$ . In this way, for any two ROIs  $p, q$ , there has an edge between  $p$  and  $q$  if and only if  $\tilde{A}_{p,q} = 1$ . Thus, we can obtain thresholded FCN  $\tilde{E}_i$  to represent the  $i$ -th subject for subsequent feature learning and classification.

##### 2) FEATURE LEARNING

Following [25] and [43], we first extract the clustering coefficient [53] from each ROI in the thresholded FCN as the feature, and then concatenate features from all 90 ROIs to a vector for representing each subject. In this way, each subject will be represented with a 90-dimensional feature vector. Based on these extracted features and thresholded FCNs, as shown in Fig. 1, we further perform our proposed gk-SFS method to select the most discriminative features for improving the classification performance.

##### 3) DISEASE CLASSIFICATION

Finally, we perform the linear support vector machine (SVM) with a default parameter (*i.e.*,  $C = 1$ ) for identifying patients

from HCs. Specifically, we employ the LIBSVM toolbox provided in [54] for SVM-based brain disease classification.

## IV. EXPERIMENTS

In this section, we first introduce the experimental settings, competing methods, and classification results. Then, we analyze experimental results, including the influence of parameters in the proposed gk-SFS method on classification performance and disease-related brain regions identified by the proposed method.

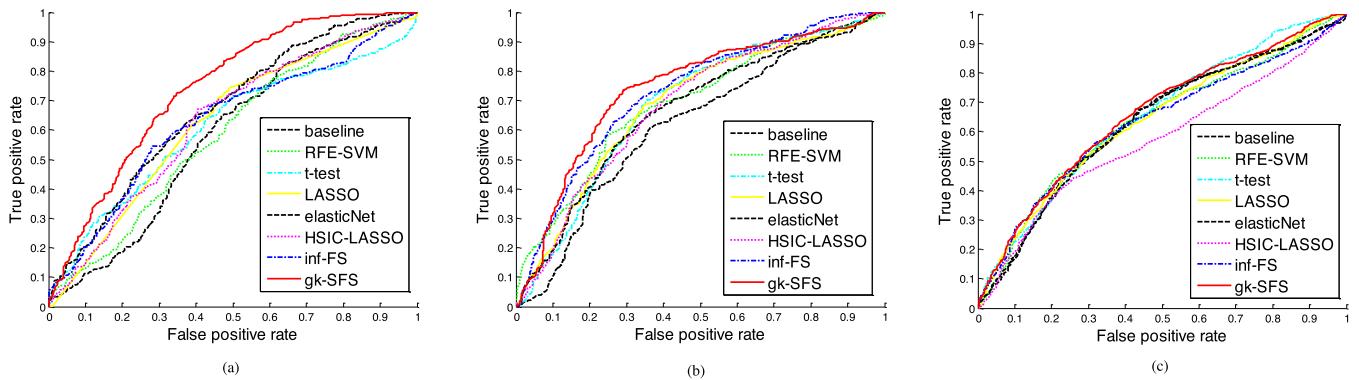
### A. EXPERIMENTAL SETTINGS

We extensively perform experiments to evaluate the performance of our proposed gk-SFS method. Specifically, we perform three tasks: 1) IMCI vs. eMCI, 2) eMCI vs. HC and 3) ADHD vs. HC classifications. A 10-fold cross validation strategy is used in the experiments. Specifically, for each classification task, all subjects are first equivalently partitioned into 10 subsets. In each fold cross validation, one subset is alternatively used as the testing data, and the remaining 9 subsets are combined as the training set. In addition, the process of data partition is independently repeated 10 times to avoid any bias.

We evaluate the performance of the proposed gk-SFS method via four evaluation metrics, including 1) accuracy, which is the proportion of subjects that are correctly classified samples in all samples, 2) sensitivity, which denotes the proportion of patients that are correctly classified, 3) specificity, which is the proportion of HCs that are correctly predicted, and 4) the area under the receiver operating characteristic (ROC) curve (AUC). We calculate the average performance in all folds of cross validation as final performance. We normalize each feature from the thresholded FCNs using its mean and standard deviation from all training subjects. In addition, we adopt the inner cross validation on training subjects to determine the optimal values of parameters used in the proposed gk-SFS.

### B. METHODS FOR COMPARISON

We compare our proposed gk-SFS feature selection method with several state-of-the-art methods, including SVM-based RFE method (called RFE-SVM) proposed in [32],  $t$ -test, LASSO, elasticNet, HSIC-LASSO [33] and inf-FS [30] methods. In all competing methods (*i.e.*, RFE-SVM,  $t$ -test, LASSO, elasticNet, HSIC-LASSO and inf-FS), the optimal parameter values are still determined via using inner cross validation on training data. In addition, we also perform the method without feature selection as the baseline for comparison. That is, in the baseline method, all features extracted from the thresholded FCNs are used for classification. For fair comparison, the linear SVM classifier with a default parameter (*i.e.*,  $C = 1$ ) is used for disease classification in all eight methods. In addition, we calculate the classification results of all eight methods with thresholds ranging from 0.1 to 0.8 with step size of 0.05, and present the best classification performance of those eight methods for fair comparison.



**FIGURE 2.** The ROC curves achieved by all eight methods in three classification tasks: (a) IMCI vs. eMCI, (2) eMCI vs. HC and (3) ADHD vs. HC.

**TABLE 2.** Classification performance of eight different methods in three classification tasks. ACC: ACCuracy; SEN: SENsitivity; SPE: SPEcificity.

Method	IMCI vs. eMCI				eMCI vs. HC				ADHD vs. HC			
	ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)	AUC
baseline	56.8	47.0	63.6	0.58	62.0	61.6	62.2	0.62	61.4	47.6	<b>73.1</b>	0.63
RFE-SVM	57.6	49.8	62.9	0.59	65.9	62.9	69.4	0.69	62.0	49.7	72.0	0.63
t-test	61.2	50.9	67.7	0.61	67.1	69.1	64.8	0.68	61.6	<b>57.0</b>	65.7	0.65
LASSO	61.2	<b>72.1</b>	52.3	0.63	66.4	67.0	65.6	0.69	61.6	49.8	71.1	0.64
elasticNet	62.8	61.2	63.4	0.67	64.6	65.2	63.6	0.66	61.8	49.2	72.5	0.63
HSIC-LASSO	62.1	60.9	62.3	0.63	64.0	61.1	66.6	0.68	61.0	41.5	76.2	0.57
inf-FS	64.4	53.3	72.1	0.63	68.6	67.5	69.8	0.73	61.9	47.2	74.2	0.63
gk-SFS (Ours)	<b>68.4</b>	64.9	<b>70.4</b>	<b>0.74</b>	<b>71.7</b>	<b>72.1</b>	<b>71.0</b>	<b>0.74</b>	<b>63.0</b>	52.1	71.6	<b>0.66</b>

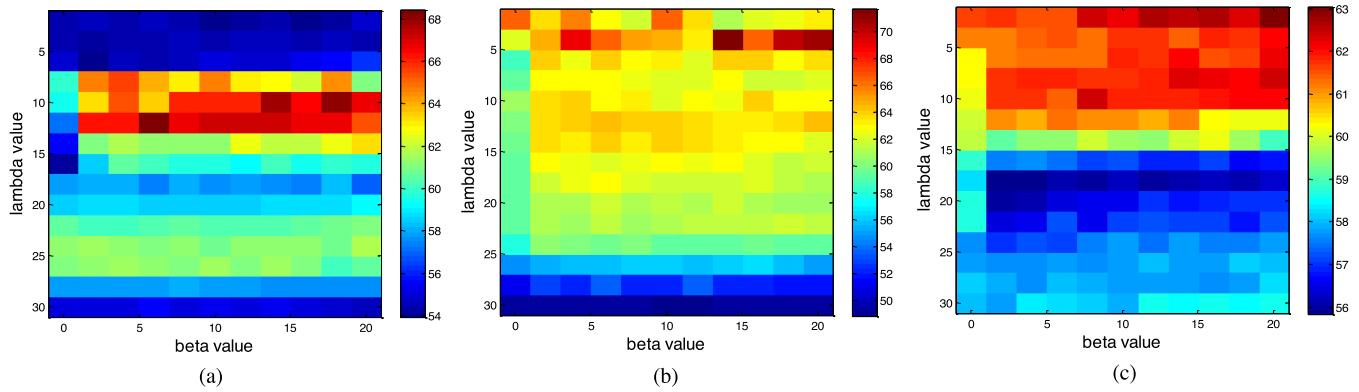
### C. CLASSIFICATION RESULTS

Table 2 summarizes the results of all eight methods in three classification tasks. Figure 2 plots the corresponding ROC curves achieved by these eight methods in three tasks. As can be seen from Table 2 and Fig. 2, compared with all competing methods, our proposed gk-SFS method can achieve better classification performance. For example, our proposed method achieves the accuracy of 68.4%, 71.7% and 63.0% for IMCI vs. eMCI, eMCI vs. HC and ADHD vs. HC classifications, respectively, while the best accuracy values achieved by the competing methods are 64.4%, 68.6% and 62.0%, respectively. Moreover, our proposed gk-SFS method obtains the AUC values of 0.74, 0.74 and 0.66 in three classification tasks, respectively. These results indicate that our proposed method can achieve good performance in AD/MCI classification and ADHD classification by selecting the most discriminative features via the proposed gk-SFS method. All these results indicate the efficacy of our proposed method in feature selection. From Table 2 and Fig. 2, we can also see that methods with feature selection (*i.e.*, RFE-SVM, *t*-test, LASSO, elasticNet, HSIC-LASSO, inf-FS and gk-SFS) can achieve better performance than the method without performing feature selection (*i.e.*, baseline), indicating the important contribution of feature selection for improving performance in brain disease classification.

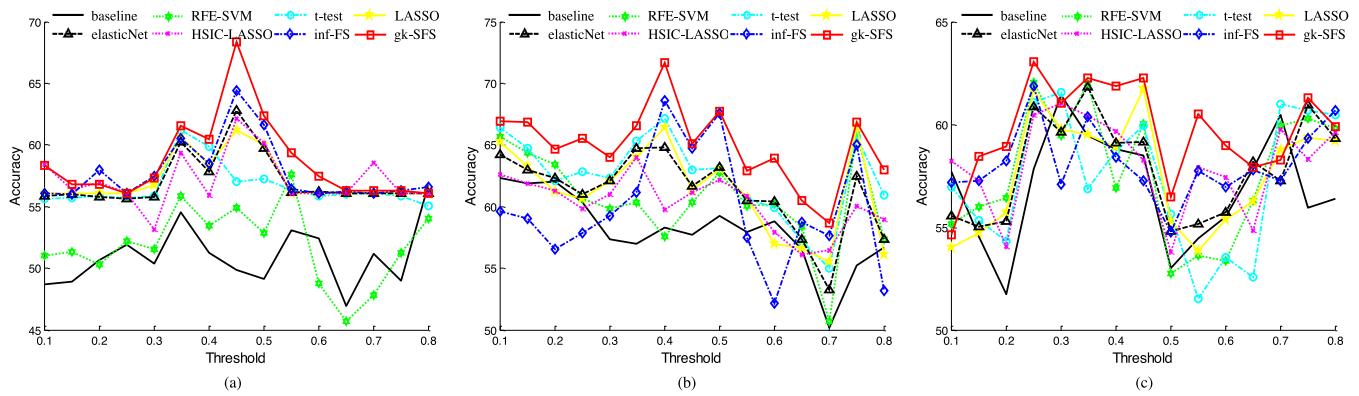
### D. EFFECT OF REGULARIZATION PARAMETERS

Our proposed gk-SFS method contains two regularization items, *i.e.*, Laplacian regularization item and sparsity regularization item. The first item is used to preserve the structural information of all networks during the mapping process, and the second item is used to control the sparsity of solution, and thus determining the number of selected features. The parameters, *i.e.*,  $\beta$  and  $\lambda$ , are used to balance the relative contributions of two items. To investigate the effects of these two parameters on performance of the proposed gk-SFS method, we perform three classification tasks with varying the value of  $\beta$  in the set of {2, 4, 6, 8, 10, 12, 14, 16, 18, 20}, varying the value of  $\lambda$  in the set of {2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30}. It is worth noting that a smaller  $\lambda$  value means that more features to be selected. For example, when  $\lambda = 0$ , all extracting features will be selected for classification. Also, when  $\beta = 0$ , the proposed gk-SFS method degenerates to a LASSO method. Figure 3 graphically shows the obtained ACC values *w.r.t.* different combinations of  $\beta$  and  $\lambda$  values in three classification tasks.

It can be seen from Fig. 3 (a)-(c) that our proposed gk-SFS method *w.r.t.* different combinations of  $\beta > 0$  and  $\lambda > 0$  consistently outperform the LASSO method (*i.e.*,  $\beta = 0$  and  $\lambda > 0$ ), indicating importance of introducing the Laplacian



**FIGURE 3.** The accuracy of the proposed gk-SFS method w.r.t. the combinations of  $\lambda$  and  $\beta$  values in the tasks of (a) IMCI vs. eMCI classification, (b) eMCI vs. HC classification and (c) ADHD vs. HC classification.



**FIGURE 4.** Results of eight different methods using different threshold values in three classification tasks.

regularization item. Also, Fig. 3 suggests that the classification performance of gk-SFS is largely affected by the  $\lambda$  value in three classification tasks. This implies that it is important to select the optimal  $\lambda$  value in our gk-SFS method. The possible reason is that the parameter  $\lambda$  controls the sparsity of solution in Eq. 4, and hence determines the number of selected features that are essential for the subsequent classification task.

### E. EFFECT OF THRESHOLD

In FCN-based studies, thresholding the fully-connected network is an essential step for exploring structural information of FCNs [23], [55]. Currently, we have no a gold standard to decide the optimal threshold. In practice, we often need to explore over a range of plausible thresholds to determine an appropriate one [6], [24]. To investigate the effect of thresholds on the performance of our proposed gk-SFS method, we also vary the values of threshold from 0.1 to 0.8 with the step size of 0.05. Figure 4 reports the classification accuracy of our proposed gk-SFS method w.r.t. different thresholds in three classification tasks. For comparison, we also show the accuracy of the competing methods (*i.e.*, baseline, RFE-SVM, *t*-test, LASSO, elasticNet, HSIC-LASSO and inf-FS) in three classification tasks.

As shown in Fig. 4, the proposed gk-SFS method generally outperforms those competing methods, suggesting the effectiveness of our method. Moreover, from Fig. 4 we can observe that our gk-SFS method usually achieves the good performance using the threshold value within the range of [0.2, 0.7] (*e.g.*, the optimal thresholds corresponding to the best performance of the proposed gk-SFS are 0.45, 0.4 and 0.25 for three classification tasks, respectively). This range corresponds to the average connection density (*i.e.*, the proportion of existing connections to all possible connections) of [80%, 30%], which is consistent with the existing study [6]. In addition, we can see from Fig. 4, the ACC values yielded by eight methods are largely affected by the threshold values in three classification tasks. The underlying reason is that FCNs with threshold values actually contain different topological properties of the networks, thus leading to different feature representations for subsequent classification tasks. Therefore, it's essential to select an optimal threshold value for FCN-based studies.

### F. IDENTIFIED BRAIN REGIONS

We now investigate which features (*i.e.*, brain regions) are selected by the proposed gk-SFS method for brain disease diagnosis. Since features selected in each fold of cross

**TABLE 3.** Important ROIs selected by our method in the tasks of eMCI vs. HC classification and ADHD vs. HC classification.

ROI Name (eMCI vs. HC)	ROI name (ADHD vs. HC)
R. Orbitofrontal cortex (superior)	L. Superior frontal gyrus (dorsal)
R. Middle frontal gyrus right	R. Superior frontal gyrus (dorsal)
R. Orbitofrontal cortex (medial)	R. Olfactory
L. Rectus gyrus	L. Superior frontal gyrus (media)
R. Posterior cingulate gyrus	L. Posterior cingulate gyrus
L. Hippocampus	L. Hippocampus
R. Hippocampus	L. ParaHippocampal gyrus
L. ParaHippocampal gyrus	L. Superior occipital gyrus
L. Amygdala	R. Superior occipital gyrus
L. Middle occipital gyrus	L. Middle occipital gyrus
L. Inferior occipital gyrus	L. Inferior occipital gyrus
L. Putamen	R. Supramarginal gyrus
L. Thalamus	L. Precuneus
R. Superior temporal gyrus	L. Paracentral lobule
L. Temporal pole (middle)	R. Paracentral lobule
L. Inferior temporal	R. Pallidum
R. Inferior temporal	R. Superior temporal gyrus
	R. Middle temporal gyrus
	R. Inferior temporal

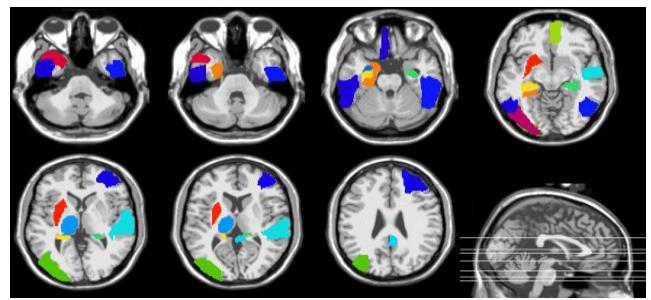
validation could be different, we select those features that occur in all 10-fold cross-validation as the most important features for classification. Table 3 lists these important brain regions identified by our gk-SFS method in eMCI vs. HC and ADHD vs. HC classifications, while Fig. 5 visually shows these important brain regions in the template space. As can be seen from Table 3 some important brain regions include *hippocampus*, *hippocampal gyrus*, *amygdala*, *cingulate*, *temporal pole* and *heschl gyrus*, which have been reported in previous AD/MCI studies [55]–[58], and some important brain regions, such as *occipital gyrus*, *paracentral lobule*, *cingulate gyrus* and *precuneus*, are also consistent with those reported in that previous ADHD studies [59]–[61].

## V. DISCUSSION

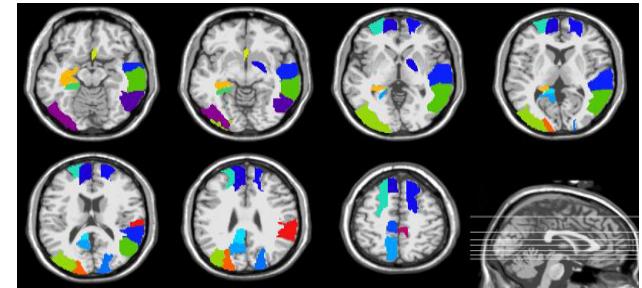
In this paper, we propose a novel graph-kernel based structural feature section (gk-SFS) for functional connectivity network based disease classification. The key of our proposed gk-SFS method is to capture the local-to-global structural information of networks via graph kernels to select the most discriminative features. We further develop a gk-SFS based learning framework for brain disease diagnosis. We validate the efficacy of our proposed method on two real fMRI datasets. The experimental results suggest that, compared with the existing methods, our proposed gk-SFS method can achieve better performance in three tasks of disease classification, thus can be potentially applied to the automated neuroimage-based diagnosis of brain diseases. Besides, the proposed method can help to identify disease-related brain regions, thus help us to better understand the pathology of brain diseases.

## A. SIGNIFICANT OF RESULTS

Feature selection has been widely applied to brain connectivity network analysis tasks. However, existing studies mainly



(a)



(b)

**FIGURE 5.** Important brain regions selected by our method in the tasks of (a) eMCI vs. HC classification and (b) ADHD vs. HC classification. (a) eMCI vs. HC classification. (b) ADHD vs. HC classification.

focus on using vector-based feature selection methods, thus ignoring the useful structural information of networks that could improve the performance of brain network classification. In this paper, we propose a novel structural feature selection method that can take account of local-to-global structural information of brain networks and thus select more discriminative features for brain disease classification. The experimental results on two real datasets show the effectiveness of our proposed method.

In addition, some important brain regions have been identified by our proposed gk-SFS method in the eMCI vs. HC classification task. These brain regions include *hippocampus* [56], [57], *hippocampal gyrus* [62], *amygdala* [63], *cingulate* [56], [58], *temporal pole* [64] and *heschl gyrus* [55], which have been reported in the previous AD/MCI studies. For example, it is widely reported that the *hippocampus* and *amygdala* formations are heavily damaged in early AD [21], [65]. These damaged brain regions may lead to changes of brain function, which could account for cognitive deficits in patients. Our results demonstrate that our proposed method can *not only* effectively identify patients with brain diseases from HCs by taking full account of structural information of brain networks, *but also* provide an evidence of altered network structure in patients with brain diseases.

## B. NETWORK THRESHOLDING

In brain connectivity network analysis, the thresholding process is usually used for exploring structural properties of networks [23], [24]. However, there is no gold standard to determine the optimal threshold currently. Previous studies usually explore properties of networks over a range of

applicable thresholds to decide the optimal one [6], [24]. Accordingly, in this paper, we calculate/compare the best classification performance of all eight methods over thresholds ranging from 0.1 to 0.8 with the step size of 0.05. The experimental results show that, compared with the competing methods, our proposed gk-SFS method can achieve the better classification accuracy. Further analysis shows that the proposed gk-SFS outperforms the competing methods over most of thresholds, as shown in Fig. 4. These results suggest the effectiveness of our proposed method. In addition, our study also shows that the FCNs with different thresholds exhibit different topological properties, and thus yield different feature representations and classification performances. This further indicates the importance of selecting the optimal threshold [6], [23], [24].

### C. LIMITATION

This study is currently limited by the following three aspects. *First*, we only use the similarity between a pair of networks as the structure information in our gk-SFS model. Actually, we could embed more prior knowledge (*e.g.*, effects of different brain regions by specific brain diseases) into the proposed method, which may further improve the performance of brain network analysis. *Second*, we separately perform feature selection and classification steps, which could lead to sub-optimal classification performance. As a further work, we will explore a unified learning framework to perform feature learning and classification jointly. *Finally*, only two datasets with rs-fMRI data are used for performance evaluation in the current study. In the future, we will further evaluate our proposed method on more datasets with a larger amount of subjects.

### VI. CONCLUSION

In this paper, we propose a novel graph-kernel based structured feature selection method for brain disease classification based on functional connectivity networks. Different from the existing methods that usually focus on vector-based features, our proposed gk-SFS method use the graph kernel to capture both the local-to-global structural information of functional connectivity networks, thus yielding more discriminative features for classification. Furthermore, we develop a gk-SFS based learning framework for automatic brain disease diagnosis. Experimental results of brain diseases on two real datasets with rs-fMRI data demonstrate that our proposed method can *not only* improve the performance of brain disease classification, *but also* identify disease-related brain regions that help to better understand the pathology of brain diseases.

### REFERENCES

- [1] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, “Functional connectivity in the resting brain: A network analysis of the default mode hypothesis,” *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 1, pp. 253–258, 2003.
- [2] C. Lian, M. Liu, J. Zhang, and D. Shen, “Hierarchical fully convolutional network for joint atrophy localization and Alzheimer’s disease diagnosis using structural MRI,” *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.
- [3] J. Zhang, Y. Gao, Y. Gao, B. C. Munsell, and D. Shen, “Detecting anatomical landmarks for fast Alzheimer’s disease diagnosis,” *IEEE Trans. Med. Imag.*, vol. 35, no. 12, pp. 2524–2533, Dec. 2016.
- [4] G. Chen et al., “Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging,” *Radiology*, vol. 259, no. 1, pp. 213–221, 2011.
- [5] M. Guye, G. Bettus, F. Bartolomei, and P. J. Cozzone, “Graph theoretical analysis of structural and functional connectivity MRI in normal and pathological brain networks,” *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 23, nos. 5–6, pp. 409–421, 2010.
- [6] M. Zanin et al., “Optimizing functional network representation of multivariate time series,” *Sci. Rep.*, vol. 2, Sep. 2012, Art. no. 630.
- [7] 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.
- [8] B. Jie, M. Liu, J. Liu, D. Zhang, and D. Shen, “Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer’s disease,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 238–249, Jan. 2017.
- [9] C.-Y. Wee et al., “Identification of MCI individuals using structural and functional connectivity networks,” *NeuroImage*, vol. 59, no. 3, pp. 2045–2056, 2012.
- [10] F. Liu et al., “Multivariate classification of social anxiety disorder using whole brain functional connectivity,” *Brain Struct. Funct.*, vol. 220, no. 1, pp. 101–115, Jan. 2015.
- [11] X. Wang, Y. Jiao, T. Tang, H. Wang, and Z. Lu, “Altered regional homogeneity patterns in adults with attention-deficit hyperactivity disorder,” *Eur. J. Radiol.*, vol. 82, no. 9, pp. 1552–1557, 2013.
- [12] B. Jie, C.-Y. Wee, D. Shen, and D. Zhang, “Hyper-connectivity of functional networks for brain disease diagnosis,” *Med. Image Anal.*, vol. 32, pp. 84–100, Aug. 2016.
- [13] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. Roy. Stat. Soc. B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
- [14] M. Liu and D. Zhang, “Sparsity Score: A novel graph-preserving feature selection method,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 28, no. 4, p. 1450009, 2014.
- [15] O. Sporns, “Contributions and challenges for network models in cognitive neuroscience,” *Nature Neurosci.*, vol. 17, no. 5, pp. 652–660, 2014.
- [16] T. Zhang et al., “Predicting functional cortical ROIs via DTI-derived fiber shape models,” *Cerebral Cortex*, vol. 22, no. 4, pp. 854–864, 2012.
- [17] M. R. Brier et al., “Functional connectivity and graph theory in preclinical Alzheimer’s disease,” *Neurobiol. Aging*, vol. 35, no. 4, pp. 757–768, 2014.
- [18] D. Zhang, J. Huang, B. Jie, J. Du, L. Tu, and M. Liu, “Ordinal pattern: A new descriptor for brain connectivity networks,” *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1711–1722, Jul. 2018.
- [19] N. Reggente et al., “Multivariate resting-state functional connectivity predicts response to cognitive behavioral therapy in obsessive-compulsive disorder,” *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 9, pp. 2222–2227, 2018.
- [20] J. Han et al., “Representing and retrieving video shots in human-centric brain imaging space,” *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, Jul. 2013.
- [21] F. Bai et al., “Abnormal functional connectivity of hippocampus during episodic memory retrieval processing network in amnestic mild cognitive impairment,” *Biol. Psychiatry*, vol. 65, no. 11, pp. 951–958, 2009.
- [22] K. Wang et al., “Altered functional connectivity in early Alzheimer’s disease: A resting-state fMRI study,” *Hum. Brain Mapping*, vol. 28, no. 10, pp. 967–978, Oct. 2007.
- [23] E. J. Sanz-Arigita et al., “Loss of ‘small-world’ networks in Alzheimer’s disease: Graph analysis of FMRI resting-state functional connectivity,” *PLoS ONE*, vol. 5, no. 11, pp. e13788-1–e13788-14, 2010.
- [24] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius, “Network analysis of intrinsic functional brain connectivity in Alzheimer’s disease,” *PLoS Comput. Biol.*, vol. 4, no. 6, pp. e1000100-1–e1000100-11, 2008.
- [25] C.-Y. Wee et al., “Enriched white matter connectivity networks for accurate identification of MCI patients,” *NeuroImage*, vol. 54, no. 3, pp. 1812–1822, 2011.
- [26] M. Liu and D. Zhang, “Feature selection with effective distance,” *Neurocomputing*, vol. 215, pp. 100–109, Nov. 2016.
- [27] F. de Vos et al., “A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer’s disease,” *NeuroImage*, vol. 167, pp. 62–72, Feb. 2017.
- [28] L. Zhu, L. Miao, and D. Zhang, “Iterative Laplacian score for feature selection,” in *Proc. Chin. Conf. Pattern Recognit.*, 2012, pp. 80–87.

- [29] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc. 20th Int. Conf. Int. Conf. Mach. Learn.*, 2003, pp. 856–863.
- [30] G. Roffo, S. Melzi, U. Castellani, and A. Vinciarelli, "Infinite latent feature selection: A probabilistic latent graph-based ranking approach," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 1398–1406.
- [31] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, 2002.
- [32] K. Yan and D. Zhang, "Feature selection and analysis on correlated gas sensor data with recursive feature elimination," *Sens. Actuators B, Chem.*, vol. 212, pp. 353–363, Jun. 2015.
- [33] M. Yamada, W. Jitkrittum, L. Sigal, E. P. Xing, and M. Sugiyama, "High-dimensional feature selection by feature-wise kernelized lasso," *Neural Comput.*, vol. 26, no. 1, pp. 185–207, 2014.
- [34] R. Chaves et al., "SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting," *Neurosci. Lett.*, vol. 461, no. 3, pp. 293–297, 2009.
- [35] C. Chu, A.-L. Hsu, K.-H. Chou, P. Bandettini, and C. Lin, "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images," *NeuroImage*, vol. 60, no. 1, pp. 59–70, Mar. 2012.
- [36] A. Y. Ng, "Feature selection,  $L_1$  vs.  $L_2$  regularization, and rotational invariance," in *Proc. 21st Int. Conf. Mach. Learn.*, Banff, AB, Canada, 2004, pp. 78–85.
- [37] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
- [38] S. V. N. Vishwanathan, N. N. Schraudolph, I. R. Kondor, and K. M. Borgwardt, "Graph kernels," *J. Mach. Learn. Res.*, vol. 11, pp. 1201–1242, Mar. 2010.
- [39] N. Shervashidze, P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *J. Mach. Learn. Res.*, vol. 12, pp. 2539–2561, Sep. 2011.
- [40] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-spectral remote sensing image classification with graph kernels," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 741–745, Oct. 2010.
- [41] Y. Zhang, H. Lin, Z. Yang, and Y. Li, "Neighborhood hash graph kernel for protein–protein interaction extraction," *J. Biomed. Inform.*, vol. 44, no. 6, pp. 1086–1092, 2011.
- [42] W. Shao, M. Liu, and D. Zhang, "Human cell structure-driven model construction for predicting protein subcellular location from biological images," *Bioinformatics*, vol. 32, no. 1, pp. 114–121, 2015.
- [43] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 576–589, Feb. 2014.
- [44] F. Mokhtari, S. K. Bakhtiari, G. A. Hosseini-Zadeh, and H. Soltanian-Zadeh, "Discriminating between brain rest and attention states using fMRI connectivity graphs and subtree SVM," *Proc. SPIE*, vol. 8314, Feb. 2012, Art. no. 83144C.
- [45] K. M. Borgwardt and H.-P. Kriegel, "Shortest-path kernels on graphs," in *Proc. 5th IEEE Int. Conf. Data Mining*, Nov. 2005, pp. 1–8.
- [46] T. Gärtner, P. Flach, and S. Wrobel, "On graph kernels: Hardness results and efficient alternatives," in *Learning Theory and Kernel Machines*. Berlin, Germany: Springer, 2003, pp. 129–143.
- [47] B. Jie, M. Liu, D. Zhang, and D. Shen, "Sub-network kernels for measuring similarity of brain connectivity networks in disease diagnosis," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2340–2353, May 2018.
- [48] N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.
- [49] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, pp. S61–S72, Mar. 2009.
- [50] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
- [51] M. Liu and D. Zhang, "Pairwise constraint-guided sparse learning for feature selection," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 298–310, Jan. 2016.
- [52] J. Liu and J. Ye, "Efficient  $L_1/L_q$  norm regularization," Arizona State Univ., Tempe, AZ, USA, Tech. Rep., 2009, pp. 1–19.
- [53] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [54] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.
- [55] Z. Liu et al., "Altered topological patterns of brain networks in mild cognitive impairment and Alzheimer's disease: A resting-state fMRI study," *Psychiatry Res., Neuroimag.*, vol. 202, no. 2, pp. 118–125, 2012.
- [56] Y. Zhou, J. H. Dougherty, Jr., K. F. Hubner, B. Bai, R. L. Cannon, and R. K. Hutson, "Abnormal connectivity in the posterior cingulate and hippocampus in early Alzheimer's disease and mild cognitive impairment," *Alzheimer's Dementia*, vol. 4, no. 4, pp. 265–270, 2008.
- [57] F. Bai et al., "Abnormal whole-brain functional connection in amnestic mild cognitive impairment patients," *Behav. Brain Res.*, vol. 216, no. 2, pp. 666–672, 2011.
- [58] S. D. Han et al., "Functional connectivity variations in mild cognitive impairment: Associations with cognitive function," *J. Int. Neuropsychol. Soc.*, vol. 18, no. 1, pp. 39–48, 2012.
- [59] J. C. Mostert et al., "Characterising resting-state functional connectivity in a large sample of adults with ADHD," *Prog. Neuropsychopharmacol. Biol. Psychiatry*, vol. 67, pp. 82–91, Jun. 2016.
- [60] A. dos S. Siqueira, C. E. Biazoli, Jr., W. E. Comfort, L. A. Rohde, and J. R. Sato, "Abnormal functional resting-state networks in ADHD: Graph theory and pattern recognition analysis of fMRI data," *Biomed. Res. Int.*, vol. 2014, no. 4, 2014, Art. no. 380531.
- [61] T. J. Silk, A. Vance, N. Rinehart, J. L. Bradshaw, and R. Cunningham, "Dysfunction in the fronto-parietal network in attention deficit hyperactivity disorder (ADHD): An fMRI study," *Brain Imag. Behav.*, vol. 2, no. 2, pp. 123–131, 2008.
- [62] B. Cheng, M. Liu, H.-I. Suk, D. Shen, and D. Zhang, "Multimodal manifold-regularized transfer learning for MCI conversion prediction," *Brain Imag. Behav.*, vol. 9, no. 4, pp. 913–926, 2015.
- [63] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiol. Aging*, vol. 32, no. 12, pp. 2322.e19–2322.e27, Dec. 2011.
- [64] F. Nobili et al., "Principal component analysis of FDG PET in amnestic MCI," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 35, no. 12, pp. 2191–2202, 2008.
- [65] H. Yao et al., "Decreased functional connectivity of the amygdala in Alzheimer's disease revealed by resting-state fMRI," *Eur. J. Radiol.*, vol. 82, no. 9, pp. 1531–1538, Sep. 2013.



**MI WANG** is currently pursuing the M.S. degree with Anhui Normal University, China. Her research interests include machine learning and medical image analysis.



**BIAO JIE** received the M.S. degree in computer science from Yunnan Normal University, Yunnan, China, in 2006, and the Ph.D. degree in computer science from the Nanjing University of Aeronautics and Astronautics, China, in 2015. In 2006, he joined the School of Computer and Information, Anhui Normal University, where he is currently a Professor. His research interests include machine learning and medical image analysis.



**WEIXIN BIAN** received the Ph.D. degree from the China University of Mining and Technology. He is currently an Associate Professor with the School of Computer and Information, Anhui Normal University. His research interests include information science, image processing, machine learning, and pattern recognition.



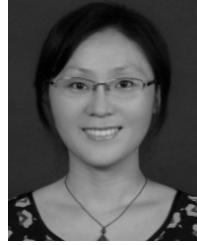
**ZHENGDONG WANG** is currently pursuing the M.S. degree with Anhui Normal University, China. Her research interests include machine learning and medical image analysis.



**XINTAO DING** received the M.S. degree in computational mathematics from East China Normal University, in 2005, and the Ph.D. degree in geography and tourism from Anhui Normal University, China, in 2015, where he is currently an Associate Professor. His research interests include machine learning and pattern recognition.



**WEN ZHOU** received the Ph.D. degree from the School of Software Engineering, Tongji University, in 2018. In 2018, he joined the School of Computer and Information, Anhui Normal University, where he is currently a Lecturer. His research interests include computer vision and machine learning.



**MINGXIA LIU** received the B.S. and M.S. degrees from Shandong Normal University, Shandong, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2015. Her current research interests include machine learning, pattern recognition, and neuroimaging analysis.

• • •