



# Machine Learning Classification Combining Multiple Features of A Hyper-Network of fMRI Data in Alzheimer's Disease

Hao Guo<sup>1</sup>, Fan Zhang<sup>1</sup>, Junjie Chen<sup>1</sup>, Yong Xu<sup>2</sup> and Jie Xiang<sup>1\*</sup>

<sup>1</sup> Department of Software Engineering, College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, China, <sup>2</sup> Department of Psychiatry, First Hospital of Shanxi Medical University, Taiyuan, China

## OPEN ACCESS

### Edited by:

Lingzhong Fan,  
Institute of Automation (CAS), China

### Reviewed by:

Veena A. Nair,  
University of Wisconsin-Madison,  
United States

Yu Zhang,  
Montreal Neurological Institute and  
Hospital, Canada

### \*Correspondence:

Jie Xiang  
xiangjie\_tyut@sina.com

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 19 June 2017

**Accepted:** 23 October 2017

**Published:** 21 November 2017

### Citation:

Guo H, Zhang F, Chen J, Xu Y and  
Xiang J (2017) Machine Learning  
Classification Combining Multiple  
Features of A Hyper-Network of fMRI  
Data in Alzheimer's Disease.  
*Front. Neurosci.* 11:615.  
doi: 10.3389/fnins.2017.00615

Exploring functional interactions among various brain regions is helpful for understanding the pathological underpinnings of neurological disorders. Brain networks provide an important representation of those functional interactions, and thus are widely applied in the diagnosis and classification of neurodegenerative diseases. Many mental disorders involve a sharp decline in cognitive ability as a major symptom, which can be caused by abnormal connectivity patterns among several brain regions. However, conventional functional connectivity networks are usually constructed based on pairwise correlations among different brain regions. This approach ignores higher-order relationships, and cannot effectively characterize the high-order interactions of many brain regions working together. Recent neuroscience research suggests that higher-order relationships between brain regions are important for brain network analysis. Hyper-networks have been proposed that can effectively represent the interactions among brain regions. However, this method extracts the local properties of brain regions as features, but ignores the global topology information, which affects the evaluation of network topology and reduces the performance of the classifier. This problem can be compensated by a subgraph feature-based method, but it is not sensitive to change in a single brain region. Considering that both of these feature extraction methods result in the loss of information, we propose a novel machine learning classification method that combines multiple features of a hyper-network based on functional magnetic resonance imaging in Alzheimer's disease. The method combines the brain region features and subgraph features, and then uses a multi-kernel SVM for classification. This retains not only the global topological information, but also the sensitivity to change in a single brain region. To certify the proposed method, 28 normal control subjects and 38 Alzheimer's disease patients were selected to participate in an experiment. The proposed method achieved satisfactory classification accuracy, with an average of 91.60%. The abnormal brain regions included the bilateral precuneus, right parahippocampal gyrus\hippocampus,

right posterior cingulate gyrus, and other regions that are known to be important in Alzheimer's disease. Machine learning classification combining multiple features of a hyper-network of functional magnetic resonance imaging data in Alzheimer's disease obtains better classification performance.

**Keywords:** fMRI, hyper-network, multi-feature, discriminative subgraph, Alzheimer's disease

## INTRODUCTION

Modern imaging techniques provide effective approaches for exploring the functional interactions among brain regions, increasing our understanding of the pathological basis of mental illnesses. Brain functional network approaches provide a simplified representation of brain interaction patterns, and have been successfully used to classify neurological disorders (Stam et al., 2009; Pievani et al., 2011; Wang et al., 2013). The application of brain functional networks to neurocognitive theory has attracted much attention and recognition from researchers (Richardson, 2010), and they are widely used in the study of brain diseases, including schizophrenia (Bassett et al., 2008; Liu et al., 2008; Lynall et al., 2010), depression (Liu F. et al., 2015), mild cognitive impairment (Liang et al., 2013), attention deficit hyperactivity disorder (Wang et al., 2009), and Alzheimer's disease (AD) (He et al., 2008; Supekar et al., 2008).

Because network structures are composed of nodes and edges, functional brain network analysis provides an important tool for systematically detecting abnormalities in several brain regions. Differences in network topology between normal controls and brain disease patients can provide useful biomarkers for diagnosis, and for understanding the pathological underpinnings of brain diseases. Thus, modeling of functional networks can play an essential role in accurate diagnosis. Many previous studies have reported that higher cognition arises from interactions among many different brain regions, rather than activities in isolated brain regions. A major symptom in many mental disorders is a sharp decline in cognitive ability, which can be related to abnormal connectivity patterns (Delbeuck et al., 2003; Horwitz, 2003) involving interactions among multiple brain regions.

So far, many functional connectivity modeling methods have been proposed, including correlation-based methods (Bullmore and Sporns, 2009), graphical models (Bullmore et al., 2000), partial correlation-based methods (Rosa et al., 2015), and sparse representation methods (Smith et al., 2011; Wee et al., 2014). However, there are some flaws in the conventional methods of constructing functional networks. Most of them use correlation-based methods, which are relatively sensitive for detecting network connections (Smith et al., 2011). Nevertheless, because most network modeling methods are based on correlations, they are only able to reflect relationships between paired brain regions, which does not fully characterize the multi-level information among multiple brain regions, and ignores the higher-order relationships that are important for disease diagnosis. Moreover, network models based on correlational methods may contain false connections, because of the arbitrary selection of thresholds (Biao et al., 2014; Jie et al., 2016). Other

methods of studying brain connectivity have been proposed, including graphical models such as structural equation models (McIntosh et al., 1994; Bullmore et al., 2000) and dynamic causal models (Marreiros et al., 2010). However, most of these methods are confirmative rather than exploratory, which makes them inadequate for studying brain connectivity in AD and mild cognitive impairment (MCI) because they often require a prior knowledge—such as which brain regions should be involved and how they are connected—that is usually unavailable (Huang et al., 2010). Partial correlation estimation can be implemented using the maximum likelihood estimation (MLE) of the inverse covariance matrix. However, using this method, the required sample size to obtain sufficient data for reliable estimation is much larger than the number of modeled brain regions (Jie et al., 2016).

Conventional methods for constructing functional networks typically model the relationships between pairwise brain regions. However, recent studies have reported the importance of interactions among multiple brain regions, in addition to the relationships between pairwise brain regions. In one study, Yu et al. (2011) demonstrated that higher-order interactions are inherent properties of cortical dynamics. Santos et al. (2010) reported that the recorded activity of units in pairwise interactions was not best described by neuronal activity patterns. To address this limitation, they constructed a hierarchical model of network interactions, using units of interactions at two spatial levels. The results suggested that hierarchical models can capture network interactions more accurately than pairwise models. Montani et al. (2009) modeled the impact of high-order interactions on the amount of somatosensory information transmitted by the rate of synchronous discharge. Taken together, these results suggest that higher-order interactions play an important role in the dynamics of neural networks. Moreover, some studies have also suggested that functional interactions among single brain regions can interact with several other brain regions (Huang et al., 2010). Therefore, correlational analysis reflecting pairwise information may not be able to characterize the higher-order interactions of many brain regions working together. However, this information may be crucial for understanding the pathological mechanisms underlying mental illness.

In view of the shortcomings of the conventional functional connectivity network models, many new methods of construction have been developed. Hyper-networks are one example. A properly constructed hyper-network can overcome the above disadvantages of conventional methods. Hyper-networks based on hyper-graph theory can represent the interactions among multiple brain regions (Biao et al., 2014; Jie et al., 2016). Recently, Jie et al. (Biao et al., 2014)

constructed a hyper-network for an MCI dataset, extracted the local brain region properties as features, and then selected the most discriminative features for classification. Jie et al. (2016) similarly constructed a hyper-network for an attention deficit hyperactivity disorder (ADHD) dataset and extracted the brain region properties as features. They compared the hyper-network results with those of the conventional functional connectivity network methods and verified the robustness of various technologies.

However, the above classification methods extracted the local brain region properties as features, so that feature selection and classification could be implemented. Brain region features, including global properties (clustering coefficient  $D_j$  and  $Sh$ , 1998, path length Saramäki et al., 2006, etc.) and local properties (degree, betweenness centrality Barthélemy, 2004, etc.), have been widely used in previous studies for the classification of diseases in connectivity networks. However, such extracted features may lose some useful information, especially global topological information (Zhou et al., 2014). Subgraph features that are extracted from the graph-structure have been widely applied in the diagnosis of brain diseases (Montani et al., 2009; Huang et al., 2010) and can effectively compensate for the defects of conventional feature-extraction methods. However, subgraph feature-based methods have the drawback of being insensitive to change in a single brain region (Zhou et al., 2014). Therefore, both types of methods can lead to the loss of sample information (Zhou et al., 2014). In addition, the brain network itself is a complex network structure and its biological features cannot be captured from the perspective of a single feature.

To solve the problems of conventional methods, we developed a novel method that uses machine learning classification to combine multiple features of hyper-network functional magnetic resonance imaging (fMRI) data in AD. Specifically, based on the resting-state fMRI time sequence, we constructed a hyper-network with a sparse representation method. In the current study, to address the limitations of conventional network modeling, we combined different types of features, including brain region features and subgraph features. Three types of clustering coefficients were extracted as features and a non-parametric test was applied for feature selection. The subgraph feature-based method extracted hyper-edges as features and selected them using the frequently scoring feature selection (FSFS) method. Finally, two types of kernels based on multi-kernel support vector machine (SVM) classification were combined. The study constructed hyper-networks for 38 AD patients and 28 normal subjects and verified them. The results showed that the proposed machine learning classification method combining multiple features of a hyper-network of fMRI data in AD achieved satisfactory classification performance.

The main work of this study was as follows. First, the hyper-network construction method was applied to construct a network model based on an AD dataset to analyze the interactions among multiple brain regions. Second, different from previous studies, this study extracted two types of hyper-network features—brain region features and subgraph features—to ensure the integrity of the network topology information and preserve the sensitivity to change in a single brain region. Third, a multi-kernel SVM was

proposed for the hyper-network, which combines two types of network features to achieve better classification performance.

## MATERIALS AND METHODS

### Method Framework

A flowchart of the proposed framework for machine learning classification combining multiple features of a hyper-network of fMRI data in AD is presented in **Figure 1**. Specifically, the framework consists of several major steps.

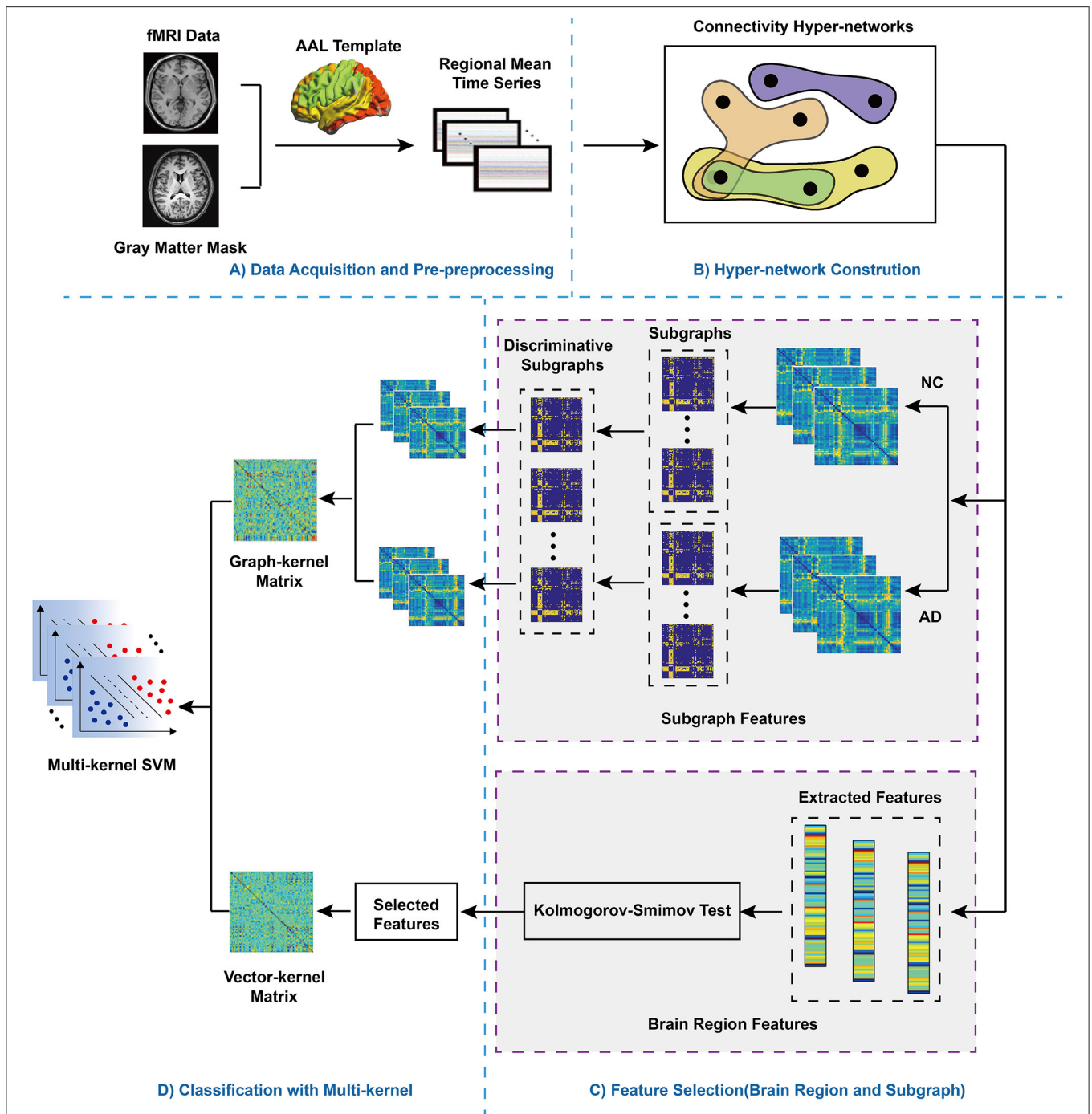
1. Data acquisition and pre-processing.
2. Construction of the hyper-network: for each subject, we constructed a hyper-network using a sparse linear regression model that estimated a region using a linear combination of the times series of other regions, and optimized the objective function by sparse learning.
3. Feature extraction and selection: non-parametric tests were performed to select the brain region features and the FSFS algorithm was used to select discriminative subgraphs; then, the corresponding kernel matrix was computed.
4. Multi-kernel SVM: multi-kernel SVM was used for classification of the kernel matrixes with brain region features and subgraph features combined.

### Data Acquisition and Pre-processing

This study was carried out in accordance with the recommendations of the medical ethics committee of Shanxi province (reference number: 2012013). All subjects gave their written informed consent in accordance with the Declaration of Helsinki. Twenty-eight healthy right-handed participants and thirty-eight major depression disorder participants underwent resting-state fMRI in a 3T MR scanner (Siemens Trio 3-Tesla scanner, Siemens, Erlangen, Germany). The subjects' demographic information and clinical characteristics are summarized in **Table 1**. Data collection was completed at the First Hospital of Shanxi Medical University. All scans were performed by radiologists who were familiar with MRI. All patients underwent a complete physical and neurological examination, standard laboratory tests, and an extensive neuropsychological assessment battery.

During the scan, participants were asked to relax with their eyes closed but not to fall asleep. The scanning parameters were set as follows: axial slices = 33, repetition time (TR) = 2,000 ms, echo time (TE) = 30 ms, thickness/skip = 4/0 mm, field of view (FOV) = 192 × 192 mm, matrix = 64 × 64 mm, flip angle = 90°, volumes = 248. The first 10 volumes of each time series were discarded to allow for magnetization stabilization. See Supplemental Text S1 for details of the scanning parameters.

Data preprocessing was performed with SPM8 (Statistical Parametric Mapping, SPM) (Friston, 2007). First, slice-timing correction and head-movement correction were carried out. Two samples exhibiting more than 3.0 mm of translation and 3.0° of rotation were discarded, which were not included in the final 28 samples. The corrected images were optimized with a 12-dimensional affine transformation and normalized to 3 × 3 × 3 mm voxels in the Montreal Neurological Institute (MNI)



**FIGURE 1 |** Flowchart of the proposed method. **(A)** Data Pre-processing. AAL: automatic anatomic labeling. fMRI: functional magnetic resonance imaging. After fMRI data pre-processing, according selected AAL template, the whole brain was divided into 90 regions. Then, the mean regional time series were extracted to divided brain regions, **(B)** Hyper-network Construction. **(C)** Feature Selection (brain region and subgraph). For subgraph features, NC, normal control; AD, Alzheimer's disease, hyperedges were extracted respectively from the hyper-networks constructed the NC group and the AD group, which were regarded as subgraph features, so two groups of subgraph features were obtained. In addition, for brain region features, the values of three different clustering coefficient respectively were computed, then Kolmogorov-Smimov test was adopted to feature selection, which obtained discriminative features. **(D)** Classification with multi-kernel SVM. Two different types of kernel matrix were combined, adopting multi-kernel SVM for classification.

standard space. Finally, linear detrending and band-pass filtering (0.01–0.10 Hz) were performed to reduce the effects of low-frequency drift and high-frequency physiological noise.

### Construction of the Hyper-Network

Most previous studies have used the simple-graph to construct network models, which only characterize information between

**TABLE 1** | Demographics and clinical characteristics of the subjects.

	NC	AD	P-value
Age(years)	(68–76) 72.6 ± 3.42	(66–76) 71.4 ± 4.68	0.44 <sup>a</sup>
Sex (male/female)	13/15	15/23	0.57 <sup>b</sup>
Handedness (R/L)	28/0	38/0	–
MMSE	(23–30) 26.1 ± 3.2	(20–25) 22.8 ± 2.1	<0.0001 <sup>a</sup>

Values are mean + standard deviation; AD, Alzheimer’s disease; NC, normal controls; MMSE, Mini-Mental State Examination; \*All tests are two-tailed; <sup>a</sup>Two-sample t-test; <sup>b</sup>Pearson Chi-square test.

pairwise brain regions. In the current study, we constructed a hyper-network connectivity model based on hyper-graph theory, which can reflect higher-order interactions among multiple brain regions. A hyper-graph is an expansion based on a simple-graph, and the approach has been widely used in numerous fields. The hyper-graph is summarized as follows.

### Hyper-Graphs

To date, hyper-graph theory has been successfully used for many applications, including image classification (Yu et al., 2012) and protein function prediction (Gallagher and Goldberg, 2013). In the field of neuroimaging, graph theory is commonly used to analyze brain connectivity (Kaiser, 2011; Sporns, 2012; Fornito et al., 2013). In the traditional graph theory approach, every edge merely links two nodes with a particular relationship, meaning that it only reflects the interactions between two nodes. In addition to paired relationships, such as functional interactions among multiple brain regions, many scenarios involve higher-order relationships, which simple graphs cannot describe. To address this limitation, some researchers have proposed the use of hyper-graphs, which are able to reflect the higher-order relationships among multiple nodes. Generally, a hyper-graph can be represented by an extension of a conventional simple graph in which one hyper-edge links two or more nodes (Schölkopf et al., 2007).

A hyper-graph is represented by  $G = (V, E)$ , where  $V$  denotes a set of nodes and  $E$  represents a set of hyper-edges. We can then use a  $|V| \times |E|$  incidence matrix  $H$  to denote  $G$ , where  $H$  is represented by the following elements:

$$H(v, e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (1)$$

where  $v \in V$  indicates a node of  $G$ , and  $e \in E$  indicates a hyper-edge of  $G$ .

Based on  $H$ , the node degree of each node  $v \in V$  can be represented as:

$$d(v) = \sum_{e \in E} H(v, e) \quad (2)$$

The edge degree of hyper-edge  $e \in E$  can be represented as:

$$\delta(e) = \sum_{v \in V} H(v, e) \quad (3)$$

Let  $D_v$  and  $D_e$  represent the diagonal matrices of node degrees  $d(v)$  and hyper-edge degrees  $\delta(e)$ :

$$A = HH^T - D_v \quad (4)$$

where  $H^T$  is the transpose of  $H$ .  $A(i, j)$  represents the number of hyper-edges that contain nodes  $v_i$  and  $v_j$ .

Notably, the traditional graph is a specific hyper-graph where one hyper-edge includes only two nodes. An example of a hyper-graph is shown in **Figure 2**: **Figure 2A** displays a conventional graph; **Figure 2B** shows a hyper-graph; and **Figure 2C** is an incidence matrix for the hyper-graph in **Figure 2B**, where 0 indicates that there is no connection between the nodes in the corresponding row and column and 1 indicates that there is such a connection.

### Construction of Hyper-Networks

Based on the sparse linear regression model, hyper-networks were constructed from rs-fMRI time series (McIntosh et al., 1994). Specifically,  $X = [x_1, \dots, x_m, \dots, x_M]^T \in R^{M \times d}$  is used to denote subjects with a total of  $M$  regions of interest (ROIs), where  $x_m$  denotes the regional mean time series of the designated  $m$ -th ROI, and  $d$  is the length of the time series. A response vector  $x_m$  is then denoted by the regional mean time series of each ROI, which can be represented by adopting a linear combination of the time series of other  $M - 1$  ROIs, as follows:

$$x_m = A_m \alpha_m + \tau_m, \quad m = 1, 2, \dots, M \quad (5)$$

where  $A_m = [x_1, \dots, x_{m-1}, 0, x_{m+1}, \dots, x_M]$  denotes a data matrix that contains all time series except the  $m$ -th ROI, in which the regional mean time series was set a vector of all zeros),  $\alpha_m$  denotes the weight vector to indicate the degree of the effect of other ROIs on the  $m$ -th ROI, and  $\tau_m \in R^d$  represents a noise term. It should be noted that the element in  $\alpha_m$  indicates that its corresponding ROI is meaningless for accurately evaluating the time series of the  $m$ -th ROI.

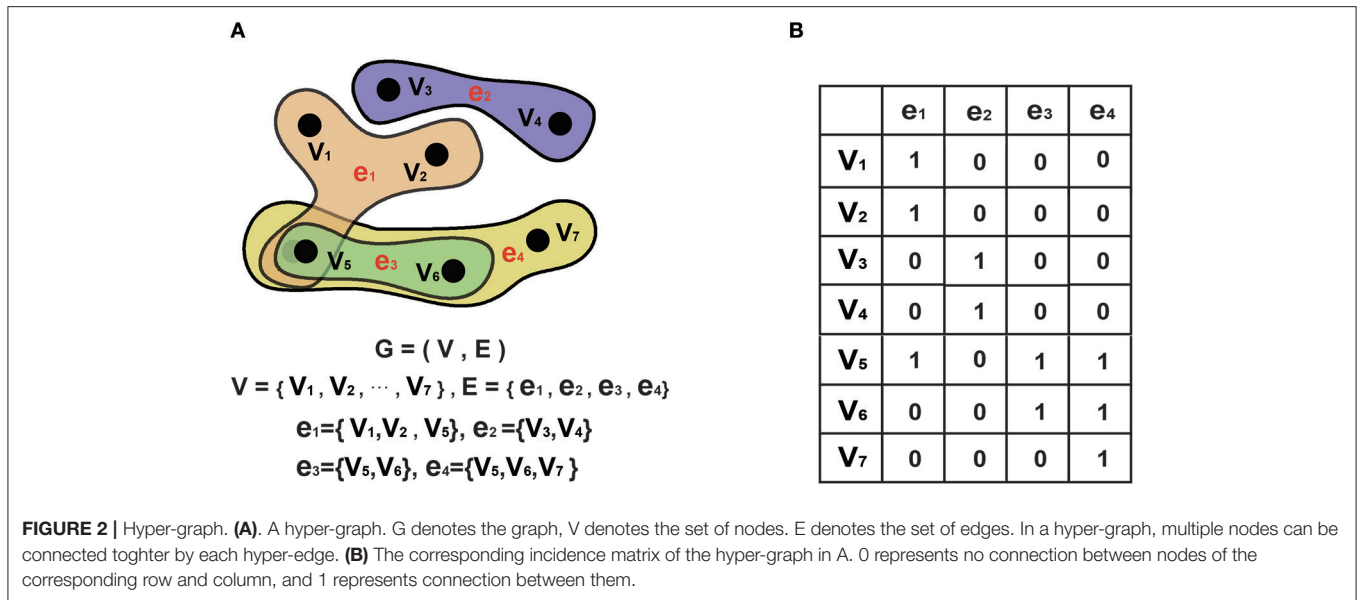
For solving sparse linear regression models, this optimization target function is as follows:

$$\min_{\alpha_m} \|x_m - A_m \alpha_m\|_2 + \lambda \|\alpha_m\|_0 \quad (6)$$

This is a well-known non-deterministic polynomial (NP) problem, owing to the  $l_0$ -norm term. Meanwhile, this method is usually approximately equal to solving a standard  $l_1$ -norm regularized optimization problem through a target function (Wright et al., 2009), as follows:

$$\min_{\alpha_m} \|x_m - A_m \alpha_m\|_2 + \lambda \|\alpha_m\|_1 \quad (7)$$

where  $\lambda > 0$  denotes a regularization parameter to control the levels of sparsity of the model. Different  $\lambda$  values correspond to different solutions of the degree of sparsity, and a larger  $\lambda$  value represents a sparser model, which indicates that there are more zero elements in  $\alpha_m$ . Most existing sparse learning algorithms can be implemented to solve the  $l_1$ -norm, such as least angle regression (Statistics, 1998). Adopting the sparse linear regression model, we can obtain one brain region’s interactions



**TABLE 2 |** The definitions and formulas of hyper-network properties.

Properties	Definitions	Formulas
HCC <sub>1</sub>	The number of adjacent nodes that have connections not facilitated by node <i>v</i>	$HCC_1(v) = \frac{2 \sum_{u,t \in N(v)} I(u,t,\neg v)}{ N(v) ( N(v)-1 )}$
HCC <sub>2</sub>	The number of adjacent nodes that have connections facilitated by node <i>v</i>	$HCC_2(v) = \frac{2 \sum_{u,t \in N(v)} I(u,t,v)}{ N(v) ( N(v)-1 )}$
HCC <sub>3</sub>	The amount of overlap among adjacent hyper-edges of node <i>v</i>	$HCC_3(v) = \frac{2 \sum_{e \in S(v)} ( e -1) -  N(v) }{ N(v) ( S(v)-1 )}$

with other brain regions, while setting the irrelevant or false connections to zero. This means that, in the weight vector  $\alpha_m$ , the brain regions corresponding to zero elements are regarded as irrelevant for estimating one region's time series. Thus, this approach provides a method for modeling interactions among one brain region and other brain regions by eliminating irrelevant connections.

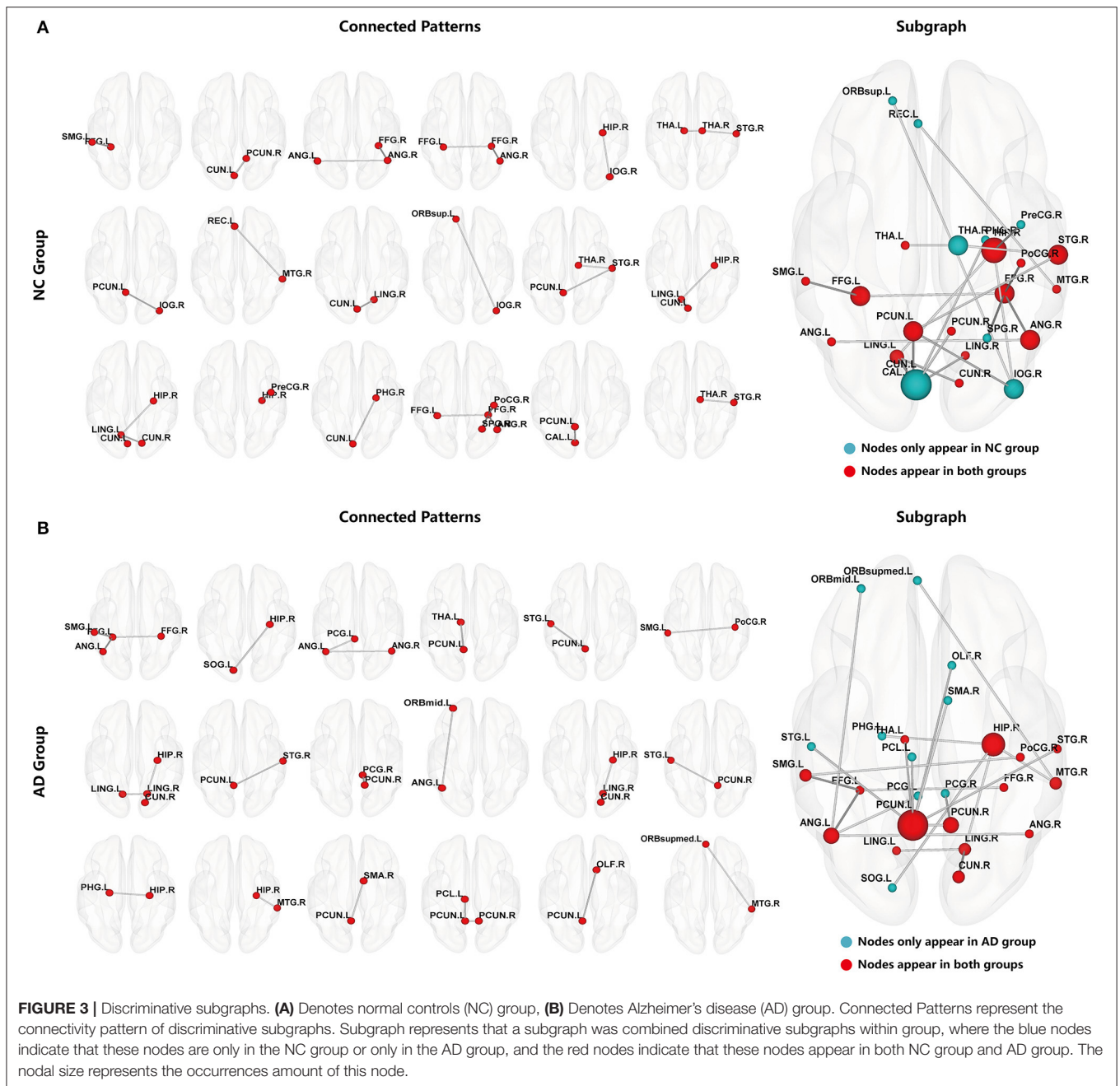
In this study, for each subject, we constructed a hyper-network by adopting a sparse linear regression model, where a node is represented by one brain region, and a hyper-edge  $e_m$  contains the  $m$ -th ROI and other ROIs with corresponding non-zero elements in the weight vector  $\alpha_m$ , which is computed by Equation (7). As a regularization parameter,  $\lambda$  controls the amount of non-zero solutions of the sparsity vector  $\alpha_m$ . In the extreme situation, when  $\alpha_m$  obtains all the zero solutions,  $\lambda$  could get the maximum value which is always denoted as  $\lambda_{max}$ . On the contrary, when  $\alpha_m$  obtains all the non-zero solutions,  $\lambda$  could get the minimum value which is a positive number close to 0, denoted as  $\lambda_{min}$ . Thus the value of  $\lambda$  should be set ranging from  $\lambda_{min}$  to  $\lambda_{max}$  (Lee et al., 2011; Li et al., 2012). One of the limitations of the above setting method is that different experimental data could achieve different  $\lambda_{min}$  and  $\lambda_{max}$ . It make

**TABLE 3 |** The abnormal brain regions and significance of brain region feature.

No	ROI	Name	P-values		
			HCC <sub>1</sub>	HCC <sub>2</sub>	HCC <sub>3</sub>
1	Right	Middle frontal gyrus	<b>0.0199</b>	0.1578	0.1690
2	Left	Inferior temporal gyrus	0.1181	0.1181	<b>0.0077</b>
3	Right	Posterior cingulate gyrus	0.3033	0.0941	<b>0.0451</b>
4	Left	Supplementary motor area	0.2075	0.0557	<b>0.0077</b>
5	Right	Parahippocampal gyrus	0.6384	<b>0.0157</b>	<b>0.0059</b>
6	Right	Inferior temporal gyrus	<b>0.0173</b>	0.2601	0.1096
7	Right	Precuneus	0.0606	<b>0.0396</b>	<b>0.0451</b>
8	Left	Fusiform gyrus	0.7929	<b>0.0478</b>	0.6121
9	Left	Supramarginal gyrus	<b>0.0481</b>	0.2145	0.1181
10	Right	Hippocampus	0.3930	<b>0.0049</b>	0.0905
11	Right	Lenticular nucleus, putamen	0.2145	0.1226	<b>0.0478</b>
12	Left	Thalamus	0.2943	<b>0.0370</b>	0.2521
13	Left	Middle temporal gyrus	0.8049	<b>0.0379</b>	0.1813

ROI denotes the region of interest. Bold indicates significance  $p < 0.05$ .

the parameter  $\lambda$  hard to be compared among different methods. Previous research standardized the range of  $\lambda$  from 0 to 1 based on  $\lambda_{min}$  and  $\lambda_{max}$  that made  $\lambda$  comparable (Liu et al., 2013). In the current study, we follow the latter setting method. When  $\lambda$  is more close to 0, there are more non-zero solutions in the  $\alpha_m$  which suggests that there are more nodes in the given hyper-edge. Otherwise, the opposite. Besides, to characterize the multi-level relationships within multiple brain regions, an array of hyper-edges can be obtained by setting different values of  $\lambda$  within a required range for each node. Thus, multi-level relationships indicate that different values of  $\lambda$  mean different levels of information within multiple brain regions. That is, a target function shown in Equation (7), which reflects a larger value of  $\lambda$ , can obtain a sparser solution and thus the hyper-edge



only includes some ROIs (i.e., nodes). We conducted tests to set different values of  $\lambda$ , ranging from 0.1 to 0.9 with a step of 0.1. Notably, in Equation (7), the values of weight vector  $\alpha_m$  are the same for brain regions within the same time series. Therefore, they will simultaneously be contained or excluded in the hyper-edge corresponding with them. In the current study, we were able to obtain the optimal solution in Equation (7) by using the SLEP package (Liu et al., 2013).

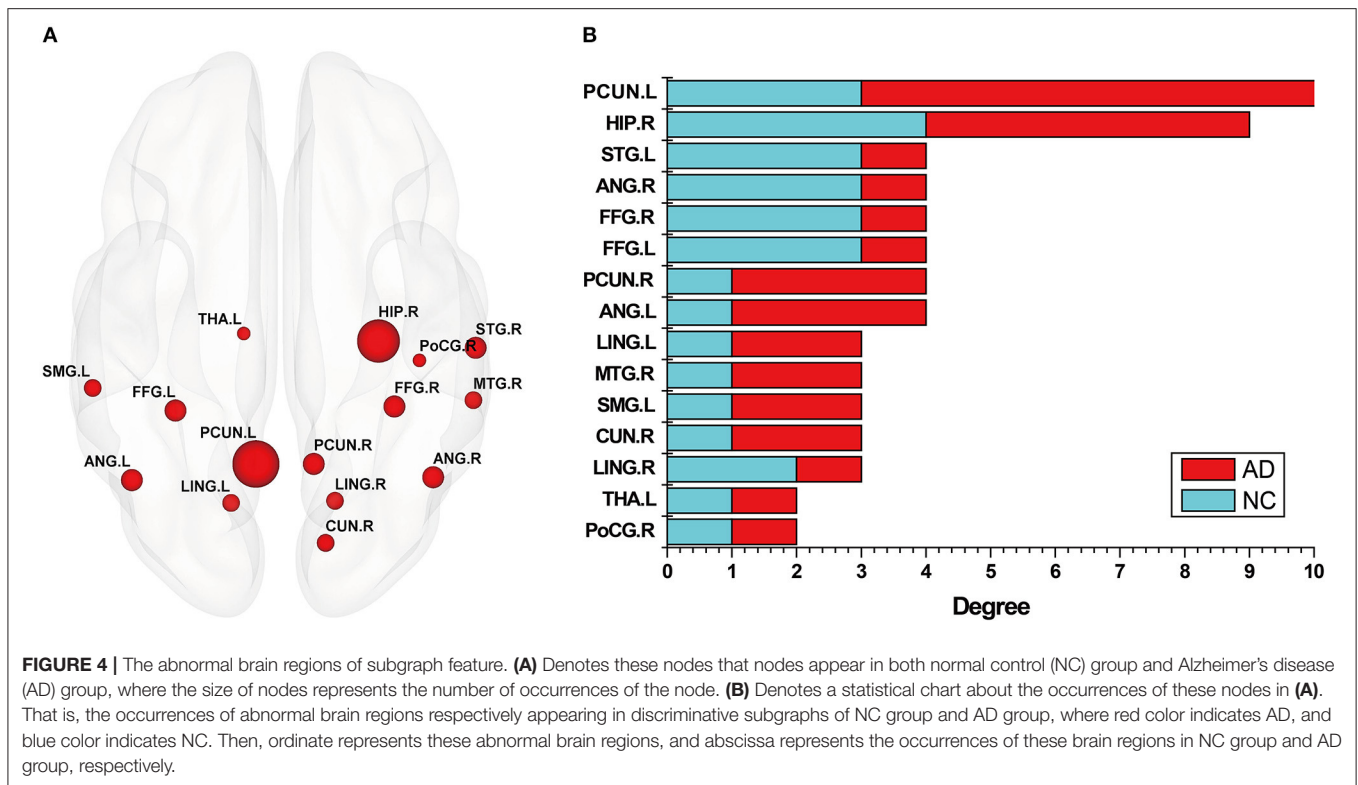
### Feature Extraction and Selection

After constructing hyper-network, we investigated two types of network features: brain region features and subgraph features.

The two types of features were then selected: the Kolmogorov-Smirnov non-parametric test was used for selecting quantifiable brain region features and the FSFS algorithm was used for selecting discriminative subgraph features.

### Brain Region Features and Feature Selection

To quantify the local brain region properties of the hyper-network, three local clustering coefficients— $HCC_1$ ,  $HCC_2$ , and  $HCC_3$  (Gallagher and Goldberg, 2013)—were adopted, as they describe the local aggregation of the hyper-network from different angles. **Table 2** shows the definitions and calculation formulas of these properties.



**FIGURE 4 |** The abnormal brain regions of subgraph feature. **(A)** Denotes these nodes that nodes appear in both normal control (NC) group and Alzheimer’s disease (AD) group, where the size of nodes represents the number of occurrences of the node. **(B)** Denotes a statistical chart about the occurrences of these nodes in **(A)**. That is, the occurrences of abnormal brain regions respectively appearing in discriminative subgraphs of NC group and AD group, where red color indicates AD, and blue color indicates NC. Then, ordinate represents these abnormal brain regions, and abscissa represents the occurrences of these brain regions in NC group and AD group, respectively.

A multiple linear regression method was adopted to eliminate the influence of confounding factors of age, gender, and educational status for every network property (independent variable: the area under the curve (AUC) value of every network property; dependent variables: age, gender, and educational status). These results indicated that the relationship between network properties and confounding factors was not significantly relevant (see Supplemental Table T1 for the detailed results).

To select the discriminative features, the Kolmogorov-Smirnov non-parametric test (Fasano and Franceschini, 1987) was used to select the quantifiable local brain region properties, corrected by the false-discovery rate (FDR) (Benjamini and Hochberg, 1995) ( $q = 0.05$ ) method. The brain region features with  $P < 0.05$  (FDR correction) were selected as discriminative features. Finally, we obtained a kernel matrix according to the above selected features.

### Subgraph Features and Feature Selection

The hyper-edges are regarded as the subgraph features of the hyper-network. The number of subgraphs is very large, but only a few features are truly discriminative. Accordingly, in this study, we selected the most discriminative subgraphs as features to be used in the classification in the next step. Detailed information on the discriminative subgraphs can be found in the Supplemental Text S2.

Discriminative subgraphs can be regarded as features for classification (Kong et al., 2013). However, because the subgraph features extracted from the normal control (NC) group and the AD group may not have discriminative ability, adopting

**TABLE 4 |** The TOP10 abnormal regions of subgraph feature.

No.	ROI	Name	Citations
1	Left	Precuneus	He et al., 2007
2	Right	Hippocampus	Liu et al., 2014
3	Right	Superior temporal gyrus	Solépadullés et al., 2009
4	Right	Angular gyrus	Liu et al., 2014
5	Right	Fusiform gyrus	Yetkin et al., 2006
6	Left	Fusiform gyrus	He et al., 2007
7	Right	Precuneus	He et al., 2007
8	Left	Angular gyrus	Liu et al., 2014
9	Right	Lingual gyrus	He et al., 2007
10	Right	Middle temporal gyrus	Berg et al., 1998

ROI denotes the region of interest.

only the extracted subgraph features would degrade the classification performance. To solve this problem, first, we used the discriminative score of the subgraph pattern (Santos et al., 2010) to complete the initial feature selection, also referred to as FSFS. This method calculates the discriminant scores of these subgraphs mined from the NC and AD groups and sorts them. The most discriminating scores  $t_1, t_2$  are selected as discriminant subgraphs.

Formally, we introduce the following notation:

$D:D = \{D_n, D_p\}$ , where  $D_n$  represents the negative samples, and  $D_p$  represents the positive samples.

$G:G = \{G_n, G_p\}$ , where  $G_p = \{g_{p1}, g_{p2}, \dots, g_{pm}\}$  denotes a set of all subgraph features in positive samples, and  $G_n =$



**TABLE 5 |** Classification performance of different methods.

Method	Research	Disease	Accuracy (%)	Sensitivity (%)	Specificity (%)
Partial-network	Rosa et al., 2015	MDD	58.33	53.33	63.33
	Wee et al., 2016	EMCI	62.71	60.00	65.52
	Guo et al., 2013	MDD	86.01	-	-
Pearson-network	Liu F. et al., 2015	MDD	63.00	40.00	83.00
	Wee et al., 2016	EMCI	66.10	76.67	55.17
	Wee et al., 2012	MCI	86.49	-	-
	Wang et al., 2011	AD	-	81.00	73.00
	Chen G. et al., 2011	AD	-	85.00	80.00
Hyper-network	Biao et al., 2014	MCI	94.60	91.70	96.00
	Jie et al., 2016	ADHD	82.90	83.90	86.10
Frequent subgraph	Du et al., 2016	ADHD	94.91	93.22	96.94
	Fei et al., 2014	MCI	97.30	-	-
Combined multiple features	Zhou et al., 2014	MCI	86.47	-	-
Hyper-network	Subgraph feature	AD	74.80	83.33	67.60
	Brain region feature	AD	83.30	84.21	82.14
	Proposed	AD	91.60	93.50	90.50

AD represents Alzheimer’s disease, MDD represents major depressive disorder, MCI represents mild cognitive impairment, EMCI represents mild cognitive impairment, and ADHD represents attention deficit hyperactivity disorder.

$\{g_{n1}, g_{n2}, \dots, g_{nk}\}$  denotes a set of all subgraph features in negative samples.

$T^*$ : The optimal set of subgraph features,  $T^* = T_1^* \cup T_2^*$  and  $T_1^* \subseteq G_p, T_2^* \subseteq G_n$ ; hence,  $T^* \subseteq G$ .

$J(T)$ : The criteria to evaluate the effectivity of subgraph feature subset  $T$ .

$S(g_s)$ : The discriminative score of a subgraph pattern  $g_s$  is defined as follows:

$$S(g_s) = |f_q(g_s|D_p) - f_q(g_s|D_n)| \tag{8}$$

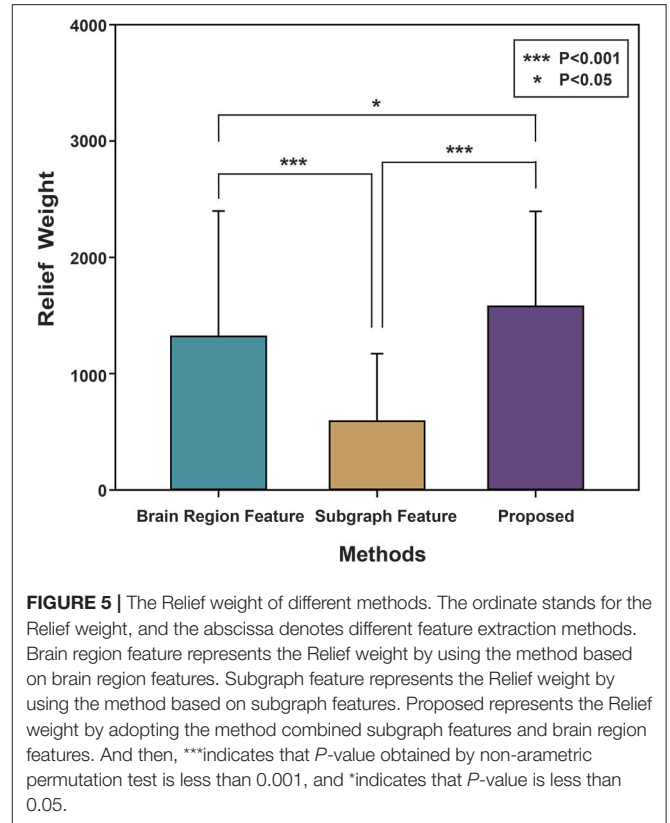
The discriminative score of subgraph  $g_s$  means its frequency difference between positive samples and negative samples, that is, the bigger the  $S(g_s)$ , the bigger the difference of these subgraphs between the AD and NC groups.  $S(g_s) = 0$  denotes that this subgraph pattern  $g_s$  was not present in any graphs in the NC group, but was present in all graphs in the MDD group, or vice versa.

In this study, we obtained the optimal set of subgraph features according to Equation (9):

$$T^* = T_{1 \subseteq G_p, T_2 \subseteq G_n} \text{ argmax } J(T) \text{ s.t. } |T_1| \leq t_1, |T_2| \leq t_2 \tag{9}$$

where  $|\cdot|$  denotes the size of the feature set, and  $t_1, t_2$  are the maximum number of features selected from the NC and MDD groups, respectively. We can then obtain the following equation:

$$J(T) = \sum_{i \leq t_1} S(g_{pi}) + \sum_{j \leq t_2} S(g_{nj}) \tag{10}$$



We can compute the discriminative score of each subgroup using Equation (8). Suppose the scores of all subgraphs are denoted as

$$S(g_p^1) \geq S(g_p^2) \dots \geq S(g_p^m), S(g_n^1) \geq S(g_n^2) \dots \geq S(g_n^k) \tag{11}$$

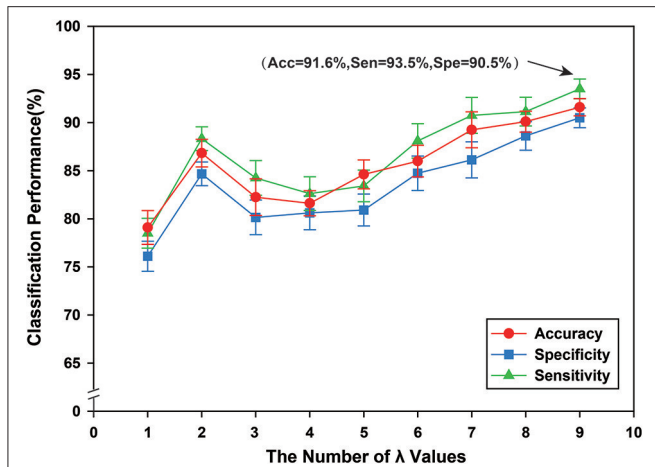
Based on Equation (11), we can obtain the optimal set of subgraph features, as

$$T^* = \{g_p^i, g_n^j | i \leq t_1, j \leq t_2\} \tag{12}$$

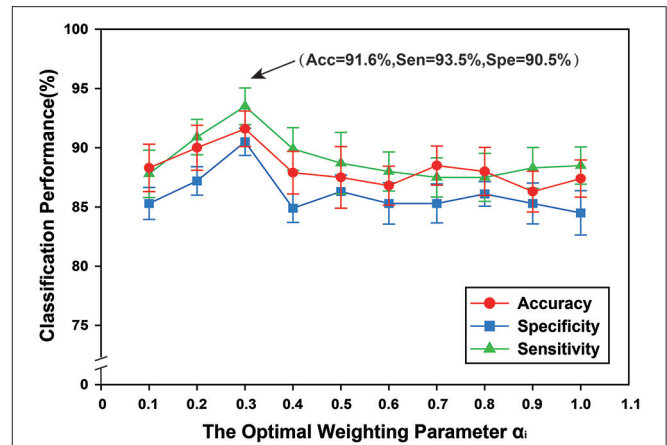
We obtained discriminative subgraphs based on the selected subgraph features by adopting the FSFS method. Due to the excessive number of discriminant subgraphs obtained by the FSFS method, we conducted a further selection using the threshold of discriminative score  $K$ .

### Construction of Classification Model

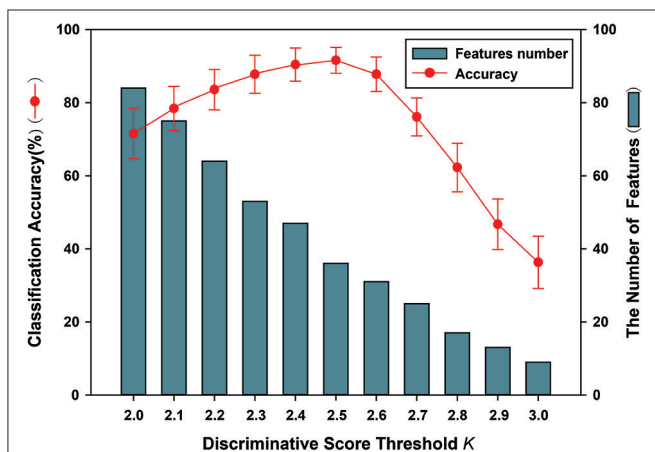
Because we used a combination of local brain region features and subgraph features as classification features, we adopted the multi-kernel SVM classifier based on the vector kernel and the graph kernel. For the vector kernel, we used the function-based RBF kernel, which is a widely used classification method (Cortes and Vapnik, 1995; Chen X. et al., 2011). The graph kernel is a common method for subgraph tests of isomorphism. It bridges the gap between graph-structured data and a large spectrum of



**FIGURE 6 |** The classification performance of different regularization parameters  $\lambda$ . The ordinate indicates accuracy, specificity and sensibility of this method, and the abscissa denotes nine groups of different regularization parameter  $\lambda$ , where 1 represents that  $\lambda$  value is {0.1}, and 2 denotes that  $\lambda$  value is {0.1, 0.2}, and 3 denotes that  $\lambda$  value is {0.1, 0.2, 0.3}, ..., 9 represents that {0.1, 0.2, ..., 0.9}. Thus, when  $\lambda$  is {0.1, 0.2, ..., 0.9}, better classification performance can be obtained, including that accuracy is 91.60%, and specificity is 93.50%, and then sensibility is 90.50%.



**FIGURE 8 |** The classification performance of different optimal weighting parameters  $\alpha_i$ . The ordinate indicates accuracy, specificity and sensibility of this method, and the abscissa denotes different optimal weighting parameter  $\alpha_i$ , ranging from 0 to 1 at a step size of 0.1, and  $\sum_{i=1}^M \alpha_i = 1$ . As shown in the figure, when  $\alpha_i = 0.3$ , better classification performance was obtained, including that accuracy is 91.60%, and specificity is 93.50%, and then sensibility is 90.50%.



**FIGURE 7 |** The classification accuracy and the number of features under different discriminative score threshold  $K$ . The ordinate indicates accuracy of this method, and the abscissa denotes different discriminative score threshold  $K$ , ranging from 2.0 to 3.0 at a step size of 0.1. As shown in the figure, when  $K = 0.25$ , the number of features is 36. Meanwhile, better classification accuracy was obtained; including that accuracy is 91.6%.

machine learning algorithms called kernel methods (Borgwardt et al., 2005), which include algorithms such as support vector machines, kernel regression, or kernel principal component analysis (Hofmann et al., 2007). The graph kernel is outlined below.

### Graph Kernel

Kernels are widely considered to be suitable indicators for evaluating the topological similarity of pairwise subjects. Kernels

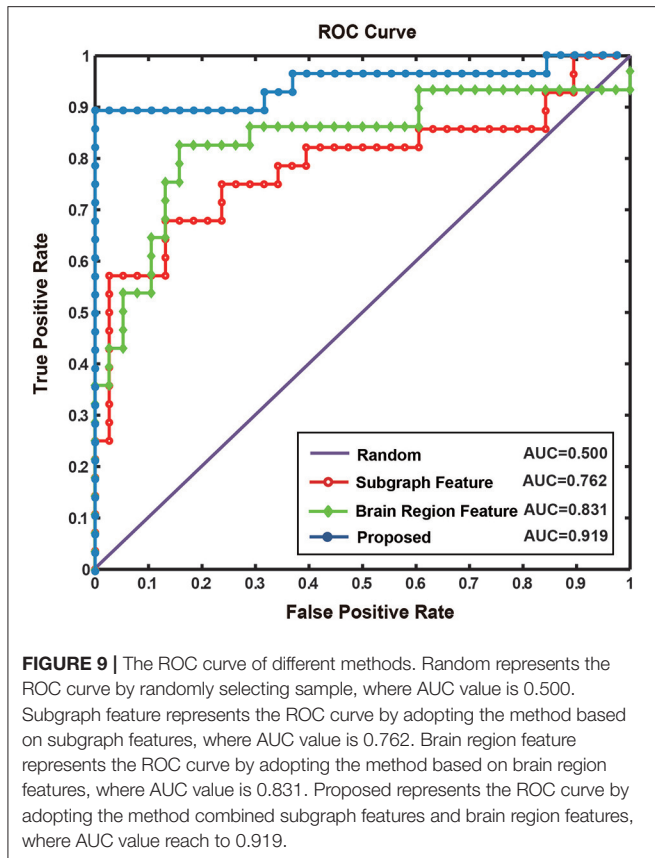
can map the data from an original space onto a higher dimensional feature space, generally causing the data to be more linearly separable. The corresponding kernel between subject  $x$  and  $y$  can be represented as follows:

$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle \tag{13}$$

where  $\varphi$  denotes a mapping function that can map data from the input space to the feature space. Many complex data types can be implemented through the kernel. The corresponding kernel of the graph is referred to as the graph kernel (Vishwanathan et al., 2008), which evaluates the topological similarity between paired graphs. Various methods have been proposed for constructing graph kernels, including walk-based (Gärtner et al., 2003), path-based (Alvarez et al., 2011), and subtree-based kernels (Shervashidze et al., 2011). Graph kernels have been successfully adopted for image classification (Harchaoui and Bach, 2007) and protein function prediction (Borgwardt et al., 2005). In the current study, we used the Weisfeiler-Lehman subtree kernel to measure the topological similarity between pairwise graphs. This method is implemented through the Weisfeiler-Lehman test of isomorphism (Shervashidze et al., 2011), which is described in detail in Supplemental Text S3.

### Multi-Kernel SVM

Recent studies have shown that a multi-kernel SVM can more effectively integrate features from different modalities than a single kernel SVM (Vishwanathan et al., 2008). The combination of multiple kernels can improve classification performance, and can also increase the interpretability of the results (Lanckriet et al., 2002). In general, given two subjects  $x$  and  $x'$ , multiple kernels



can be integrated by a linear combination method, as follows:

$$k(x, x') = \sum_{i=1}^M \alpha_i k_i(x, x') \quad s.t \sum_{i=1}^M \alpha_i = 1 \quad (14)$$

where  $k_i(x, x')$  denotes a basic kernel between  $x$  and  $x'$ ,  $\alpha_i$  denotes a weighting parameter ( $\alpha_i > 0$ ), and  $M$  denotes the number of combined kernel matrices.

In the current study, two types of kernel, based on a vector kernel and a graph kernel, were combined to construct the multi-kernel SVM classification model. However, when using two types of kernel for the classification, it was necessary to first implement one step separately to achieve normalization by computing Equation (15), then combining them.

$$k^*(x, x') = k(x, x') / \sqrt{k(x, x)k(x', x')} \quad (15)$$

Notably, in most studies of the multi-kernel method, the optimal weighting parameter  $a_i$  was simultaneously obtained with some other parameters. However, we adopted a grid search method to obtain  $a_i$ . When  $a_i$  was determined, the multi-kernel SVM can be achieved by embedding the multi-kernel method into the conventional single-kernel SVM classifier.

**TABLE 6 |** Demographics and clinical characteristics of the subjects.

	EMCI	LMCI	AD	P-value
Number(n)	33	32	29	–
Female/male(n)	16/17	13/19	18/11	0.53 <sup>a</sup>
Age (mean ± SD, year)	72.0 ± 6.0	72.6 ± 8.3	72.3 ± 7.4	0.48 <sup>a</sup>
MMSE	27.6 ± 2.1	26.9 ± 2.7	21.0 ± 3.6	<0.0001 <sup>b</sup>

EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, Alzheimer’s disease; MMSE, Mini-Mental State Examination. <sup>a</sup>chi-square test. <sup>b</sup>one-way analysis of variance tests.

**TABLE 7 |** The abnormal brain regions and significance of brain region feature between the normal control group and the early mild cognitive impairment group.

No	ROI	Name	P-values		
			HCC <sub>1</sub>	HCC <sub>2</sub>	HCC <sub>3</sub>
1	Left	Superior frontal gyrus, medial	<b>0.0396</b>	0.5082	0.6102
2	Right	Middle temporal gyrus	0.7149	<b>0.0442</b>	0.9990
3	Right	Olfactory cortex	0.1616	<b>0.0396</b>	0.6522
4	Right	Temporal pole: superior temporal gyrus	0.3013	<b>0.0050</b>	0.2474
5	Right	Parahippocampal gyrus	0.4322	<b>0.0352</b>	<b>0.0272</b>
6	Right	Postcentral gyrus	0.2351	0.1286	<b>0.0352</b>
7	Right	Angular gyrus	0.6102	0.1363	<b>0.0129</b>

ROI denotes the region of interest. Bold indicates significance  $p < 0.05$ .

In the current study, the multi-kernel SVM was used to implement the classification. We adopted the multi-kernel SVM method to effectively integrate multiple features, which fully described the overall interactive information of the brain network. Specifically, the vector kernel characterizes interactions among multiple brain regions using three different local cluster coefficients. Moreover, the graph kernel characterizes information about topological structure within the connectivity network.

**Cross-Validation**

In the current experiment, we adopted 10-fold cross-validation (Chang and Lin, 2011) to evaluate the performance of our proposed classification method. Specifically, the subject dataset was randomly divided into 10 parts, one of which was left as the testing set, while the remaining nine were regarded as training sets. In this study, 10-fold cross validation was performed 100 times to obtain more accurate results. Finally, we computed the arithmetic mean of the 100 repetitions as the final result.

**RESULTS**

Two types of features were extracted and selected from the constructed networks, including brain region features and subgraph features. Brain region features computed and selected for the HCC<sub>1</sub>, HCC<sub>2</sub> and HCC<sub>3</sub>. Subgraph features were selected by SFSS algorithm.

**TABLE 8** | The abnormal brain regions and significance of brain region feature between the normal control group and the late mild cognitive impairment group.

No	ROI	Name	P-values		
			HCC <sub>1</sub>	HCC <sub>2</sub>	HCC <sub>3</sub>
1	Left	Posterior cingulate gyrus	<b>0.0371</b>	0.8474	0.3563
2	Left	Middle frontal gyrus, orbital part	0.8915	<b>0.0134</b>	0.9569
3	Left	Insula	0.7154	<b>0.0338</b>	0.6870
4	Right	Amygdala	0.5039	<b>0.0134</b>	0.6727
5	Right	Temporal pole: middle temporal gyrus	0.9109	<b>0.0291</b>	0.3563
6	Right	Rolandic operculum	0.6727	0.8474	<b>0.0353</b>
7	Left	Supplementary motor area	0.6727	0.1936	<b>0.0250</b>
8	Left	Anterior cingulate and paracingulate gyri	0.1487	0.1668	<b>0.0405</b>
9	Right	Calcarine fissure and surrounding cortex	0.3675	0.9015	<b>0.0337</b>
10	Right	Postcentral gyrus	0.1376	0.1606	<b>0.0216</b>

ROI denotes the region of interest. Bold indicates significance  $p < 0.05$ .

## Brain Region Features

After constructing the hyper-network, three local brain region properties, HCC<sub>1</sub>, HCC<sub>2</sub>, HCC<sub>3</sub>, were extracted and selected. Specially, HCC<sub>1</sub> calculates the amount of adjacent nodes that have connections not facilitated by node  $v$ . HCC<sub>2</sub> computes the amount of adjacent nodes that have connections facilitated by node  $v$ . HCC<sub>3</sub> computes the number of overlap among adjacent hyper-edges of node  $v$ . The local brain region features and abnormal brain regions were then analyzed. **Table 3** lists the abnormal brain regions and the significance of the brain region features. We used HCC<sub>1</sub>, HCC<sub>2</sub>, and HCC<sub>3</sub>, three local clustering coefficients, to indicate a significant difference ( $p < 0.05$ , FDR correction) in abnormal brain regions. **Table 3** shows a total of 13 abnormal brain regions: the right middle frontal gyrus (MFG), left inferior temporal gyrus (ITG), right posterior cingulate gyrus (PCG), left supplementary motor area (SMA), right parahippocampal gyrus (PHG), right ITG, right precuneus (PCUN), left fusiform gyrus (FFG), left supramarginal gyrus (SMG), right hippocampus (HIP), right putamen (PUT), left thalamus (THA), and left middle temporal gyrus (MTG).

## Subgraph Features

After constructing the hyper-network, hyper-edges were extracted as subgraph features from the AD and NC groups. Subgraph features that were repeated within the group were removed to ensure their uniqueness. Then, the FSFS algorithm and the threshold of discriminative score  $K$  were used to select the most discriminative subgraphs. With the discriminative score  $K$  threshold set to 0.25, we obtained 18 subgraphs in the NC group and 32 subgraphs in the AD group. To ensure a balanced number of features, the 18 subgraph features with the highest discriminative scores were selected from the AD group. **Figure 3** shows the distribution of the discriminative subgraph features in the brain.

**TABLE 9** | The abnormal brain regions and significance of brain region feature between the normal control group and the Alzheimer's disease group.

No	ROI	Name	P-values		
			HCC <sub>1</sub>	HCC <sub>2</sub>	HCC <sub>3</sub>
1	Right	Calcarine fissure and surrounding cortex	<b>0.0060</b>	0.9512	0.9580
2	Right	Cuneus	<b>0.0058</b>	0.9641	0.7833
3	Left	Lingual gyrus	<b>0.0067</b>	0.4542	0.2203
4	Left	Fusiform gyrus	<b>0.0155</b>	0.8045	0.5478
5	Right	Superior parietal gyrus	<b>0.0343</b>	0.1704	<b>0.0390</b>
6	Left	Thalamus	<b>0.0138</b>	0.6394	0.8114
7	Right	Inferior frontal gyrus, triangular part	0.5780	<b>0.0301</b>	0.6009
8	Right	Amygdala	0.1272	<b>0.0160</b>	0.5629
9	Right	Temporal pole: middle temporal gyrus	0.2514	<b>0.0317</b>	0.2289
10	Left	Inferior frontal gyrus, opercular part	0.6086	0.2079	<b>0.0334</b>
11	Right	Olfactory cortex	0.8695	0.2908	<b>0.0179</b>
12	Left	Superior occipital gyrus	0.2246	0.0912	<b>0.0325</b>
13	Right	Postcentral gyrus	0.5553	<b>0.0189</b>	<b>0.0074</b>
14	Left	Inferior parietal, but supramarginal and angular gyri	0.0642	0.2706	<b>0.0361</b>

ROI denotes the region of interest. Bold indicates significance  $p < 0.05$ .

To make it easier to analyze the differences between groups, the 18 subgraphs in each group were combined, as shown in **Figure 3**. The subgraph of group A and the subgraph of group B in **Figure 3** were further analyzed. The results showed that the majority of the discriminative regions were those brain regions that appeared together in both groups; however, those that showed significant differences indicated abnormal regions. **Figure 3** shows that these abnormal brain regions were mainly distributed in the left PCUN, right HIP, right superior temporal gyrus (STG), right angular gyrus (ANG), right FFG, left FFG, right PCUN, left ANG, left lingual gyrus (LING), right MTG, left SMG, right cuneus (CUN), right LING, left THA, and right postcentral gyrus (PoCG).

In addition, in order to better analyze the abnormal brain regions, and the differences among them in the NC group and AD group, this study examined the distribution of these abnormal regions in the brain. The number of occurrences of abnormal brain regions was summed in the NC and AD groups, and the regions were then sorted according to their sums. **Figure 4A** shows the distribution of the abnormal regions, and **Figure 4B** shows the distribution of the sum of occurrences in the NC and AD groups. We counted the sum of occurrences of these abnormal brain regions, and then chose the 10 highest for further analysis: the left PCUN, right HIP, right STG, right ANG, right FFG, left FFG, right PCUN, left ANG, left LING, and right MTG. **Table 4** shows the detailed information of the top 10 abnormal regions with significant differences.

**TABLE 10** | Top 10 abnormal brain regions of subgraph features of the Alzheimer's disease patients in our collected dataset and ADNI dataset.

No	Our Collected Dataset		Citations	ADNI Dataset		Citations
1	Left	Precuneus <sup>+</sup>	He et al., 2007	Right	Hippocampus <sup>+</sup>	Liu et al., 2014
2	Right	Hippocampus <sup>+</sup>	Liu et al., 2014	Right	Lingual gyrus <sup>+</sup>	He et al., 2007
3	Right	Superior temporal gyrus <sup>+</sup>	Solépadullés et al., 2009	Right	Fusiform gyrus <sup>+</sup>	Yetkin et al., 2006
4	Right	Angular gyrus	Liu et al., 2014	Left	Lingual gyrus	He et al., 2007
5	Right	Fusiform gyrus <sup>+</sup>	Yetkin et al., 2006	Left	Fusiform gyrus <sup>+</sup>	He et al., 2007
6	Left	Fusiform gyrus <sup>+</sup>	He et al., 2007	Left	Precuneus <sup>+</sup>	He et al., 2007
7	Right	Precuneus	He et al., 2007	Left	Posterior cingulate gyrus	He et al., 2007
8	Left	Angular gyrus <sup>+</sup>	Liu et al., 2014	Left	Angular gyrus <sup>+</sup>	Liu et al., 2014
9	Right	Lingual gyrus <sup>+</sup>	He et al., 2007	Right	Superior temporal gyrus <sup>+</sup>	Solépadullés et al., 2009
10	Right	Middle temporal gyrus	Berg et al., 1998	Right	Cuneus	He et al., 2007

<sup>+</sup>Indicates the abnormal brain region appears in both datasets. ADNI, Alzheimer's Disease Neuroimaging Initiative.

**TABLE 11** | Top 10 abnormal brain regions of subgraph features in the early mild cognitive impairment group, the late mild cognitive impairment group and Alzheimer's disease group in ADNI dataset.

No	EMCI Group		Citations	LMCI Group		Citations	AD Group	
1	Right	Superior frontal gyrus, dorsolateral <sup>x</sup>	Yetkin et al., 2006	Right	Superior frontal gyrus, dorsolateral <sup>x</sup>	Yetkin et al., 2006	Right	Hippocampus <sup>*</sup>
2	Left	Middle frontal gyrus <sup>x</sup>	Zhao et al., 2015	Left	Superior frontal gyrus, medial <sup>x</sup>	Yetkin et al., 2006	Right	Lingual gyrus
3	Right	Middle frontal gyrus <sup>x</sup>	Zhao et al., 2015	Left	Parahippocampal gyrus <sup>x</sup>	Liu J. et al., 2015	Right	Fusiform gyrus
4	Right	Olfactory cortex	Zhao et al., 2015	Right	Precentral gyrus	Devanand et al., 2006	Left	Lingual gyrus
5	Left	Superior frontal gyrus, medial <sup>x</sup>	Yetkin et al., 2006	Left	Middle frontal gyrus <sup>x</sup>	Zhao et al., 2015	Left	Fusiform gyrus
6	Right	Superior frontal gyrus, medial <sup>x</sup>	Yetkin et al., 2006	Right	Middle frontal gyrus <sup>x</sup>	Zhao et al., 2015	Left	Precuneus
7	Left	Insula <sup>x</sup>	Devanand et al., 2006	Right	Superior frontal gyrus, medial <sup>x</sup>	Yetkin et al., 2006	Left	Posterior cingulate gyrus
8	Left	Superior occipital gyrus <sup>x</sup>	Morgen et al., 2013	Left	Insula <sup>x</sup>	Devanand et al., 2006	Left	Angular gyrus
9	Left	Parahippocampal gyrus <sup>x</sup>	Liu J. et al., 2015	Left	Superior occipital gyrus <sup>x</sup>	Morgen et al., 2013	Right	Superior temporal gyrus
10	Left	Fusiform gyrus	Lim et al., 2011	Right	Hippocampus <sup>*</sup>	Morgen et al., 2013	Right	Cuneus

<sup>x</sup>Indicates the abnormal brain region appears in both EMCI and LMCI groups. <sup>\*</sup>Indicates the abnormal brain region appears in both LMCI and AD groups. EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative. The citations of AD group are shown in **Table 10**.

## Classification Results

The classification accuracy, specificity, sensitivity, and AUC under the ROC curve were used as a quantitative measure to evaluate the experimental results. To demonstrate the classification performance of the proposed method, we compared the accuracy, sensitivity, and specificity of different classification methods, and analyzed the differences among different network construction methods and feature extraction methods. As can be seen from **Table 5**, the proposed method performed better than the conventional methods of constructing the functional network by partial correlations or Pearson correlations.

To accurately compare the different methods of feature extraction, we used the same dataset and constructed the same network, and the brain region features, subgraph features, and multi-features method were used for the classification, respectively. The classification results are shown in **Table 5**. The experimental results showed that the proposed method's accuracy of 91.60%, specificity of 90.50%, and sensitivity of 93.5% were significantly better than the classification results using only a single feature.

Finally, the Relief algorithm was used to evaluate the importance of features. The Relief algorithm was first proposed by Kira (Kira and Rendell, 1992) and has been widely applied in selecting features for classification (Rosario and Thangadurai,

2015). As shown in **Figure 5**, to verify the validity of the proposed method, the brain region feature, subgraph feature, and multi-feature methods were each evaluated by the Relief algorithm. The weight of every feature was obtained according to the correlation between the feature and its class. The greater the relief weight, the stronger the correlation between the feature and the class, indicating that the feature is more important for classification. **Figure 5** shows that the relief weight of the multi-feature method was significantly higher than that of the single feature method.

In conclusion, the proposed method of machine learning classification combining multiple features of a hyper-network of fMRI data in AD could be used to effectively classify healthy people and AD patients.

## Effect of Regularization Parameter $\lambda$

The proposed classification model involves setting some parameters, which would be expected to affect the final results. Here, we tested the classification performance with different parameters, including the regularization parameter  $\lambda$  of the sparse target optimization function, the threshold of the discriminative score  $K$ , and the optimal weighting parameter  $\alpha^i$ , and attempted to determine the optimal parameter settings.

We constructed the hyper-networks by adopting a sparse representation method, where  $\lambda$  indicates a regularization

**TABLE 12** | The abnormal brain regions of brain region features of the Alzheimer's disease patients in our collected dataset and ADNI dataset.

No	Our Collected Dataset	Citations	ADIN Dataset	Citations
1	Right Middle frontal gyrus	Wang et al., 2011	Right Calcarine fissure and surrounding cortex	Hampel et al., 2002
2	Left Inferior temporal gyrus	Wang et al., 2011	Right Cuneus	He et al., 2007
3	Right Posterior cingulate gyrus	He et al., 2007	Left Lingual gyrus	He et al., 2007
4	Left Supplementary motor area	Wang et al., 2011	Left Fusiform gyrus <sup>+</sup>	He et al., 2007
5	Right Parahippocampal gyrus	Wang et al., 2011	Right Superior parietal gyrus	Zhou and Jin, 2008
6	Right Inferior temporal gyrus	Wang et al., 2011	Left Thalamus <sup>+</sup>	Solépadullés et al., 2009
7	Right Precuneus	He et al., 2007	Right Inferior frontal gyrus, triangular part	He et al., 2007
8	Left Fusiform gyrus <sup>+</sup>	He et al., 2007	Right Amygdala	Hampel et al., 2002
9	Left Supramarginal gyrus	Grady et al., 2003	Right Temporal pole: middle temporal gyrus	Zhou and Jin, 2008
10	Right Hippocampus	Liu et al., 2014	Left Inferior frontal gyrus, opercular part	He et al., 2007
11	Right Lenticular nucleus, putamen	De-Jong et al., 2008	Right Olfactory cortex	Vasavada et al., 2015
12	Left Thalamus <sup>+</sup>	Solépadullés et al., 2009	Left Superior occipital gyrus	Grady et al., 2003
13	Left Middle temporal gyrus	Grady et al., 2003	Right Postcentral gyrus	Wang et al., 2011
14	–	–	Left Inferior parietal, but supramarginal and angular gyri	Grady et al., 2003

<sup>+</sup>indicates the abnormal brain region appears in both groups. ADNI, Alzheimer's Disease Neuroimaging Initiative.

parameter for controlling the sparsity of the network ( $\lambda > 0$ ). By setting different values of  $\lambda$  within a required range, we obtained an array of hyper-edges. To research the classification performance of this method with different  $\lambda$  values, nine groups of different  $\lambda$  values were tested,  $\{0.1\}$ ,  $\{0.1, 0.2\}$ ,  $\{0.1, 0.2, 0.3\}$ ,  $\dots$ ,  $\{0.1, 0.2, \dots, 0.9\}$ . The classification results indicated that a greater number of  $\lambda$  values corresponded with better classification performance. In a previous study (Jie et al., 2016), the  $\lambda$  value was set to  $\{0.1, 0.2, \dots, 0.9\}$ , as shown in **Figure 6**, and was confirmed experimentally. Therefore, in this study,  $\lambda$  was set to  $\{0.1, 0.2, \dots, 0.9\}$ . **Figure 6** shows the classification performance under different regularization parameters.

### Effect of Discriminative Score Threshold $K$

The FSFS algorithm was adopted to select discriminative subgraphs. Because of the excessive number of selected subgraphs, a threshold value was set (the discriminative score threshold  $K$ ). The other parameters were controlled to select a more accurate discriminative score threshold  $K$ . The threshold  $K$  ranged from 0.20 to 0.30 and the interval was 0.01. **Figure 7** shows the classification accuracy and the number of features under different discriminative score threshold  $K$ -values. The experimental results showed that when the discriminative score threshold  $K = 0.25$ , the number of features was 36, and the classification accuracy was optimal. One potential explanation is that when the threshold was too small, features that contributed little to the classification were also chosen, but when it was too large, features that made large contributions were removed, leading to lower classification accuracy in both cases.

### Effect of Optimal Weighting Parameter $\alpha_i$

A multi-kernel SVM was used for classification, which involved finding the optimal weighting parameter  $\alpha_i$ . To examine the effects of different values of  $\alpha_i$  on classification performance, the range was set from 0 to 1, with a step size of 0.1 and  $\sum_{i=1}^M \alpha_i = 1$ . **Figure 8** shows the classification performance

under different optimal weighting parameters  $\alpha_i$ . The best classification performance was obtained when  $\alpha_i = 0.3$ , with accuracy of 91.60%, sensitivity of 93.50%, and specificity of 90.5%. The experimental results showed that different values of optimal weighting parameters  $\alpha_i$  influenced the classification results.

## DISCUSSION

In this study, we proposed a method of machine learning classification combining multiple features of a hyper-network of fMRI data in AD. Hyper-networks were constructed on the AD dataset to analyze the interactions among multiple brain regions. Then, two types of features were used for feature extraction and selection: brain region features were selected using a non-parametric test method, and subgraph features were selected using the FSFS algorithm. Finally, two types of kernel (vector kernel and graphkernel) were fused, and a multi-kernel SVM classifier was used for classification. The experimental results verified the validity of the proposed method.

### The most Discriminative Brain Regions Using Brain Region Features

Two methods were used to discriminate significantly abnormal brain regions between groups. The results using only brain region features showed 13 abnormal regions, as shown in brain region features of experiments and results section. Many previous researches have found that these brain regions are abnormal in AD patients. Specifically, the posterior cingulate cortex (PCC) is mainly involved in episodic memory and short-term memory processing (Gusnard and Raichle, 2001; Buckner et al., 2008) and is a critical region in human brain structural and functional networks (Greicius et al., 2004; Cavanna and Trimble, 2006; Zhang et al., 2009; Binnewijzend et al., 2012). Studies have shown that the PCC is one of the most robust brain regions in the resting state. The PCUN is also an important component of the

**TABLE 13** | The abnormal brain regions differences of brain region feature among the early mild cognitive impairment group, the late mild cognitive impairment group and Alzheimer's disease group in ADNI dataset.

No	EMCI Group	Citations	LMCI Group	Citations	AD Group
1	Left Superior frontal gyrus, medial	Yetkin et al., 2006	Left Posterior cingulate gyrus	Zhao et al., 2015	Right Calcarine fissure and surrounding cortex*
2	Right Middle frontal gyrus	Zhao et al., 2015	Left Middle frontal gyrus, orbital part	Zhao et al., 2015	Right Cuneus
3	Right Olfactory cortex <sup>Δ</sup>	Zhao et al., 2015	Left Insula	Devanand et al., 2006	Left Lingual gyrus
4	Right Temporal pole: superior temporal gyrus	Fei et al., 2014	Right Amygdala*	Yao et al., 2014	Left Fusiform gyrus
5	Right Parahippocampal gyrus	Liu J. et al., 2015	Right Temporal pole: middle temporal gyrus*	Morgen et al., 2013	Right Superior parietal gyrus
6	Right Postcentral gyrus <sup>++</sup>	Devanand et al., 2006	Right Rolandic operculum	Fei et al., 2014	Left Thalamus
7	Right	Liu J. et al., 2015	Left Supplementary motor area	Poettrich et al., 2009	Right Inferior frontal gyrus, triangular part
8	-	-	Left Anterior cingulate and paracingulate gyri	Zhao et al., 2015	Right Amygdala*
9	-	-	Right Calcarine fissure and surrounding cortex*	Zhao et al., 2015	Right Temporal pole: middle temporal gyrus*
10	-	-	Right Postcentral gyrus <sup>++</sup>	Devanand et al., 2006	Left Inferior frontal gyrus, opercular part
11	-	-	-	-	Right Olfactory cortex <sup>Δ</sup>
12	-	-	-	-	Left Superior occipital gyrus
13	-	-	-	-	Right Postcentral gyrus <sup>++</sup>
14	-	-	-	-	Left Inferior parietal, but supramarginal and angular gyri

\*Indicates the abnormal brain region appears in both LMCI and AD groups. <sup>Δ</sup>Indicates the abnormal brain region appears in both EMCI and AD groups. <sup>++</sup>Indicates the abnormal brain region appears in all the three groups. EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative. The citations of AD group are shown in Table 12.

default mode network, and is closely related to the extraction of episodic memory (Fransson and Marrelec, 2008). Using r-fMRI, several recent studies have suggested that the PCC/PCu exhibits reduced regional activity in AD patients (He et al., 2007). In addition, using resting-state fMRI to measure the amplitude of low-frequency fluctuations (ALFF) of intrinsic brain activity in 23 patients with moderate AD and 27 age- and gender-matched healthy controls, Liu et al. (2014) found that AD patients also showed increased ALFF in the bilateral Hip/PHG. The Hip/PHG is considered to be critical to memory function. Compared with normal controls, the AD patients showed decreased ALFF values in the bilateral PCC/PCu, MTG, and STG. Yetkin et al. (2006) proved that AD patients showed more activation than controls in the right MFG, left ITG, left THA, and right PUT and so on. Wang et al. (2011) used resting-state functional MRI to investigate spatial patterns of spontaneous brain activity in 22 healthy elderly subjects, 16 MCI, and 16 AD patients. The results showed that ALFF differences between AD patients and healthy elderly subjects were mainly found in the bilateral PHG/Hip, bilateral SMA, and left FFG. The results obtained in this study are consistent with those of previous studies.

### The Most Discriminative Brain Regions Using Subgraph Features

The results using only subgraph features showed that the abnormal brain regions included the left PCUN, right HIP, right STG, right ANG, right FFG, left FFG, right PCUN, left ANG, left LING, right MTG, left SMG, right CUN, right LING, left THA, and right PoCG. These abnormal brain regions have been shown to be associated with AD in previous studies. Both structural MRI and resting-state fMRI scans were collected from 14 AD subjects and 14 age-matched normal controls. He et al. (2007) found that regional coherence was significantly decreased in the PCUN in the AD patients compared with the normal controls. Recent functional imaging studies have indicated that the pathophysiology of AD may be associated with changes in spontaneous low-frequency (<0.08 Hz) blood oxygenation level dependent fluctuations (LFBF) measured during a resting state (He et al., 2007). He et al. also found that AD patients showed increased LFBF coherence in the bilateral CUN, right LING, and left FFG. Neuropathological studies indicate that brain lesions are already present in the inferior parietal lobule (IPL) (including the ANG and left SMG) in incipient AD, although they are observed less frequently than in medial temporal areas (Berg et al., 1998; Markesbery et al., 2006; Haroutunian et al., 2007). Liu et al. (2014) found that AD patients also showed increased ALFF in the IPL. Yetkin et al. (2006) evaluated brain activation in patients with probable AD, MCI, and controls while performing a working memory task. The AD group showed more right FFG and left THA activation than the control group. In this study, the right HIP (Liu et al., 2014) and right STG (Solépadullés et al., 2009) are consistent with previous studies.

### Classification Performance

Conventional methods of constructing functional networks cannot reflect the interactions among multiple brain regions and thus ignore the higher-order information among them.

To study the complex interaction information among multiple brain regions, Jie et al. (2016) proposed to construct a hyper-network model. In the Jie et al. study, the local brain region properties were extracted from the hyper-network as features, and the most discriminative features were selected. Finally, the multi-kernel SVM was adopted for classification. The construction of the hyper-networks enabled us to identify the interaction information among brain regions. In addition, to show that the classification method based on subgraph features can better capture the topological information among brain regions, Fei et al. (2014) adopted frequent subgraph mining technology to mine frequent sub-networks in an MCI dataset, then used a discriminative subgraph mining algorithm to mine discriminative sub-networks. Finally, they used SVM based on a graph kernel for the classification. Du et al. (2016) used the frequent subgraph mining technique to mine frequent sub-networks in an ADHD dataset, then the FSFS method to select the sub-networks, and graph kernel principal component analysis to extract the features. Finally, SVM was used to classify the data. Wang et al. (Zhou et al., 2014) adopted the same technology to mine frequent sub-networks in an MCI dataset, and then combined the discriminative sub-networks with conventional quantitative properties to select features. Finally, they used multi-kernel SVM for classification. The above results show that classification methods based on subgraph features can effectively improve classification performance.

The results of this study were compared with those obtained by conventional methods of functional connectivity network construction based on partial or Pearson correlations (Table 5). The results showed that the proposed method does not just identify the interactive information between brain regions, but can effectively represent the higher-order information among them. In addition, Jie et al. (2016) experimented in the MCI dataset. In comparison, the classification performance of the proposed method was similar and the difference might have been due to the use of different datasets. The same method using different datasets may obtain different classification results, and different methods also differ in the way they construct the network and extract and classify features. To accurately compare the different methods of feature extraction, we used the same dataset and constructed the same hyper-network, and compared the classification results using the brain region feature method, subgraph feature method, and multi-feature method, respectively (Table 5). The diagnostic accuracy of the multi-feature method was 8.3% better than that obtained using only single features. Furthermore, Figure 9 shows the ROC curves for the different classification methods. The AUC value was 0.762 for the subgraph feature method and 0.831 for the brain region feature method, compared with 0.919 for the multi-feature classification method, an increase of at least 0.088. The results show that the proposed classification method combining subgraph features and brain region features preserved not only the global topological information of the brain region, but also the sensitivity to change in a single brain region. The multi-feature classification method can effectively improve the diagnosis accuracy of AD.

**TABLE 14** | Classification performance of different methods.

Method	Research	Diseases	Accuracy (%)	Sensitivity (%)	Specificity (%)
Hyper-network	Subgraph feature	EMCI	64.90	70.10	61.60
	Brain region feature proposed	EMCI	69.21	73.27	65.67
		EMCI	72.80	78.25	67.13
	Subgraph feature proposed	LMCI	71.10	76.33	66.95
		LMCI	74.33	79.21	69.00
	Subgraph feature proposed	LMCI	78.63	82.54	72.18
		AD	75.60	78.33	71.05
	Brain region feature proposed	AD	80.23	81.25	77.90
		AD	88.91	91.73	85.66

EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; AD, Alzheimer's disease.

The Relief algorithm was used to verify the importance of the underlying features for classification accuracy, with significance analyzed by a non-parametric permutation test. As shown in Figure 5, the average Relief weight of the multi-feature method was significantly higher than that of the single-feature method, indicating that the multi-feature method was better for assessing the importance of features. However, the underlying reason was that the multi-feature based method effectively fused two different yet complementary interaction information: brain region features and subgraph features. Therefore, it not only reflected the information from a single brain region, but also captured the global topological information among brain regions. All of the above experimental results demonstrate the validity of the proposed method.

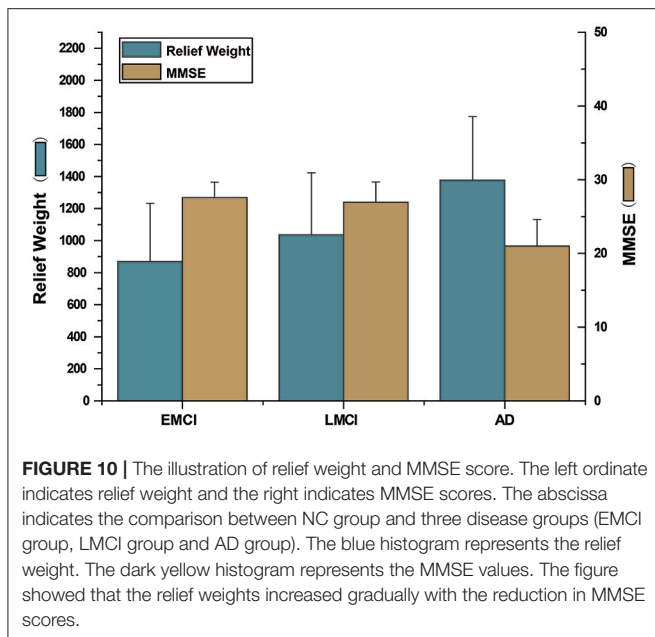
## Features Selection

The current findings demonstrated that the multi-feature combination method effectively integrated multiple network properties, further improving classification performance. The relief analysis method was performed to evaluate the contributions of selected features during the classification process. The relief weight obtained with multiple combined features was significantly greater than the weight obtained when only brain region features or subgraph features were adopted. Regarding the underlying mechanisms, this is likely to be because adopting multiple features can integrate complementary network information, combining local brain region features and subgraph features, thus further improving classification accuracy. Some studies have also demonstrated that multiple features can effectively combine multiple different complementary network properties for classification (Jie et al., 2014; Zhou et al., 2014). Global network topology information will be lost only from the perspective of the brain region features. In addition, subgraph features can also result in the loss of sensitivity of a single brain region.

## Regularization Parameter $\lambda$

Previous researches had demonstrated that the parameter  $\lambda$  had a great effect on the hyper-network structure. The parameter  $\lambda$  determined the sparsity and scale of network regions. If  $\lambda$  was





too small, the network would be too coarse and involve much noise; if  $\lambda$  was too large, the network would be too sparse (Lv et al., 2015). Besides, it was found that the reliability of network structure, especial modularity, was sensitive to the sparsity which was controlled by  $\lambda$  (Xuan and Wang, 2015). Furthermore, the parameter  $\lambda$  also impacted on the classification performance. The ultimate classification accuracy was extremely sensitive to the network model parameters, especial  $\lambda$  (Qiao et al., 2016). As the authors known there was no golden criterion for selection of  $\lambda$ . How to find a suitable  $\lambda$  was important for the construction of hyper-network and classifier. Some optimization methods were proposed. Qiao et al. conducted parameter selection in a large range by computing the classification accuracy based on leave one out test on all the subjects, choosing the corresponding parametric of the best classification accuracy (Qiao et al., 2016). Xuan et al. chose the parameter  $\lambda$  by computing the value of intra-class correlation coefficient (Xuan and Wang, 2015), which could describe the reliability of network structure (Braun et al., 2012). However, It was found that it was hard to achieve a high reliable network structure by setting a single  $\lambda$ . The research showed that the network achieved a relatively high reliability only when  $\lambda$  took 0.01 (it was very close to 0, which suggested that almost all the nodes in the network were connected in the given hyper-edge). In other cases, it performed moderately (Xuan and Wang, 2015). Multi-level  $\lambda$  setting method was proposed (Jie et al., 2016). Different from single  $\lambda$  setting, multi-level  $\lambda$  setting method set a combination of several  $\lambda$  which provided more network structure topology information than the former method. Multi-level  $\lambda$  setting method could avoid the arbitrary decision of single  $\lambda$  setting method and reduce the influence of the low reliability caused by single network structure.

How to get the most optimizing combination of  $\lambda$  values was one of the important thing in the multi-level  $\lambda$  setting method. Enumeration method was not suitable because of the

huge computation consumption caused by the large amount of random combinations. For a nine intervals setting which was adopted in the current study, there were 511 different combinations in total ( $C_9^1 + C_9^2 + C_9^3 + C_9^4 + C_9^5 + C_9^6 + C_9^7 + C_9^8 + C_9^9 = 511$ ). More intervals could result in more combinations. In the current study, a series of serial ascending order combinations was adopted, embodying as  $\{0.1\}$ ,  $\{0.1, 0.2\}$ ,  $\dots$ ,  $\{0.1, 0.2, \dots, 0.9\}$  (nine combinations in total). The method remained small  $\lambda$  values in the combinations as many as possible that means more nodes were connected in the constructed hyper-edges. It was thought that the hyper-edges with many nodes could describe the underlying relationship among several nodes. Reverse order combination was not taken into account because many large  $\lambda$  values was remained in the combinations. Strict  $\lambda$  setting could result in a few nodes in the constructed hyper-edges. In our experiments, it was found that almost all the hyper-edges connected only two nodes when  $\lambda$  was 0.9. It suggested the hyper-network had degenerated into the conventional network. Admittedly, ascending order method was still arbitrary. It was one of the limitations of the current study. A feasible optimizing combination selection method should be researched in the future.

To characterize multi-level relationships within multiple brain regions, it is necessary to set a range of different  $\lambda$  values. The more  $\lambda$  values, the more interaction information among multiple brain regions contained in the hyper-edges. In the current study, we set nine groups of different  $\lambda$  values and the classification results showed that a greater number of  $\lambda$  values corresponded to better classification performance. This result suggests that when a hyper-edge contains more multi-level interaction information, the hyper-network reflects greater structural differences among the different samples. These structural differences could be embodied both by node metrics and hyper-edge connection patterns. Therefore, the advantages of multiple sparse levels indicates the superiority of hyper-networks compared with conventional simple networks.

## Limitations

We proposed a method of machine learning classification combining multiple features of a hyper-network of fMRI data in AD, which could be used to effectively classify normal controls and AD patients. However, there were some limitations. A sparse linear regression model was used to construct the hyper-networks. However, when constructing the hyper-edges, for a chosen brain region, if the pairwise correlations between other brain regions were very high, then this method tended to select only one region with a grouping effect from the group, but did not care which one was selected. It is possible that some related brain regions were not selected, which means the grouping effect information could not be explained. Constructing the hyper-network based on sparse representation confirmed the stability of constructed hyper-edges, which is also an important step. To address this limitation in future studies, we plan to adopt other effective methods, such as the robust least absolute shrinkage and selection operator (LASSO) (Xu et al., 2008) and group LASSO (Yuan and Lin, 2006).

## REPEATABILITY VERIFICATION

To further verify the repeatability of the proposed method, we tested it with the public Alzheimer's Disease Neuroimaging Initiative (ANDI) data set. The data set included data from 94 subjects, including 33 early mild cognitive impairment (EMCI) patients, 32 late mild cognitive impairment (LMCI) patients and 29 AD patients. There were no significant differences in gender or age among the four groups, but mini-mental state examination (MMSE) scores were significantly different among the groups. Demographics and clinical characteristics of the subjects are listed in **Table 6**.

We adopted the same method as described above. First, a sparse linear model was used to construct the hyper-network with the EMCI data set. Local brain region properties (HCC1, HCC2, and HCC3) and subgraphs were then extracted as features. Finally, two different features were combined, and multi-kernel SVM was adopted to perform classification. Ten-fold cross-validation was repeated 100 times. The experimental protocol was repeated with both the LMCI and AD data sets.

The local brain region properties HCC1, HCC2, and HCC3 were further analyzed, and significant differences ( $p < 0.05$ , FDR correction) in abnormal brain regions were selected for the three groups. **Tables 7–9** list the abnormal brain regions between NC group and EMCI group, LMCI group, AD group respectively. As above, the FSFS algorithm was adopted to select discriminative subgraphs in the three data sets. To achieve a similar number of discriminative subgraph features to that in our previous study, the discriminative score  $K$  threshold was set at 0.2. We obtained 24 discriminative subgraphs between the NC and EMCI groups, and 30 discriminative subgraphs between the NC and LMCI groups, and 32 discriminative subgraphs between the NC and AD groups. We further analyzed the subgraph patterns for each group pair, revealing the abnormal brain regions between them. Detailed results are shown in the Supplemental Figure S1.

We compared the abnormal regions of subgraph features of AD patients in our collected dataset and ADNI dataset. It was found that there were many overlapped abnormal regions between two datasets (**Table 10**). In the top 10 abnormal regions, there were seven regions found both in our collected dataset and ADNI dataset, including right hippocampus, right lingual gyrus, right fusiform gyrus, left fusiform gyrus, left precuneus, left angular gyrus and right superior temporal gyrus. In addition, we compared the abnormal brain regions of the EMCI, LMCI, and AD groups in ADNI dataset (**Table 11**). The results showed that there were a large number of overlapped abnormal brain regions between EMCI and LMCI, which included right dorsolateral superior frontal gyrus, left middle frontal gyrus, right middle frontal gyrus, left medial superior frontal gyrus, right medial superior frontal gyrus, left insula, left superior occipital gyrus and left parahippocampal gyrus. The above overlapped regions showed that the subgraph features had good repeatability and stability. Analysis showed that many abnormal brain regions of EMCI and LMCI groups were located in the frontal lobe and limbic system. It was noted that only one overlapped abnormal brain region (right hippocampus) was found between LMCI and AD groups. The results indicated that there were obvious

differences in the abnormal brain regions obtained by subgraph features of different diseases.

Meanwhile, we compared the abnormal regions of brain region features of AD patients in our collected dataset and ADNI dataset. Differentiating from subgraph features, the results of brain region features showed that there were only two overlapped abnormal brain regions, including left fusiform gyrus and left thalamus (**Table 12**). In addition, after compared among EMCI, LMCI and AD groups in ADNI dataset, the results of abnormal regions showed a great deal of difference (**Table 13**). There was only one brain region (right postcentral gyrus) appearing in the all three groups. Only one brain region (right olfactory cortex) overlapped in EMCI and AD groups and three brain regions (right amygdala, right temporal pole: middle temporal gyrus and right calcarine fissure and surrounding cortex) overlapped in LMCI and AD groups. There was not any region overlapped in EMCI and LMCI groups. The results showed that, compared the subgraph features, the brain region features were not stable. The abnormal brain regions obtained by brain region features were significantly different in different datasets or different diseases. Furthermore, the contrast analysis found a result consistent with subgraph features. The result showed that the abnormal brain regions mainly were located in the frontal lobe and limbic lobe in EMCI and LMCI groups (Pennanen et al., 2005; Whitwell et al., 2008; Schroeter et al., 2009; Wang et al., 2011).

In this paper, the subgraph features and the brain region features were used as the features of classification, which described the different network properties from different perspectives. Subgraph features were represented as connected patterns while brain region features were represented as quantifiable values. Compared with brain region features, subgraph features showed better repeatability and stability. To be specific, there were more overlapped abnormal brain regions of subgraph features, no matter in different datasets (AD group in our collected dataset vs. AD group in ADNI dataset) or different diseases (EMCI group in ADNI dataset vs. LMCI group in ADNI dataset). The result suggested that the differences of network structure, which were embodied by connected pattern, were not susceptible to the different datasets. It should be noted that the characteristic of subgraph features also implied that it was insensitive to the changes of samples. On the contrary, the brain region features were sensitive to the changes of samples. Different datasets showed significant differences in abnormal brain regions. Therefore, the abnormal brain regions obtained by brain region features in a dataset were difficult to apply to other datasets. Although the repeatability of the brain region features was not strong, the quantifiable local property could capture the specific inter-group differences. These differences could distinguish between the given groups, despite they were not repeatable in other datasets. The direct evidence of this conclusion was that the classification accuracy of brain region features was higher than that of subgraph features in the both datasets (**Tables 5, 14**).

We performed classification in the three different group pairs separately, including NC vs. EMCI, NC vs. LMCI and NC vs. AD. In addition, to compare the different feature extraction methods, we also performed classification using only brain region

features and subgraph features (Table 14). In the three different group pairs, the multi-feature method consistently showed better classification performance than the single feature method. This result is consistent with the findings of our previous study. In addition, the classification accuracy between the NC and AD groups reached 88.91%, which was closed to the results of our self-collection data set (91.60%). These findings suggest that our proposed method is stable and repeatable.

It should be noted that the performance of the proposed method with EMCI and LMCI patients was relatively low (72.80 and 78.63%, respectively). These findings suggest that the hyper-network method was unable to reveal differences in network structure in the early stages of disease development. Thus, the selected network features, brain regions, or subgraph features, appear to have been insufficient for describing between-group differences effectively (the classification accuracy with single feature types was below 80%, in both the EMCI and LMCI groups). In addition, we analyzed the relief weights of features and MMSE scores among three disease groups (Figure 10). The analysis revealed that the relief weights increased gradually with the reduction in MMSE scores. This result demonstrates that the classification accuracy gradually increased with the development of the disease. The severity of the disease would be expected to enhance the differences in network structure between patients and normal controls. However, in the early stage of illness, particularly in the EMCI group, the hyper-network method was unable to reveal differences in network structure, compared with normal controls. This represents a potential limitation of the proposed method.

## CONCLUSION

Compared with the conventional methods of constructing functional connectivity networks, a hyper-network can reflect the interaction information among multiple brain regions and improve the classification of disease using higher-order information. However, existing methods use brain region features for classification, but an obvious deficiency of this method is that some useful topology information might be lost. To address the current limitations of conventional network modeling approaches, we proposed a method of machine learning classification combining multiple features of a hyper-network using fMRI data in AD. The proposed method has two important advantages. First, the method considers the interactions among brain regions and thus reflects more complex interactions. Second, it combines two types of complementary features for feature extraction, which ensures the integrity of the structural information and the sensitivity to change in a single brain region. The results of analyses with two different data sets showed that the proposed method improved classification

## REFERENCES

Alvarez, M. A., Qi, X., and Yan, C. (2011). A shortest-path graph kernel for estimating gene product semantic similarity. *J. Biomed. Semant.* 2, 1–9. doi: 10.1186/2041-1480-2-3

performance of AD, compared with conventional methods. However, it should be noted that the proposed method was unable to identify EMCI patients because of the lack of significant structural differences of hyper-networks in these patients.

## AUTHOR CONTRIBUTIONS

HG was responsible for the study design and writing the manuscript. FZ performed data analysis and statistical processing. YX provided and integrated experimental data. JC supervised the paper. JX was the head of the funds and supervised the paper. All authors approved the final version of the manuscript.

## FUNDING

This study was supported by research grants from the National Natural Science Foundation of China (61373101, 61472270, 61402318, 61672374), Natural Science Foundation of Shanxi Province (201601D021073) and Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2016139). The sponsors had no role in the design or execution of the study, the collection, management, analysis, and interpretation of the data, or preparation, review, and approval of the manuscript. This manuscript has not been published or presented elsewhere in part or in entirety, and is not under consideration by any another journal. This study was approved by the medical ethics committee of Shanxi Province, and the approved certification number is 2012013. All subjects have been given written informed consent in accordance with the Declaration of Helsinki. Meanwhile, all the authors have read through the manuscript, approved it for publication, and declared no conflict of interest. JX had full access to all of the data in the study and takes responsibility for its integrity and the accuracy of data analysis.

## ACKNOWLEDGMENTS

This research was supported by research grants from the National Natural Science Foundation of China (61373101, 61472270, and 61402318), Natural Science Foundation of Shanxi Province (201601D021073) and Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2016139).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2017.00615/full#supplementary-material>

Barthélemy, M. (2004). Betweenness centrality in large complex networks. *Eur. Phys. J. B* 38, 163–168. doi: 10.1140/epjb/e2004-00111-4

Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., Meyerlindenberg, A., et al. (2008). Hierarchical organization of human cortical

- networks in health and schizophrenia. *J. Neurosci. Offic. J. Soc. Neurosci.* 28, 9239–9248. doi: 10.1523/JNEUROSCI.1929-08.2008
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300.
- Berg, L., McKeel, D. W., Miller, J. P., Storandt, M., Rubin, E. H., Morris, J. C., et al. (1998). Clinicopathologic studies in cognitively healthy aging and Alzheimer's disease: relation of histologic markers to dementia severity, age, sex, and apolipoprotein E genotype. *Arch. Neurol.* 55:326. doi: 10.1001/archneur.55.3.326
- Biao, J., Dinggang, S., and Daoqiang, Z. (2014). Brain connectivity hyper-network for MCI classification. *Med. Image Comput. Comput. Assist. Interv.* 17, 724–732.
- Binnewijzend, M. A., Schoonheim, M. M., Sanz-Arigitia, E., Wink, A. M., Wm, V. D. F., Tolboom, N., et al. (2012). Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment. *Neurobiol. Aging* 33, 2018–2028. doi: 10.1016/j.neurobiolaging.2011.07.003
- Borgwardt, K. M., Cheng, S. O., Schönauer, S., Vishwanathan, S. V. N., Smola, A. J., Kriegel, H. P., et al. (2005). Protein function prediction via graph kernels. *Bioinformatics* 21(Suppl 1):i47.
- Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., et al. (2012). Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage* 59:1404. doi: 10.1016/j.neuroimage.2011.08.044
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N.Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10:186. doi: 10.1038/nrn2575
- Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S., Sharma, T., et al. (2000). How good is good enough in path analysis of fMRI data? *Neuroimage* 11, 289–301. doi: 10.1006/nimg.2000.0544
- Cavanna, A. E., and Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129:564. doi: 10.1093/brain/awl004
- Chang, C. C., and Lin, C. J. (2011). LIBSVM: a library for support vector machines. *Acm. Trans. Intell. Sys. Technol.* 2:27. doi: 10.1145/1961189.1961199
- Chen, G., Ward, B. D., Xie, C., Li, W., Wu, Z., Jones, J. L., et al. (2011). Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. *Radiology* 259, 213–231. doi: 10.1148/radiol.10100734
- Chen, X., Yang, J., Ye, Q., and Liang, J. (2011). Recursive projection twin support vector machine via within-class variance minimization. *Patt. Recog.* 44, 2643–2655. doi: 10.1016/j.patcog.2011.03.001
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- De-Jong, L., Der-Hiele, K. V., Veer, I. M., Ji H, Westendorp, R., Bollen, E., et al. (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain A J. Neurol.* 131, 3277–3285. doi: 10.1093/brain/awn278
- Delbeuck, X., Linden, M. V. D., and Collette, F. (2003). Alzheimer' disease as a disconnection syndrome? *Neuropsychol. Rev.* 13, 79–92. doi: 10.1023/A:1023832305702
- Devanand, D. P., Habeck, C. G., Tabert, M. H., Scarmeas, N., Pelton, G. H., Moeller, J. R., et al. (2006). PET network abnormalities and cognitive decline in patients with mild cognitive impairment. *Neuropsychopharmacol. Offic. Public. Am. Coll. Neuropsychopharmacol.* 31, 1327–1334. doi: 10.1038/sj.npp.1300942
- Dj, W., and Sh, S. (1998). Collectivedynamics of 'small-world' networks. *Nature* 393, 440–442.
- Du, J., Wang, L., Jie, B., and Zhang, D. (2016). Network-based classification of ADHD patients using discriminative subnetwork selection and graph kernel PCA. *Comput. Med. Imaging Graph. Offic. J. Comput. Med. Imaging Soc.* 52:82. doi: 10.1016/j.compmedimag.2016.04.004
- Fasano, G., and Franceschini, A. (1987). A multidimensional version of the Kolmogorov–Smirnov test. *Month. Notic. R. Astron. Soc.* 50, 9–20. doi: 10.1093/mnras/225.1.155
- Fei, F., Jie, B., and Zhang, D. (2014). Frequent and discriminative subnetwork mining for mild cognitive impairment classification. *Brain Connect.* 4:347. doi: 10.1089/brain.2013.0214
- Fornito, A., Zalesky, A., and Breakspear, M. (2013). Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage* 80, 426–444. doi: 10.1016/j.neuroimage.2013.04.087
- Fransson, P., and Marrelec, G. (2008). The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: evidence from a partial correlation network analysis. *Neuroimage* 42, 1178–1184. doi: 10.1016/j.neuroimage.2008.05.059
- Friston, K. J. (2007). Statistical parametric mapping: the analysis of functional brain images. *Neurosurgery* 61, 216–216. doi: 10.1016/B978-012372560-8/50002-4
- Gallagher, S. R., and Goldberg, D. S. (2013). "Clustering coefficients in protein interaction hypernetworks," in *International Conference on Bioinformatics, Computational Biology and Biomedical Informatics* (Washington, DC).
- Gärtner, T., Flach, P., and Wrobel, S. (2003). "On graph kernels: hardness results and efficient alternatives," in *Computational Learning Theory and Kernel Machines, Conference on Computational Learning Theory and Kernel Workshop, Colt/kernel 2003* (Washington, DC).
- Grady, C. L., McIntosh, A. R., Beigal, S., Keightley, M. L., Burian, H., and Sandra, E. B. (2003). Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. *J. Neurosci. Offic. J. Soc. Neurosci.* 23:986.
- Greicius, M. D., Srivastava, G., Reiss, A. L., and Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4637–4642. doi: 10.1073/pnas.0308627101
- Guo, H., Cao, X., Liu, Z., and Chen, J. (2013). Abnormal functional brain network metrics for machine learning classifier in depression patients identification. *Res. J. Appl. Sci. Eng. Technol.* 5, 3015–3020.
- Gusnard, D. A., and Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2, 685–694. doi: 10.1038/35094500
- Hampel, H., Teipel, S. J., Bayer, W., Alexander, G. E., Schwarz, R., Schapiro, M. B., et al. (2002). Age transformation of combined hippocampus and amygdala volume improves diagnostic accuracy in Alzheimer's disease. *J. Neurol. Sci.* 194, 15–19. doi: 10.1016/S0022-510X(01)00669-4
- Harchaoui, Z., and Bach, F. (2007). "Image classification with segmentation graph kernels," in *Computer Vision and Pattern Recognition, CVPR '07. IEEE Conference 2007* (Minneapolis).
- Haroutunian, V., Davies, P., Vianna, C., Buxbaum, J. D., and Purohit, D. P. (2007). Tau protein abnormalities associated with the progression of alzheimer disease type dementia. *Neurobiol. Aging* 28:1. doi: 10.1016/j.neurobiolaging.2005.11.001
- He, Y., Chen, Z., and Evans, A. (2008). Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease. *J. Neurosci.* 4, 4756–4766. doi: 10.1523/JNEUROSCI.0141-08.2008
- He, Y., Wang, L., Zang, Y., Tian, L., Zhang, X., Li, K., et al. (2007). Regional coherence changes in the early stages of Alzheimer's disease: a combined structural and resting-state functional MRI study. *Neuroimage* 35:488. doi: 10.1016/j.neuroimage.2006.11.042
- Hofmann, T., Schölkopf, B., and Smola, A. J. (2007). Kernel methods in machine learning. *Ann. Statist.* 36, 1171–1220. doi: 10.1214/00905360700000677
- Horwitz, B. (2003). The elusive concept of brain connectivity. *Neuroimage* 19, 466–470. doi: 10.1016/S1053-8119(03)00112-5
- Huang, S., Li, J., Sun, L., Ye, J., Fleisher, A., Wu, T., et al. (2010). Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *Neuroimage* 50, 935–949. doi: 10.1016/j.neuroimage.2009.12.120
- Jie, B., Wee, C. Y., Shen, D., and Zhang, D. (2016). Hyper-connectivity of functional networks for brain disease diagnosis. *Med. Image Anal.* 32:84. doi: 10.1016/j.media.2016.03.003
- Jie, B., Zhang, D., Gao, W., Wang, Q., Wee, C. Y., and Shen, D. (2014). Integration of network topological and connectivity properties for neuroimaging classification. *IEEE Trans. Bio Med. Eng.* 61, 576–589. doi: 10.1109/TBME.2013.2284195
- Kaiser, M. (2011). A tutorial in connectome analysis: topological and spatial features of brain networks. *Neuroimage* 57, 892–907. doi: 10.1016/j.neuroimage.2011.05.025

- Kira, K., and Rendell, L. A. (1992). "The feature selection problem: traditional methods and a new algorithm," in *National Conference on Artificial Intelligence* (San Jose, CA).
- Kong, X., Yu, P. S., Wang, X., and Ragin, A. B. (2013). Discriminative feature selection for uncertain graph classification. *arXiv:1301.6626*.
- Lanckriet, G. R. G., Cristianini, N., Bartlett, P., El Ghaoui, L., and Jordan, M. I. (2002). "Learning the kernel matrix with semi-definite programming," in *Proceedings of the Nineteenth International Conference Machine Learning* (Sydney).
- Lee, H., Lee, D. S., Kang, H., Kim, B. N., and Chung, M. K. (2011). Sparse brain network recovery under compressed sensing. *IEEE Trans. Med. Imaging* 30, 1154–1165. doi: 10.1109/TMI.2011.2140380
- Li, Y., Long, J., He, L., Lu, H., Gu, Z., and Sun, P. (2012). A sparse representation-based algorithm for pattern localization in brain imaging data analysis. *PLoS ONE* 7:e50332. doi: 10.1371/journal.pone.0050332
- Liang, P., Li, Z., Deshpande, G., Wang, Z., Hu, X., and Li, K. (2013). Altered causal connectivity of resting state brain networks in amnesic MCI. *PLoS ONE* 9:e88476. doi: 10.1371/journal.pone.0088476
- Lim, T. S., Lee, H. Y., Barton, J., and Moon, S. (2011). Deficits in face perception in the amnesic form of mild cognitive impairment. *J. Neurol. Sci.* 309:123. doi: 10.1016/j.jns.2011.07.001
- Liu, F., Guo, W., Fouche, J. P., Wang, Y., Wang, W., Ding, J., et al. (2015). Multivariate classification of social anxiety disorder using whole brain functional connectivity. *Brain Struct. Funct.* 220, 101–115. doi: 10.1007/s00429-013-0641-4
- Liu, J., Ji, S., and Ye, J. (2013). *SLEP: Sparse Learning with Efficient Projections*. Phoenix: Arizona State University.
- Liu, J., Zhang, X., Yu, C., Duan, Y., Zhuo, J., Cui, Y., et al. (2015). Impaired parahippocampus connectivity in mild cognitive impairment and Alzheimer's disease. *J. Alzheimers Dis.* 49:1051. doi: 10.3233/JAD-150727
- Liu, X., Wang, S., Zhang, X., Wang, Z., Tian, X., He, Y., et al. (2014). Abnormal amplitude of low-frequency fluctuations of intrinsic brain activity in Alzheimer's disease. *J. Alzheim. Dis.* 40, 387–397. doi: 10.3233/JAD-131322
- Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., et al. (2008). Disrupted small-world networks in schizophrenia. *Brain* 131, 945–961. doi: 10.1093/brain/awn018
- Lv, J., Jiang, X., Zhu, D., Chen, H., Zhang, T., Zhang, S., et al. (2015). Sparse representation of whole-brain fMRI signals for identification of functional networks. *Med. Image Anal.* 20:112. doi: 10.1016/j.media.2014.10.011
- Lynall, M. E., Bassett, D. S., Kerwin, R., Mckenna, P. J., Kitzbichler, M., Müller, U., et al. (2010). Functional connectivity and brain networks in schizophrenia. *J. Neurosci.* 30, 9477–9487. doi: 10.1523/JNEUROSCI.0333-10.2010
- Markesbery, W. R., Schmitt, F. A., Kryscio, R. J., Davis, D. G., Smith, C. D., Wekstein, D. R., et al. (2006). Neuropathologic substrate of mild cognitive impairment. *Arch. Neurol.* 63, 38–46. doi: 10.1001/archneur.63.1.38
- Marreiros, A. C., Stephan, K. E., and Friston, K. J. (2010). Dynamic causal modeling. *Neuroimage* 19, 1273–1302. doi: 10.4249/scholarpedia.9568
- Mcintosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I., Horwitz, B., et al. (1994). Network analysis of cortical visual pathways mapped with PET. *J. Neurosci. Offic. J. Soci. Neurosci.* 14, 655–666.
- Montani, F., Ince, R. A., Senatore, R., Arabzadeh, E., and Diamond, M. E., Panzeri, S., et al. (2009). The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. *Philos. Trans.* 367, 3297–3310. doi: 10.1098/rsta.2009.0082
- Morgen, K., Frolich, L., Tost, H., Plichta, M. M., Kolsch, H., Rakebrandt, F., et al. (2013). APOE-dependent phenotypes in subjects with mild cognitive impairment converting to Alzheimer's disease. *J. Alzheim. Dis.* 37, 389–401. doi: 10.3233/JAD-130326
- Pennanen, C., Testa, C., Laakso, M. P., Hallikainen, M., Helkala, E. L., Hanninen, T., et al. (2005). A voxel based morphometry study on mild cognitive impairment. *J. Neurol. Neurosurg. Psychiatry* 76:11. doi: 10.1136/jnnp.2004.035600
- Pievani, M., Agosta, F., Galluzzi, S., Filippi, M., and Frisoni, G. B. (2011). Functional networks connectivity in patients with Alzheimer's disease and mild cognitive impairment. *J. Neurol.* 258, 170.
- Poettrich, K., Weiss, P. H., Werner, A., Lux, S., Donix, M., Gerber, J., et al. (2009). Altered neural network supporting declarative long-term memory in mild cognitive impairment. *Neurobiol. Aging* 30:284. doi: 10.1016/j.neurobiolaging.2007.05.027
- Qiao, L., Zhang, H., Kim, M., Teng, S., Zhang, L., Shen, D., et al. (2016). Estimating functional brain networks by incorporating a modularity prior. *Neuroimage* 141:399. doi: 10.1016/j.neuroimage.2016.07.058
- Richardson, M. (2010). Current themes in neuroimaging of epilepsy: brain networks, dynamic phenomena, and clinical relevance. *Clin. Neurophysiol. Offic. J. Int. Feder. Clin. Neurophysiol.* 121, 1153–1175. doi: 10.1016/j.clinph.2010.01.004
- Rosa, M. J., Portugal, L., Shawe-Taylor, J., and Mourao-Miranda, J. (2015). Sparse network-based models for patient classification using fMRI. *Neuroimage* 105:493. doi: 10.1016/j.neuroimage.2014.11.021
- Rosario, S. F., and Thangadurai, K. (2015). *RELIEF: Feature Selection Approach*.
- Santos, G. S., Gireesh, E. D., Plenz, D., and Nakahara, H. (2010). Hierarchical Interaction Structure of Neural Activities in Cortical Slice Cultures. *J. Neurosci. Offic. J. Soc. Neurosci.* 30, 8720–8733. doi: 10.1523/JNEUROSCI.6141-09.2010
- Saramäki, J., Kivela, M., Onnela, J. P., Kaski, K., and Kertész, J. (2006). Generalizations of the clustering coefficient to weighted complex networks. *Phys. Rev. E Stat. Nonlinear Soft Matt. Phys.* 75(2 Pt 2):027105.
- Schölkopf, B., Platt, J., and Hofmann, T. (2007). "Learning with hypergraphs: clustering, classification, and embedding," in *Conference on Advances in Neural Information Processing Systems* (Vancouver, BC).
- Schroeter, M. L., Stein, T., Maslowski, N., and Neumann, J. (2009). Neural correlates of Alzheimer's disease and mild cognitive impairment: a systematic and quantitative meta-analysis involving 1351 patients. *Neuroimage* 47:1196. doi: 10.1016/j.neuroimage.2009.05.037
- Shervashidze, N., Schweitzer, P., Van Leeuwen, E. J., Mehlhorn, K., and Borgwardt, K. M. (2011). Weisfeiler-Lehman Graph Kernels. *J. Machine Learn. Res.* 12, 2539–2561.
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for FMRI. *Neuroimage* 54:875. doi: 10.1016/j.neuroimage.2010.08.063
- Solépadullés, C., Bartrésfàz, D., Junqué, C., Vendrell, P., Rami, L., Clemente, I. C., et al. (2009). Brain structure and function related to cognitive reserve variables in normal aging, mild cognitive impairment and Alzheimer's disease. *Neurobiol. Aging* 30, 1114–1124. doi: 10.1016/j.neurobiolaging.2007.10.008
- Sporns, O. (2012). From simple graphs to the connectome: networks in neuroimaging. *Neuroimage* 62:881. doi: 10.1016/j.neuroimage.2011.08.085
- Stam, C. J., Haan, W. D., Daffertshofer, A., Jones, B. F., Manshanden, I., and van Cappellen van Walsum, A. M. (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. *Brain* 132, 213–224. doi: 10.1093/brain/awn262
- Statistics, I. O. M. (1998). The annals of statistics. *Probab. Theory Relat. Fields* 26, 2369–2394.
- Supekar, K., Menon, V., Rubin, D., Musen, M., and Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Comput. Biol.* 4:e1000100. doi: 10.1371/journal.pcbi.1000100
- Vasavada, M. M., Wang, J., Eslinger, P. J., Gill, D., Sun, X., Karunanayak, P., et al. (2015). Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment. *J. Alzheimers Dis.* 45, 947–958. doi: 10.3233/JAD-141947
- Vishwanathan, S. V. N., Schraudolph, N. N., Kondor, R., and Borgwardt, K. M. (2008). Graph Kernels. *J. Mach. Learn. Res.* 11, 1201–1242.
- Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., et al. (2013). Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. *Biol. Psychiatry* 73, 472–481. doi: 10.1016/j.biopsych.2012.03.026
- Wang, L., Zhu, C., He, Y., Zang, Y., Cao, Q., Zhang, H., et al. (2009). Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. *Hum. Brain Map.* 30, 638–649. doi: 10.1002/hbm.20530
- Wang, Z., Yan, C., Cheng, Z., Qi, Z., Zhou, W., Jie, L., et al. (2011). Spatial patterns of intrinsic brain activity in mild cognitive impairment and Alzheimer's disease: a resting-state functional MRI study. *Hum. Brain Map.* 32, 1720. doi: 10.1002/hbm.21140
- Wee, C. Y., Yang, S., Yap, P. T., and Shen, D. (2016). Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. *Brain Imaging Behav.* 10, 342–356. doi: 10.1007/s11682-015-9408-2
- Wee, C. Y., Yap, P. T., Denny, K., Brownnyke, J. N., Potter, G. G., Welsh-Bohmer, K. A., et al. (2012). Resting-state multi-spectrum functional

- connectivity networks for identification of MCI patients. *PLoS ONE* 7:e37828. doi: 10.1371/journal.pone.0037828
- Wee, C. Y., Yap, P. T., Zhang, D., Wang, L., and Shen, D. (2014). Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Struct. Funct.* 219:641. doi: 10.1007/s00429-013-0524-8
- Whitwell, J. L., Shiung, M. M., Przybelski, S. A., Weigand, S. D., Knopman, D. S., Boeve, B. F., et al. (2008). MRI patterns of atrophy associated with progression to AD in amnesic mild cognitive impairment. *Neurology* 70, 512–520. doi: 10.1212/01.wnl.0000280575.77437.a2
- Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Trans. Patt. Anal. Mach. Intell.* 31:210. doi: 10.1109/TPAMI.2008.79
- Xu, H., Caramanis, C., and Mannor, S. (2008). Robust regression and Lasso. *Inform. Theory IEEE Trans.* 56, 3561–3574. doi: 10.1109/TIT.2010.2048503
- Xuan, L., and Wang, H. (2015). Identification of functional networks in resting state fMRI data using adaptive sparse representation and affinity propagation clustering. *Front. Neurosci.* 9:383. doi: 10.3389/fnins.2015.00383
- Yao, H., Zhou, B., Zhang, Z., Wang, P., Guo, Y., Shang, Y., et al. (2014). Longitudinal alteration of amygdalar functional connectivity in mild cognitive impairment subjects revealed by resting-state FMRI. *Brain Connect.* 4, 361–370. doi: 10.1089/brain.2014.0223
- Yetkin, F. Z., Rosenberg, R. N., Weiner, M. F., Purdy, P. D., and Cullum, C. M. (2006). FMRI of working memory in patients with mild cognitive impairment and probable Alzheimer's disease. *Eur. Radiol.* 16, 193–206. doi: 10.1007/s00330-005-2794-x
- Yu, J., Tao, D., and Wang, M. (2012). Adaptive hypergraph learning and its application in image classification. *IEEE Trans. Image Proces.* 21, 3262–3272. doi: 10.1109/TIP.2012.2190083
- Yu, S., Yang, H., Nakahara, H., Santos, G. S., Nikolić, D., and Plenz, D. (2011). Higher-order interactions characterized in cortical activity. *J. Neurosci. Offic. J. Soc. Neurosci.* 31, 17514–17526. doi: 10.1523/JNEUROSCI.3127-11.2011
- Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc.* 68, 49–67. doi: 10.1111/j.1467-9868.2005.00532.x
- Zhang, H., Wang, S., J., Liu, B., Ma, Z., Yang, M., and Zhang, Z. (2009). Detection of PCC functional connectivity characteristics in resting-state fMRI in mild Alzheimer's disease. *Behav. Brain Res.* 197, 103–108. doi: 10.1016/j.bbr.2008.08.012
- Zhao, H., Li, X., Wu, W., Li, Z., Qian, L., Li, S. S., et al. (2015). Atrophic patterns of the frontal-subcortical circuits in patients with mild cognitive impairment and Alzheimer's disease. *PLoS ONE* 10:e0130017. doi: 10.1371/journal.pone.0130017
- Zhou, L., Wang, L., and Ogunbona, P. (2014). "Discriminative sparse inverse covariance matrix: application in brain functional network classification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers), 3097–3104.
- Zhou, Y., and Jin, J. (2008). Effect of acupuncture given at the HT 7, ST 36, ST 40 and KI 3 acupoints on various parts of the brains of Alzheimer's disease patients. *Acupunct. Electro Therapeut. Res.* 33:9. doi: 10.3727/036012908803861186

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Guo, Zhang, Chen, Xu and Xiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.