



# Resting-State Brain Functional Hyper-Network Construction Based on Elastic Net and Group Lasso Methods

Hao Guo<sup>1,2</sup>, Yao Li<sup>1</sup>, Yong Xu<sup>3</sup>, Yanyi Jin<sup>1</sup>, Jie Xiang<sup>1</sup> and Junjie Chen<sup>1\*</sup>

<sup>1</sup> Department of Software Engineering, College of Information and Computer, Taiyuan University of Technology, Taiyuan, China, <sup>2</sup> National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences, Beijing, China, <sup>3</sup> Department of Psychiatry, First Hospital of Shanxi Medical University, Taiyuan, China

## OPEN ACCESS

### Edited by:

Pedro Antonio Valdes-Sosa,  
Clinical Hospital of Chengdu Brain  
Science Institute, China

### Reviewed by:

Jan Casper De Munck,  
VU University Amsterdam,  
Netherlands  
Zhichao Lian,  
Nanjing University of Science and  
Technology, China

### \*Correspondence:

Junjie Chen  
feiyu\_guo@sina.com

**Received:** 27 January 2018

**Accepted:** 25 April 2018

**Published:** 15 May 2018

### Citation:

Guo H, Li Y, Xu Y, Jin Y, Xiang J and  
Chen J (2018) Resting-State Brain  
Functional Hyper-Network  
Construction Based on Elastic Net  
and Group Lasso Methods.  
*Front. Neuroinform.* 12:25.  
doi: 10.3389/fninf.2018.00025

Brain network analysis has been widely applied in neuroimaging studies. A hyper-network construction method was previously proposed to characterize the high-order relationships among multiple brain regions, where every edge is connected to more than two brain regions and can be represented by a hyper-graph. A brain functional hyper-network is constructed by a sparse linear regression model using resting-state functional magnetic resonance imaging (fMRI) time series, which in previous studies has been solved by the lasso method. Despite its successful application in many studies, the lasso method has some limitations, including an inability to explain the grouping effect. That is, using the lasso method may cause relevant brain regions be missed in selecting related regions. Ideally, a hyper-edge construction method should be able to select interacting brain regions as accurately as possible. To solve this problem, we took into account the grouping effect among brain regions and proposed two methods to improve the construction of the hyper-network: the elastic net and the group lasso. The three methods were applied to the construction of functional hyper-networks in depressed patients and normal controls. The results showed structural differences among the hyper-networks constructed by the three methods. The hyper-network structure obtained by the lasso was similar to that obtained by the elastic net method but very different from that obtained by the group lasso. The classification results indicated that the elastic net method achieved better classification results than the lasso method with the two proposed methods of hyper-network construction. The elastic net method can effectively solve the grouping effect and achieve better classification performance.

**Keywords:** depression, hyper-network, elastic net, group lasso, classification

## INTRODUCTION

Evidence from numerous anatomical and physiological studies suggests that cognitive processing depends on the interaction among distributed brain regions (Sporns, 2014). A brain functional network is a simplified representation of brain interactions and has been widely applied in studies of mental disorders, including epilepsy (Zhang et al., 2012), major depressive disorder (MDD) (Kaiser et al., 2016), Alzheimer's disease (Pievani et al., 2011), and schizophrenia (Lynall et al., 2010). The

rapid development of neuroimaging technology provides a good foundation for research on brain functional networks. In recent years, the use of resting-state functional magnetic resonance imaging (fMRI) for mapping neural functional networks has attracted increasing attention. A low frequency blood oxygen level-dependent (BOLD) signal is associated with spontaneous neuronal activity in the brain (Zeng et al., 2012), and the interaction among brain regions in a resting state can be represented by the functional network constructed by the BOLD signal.

Various analytical methods have been proposed for modeling brain functional connectivity from fMRI data, including correlation (Bullmore and Sporns, 2009; Sporns, 2011; Wee et al., 2012; Jie et al., 2014), graphical modeling (Bullmore et al., 2000; Chen and Herskovits, 2007), partial correlation (Salvador et al., 2005; Marrelec et al., 2006, 2007), and sparse representation methods (Lee et al., 2011; Wee et al., 2014). The correlation method is the most common and has been successfully applied to the classification of patients and normal controls (Zeng et al., 2012; Ye et al., 2015). However, one of its main limitations is that it can only capture pairwise information and therefore cannot fully reflect the interactions among multiple brain regions (Huang et al., 2010). Moreover, a network based on correlations may include a number of false connections due to the arbitrary selection of thresholds (Wee et al., 2014). The other methods also have their shortcomings. When used to study brain connections, graphical models lack prior knowledge (Huang et al., 2010), such as which brain regions should be involved and how they are connected. Partial correlation estimation is usually achieved using the maximum likelihood estimation (MLE) of the inverse covariance matrix. However, a limitation of this method is that the data sample size required for a reliable estimate is much larger than the number of modeled brain regions (Huang et al., 2010). Sparse inverse covariance estimation (Zhou et al., 2014; Fu et al., 2015) resolves the deficiencies of MLE to some extent, but there are still problems with this approach. Although it is effective for learning sparse connection networks, it is not suitable for evaluating the connections due to shrinkage effects (Smith et al., 2011). Sparse representation can filter out false or insignificant connections by applying regularization parameters to produce sparse networks. However, apart from sparse structures, brain functional networks usually include other types, such as small world, scale-free topology, hierarchical, and modular structures (Sporns, 2010). Wee et al. (2014) adopted the group lasso method using  $l_{2,1}$  regularization for functional connectivity modeling to estimate networks with the same topology but different connection strengths, while ignoring the network topology patterns of specific groups.

The traditional methods describe the relationship between two regions. However, later studies indicate that interactions take place not only between two regions, but among multiple regions. Recent neuroscience studies have identified significant higher-order interactions in neuronal spiking, local field potentials, and cortical activities (Montani et al., 2009; Ohiorhenuan et al., 2010; Yu et al., 2011). In particular, studies suggest that one brain region predominantly interacts directly with a few other brain

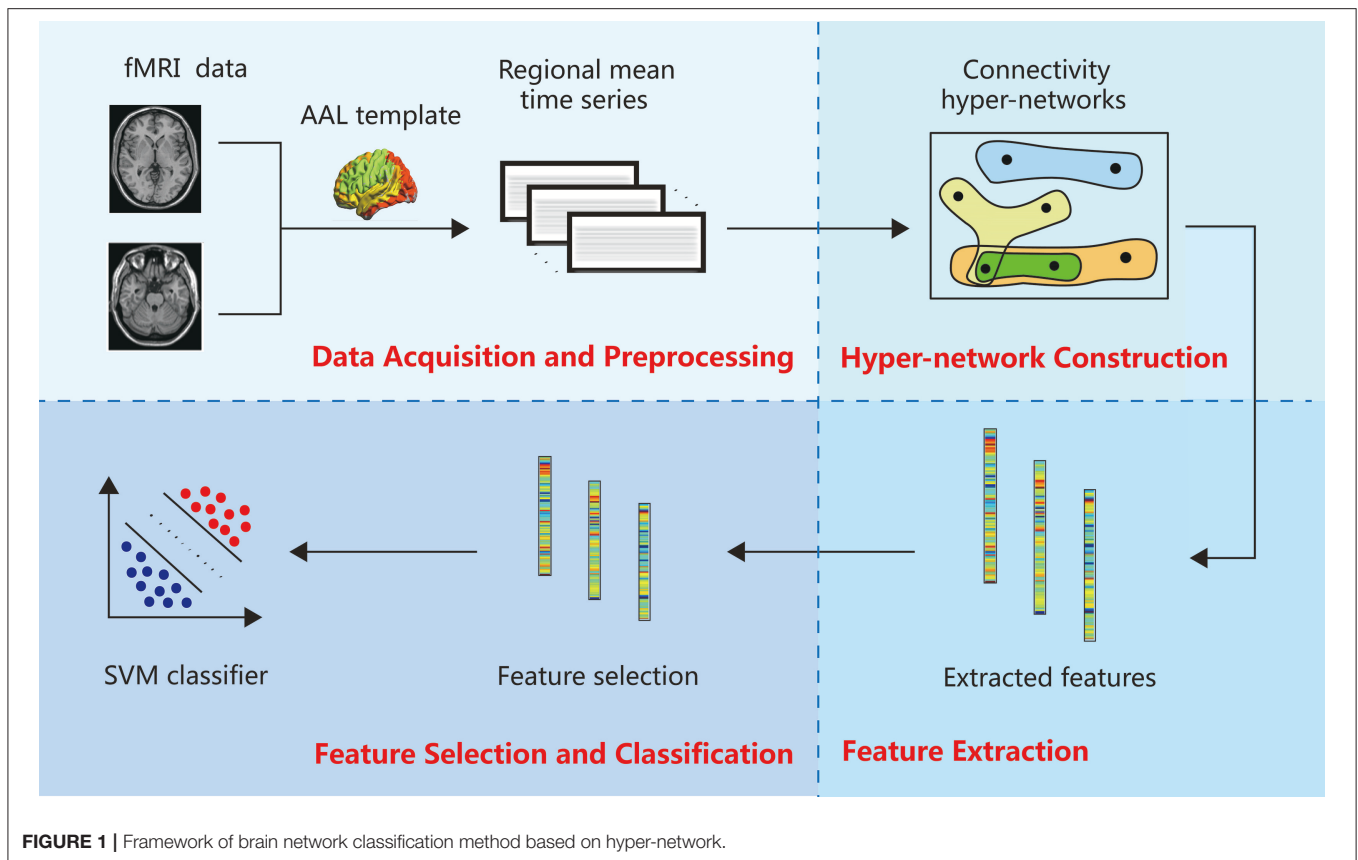
regions in neurological processes (Huang et al., 2010). Functional networks based on pairwise relationships can thus only reflect the second-order relationships between brain regions, ignoring the high-order relationships that may be crucial for studies of underlying pathology.

The hyper-network (Jie et al., 2016) method was proposed to address this issue. Hyper-networks (Jie et al., 2016) are based on hyper-graph theory. Each node represents a brain region and each hyper-edge includes many nodes that represent the interactions among multiple brain regions. The existing method of constructing brain functional hyper-network is to use sparse regression model. According to the model, the sparse solution could be produced, and the nonzero elements in sparse solution represent correlation. Using the sparse linear regression model, a region can be represented as a linear combination of other regions and its interactions with a few other regions can be obtained. Insignificant and false interactions are forced to zero. In hyper-network construction, the process of obtaining sparse solution is solved by the least absolute shrinkage and selection operator (lasso) method (Jie et al., 2016). However, the limitation of using the lasso method to solve the sparse linear regression model is that when constructing the hyper-edges of a designated brain region, if the pairwise correlations among other brain regions are very high, then the lasso tends to select only one region from the group with a grouping effect (Zou and Trevor, 2005). As this may mean that some related areas cannot be selected, the method lacks the ability to explain the grouping information.

To solve the problem of the grouping effect among brain regions, we propose two alternative methods to improve the construction of a hyper-network: (1) the elastic net (De Mol et al., 2008; Furqan and Siyal, 2016; Teipel et al., 2017) and (2) the group lasso method (Friedman et al., 2010a; Yu et al., 2015; Souly and Shah, 2016). Then we extracted features using the different clustering coefficients defined by hyper-network to depict the functional brain network topology and performed non-parametric test to select those features with significant difference. Finally, we applied a multi-kernel support vector machine (SVM) technique on the selected features for classification. Besides, we analyzed network topology based on three methods using hyper-edges and average clustering coefficients. Furthermore, the comparative analysis of depressed patients was performed via using the classification features with significant differences between groups. The classification results showed that the elastic net method achieved better classification results than the lasso method. In addition, we further analyzed the influence of the model parameters and the classifier parameters.

## MATERIALS AND METHODS

The hyper-network method of brain network classification involves data acquisition and preprocessing, construction of the hyper-network, feature extraction, feature selection and classification. In this section, we describe each of these steps in detail. **Figure 1** illustrates the whole process.



## Data Acquisition and Preprocessing

Before starting the study, written agreement was obtained from each participant in accordance with the recommendations of the Shanxi Medical Ethics Committee (reference number: 2012013). All of the participants signed written informed consent according to the Helsinki Declaration. Seventy subjects were recruited: 38 first-episode, drug-naïve patients with MDD (15 male; mean age  $28.4 \pm 9.68$  years, range 17–49) and 28 healthy right-handed volunteers (13 male; mean age  $26.6 \pm 9.4$  years, range 17–51). Data from four participants were discarded due to problems with the data. Resting-state fMRI was performed on all subjects using a 3T magnetic resonance scanner (Siemens Trio 3-Tesla scanner, Siemens, Erlangen, Germany). The detailed demographics and Clinical Characteristics of the Subjects were illustrated by **Table 1**.

Data acquisition was completed at the First Hospital of Shanxi Medical University. All scans were performed by radiologists familiar with magnetic resonance imaging. During the scan, subjects were asked to stay awake, and to relax and close their eyes while not thinking about anything in particular. Each scan consisted of 248 contiguous EPI functional volumes with the scan parameters set as follows: axial slices = 33, repetition time = 2,000 ms, echo time = 30 ms, thickness/skip = 4/0 mm, field of view =  $192 \times 192$  mm, matrix =  $64 \times 64$  mm, flip angle =  $90^\circ$ . Due to the instability of the initial magnetic resonance signal and the adaptability of the subjects to the environment, the time series of the first 10 functional volumes were discarded. The detailed scanning parameters are provided in the Supplemental Text S1.

**TABLE 1 |** Demographics and Clinical Characteristics of the Subjects.

	CON	MDD	P-value
Age(years)	17–51 ( $26.6 \pm 9.4$ )	17–49 ( $28.4 \pm 9.68$ )	0.44 <sup>a</sup>
Sex (male/female)	13/15	15/23	0.57 <sup>b</sup>
Handedness (R/L)	28/0	38/0	–
HAMD	N/A	15–42 ( $22.8 \pm 13.3$ )	–

Data are presented as the range of minimum to maximum (mean  $\pm$  SD). HAMD, Hamilton Depression Rating Scale; MDD, major depressive disorder; NA, nonapplicable; CON, normal controls.

<sup>a</sup>The P-values were obtained by two-sample two-tailed t-test.

<sup>b</sup>The P-value were obtained by two-tailed Pearsons  $\chi^2$ -test.

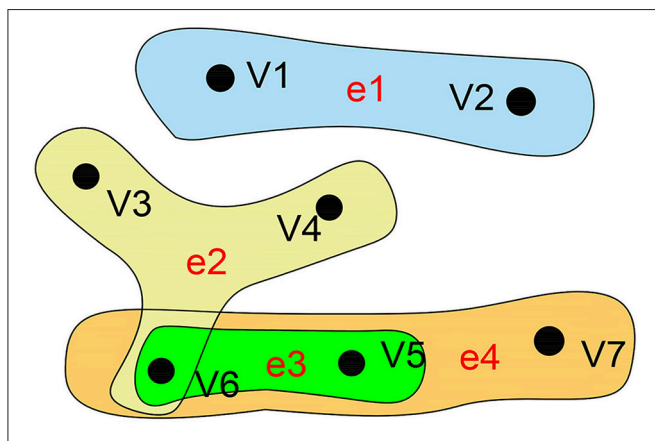
Functional data preprocessing was performed using the statistical parametric mapping (SPM8) software package (<http://www.fil.ion.ucl.ac.uk/spm>). First, the datasets were corrected for slice timing and head motion. Two samples from depressed patients and two from controls that exhibited more than 3.0 mm of translation and 3 degrees of rotation were discarded, and thus were not included in the 66 samples for data analysis. The corrected images were optimized using 12-dimensional affine transformation and spatially normalized to  $3 \times 3 \times 3$  mm voxels in the Montreal Neurological Institute 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.

## Hyper-Graph Theory

As a branch of mathematics, graph theory has been widely applied to analyze the functional interaction between brain regions, mainly by unambiguously discretizing the brain into distinct nodes and their interconnecting edges (Fornito et al., 2013). A traditional graph only characterizes two correlated nodes but ignores the high-order information, which can be represented by a hyper-graph. The biggest difference between a hyper-graph and a traditional graph is that one hyper-edge of the hyper-graph can connect to more than two nodes. Compared with the traditional graph, the hyper-graph pays more attention to the relationships than to the nodes. In short, the hyper-edge contains an unfixed number of points, which means it is a kind of intuitive mathematical expression reflecting multivariate relational data. Hyper-graphs are widely applied in many fields of computer science (Mäkinen, 1990), especially the Internet of Things, social networks, large-scale integrated design, relational databases, biomedicine, and many other applications where there are complex associations among large numbers of non-independent data points. Researchers are increasingly finding that multivariate relationships can more naturally express the internal relations and patterns hiding in information. In previous similar studies, hyper-graphs have been successfully applied to image classification (Yu et al., 2012), protein function prediction (Gallagher and Goldberg, 2013), and pattern recognition (Ren et al., 2011). **Figure 2** illustrates an example of a hyper-graph.

The mathematical expression of a hyper-graph can be represented by  $H = (V, E)$  (Kaufmann et al., 2009), where  $V$  represents a set of nodes and  $E$  represents a set of hyper-edges, and hyper-edge  $e \in E$  is a subset of  $V$ . If any two hyper-edges of a hyper-graph are not contained in each other, then the hyper-graph is called an irreducible or simple hyper-graph (Berge, 1989). We can use a  $|V| \times |E|$  incidence matrix to represent  $H$ :

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



**FIGURE 2 |** Hyper-graph. A hyper-graph in which each hyper-edge can connect more than two nodes. Here, the hyper-graph contains 7 nodes and 4 edges.  $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ ,  $E = \{e_1, e_2, e_3, e_4\}$ ,  $e_1 = \{v_1, v_2\}$ ,  $e_2 = \{v_3, v_4, v_6\}$ ,  $e_3 = \{v_5, v_6\}$ ,  $e_4 = \{v_5, v_6, v_7\}$ .

$H(v, e)$  represents the corresponding element in the incidence matrix,  $v \in V$  represents the node, and  $e \in E$  represents the hyper-edge. Nodes are the column elements of the incidence matrix and hyper-edges are the row elements. If node  $v$  belongs to the hyper-edge  $e$ ,  $H(v, e) = 1$ , and if it does not,  $H(v, e) = 0$ .

Based on  $H$ , the node degree of each vertex  $v$  is represented as

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

The edge degree of hyper-edge  $e$  is represented as

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

$D_v$  and  $D_e$  denote the diagonal matrices of node degrees  $d(v)$  and hyper-edge degrees  $\delta(e)$ , respectively. The adjacency matrix  $A$  of the hyper-graph is defined as

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

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

## Hyper-Network Construction

According to the anatomical automatic labeling (Tzourio-Mazoyer et al., 2002) template, the brain can be parcellated into 90 anatomical regions of interest (ROIs; 45 in each hemisphere), with each ROI representing a node of the functional brain network. Each regional mean time series was regressed against the average cerebral spinal fluid and white matter signals as well as the six parameters from motion correction. According to the sparse linear regression method, the residuals were used to construct the hyper-network (Jie et al., 2016). Using the sparse linear regression model, a region can be represented as a linear combination of other regions, and its interactions with a few other regions can be obtained. Insignificant and false interactions are forced to zero.

The sparse linear regression model is represented as follows:

$$x_m = A_m \alpha_m + \tau_m \quad (5)$$

$x_m$  denotes the average time series of the designated  $m$ -th ROI.  $A_m = [x_1, \dots, x_{m-1}, 0, x_{m+1}, \dots, x_M]$  denotes the data matrix that includes the mean time series of the ROIs, except for the  $m$ -th ROI, which is set to 0.  $\alpha_m$  denotes the weight vector of the degree of influence on other ROIs to the  $m$ -th ROI, and  $\tau_m$  denotes a noise term. The corresponding ROIs of the nonzero element in  $\alpha_m$  are the ROIs that interact with the designated brain region. The zero element indicates that the corresponding ROI is meaningless for accurately estimating the time series of the  $m$ -th ROI.

## Solving the Sparse Linear Regression Model Based on the Lasso Method

In the literature, the brain functional hyper-network is constructed using a sparse linear regression model, which

is commonly solved by the lasso method. The optimization objective function is

$$\min_{\alpha_m} \|x_m - A_m \alpha_m\|_2 + \lambda \|\alpha_m\|_1 \quad (6)$$

This solves the  $l_1$  norm, and  $x_m$ ,  $A_m$ , and  $\alpha_m$  have the same meaning as in Equation (5).  $\|\cdot\|_2$  denotes the  $l_2$  norm, and  $\|\cdot\|_1$  denotes the  $l_1$  norm.  $\lambda$  is the regularization parameter that controls the sparsity of the model, and different  $\lambda$  values correspond to different sparse solutions. The larger the value of  $\lambda$ , the sparser the model; that is, there are more zeros in the  $\alpha_m$ . The smaller the value of  $\lambda$ , the more dense the model; that is, there are more non-zeros in the  $\alpha_m$ . Therefore, the value of  $\lambda$  had a range. However, different experimental data will have different  $\lambda$  ranges and 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., 2011). Thus, in current studies, the value of  $\lambda$  was set from 0.1 to 0.9 in increment of 0.1. The SLEP package (Liu et al., 2011) was adopted to solve the optimization problem. For each subject, a hyper-network was constructed by using sparse representation based on lasso method. A node was a brain region and a hyper-edge  $e_m$  included a centroid ROI (i.e.,  $m$ -th ROI) and a few of other brain regions with the corresponding non-zero elements in the weight vector  $\alpha_m$  computed in Formula (6). As mentioned above, a centroid ROI should generate a corresponding weight vector  $\alpha_m$ , which generated a corresponding hyper-edge. But to reflect the multi-level interactions among brain regions, for each ROI, a group of hyper-edges was generated by varying the value of  $\lambda$  in a range from 0.1 to 0.9. Thereby, for each subject, a  $90 \times 810$  matrix was generated as a subject included 90 ROIs, which represented a hyper-network is constructed by lasso method.

### Solving the Sparse Linear Regression Model Based on the Elastic Net Method

Although the lasso method has been successfully applied in many studies, it also has limitations. The lasso is not robust when the variables are high correlations and will select one randomly and neglect the others (Friedman et al., 2010b). Obviously, the lasso method cannot solve the grouping effect, and could cause some relevant brain areas to be missed in the process of selecting related areas within a designated region. If there is a group of brain regions among which the pairwise correlations are very high, then the lasso tends to select only one region from the group and does not care which one is selected (Zou and Trevor, 2005). The ideal hyper-edge construction method should be able to select the interacting brain regions as accurately as possible. To solve the problem of the grouping effect among brain regions, we propose two methods to improve the construction of a hyper-network: the elastic net method and the group lasso method.

The elastic net is an extension of the lasso that is robust to extreme correlations among the predictors (Friedman et al., 2010b). Similar to the lasso, the elastic net can also solve the problem of sparse representation. The difference is that the elastic net can overcome the limitation of the lasso to select related variables in a group when solving the linear regression model to construct the hyper-edges, thus addressing the grouping effect. The elastic net uses the mixed penalty term of  $l_1$  norm (lasso)

and  $l_2$  norm (ridge regression), which can be represented as the following regularized objective function optimization problem:

$$\min_{\alpha_m} \|x_m - A_m \alpha_m\|_2 + \lambda_1 \|\alpha_m\|_1 + \lambda_2 \|\alpha_m\|_2^2 \quad (7)$$

$x_m$ ,  $A_m$ , and  $\alpha_m$  have the same meaning as in Equation (5).  $\lambda_1$  is the  $l_1$ -norm regularization parameter, and  $\lambda_2$  is the regularization parameter for the squared  $l_2$ -norm. Despite elastic net has one more parameter ( $l_2$ -norm) than lasso, it greatly effects the calculation result and works well in solving the grouping effect (Furqan and Siyal, 2016). Because  $l_2$ -norm (Hoerl and Kennard, 2000) performs well with many variables that are highly correlated and can effectively adjust the high correlation between independent variables so that the model can automatically choose related features in a group with grouping effect (Friedman et al., 2010b). Therefore, with  $l_1$  being for automatic variable selection and  $l_2$  encouraging grouped selection (Ogutu et al., 2012), the integration of  $l_1$  and  $l_2$  should greatly improve the construction of hyper-network. Similar to the lasso, we constructed a hyper-network for each subject, in which ROIs were the nodes and the hyper-edge  $e_m$  included a centroid ROI (i.e.,  $m$ -th ROI) and the corresponding ROIs of the nonzero elements in  $\alpha_m$  computed in formula (7). Like the lasso method, a centroid ROI corresponded to a hyper-edge. For each ROI, with the value of  $\lambda_2$  fixed, a group of hyper-edges was generated by varying the value of  $\lambda_1$  in a range from 0.1 to 0.9 in increments of 0.1. In the experiment, the value of  $\lambda_2$  was chosen as 0.2 because this was found to provide the highest classification accuracy (see the Methodology section for a detailed description of the analysis).

### Solving the Sparse Linear Regression Model Based on the Group Lasso Method

The clustering method was used to group the strongly correlated brain regions, then the group lasso method was adopted to construct the hyper-edges, which can also help to solve the grouping effect among brain regions. The lasso and elastic net methods are used to select single variables (Yuan and Lin, 2006), whereas the group lasso can select groups of variables based on predefined variable groups (Meier et al., 2008). We first clustered the 90 brain regions according to the average time series of ROIs when constructing the hyper-network. Here, the  $k$ -medoids (Park and Jun, 2009) algorithm was adopted. First, the pairwise similarity values among brain regions were calculated: higher values indicate greater similarity between the two samples. The clustering classified the brain areas into  $k$  groups, each of which represented a class of objects, under two conditions: (1) each group must contain at least one object and (2) each object must belong to a group. To stabilize the clustering as much as possible,  $k$ -means [Arthur and Vassilvitskii, 2007] was used to select the  $k$  initial cluster centers. A point was randomly selected as the first initial cluster center, and each subsequent center was chosen randomly from the remaining data points with a probability proportional to its distance from the point closest to the existing cluster center. We repeated the clustering 10 times to select the best clustering result. The  $k$ -value used in the experiment can affect the network structure and classification performance. The highest classification accuracy was obtained when  $k$  was equal

to 48 (see the Methodology section). Then, the group lasso was adopted to select the brain areas for the construction of the hyper-edge, with the following optimization objective function:

$$\min_{\alpha_m} \|x_m - A_m \alpha_m\|_2 + \beta \sum_{i=1}^k \|\alpha_m G_i\|_{2,1} \quad (8)$$

$\beta$  is the  $l_{2,1}$ -norm regularization parameter, which is regarded as an intermediate between the  $l_1$ -norm and  $l_2$ -norm penalty, with different values corresponding to different degrees of sparsity. The greater the value of  $\beta$ , the sparser the model, and the fewer the number of groups selected. It can perform variable selection effectively in the group level (Yuan and Lin, 2006). That is, if there is a group of brain regions in which the pairwise correlations are relatively high, the group lasso considers this group as a whole and determines whether it is important to the problem. The  $\alpha_m$  is divided into  $k$  non-overlapping groups by clustering, and  $\alpha_m G_i$  represents the  $i$ -th group. Similarly, we constructed the hyper-network with ROIs as the nodes, the hyper-edge including the  $m$ -th ROI and the corresponding ROIs of the nonzero elements in  $\alpha_m$  computed in Formula (8). A hyper-edge was generated by a centroid ROI. For each ROI, a group of hyper-edges was generated by varying the value of  $\beta$  in a range from 0.1 to 0.9 in increments of 0.1.

## Feature Extraction

The feature extraction was carried out on the three hyper-networks constructed by the three methods. The clustering coefficient is widely applied as a metric to measure the local characteristics of a network. However, the clustering coefficient in a hyper-network is not always defined in exactly the same way. Here, the feature extraction selected three different definitions of the clustering coefficient (Gallagher and Goldberg, 2013), reflecting different angles. The first type of clustering coefficient,  $HCC^1$ , computes the number of adjacent nodes that have connections not facilitated by node  $v$ . The second type,  $HCC^2$ , calculates the number of adjacent nodes that have connections facilitated by node  $v$ . The third type,  $HCC^3$ , calculates the amount of overlap amongst adjacent hyper-edges of node  $v$ . The formula is as follows:

$$HCC^1(v) = \frac{2 \sum_{u,t \in N(v)} I(u,t, \neg v)}{|N(v)| (|N(v)| - 1)} \quad (9)$$

$HCC^1(v)$  represents the first type of clustering coefficient, and  $u, t, v$  represent the nodes.  $N(v) = \{u \in V : \exists e \in E, u, v \in e\}$ , where  $V$  represents a set of nodes,  $E$  represents a set of hyper-edges,  $e$  represents a hyper-edge, and  $N(v)$  represents the set of nodes of other hyper-edges containing node  $v$ . If  $\exists e_i \in E$ , such as  $u, t \in e_i$ , but  $v \notin e_i$ , then  $I(u, t, \neg v) = 1$ ; otherwise,  $I(u, t, \neg v) = 0$ .

$$HCC^2(v) = \frac{2 \sum_{u,t \in N(v)} I'(u, t, v)}{|N(v)| (|N(v)| - 1)} \quad (10)$$

$HCC^2(v)$  represents the second type of clustering coefficient.  $u, t, v$  and  $N(v)$  have the same meaning as in Equation (9). If  $\exists e_i \in E$ ,

such as  $u, t, v \in e_i$ , then  $I'(u, t, v) = 1$ ; otherwise,  $I'(u, t, v) = 0$ .

$$HCC^3(v) = \frac{2 \sum_{e \in S(v)} (|e| - 1) - |N(v)|}{|N(v)| (|S(v)| - 1)} \quad (11)$$

$HCC^3(v)$  represents the third type of clustering coefficient.  $|e|$  represents the number of nodes in a hyper-edge.  $v$  and  $N(v)$  have the same meaning as in Equation (9).  $S(v) = \{e_i \in E : v \in e_i\}$ , where  $v$  represents a node,  $e_i$  represents a hyper-edge, and  $S(v)$  represents the set of hyper-edges containing node  $v$ .

The three clustering coefficients reflect the local clustering properties of the hyper-network from different perspectives. The definition of each clustering coefficient was extracted from the hyper-network as features. Multiple linear regression analyses were applied to evaluate the confounding effects of age, gender, and educational attainment for each network property. Because the three clustering coefficients define local properties, for simplicity, we calculated the average clustering coefficient (average  $HCC^1$ ,  $HCC^2$ , and  $HCC^3$ ) for each subject (using the average metric values for 90 brain areas), and added it to the multiple linear regression as the independent variable. The results showed that no significant correlations were found between the network properties and the potential confounding variables. (The results are presented in detail in Supplemental Text S2).

## Feature Selection and Classification

Some of the features extracted from the hyper-network may be irrelevant or redundant. To select the key features for classification, the most discriminative features were chosen according to the statistical difference analysis. Specifically, Kolmogorov and Smirnov's nonparametric permutation test (Fasano and Franceschini, 1987) was performed to compare 270 node properties between the MDD group and the control group, corrected using Benjamini and Hochberg's false-discovery rate (FDR) method ( $q = 0.05$ ) (Benjamini and Hochberg, 1995). The local properties with significant between-group differences according to the nonparametric permutation test were used as the classification features to construct the classification model.

Classifier training was performed using the libsvm classification package (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). The radial basis function (RBF) kernel was adopted for the classification, and leave-one-out cross-validation was used to evaluate the performance. Suppose there are  $N$  features, with each feature taking a turn as the testing set, and the remaining  $N-1$  features as the training set. The parameter optimizations of  $c$  and  $g$  were carried out by the K-fold cross-validation method using the training set (Hsu et al., 2003). The optimal parameters of  $c$  and  $g$  which can achieve the highest training validation classification accuracy were chosen to establish  $N$  different models. The classification test was performed and the average classification accuracy of the  $N$  models was selected. The mean and standard deviation of the classification features needed to be calculated for standardization. Because the random selection of the initial seed point in the group lasso method affects the classification result, the mean of 50 experiments was calculated as the final classification result. In addition, the Relief algorithm (Kira and Rendell, 1992) was selected to compare the importance

of the chosen features (i.e., the extent of their contribution to the classification) of the three methods in section Classification Performance.

## EXPERIMENTS AND RESULTS

### Comparison of Network Structure Based on Three Methods

We performed the following analyses to determine whether there were significant differences among the hyper-networks constructed by the three methods.

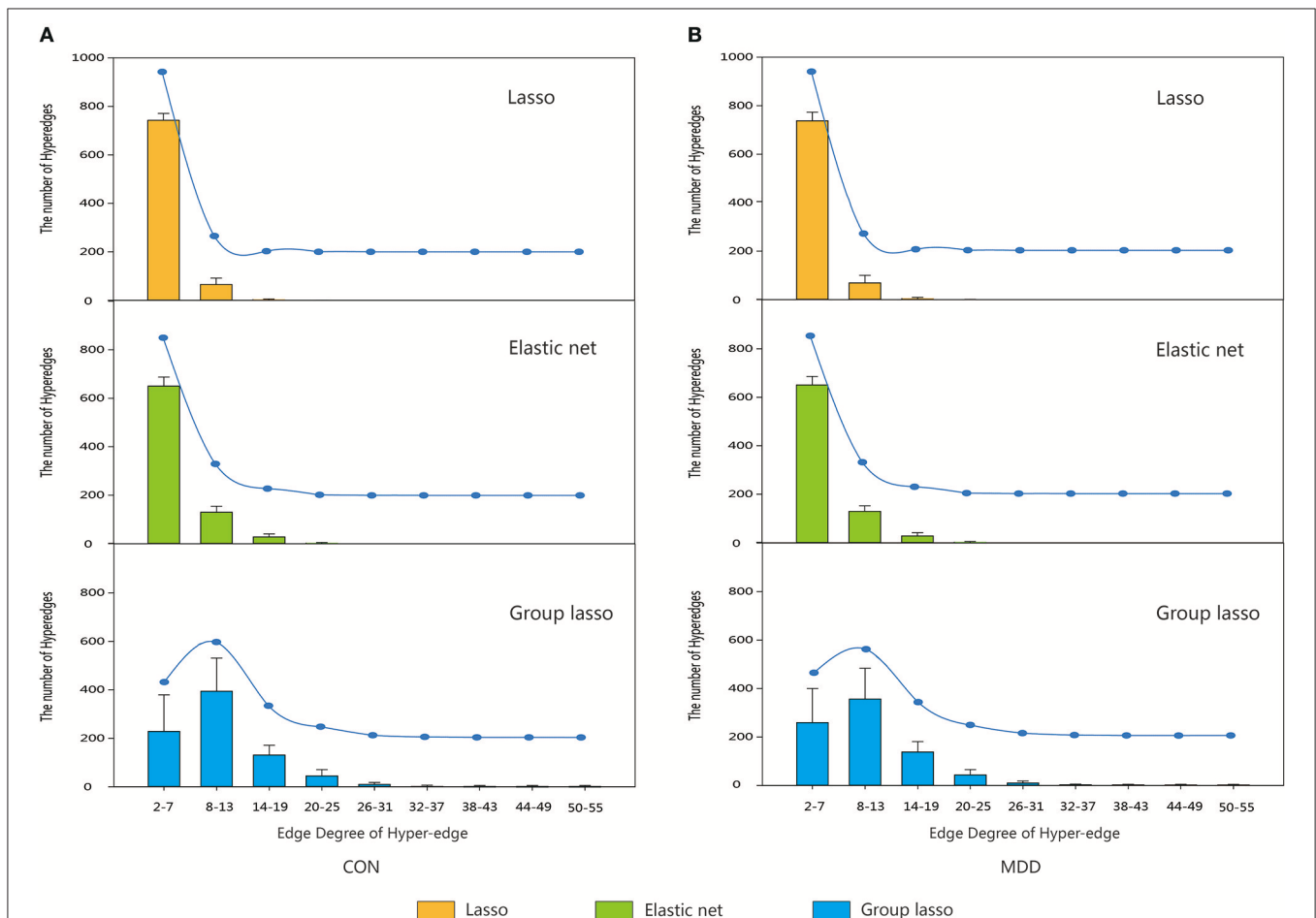
Hyper-networks were constructed for subjects in the control and MDD groups and the hyper-edges were analyzed. The edge degrees of the hyper-edges constructed by the three methods were computed and their distribution is shown in **Figure 3**.

The results indicated that for both the MDD and the control group, the edge degrees of hyper-edges constructed by lasso and elastic net methods most lied in Equations (2)–(7), and the distributions were also close. The hyper-edges constructed by

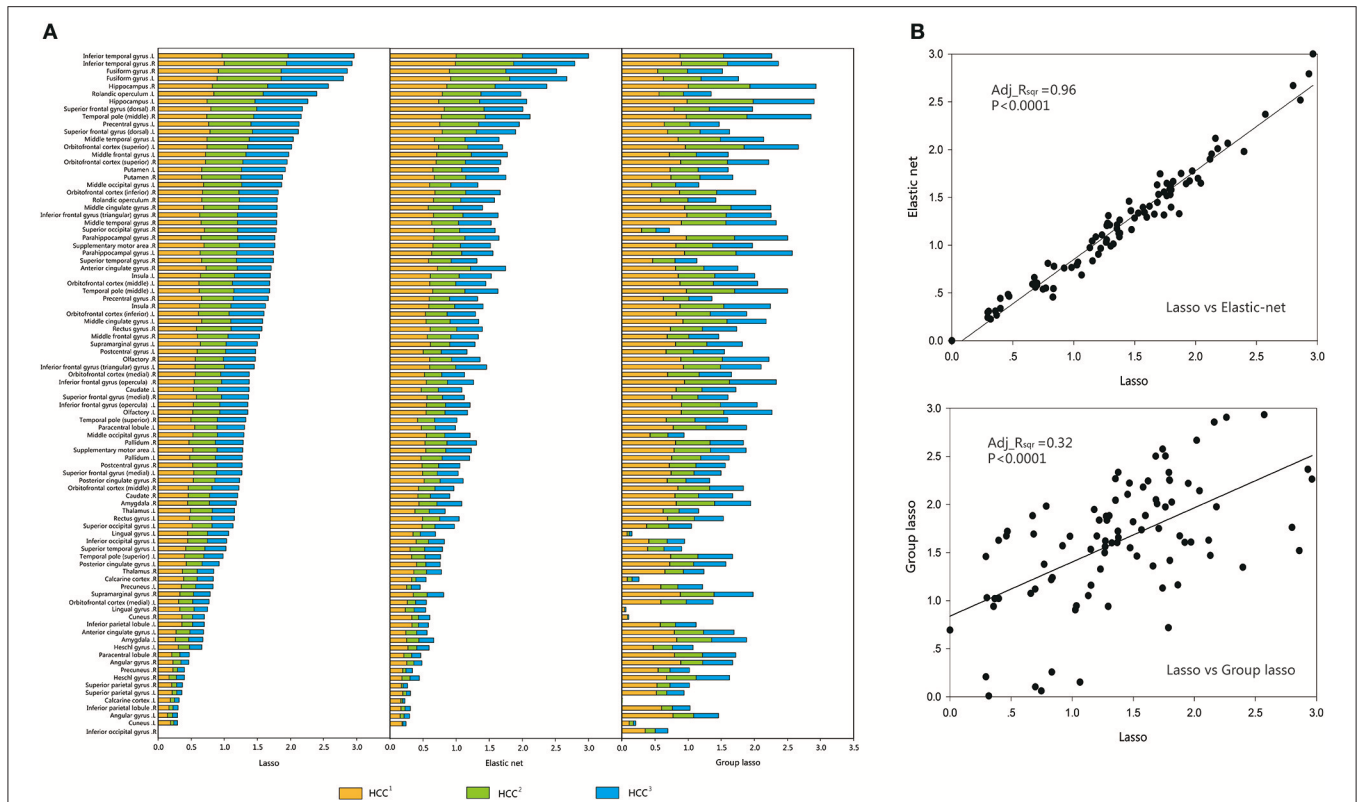
the group lasso method were different, with a broader range of edge degrees than the other two methods and a relatively discrete distribution.

For each brain region, the mean value of each group (CON or MDD) was computed for each type of clustering coefficient, and the acquired data were standardized. The brain regions were sorted according to the size of the standardized metric value for the lasso method, and according to the sequence of comparison for the other two methods. A regression analysis was performed to verify the associations between the network metrics obtained by the two proposed methods and the original method. The results showed that the lasso method was strongly related to the elastic net method and weakly related to the group lasso method (**Figures 4, 5**).

The average clustering coefficients (average  $HCC^1$ ,  $HCC^2$ , and  $HCC^3$ ) of each subject (using the average metric values for 90 brain areas) were calculated. Nonparametric permutation tests were used to compare the differences in the average  $HCC^1$ ,  $HCC^2$ , and  $HCC^3$  among the hyper-networks constructed by the three methods in the depression group and the normal control



**FIGURE 3 |** The distribution of the edge degree of hyper-edge obtained by three methods. Orange indicate the distribution of the edge degree of hyper-edge based on lasso method. Green indicate the distribution of the edge degree of hyper-edge based on elastic net method. Blue indicate the distribution of the edge degree of hyper-edge based on group lasso method. Error bars show standard deviation. **(A)** CON group. **(B)** MDD group. CON, normal control; MDD, major depressive disorder.



**FIGURE 4 |** Comparison and correlation analysis about the standardized metric values in CON Group. **(A)** Comparison that the brain regions were sorted according to the size of the standardized metric value in the lasso method, and according to the corresponding sequence about the other two methods for comparison in CON group. **(B)** Correlation analysis. Correlation analysis was performed to verify the associations between the network metrics obtained by the two proposed methods and the original method. HCC<sup>1</sup> indicate the first type of clustering coefficient; HCC<sup>2</sup> indicate the second type of clustering coefficient; HCC<sup>3</sup> indicate the third type of clustering coefficient. *P* indicate the significance of correlation analysis. Adj. R<sub>Sqr</sub>, adjusted R square.

group, respectively, and the result was corrected using the FDR method. **Figure 6** shows the average clustering coefficients of the hyper-networks for the two groups; the results suggest that the hyper-networks obtained by the three methods contained structural differences.

### Brain Regions With Significant Differences

After constructing the hyper-network and extracting the features based on the three methods, for each feature we carried out a nonparametric permutation test to evaluate the difference between the MDD and control groups for all subjects, corrected using the FDR method. **Table 2** and **Figure 7** illustrate the brain regions that showed significant between-group differences in the three hyper-network construction methods.

Comparison of the regions obtained by the elastic net and lasso methods indicated that there were more overlapping than non-overlapping regions, including partial frontal areas [left superior frontal gyrus (medial), right middle frontal gyrus, right paracentral lobule], the bilateral rolandic operculums, partial limbic lobe (right parahippocampal gyrus), partial subcortical gray nucleus (left putamen), and partial occipital area (left lingual gyrus). The hyper-networks produced by the lasso and elastic net methods were similar, so there were more overlapping than non-overlapping regions between the lasso and elastic net methods.

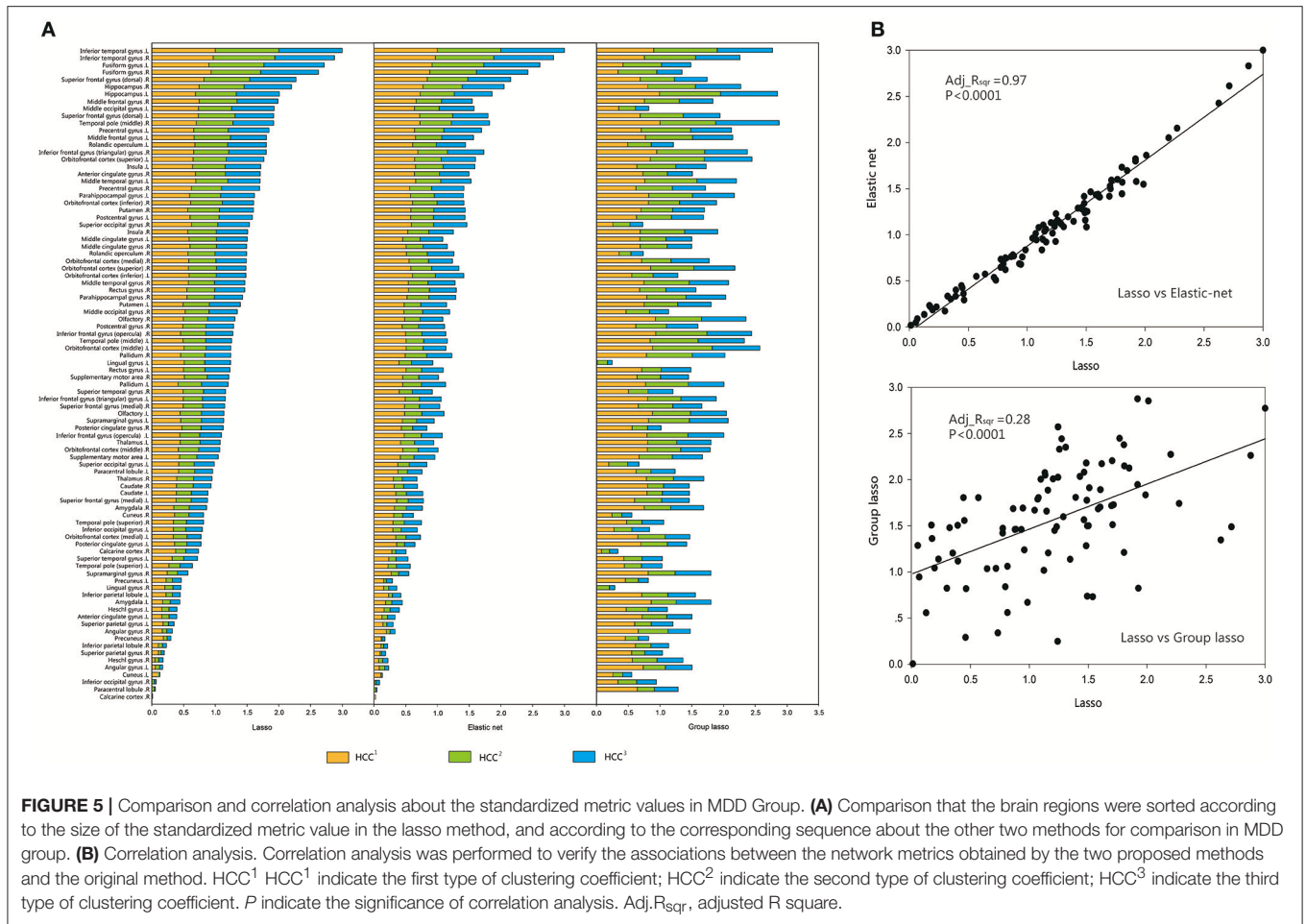
The group lasso result was much different from that of the lasso method; there were more non-overlapping than overlapping regions, including the left inferior frontal gyrus (triangular part), left inferior frontal gyrus (orbital part), left paracentral lobule, left median cingulate and paracingulate gyri, right median cingulate and paracingulate gyri, right posterior cingulate gyrus, right hippocampus, right angular gyrus, left precuneus, and right lingual gyrus. These brain regions are mainly concentrated in the frontal lobe and limbic system.

### Classification Performance

We assessed the classification performance by measuring the accuracy (the proportion of subjects correctly identified), the sensitivity (the proportion of patients correctly identified), the specificity (the proportion of controls correctly identified), and balanced accuracy. Balanced accuracy is defined as the arithmetic mean of sensitivity and specificity, and can avoid inflated performance on imbalanced datasets (Velez et al., 2007).

We evaluated the classification performance of the three brain network classification methods based on the lasso, elastic net, and group lasso, and compared them with the traditional network classification method (denoted as TCN). The traditional network, based on Pearson’s correlations, was constructed with sparsity set from 5 to 40%. Three local properties were calculated: degree,





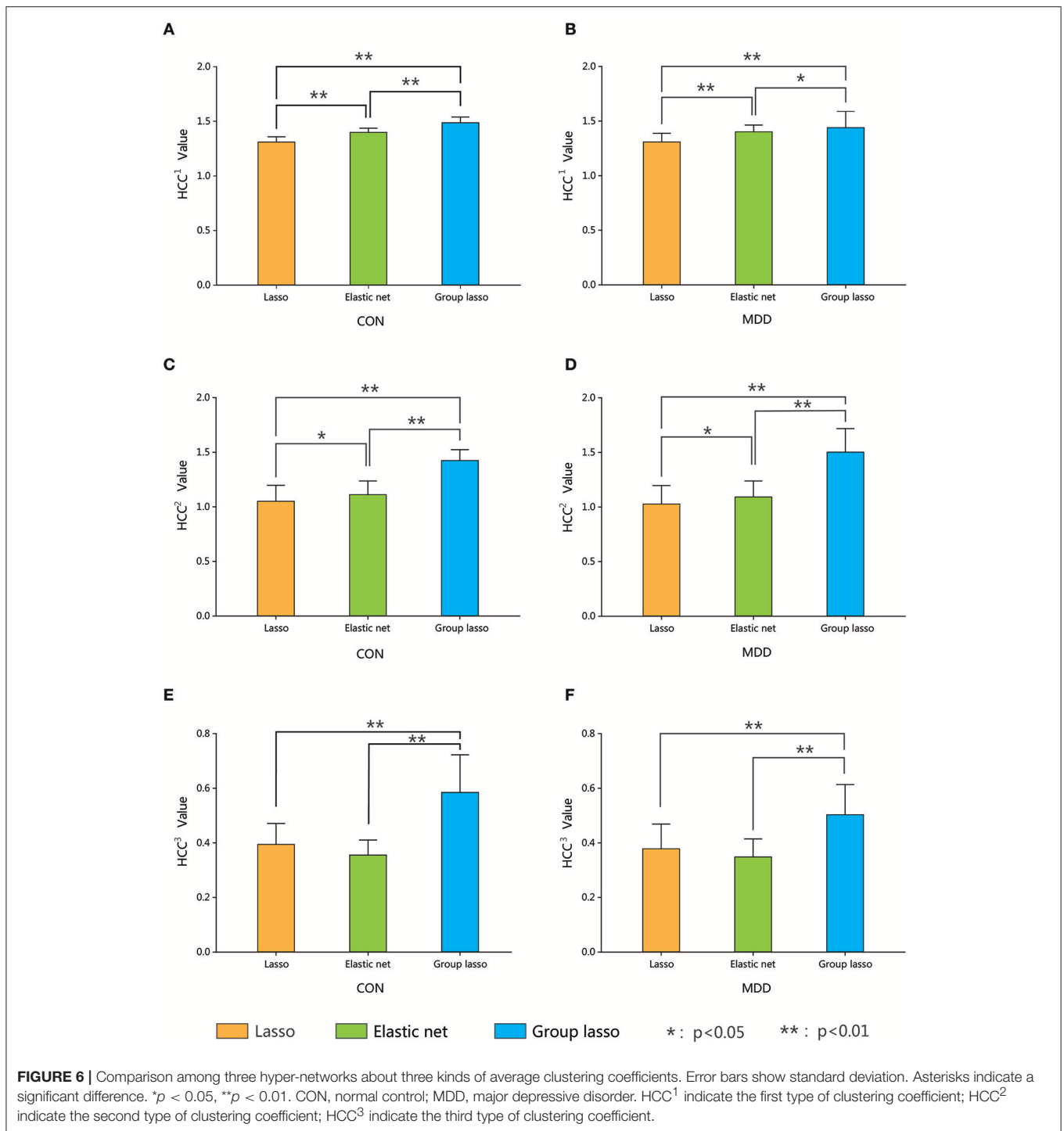
centrality degree, and node efficiency. To characterize the overall characteristics of the metrics in the selected threshold space, the AUC value of each property was calculated. We selected the AUC values of the local properties that showed a significant between-group difference in the KS nonparametric permutation test as the classification features. The classification results of the comparison methods are summarized in **Table 3**.

To compare the importance of the chosen features (i.e., the extent of their contribution to the classification) of the three methods, the brain areas with significant between-group differences based on the lasso method were compared with the results obtained by the other two methods. We combined the brain areas with no overlap between the two methods (the lasso and the elastic net, and the lasso and the group lasso). The Relief algorithm (Kira and Rendell, 1992) was adopted to calculate the weights of the corresponding features of these brain areas. The results in **Figure 8** indicate that the classification weight of the elastic net method was higher than that of the lasso method and that of the group lasso was lower.

For all four methods, the best classification accuracy was obtained using the hyper-network construction method based on elastic net, so we analyzed the differences in the connection patterns between the MDD and control groups with the elastic net method. The average hyper-edges of the MDD and control groups were calculated based on the brain areas with significant

between-group differences in the elastic net method. First, for each group of subjects, we constructed the hyper-edges according to the fixed value of  $\lambda_1$  and Equation (7), computing the number of occurrences of each brain area in each group of hyper-edges. Then, the average of edge degrees of the hyper-edges for all subjects was calculated, denoted as  $d$ . If it was not an integer, then  $d$  was rounded to the nearest integer (greater than or equal to). Finally, the number of occurrences of the brain area in each group of hyper-edges was sorted from high to low, and the first  $d$  brain areas were chosen to construct the corresponding average hyper-edges. **Figure 9** illustrates the average hyper-edges based on the 12 ROIs listed in **Table 2** (elastic net) with  $\lambda_1 = 0.3$ . The left side of each subgraph in **Figure 9** shows the average hyper-edges of the control group and the right side shows those of the depression group. The green nodes in each subgraph represent the brain areas with significant between-group differences.

**Figure 9** shows that the hyper-edges of the MDD and control groups were discrepant. For example, **Figure 9D** shows that in the MDD group, the right Rolandic operculum (ROL.R) mainly interacted with the right paracentral lobule (PCL.R), left cuneus (CUN.L), left heschl gyrus (HES.L), right heschl gyrus (HES.R), left calcarine fissure and surrounding cortex (CAL.L), and the left superior temporal gyrus (STG.L); however, in the control group, the right Rolandic operculum (ROL.R) mainly interacted with the right postcentral gyrus (POCG.R), right supramarginal gyrus



(SMG.R), left superior temporal gyrus (STG.L), left heschl gyrus (HES.L), and right heschl gyrus (HES.R).

## DISCUSSION

The accuracy of brain network classification based on a hyper-network is strongly dependent on the network construction.

A new method of hyper-network construction was recently proposed; however, the method was unable to obtain some relevant brain regions when establishing the hyper-edges because of the grouping effect among regions. To solve this problem, we proposed two new methods for constructing a hyper-network based on the elastic net and the group lasso. In the lasso method proposed by Jie et al. (2016), the optimal objective function for solving the sparse linear regression model

**TABLE 2 |** Abnormal brain regions and significance.

ROIs	P-value		
	Clustering coefficient	Clustering coefficient	Clustering coefficient
<b>LASSO</b>			
Left supramarginal gyrus	<b>0.048</b>	0.214	0.118
Left rolandic operculum	0.118	0.118	<b>0.007</b>
Right rolandic operculum	0.303	0.094	<b>0.045</b>
Left superior frontal gyrus, medial	0.207	0.055	<b>0.007</b>
Right parahippocampal gyrus	0.638	<b>0.015</b>	<b>0.005</b>
Left thalamus	0.294	<b>0.049</b>	0.252
Left putamen	0.214	0.122	<b>0.047</b>
Right middle frontal gyrus	<b>0.019</b>	0.157	0.169
Left lingual gyrus	<b>0.017</b>	0.260	0.109
Right inferior occipital gyrus	0.060	<b>0.039</b>	<b>0.045</b>
Right fusiform gyrus	0.792	<b>0.047</b>	0.612
Right Paracentral lobule	0.393	<b>0.049</b>	0.090
Left middle temporal gyrus	0.804	<b>0.037</b>	0.181
<b>ELASTIC NET</b>			
Right middle frontal gyrus	<b>0.009</b>	<b>0.066</b>	0.573
Left inferior frontal gyrus, orbital part	<b>0.043</b>	<b>0.043</b>	0.303
Left rolandic operculum	<b>0.040</b>	0.087	<b>0.038</b>
Right rolandic operculum	<b>0.045</b>	<b>0.045</b>	0.053
Right supplementary motor area	<b>0.024</b>	0.091	<b>0.049</b>
Left superior frontal gyrus, medial	0.332	0.286	<b>0.047</b>
Left median cingulate and paracingulate gyri	0.850	<b>0.040</b>	<b>0.021</b>
Right parahippocampal gyrus	0.313	<b>0.019</b>	0.080
Left lingual gyrus	0.090	0.098	<b>0.008</b>
Right Paracentral lobule	0.612	0.229	<b>0.017</b>
Left putamen	<b>0.047</b>	<b>0.011</b>	<b>0.041</b>
Left inferior temporal gyrus	0.665	<b>0.008</b>	0.994
<b>GROUP LASSO</b>			
Left inferior frontal gyrus, triangular part	<b>0.007</b>	0.968	0.063
Left inferior frontal gyrus, orbital part	<b>0.017</b>	0.817	<b>0.007</b>
Right rolandic operculum	0.265	0.991	<b>0.003</b>
Left median cingulate and paracingulate gyri	<b>0.038</b>	0.461	<b>0.005</b>
Right median cingulate and paracingulate gyri	<b>0.012</b>	0.201	<b>0.001</b>
Right posterior cingulate gyrus	0.303	0.341	<b>0.001</b>
Right hippocampus	<b>0.001</b>	0.058	<b>0.017</b>
Right parahippocampal gyrus	<b>0.006</b>	0.586	<b>0.016</b>
Right lingual gyrus	0.351	<b>0.006</b>	0.665
Right angular gyrus	<b>0.004</b>	<b>0.045</b>	0.080
Left precuneus	0.252	0.322	<b>0.005</b>
left paracentral lobule	0.094	0.147	<b>0.009</b>
Right paracentral lobule	0.252	0.586	<b>0.002</b>
Left thalamus	0.087	<b>0.002</b>	0.404

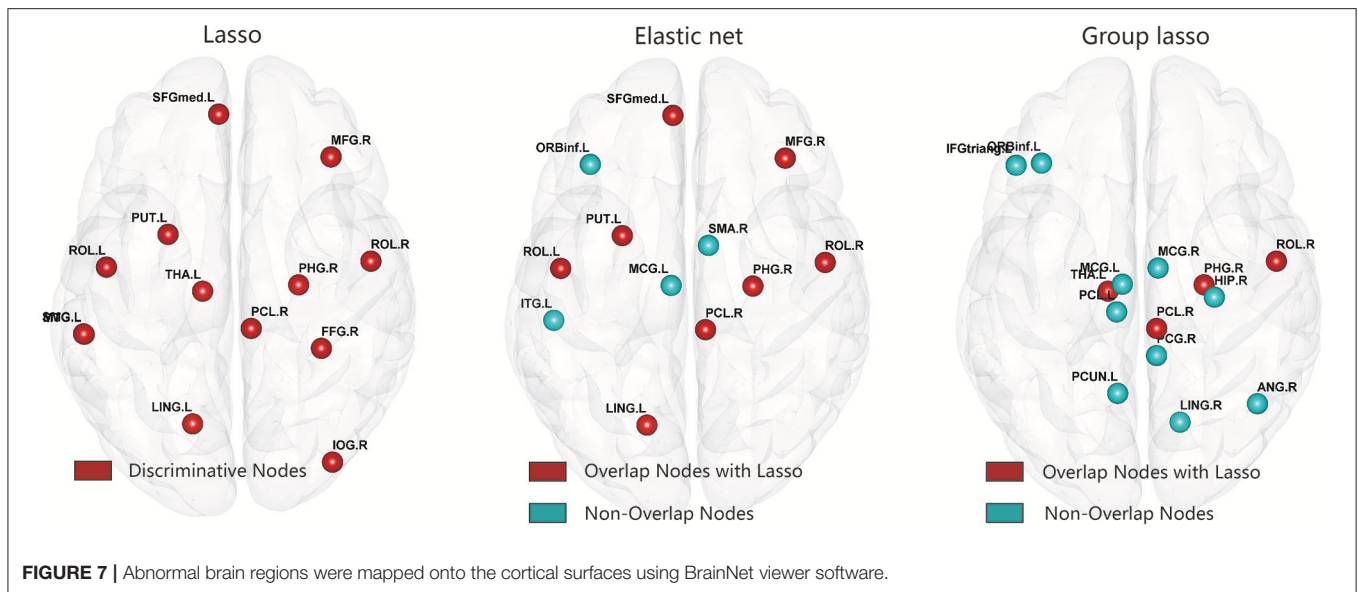
Bold indicates significance  $p < 0.05$ .

includes the loss function and the  $l_1$  norm penalty term. The penalty term makes it possible to continuously compress and select variables automatically and simultaneously. The optimal objective function of elastic net method adds an  $l_2$  norm penalty term to the lasso. Recent studies (Zou and Trevor, 2005; De Mol et al., 2008) have shown that the  $l_2$  norm can effectively adjust the high correlations among the independent variables so that the model can automatically select the relevant variables in a group with a grouping effect. The group lasso method introduces the penalty function for the variable grouping factor to select variables using the  $l_{2,1}$  norm, based on predefined variable groups.

There were differences among the hyper-networks constructed by the three methods. The MDD and control groups had similar distributions according to the analysis of the hyper-edge degrees. The lasso and elastic net methods produced very similar results for the range of edge degrees and the number of hyper-edges within each range. However, the results for the group lasso method were quite different. The range of edge degrees was 2–55, indicating the looseness of the constraint, and there were some hyper-edges with more nodes. We conclude that the structure of the hyper-network obtained by the lasso was similar to that obtained by the elastic net, but very different from that obtained by the group lasso. This conclusion was confirmed by the comparison of metrics. When constructing the hyper-edge using the lasso method, only one brain region was selected from a group because of the grouping effect. The elastic net method helped to select the related brain regions by adding the  $l_2$  norm, which can select some brain regions from that group. When the group lasso method selected a brain region from the group, all of the brain regions in the group were considered relevant. Therefore, the lasso method was the strictest, the group lasso method was the loosest, and the elastic net method was moderate.

We obtained a similar conclusion from the correlation analysis of the metrics. The metrics of all subjects in each group were averaged across brain regions, which were sorted after standardization. Linear regression analyses were performed on the metrics obtained by the lasso and the other two methods. The standardized metric values based on brain regions showed significant associations between the elastic net and the lasso methods (CON group:  $\text{adj\_R}_{\text{sqr}} = 0.96$ , MDD group:  $\text{adj\_R}_{\text{sqr}} = 0.97$ ), but significant differences between the group lasso and the lasso methods (CON group:  $\text{Adj\_R}_{\text{sqr}} = 0.32$ , MDD group:  $\text{Adj\_R}_{\text{sqr}} = 0.28$ ).

Furthermore, analysis of the average clustering coefficients (average  $HCC^1$ ,  $HCC^2$ , and  $HCC^3$ ) showed significant differences in the average  $HCC^1$  and average  $HCC^2$  among the three hyper-networks for both the MDD and control groups. The average  $HCC^3$  showed significant differences between the group lasso and the other two methods. No significant difference ( $p > 0.05$  FDR corrected,  $q = 0.05$ ) was found between the lasso and elastic net methods. These results indicated that there were structural differences among the hyper-networks constructed by the three methods. The lasso and elastic net produced similar hyper-network structures, but the structure produced by the lasso was very different from that produced by the group lasso. In terms



**FIGURE 7 |** Abnormal brain regions were mapped onto the cortical surfaces using BrainNet viewer software.

**TABLE 3 |** The classification performance of the four methods.

Methods	Accuracy	Sensitivity (%)	Specificity (%)	BAC (%)
TCN	71	79	64	71.5
Lasso	83.33	84.21	82.14	83.175
Group lasso	80.30	84.21	75	79.60
Elastic net	86.36	92.10	81.57	86.83

TCN, traditional connectivity network method; lasso, hyper-network method based on lasso; elastic net, hyper-network method based on elastic net; group lasso, hyper-network method based on group lasso; BAC, balanced accuracy.

of network construction constraints, the lasso was the strictest, the group lasso was the loosest, and the elastic net method was in-between. We believe that these results are attributable to differences in the different methods' ability to resolve the grouping effect.

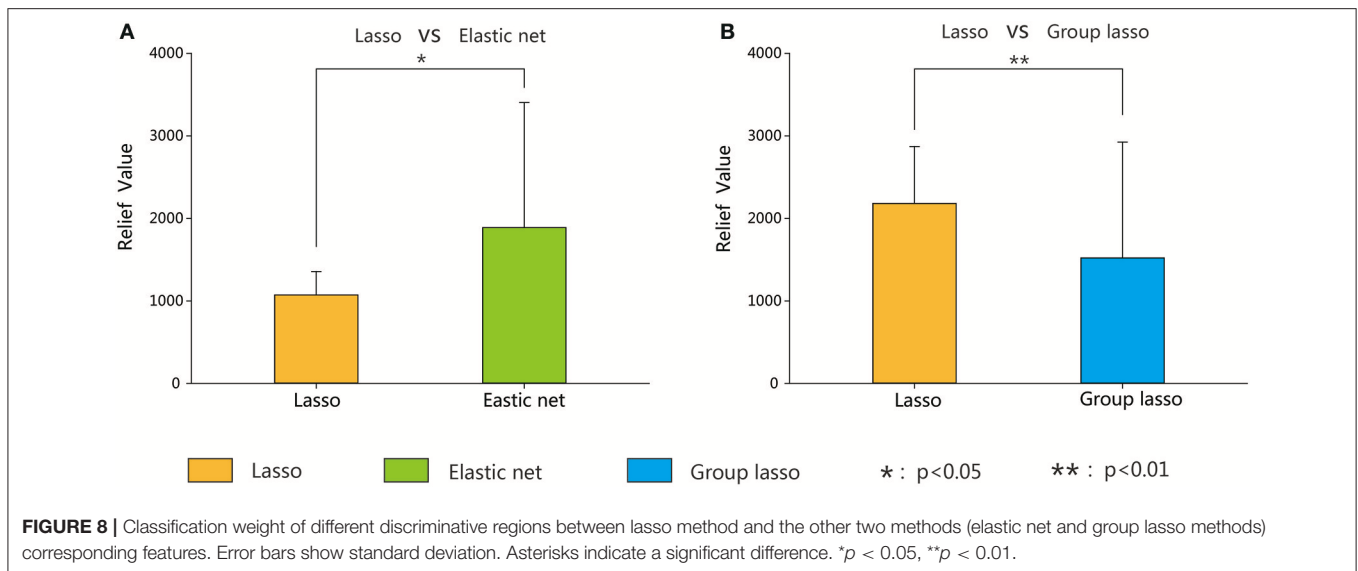
The areas with statistically significant between-group differences were not identical among the three methods. The lasso and elastic net methods had more overlapping than non-overlapping regions, while the group lasso and lasso methods showed the opposite result. This result also verified the conclusion we obtained from the other analyses.

The best classification accuracy was obtained based on the elastic net method, so we used the hyper-network based on this method to analyze the abnormal brain regions. Statistical analysis of this hyper-network identified 12 abnormal regions: some partial frontal areas (right middle frontal gyrus, left inferior frontal gyrus (orbital part), right supplementary motor area, left superior frontal gyrus (medial), right paracentral lobule), bilateral rolandic operculums, partial limbic lobes (left median cingulate and paracingulate gyri, right parahippocampal gyrus), partial subcortical gray nucleus (left putamen), partial temporal lobe (left inferior temporal gyrus), and partial occipital area (left lingual gyrus). These regions are consistent with previous results reported in the literature (Table 4). The value

of brain network research is in identifying changes in the brain network. Analysis of the connection patterns among the brain regions with between-group differences identified by the elastic net method and the other regions revealed different interaction patterns between the MDD and control groups.

Three methods of hyper-network construction and a correlation method were used to classify 38 patients with depression and 28 control subjects. The results suggest that hyper-network methods can improve brain network classification performance. The classification accuracy of our two proposed methods exceeded 80%. The elastic net hyper-network construction method outperformed the others, and the accuracy reached 86.36% when the value of  $\lambda_2$  was set to 0.2. The classification result obtained by the group lasso method was not as good as the original method. The underlying reasons are that the k-medoids clustering method introduces uncertain parameters, and it easily falls into a local optimum. Although we used the idea of k-means++ to optimize the selection of the initial points, the random selection of the first initial cluster center still led to unstable results. Moreover, the group lasso cannot choose variables flexibly within a group.

Based on the lasso and elastic net methods, we obtained two groups of brain regions with between-group differences, and then calculated the classification weight of the corresponding features of the non-overlapping brain regions between two group regions. The Relief algorithm is used to weight features according to the correlation between a feature and a category. Features with stronger weights have better classification ability (Kira and Rendell, 1992). The weights of the features obtained by the elastic net method were significantly larger than those obtained by the lasso method, which in turn were significantly greater than those obtained by the group lasso method. The results imply that a moderate connection constraint (elastic net) can acquire classification features more effectively than a constraint that is too strict (lasso) or too loose (group lasso).



## METHODOLOGY

The construction of hyper-network is based on sparse regression model with penalties. By using linear sparse regression model, a brain region can be characterized by a linear combination of a few other brain regions, which may be obtained with pairwise correlation from the view of math method. However, by introducing the penalties control, one brain region can be interacted with a few other brain regions while forcing insignificant interactions to zero, and regard these brain regions as a hyper-edge to construct the hyper-network, which was a multivariate expression that described the multiple interactions of brain regions to interpret the hyper-network topology. For more precisely diagnosis of brain diseases, we proposed elastic net and group lasso sparse regression methods to construct the hyper-network taking into the grouping effect.

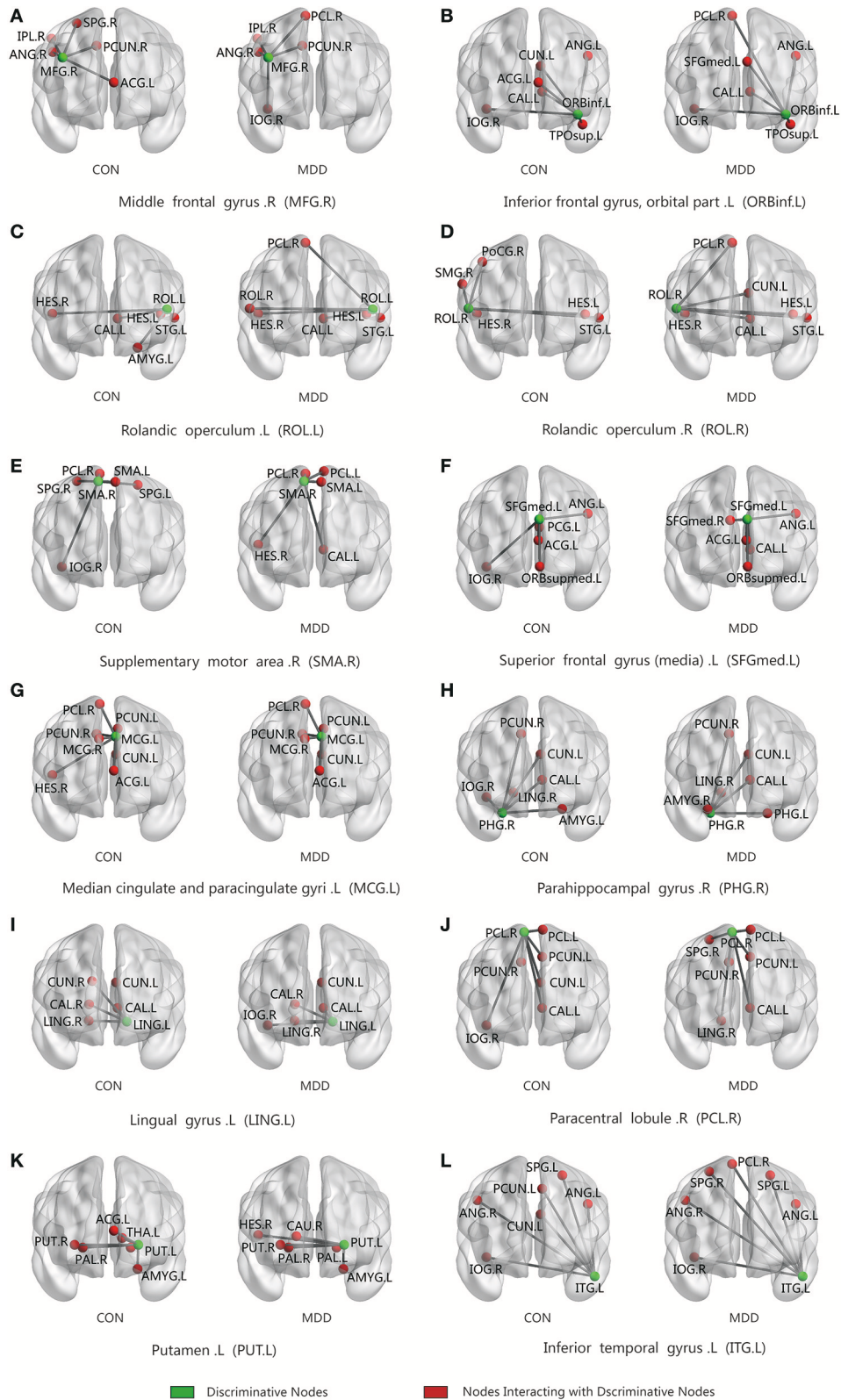
The performances of the proposed classification methods depend on the selection of certain parameters, such as the cluster number  $k$ , the hyper-network construction model parameters  $\lambda_1$  and  $\lambda_2$ , the SVM model parameters  $c$  and  $g$ . We compared brain network classification methods based on the group lasso and elastic net methods to explore this problem.

### Effect of Cluster Number $k$

Parameter  $k$  is the number of clusters in the group lasso method, and varying the value of  $k$  produces different network structures and classification results. To assess the effect on classification performance, the range of  $k$  was set at (6, 90) with a step size of 6. Because the random selection of the first initial seed point can affect the results, 50 experiments were performed for each  $k$  value, and the mean accuracy was chosen as the final classification result. **Figure 10** shows that the highest accuracy of 80.30% was obtained when  $k$  equaled 48.

### Effect of Regularization Parameters $\lambda_1$ and $\lambda_2$

It has been shown in previous studies that the parameter  $\lambda$  has a significant impact to hyper-network topology. The parameter  $\lambda$  determined the sparsity and scale of the network. When  $\lambda$  was too small, the network would be too coarse and involve much noise. When  $\lambda$  was too large, the network would be too sparse (Lv et al., 2015). Moreover, the reliability of the network topology, especially the modularity, was sensitive to the sparsity controlled by the parameter  $\lambda$  (Li and Wang, 2015). Besides, the parameter  $\lambda$  had an effect on the classification performance. The classification model parameter, especially  $\lambda$ , was sensitive to classification accuracy. Since the selection of  $\lambda$  was not specified formulation, the proper choice of the parameter  $\lambda$  is very important for constructing hyper-networks and for classification. There were also some methods to choose  $\lambda$  to optimize the network topology reliability and classification performance in previous studies (Braun et al., 2012; Li and Wang, 2015; Qiao et al., 2016). However, there are the same limitations that it had low reliable network topology. This research showed that only when  $\lambda$  is 0.01 (indicating that all the nodes on the network are on a hyper-edge), the network had a high reliability (Li and Wang, 2015). Thus, multi-level  $\lambda$  setting method was proposed (Jie et al., 2016). Different from single  $\lambda$  setting, multi-level  $\lambda$  can combine several  $\lambda$  to provide more network topology information. And the multi-level  $\lambda$  setting can avoid random selection leading to single  $\lambda$  setting and reduce the impact of low reliability caused by a single network topology. In current studies,  $\lambda_1$  is the regularization parameter of the  $l_1$  norm, which is biased to control the model sparsity. Multi-level  $\lambda_1$  setting was introduced and the interval was set 0.1.  $\lambda_2$  was the regularization parameter of the  $l_2$  norm, which encouraged the influence of grouping. As different values of  $\lambda_2$  had different grouping effects and lead different classification performance, the interval was set 0.1. Different values of  $\lambda_1$  and  $\lambda_2$  produce different solutions.



**FIGURE 9 |** Connected patterns of abnormal brain regions in hyper-network constructed by elastic net method; The average hyper-edges for CON(left) and MDD(right) groups based on 12 ROIs (the green node) listed in **Table 2** (elastic net) with  $\lambda_1 = 0.3$ . Here, each sub-figure denotes a hyper-edge constructed by the corresponding ROI, and the green nodes in each subgraph represent the brain areas with significant differences between-group. L, left; R, right; CON, normal control. MDD, major depressive disorder.

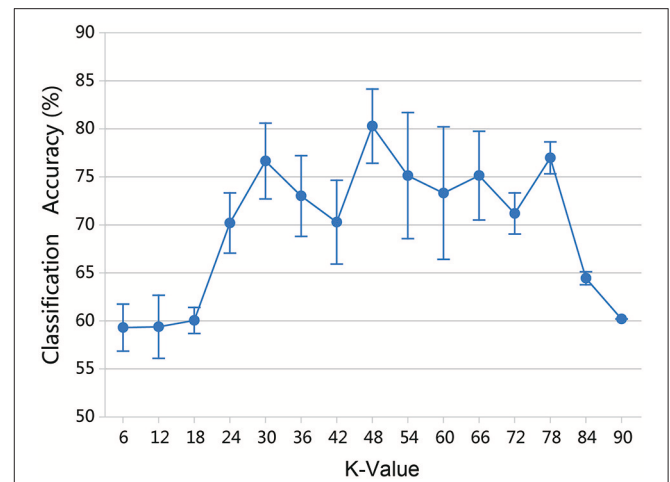
**TABLE 4** | ROIs selected from the other literature about depression.

ROI	References
Right middle frontal gyrus	Guo et al., 2012
Left inferior frontal gyrus, orbital part	Jin et al., 2011; Guo et al., 2012; Lord et al., 2012
Rolandic operculum	Zhu et al., 2016
Right supplementary motor area	Liu et al., 2012
Left superior frontal gyrus, medial	Jin et al., 2011
Left median cingulate and paracingulate gyri	Guo et al., 2012; Zhu et al., 2016
Right parahippocampal gyrus	Qiu et al., 2014
Left lingual gyrus	Lord et al., 2012; Qiu et al., 2014
Right paracentral lobule	Qiu et al., 2014
Left putamen	Lord et al., 2012; Gong et al., 2014
Left inferior temporal gyrus	Gong et al., 2014

In the multi-level  $\lambda$  setting, it is important that how to get the optimizing combination of  $\lambda$ . If Enumeration method was adopted, the computation consumption was too huge. Thus, in current studies,  $\lambda_1$  used a series of ascending order combination, which is  $\{0.1\}$ ,  $\{0.1, 0.2\}$ ,  $\{0.1, 0.2, 0.3\}$ , ...,  $\{0.1, 0.2, \dots, 0.9\}$ , while  $\lambda_2$  are adopted, which is  $0.1, 0.2, \dots, 0.9$ . For  $\lambda_2$  value, nine sets of  $\lambda_1$  were used to construct the hyper-edges,  $\{0.1\}$ ,  $\{0.1, 0.2\}$ ,  $\{0.1, 0.2, 0.3\}$ , ...,  $\{0.1, 0.2, \dots, 0.9\}$ , which formed the hyper-network. Keep the small  $\lambda$  values in the combinations as more as possible, which indicated that more nodes are connected in the constructed hyper-edges. It was thought that the hyper-edges with many nodes could depict the underlying relationship among several brain regions. Then, the features were extracted for classification. The classification results in **Figure 11** indicated that the highest accuracy of 86.36% was obtained when  $\lambda_2 = 0.2$  and  $\lambda_1 = \{0.1, 0.2, \dots, 0.9\}$ . When  $\lambda_1 = \{0.1\}$ , the classification accuracies were  $<60\%$  because when  $\lambda_1$  used only one value, some nodes were contained in only one hyper-edge. The denominator of  $HCC^3$  is then zero, which makes it impossible to create an effective model for classification.

## Effect of SVM Classification Parameters $c$ and $g$

SVM classification is widely applied in various fields. It is the key to selecting the kernel function in the classification. The RBF kernel function was chosen in this experiment. The two parameters of the SVM model, the penalty factor  $c$  and the kernel parameter  $g$ , strongly influence the classification, and thus it is important to finding the optimal values. The penalty factor  $c$  is used to control the compromise between the model complexity and the approximation error. If  $c$  is too large, the data fitting and the complexity of the learning machine will be too high. There is an necessary process to avoid overfitting when designing the classifier. on the contrary, if  $c$  is too small, the punishment for the empirical error will be small, the learning complexity of the machine and the data fitting will be low. When the overfitting or underfitting occurs, the generalization ability of the classifier will be reduced to influence the classification performance. The value

**FIGURE 10** | Classification accuracy of different  $k$  values. The results were obtained by 50 experiments at each  $k$ -value. Error bars show standard deviation.

of  $g$  of the RBF kernel function is also important to directly affect the classification accuracy of the model.

For given values of  $c$  and  $g$ , the  $K$ -fold cross-validation method was used to acquire the training set validation accuracy. The values of  $c$  and  $g$  that generated the highest validation classification accuracy were chosen as the optimum parameters. The parameters of  $c$  and  $g$  were set at  $[2^{-8}, 2^8]$ , with a step of 1. **Figure 12** displays the result when using the classification features as a training set to conduct the parameter optimization of  $c, g$ , which indicates that the highest training set validation accuracy of 90.90% was achieved when  $c = 256$  and  $g = 0.0078$ .

## Limitations

The current study had three major limitations. First, the parameters of the hyper-network model in the experiment were a proportion of the corresponding parameters for the sparsest solution. The precise values were difficult to determine due to technical limitations. Furthermore, in the group lasso method, the random selection of the initial seed points and differences in the number of clusters  $k$  can change the network structure and the classification result. The construction of stable hyper-edges is expected to further improve the hyper-network. In addition, excessive analysis was not performed for comparing the current metrics ( $HCC1, 2, 3$ ) and other clustering metrics in the experiment. The selection of better clustering metrics can be tried in follow-up research.

## CONCLUSION

The original method of hyper-network construction used the lasso to solve the sparse linear regression model. The method was limited because some related regions could not be chosen because of the grouping effect among brain regions in the process of establishing hyper-edges. To solve this problem, elastic net

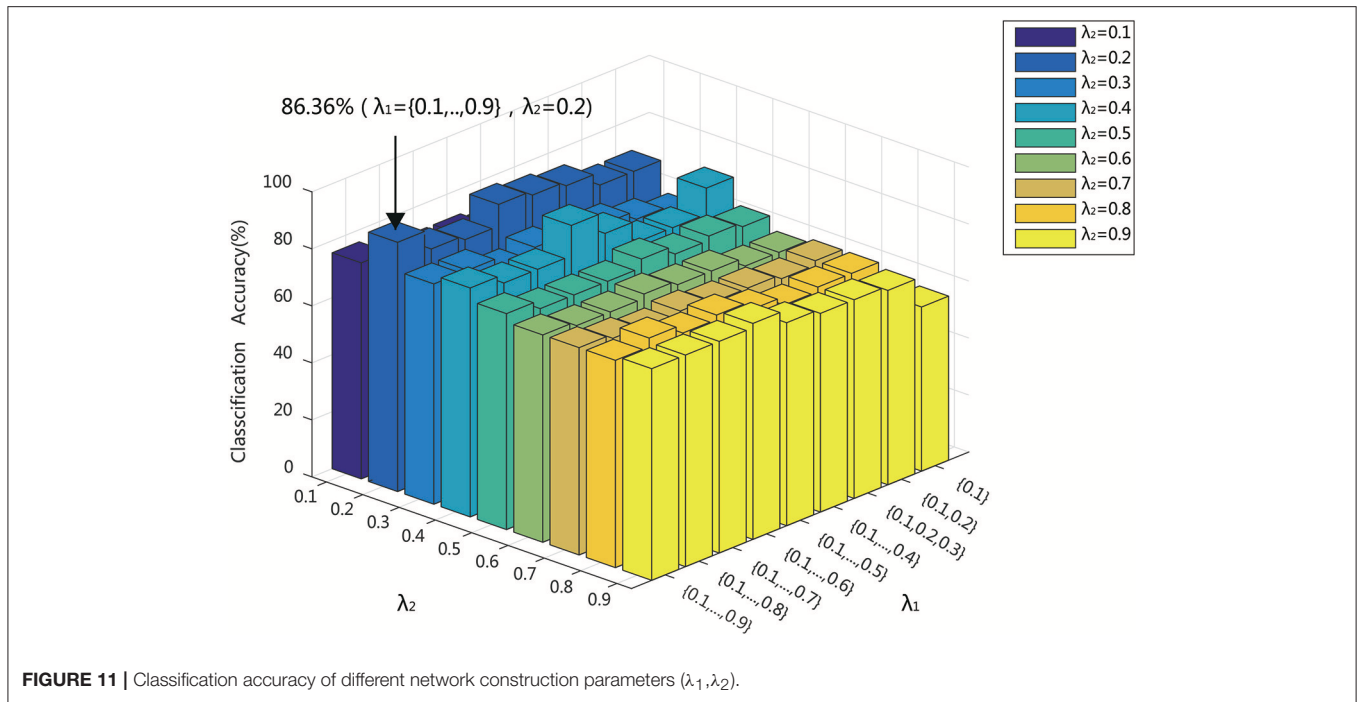


FIGURE 11 | Classification accuracy of different network construction parameters ( $\lambda_1, \lambda_2$ ).

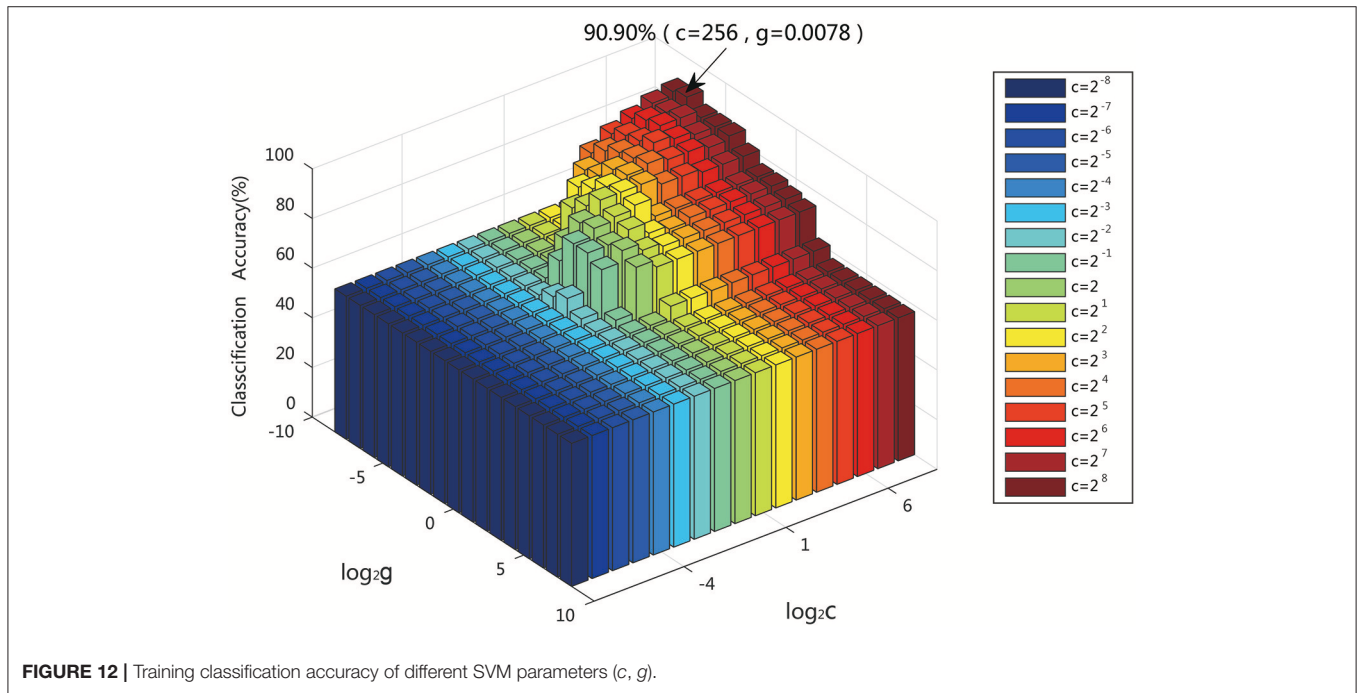


FIGURE 12 | Training classification accuracy of different SVM parameters ( $c, g$ ).

and group lasso methods were used to construct the hyper-networks. Analyses of the hyper-edges, brain area metrics, and average metrics implied that there were structural differences among the hyper-networks constructed by the three methods. The hyper-network obtained by the lasso was similar to that obtained by the elastic net but very different from that obtained by the group lasso. The lasso imposed a strict constraint on the network construction, the group lasso a loose constraint, and

the elastic net a moderate constraint. Considering the potential reasons, we concluded that the existence of the grouping effect and differences in the methods' ability to resolve it led to these consequences. Different constraint conditions resulted in varying classification accuracies. The elastic net method outperformed the others, and the group lasso method was not as good as the original method. Meanwhile, the elastic net method had a higher classification weight than the lasso method, and the group lasso



method had a lower classification weight. The results implied that a moderate connection constraint (elastic net) produced the most effective classification features, whereas stricter (lasso) and looser (group lasso) construction strategies were unable to achieve promising outcomes.

## AUTHOR CONTRIBUTIONS

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.

## REFERENCES

- Arthur, D., and Vassilvitskii, S. (2007). “k-means++: the advantages of careful seeding,” in *Eighteenth ACM-SIAM Symposium on Discrete Algorithms, SODA 2007* (New Orleans, LA).
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* 57, 289–300.
- Berge, C. (1989). *Hypergraphs*. Paris: Bordas.
- 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, 1404–1412. doi: 10.1016/j.neuroimage.2011.08.044
- Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S., and Sharma, T. (2000). How good is good enough in path analysis of fMRI data? *Neuroimage* 11, 289–301. doi: 10.1006/nimg.2000.0544
- Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575
- Chen, R., and Herskovits, E. H. (2007). Graphical-model-based multivariate analysis of functional magnetic resonance data. *Neuroimage* 35, 635–647. doi: 10.1016/j.neuroimage.2006.11.040
- De Mol, C., De Vito, E., and Rosasco, L. (2008). Elastic-net regularization in learning theory. *J. Comp.* 25, 201–230. doi: 10.1016/j.jco.2009.01.002
- Fasano, G., and Franceschini, A. (1987). A multidimensional version of the Kolmogorov–Smirnov test. *Mon. Not. R. Astron. Soc.* 50, 9–20.
- 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
- Friedman, J., Hastie, T., and Tibshirani, R. (2010a). *A Note on the Group Lasso and a Sparse Group Lasso*. Technical report, Department of Statistics, Stanford University.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010b). Regularization paths for generalized linear models via coordinate descent. *J. Statist. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01
- Fu, Z., Han, S., Tan, A., and Tu, Y. (2015). L0-regularized time-varying sparse inverse covariance estimation for tracking dynamic fMRI brain networks. *Conf. Proc. IEEE. Eng. Med. Biol. Soc.* 2015, 1496–1499. doi: 10.1109/EMBC.2015.7318654
- Furqan, M. S., and Siyal, M. Y. (2016). Elastic-net copula granger causality for inference of biological networks. *PLoS ONE* 11:e0165612. doi: 10.1371/journal.pone.0165612
- 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: ACM).

JC 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 study was supported by research grants from the National Natural Science Foundation of China (61373101, 61472270, 61402318, 61672374, and 61741212), the Natural Science Foundation of Shanxi Province (201601D021073), and the 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/fninf.2018.00025/full#supplementary-material>

- Gong, Y., Hao, L., Zhang, X., Zhou, Y., Li, J., Zhao, Z., et al. (2014). Case-control resting-state fMRI study of brain functioning among adolescents with first-episode major depressive disorder. *Shanghai Archiv. Psychiatry* 26, 207–215. doi: 10.3969/j.issn.1002-0829.2014.04.004
- Guo, H., Cao, X., Liu, Z., Li, H., Chen, J., and Zhang, K. (2012). Machine learning classifier using abnormal brain network topological metrics in major depressive disorder. *Neuroreport* 23, 1006–1011. doi: 10.1097/WNR.0b013e32835a650c
- Hoerl, A. E., and Kennard, R. W. (2000). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 42, 80–86. doi: 10.2307/1271436
- Hsu, C. W., Chang, C. C., and Lin, C. J. (2003). *A Practical Guide to Support Vector Classification*. Available online at: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- 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
- Liu, J., Ji, S., and Ye, J. (2011). *SLEP: Sparse Learning with Efficient Projections*. Available online at: <http://www.public.asu.edu/~jye02/Software/SLEP>
- Jie, B., Wee, C. Y., Shen, D., and Zhang, D. (2016). Hyper-connectivity of functional networks for brain disease diagnosis. *Med. Image Anal.* 32, 84–100. 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. Biomed. Eng.* 61, 576–589. doi: 10.1109/TBME.2013.2284195
- Jin, C., Gao, C., Chen, C., Ma, S., Netra, R., Wang, Y., et al. (2011). A preliminary study of the dysregulation of the resting networks in first-episode medication-naïve adolescent depression. *Neurosci. Lett.* 503, 105–109. doi: 10.1016/j.neulet.2011.08.017
- Kaiser, R. H., Whitfieldgabrieli, S., Dillon, D. G., Goer, F., Beltzer, M., Minkel, J., et al. (2016). Dynamic resting-state functional connectivity in major depression. *Neuropsychopharmacology* 41, 1822–1830. doi: 10.1038/npp.2015.352
- Kaufmann, M., van Kreveld, M., and Speckmann, B. (2009). “Subdivision drawings of hypergraphs,” in *Graph Drawing*, eds I. G. Tollis and M. Patrignani (Berlin; Heidelberg: Springer Berlin Heidelberg), 396–407. doi: 10.1007/978-3-642-00219-9\_39
- Kira, K., and Rendell, L. A. (1992). “The feature selection problem: traditional methods and a new algorithm,” in *Tenth National Conference on Artificial Intelligence* (San Jose, CA).
- 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, X., 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

- Liu, F., Hu, M., Wang, S., Guo, W., Zhao, J., Li, J., et al. (2012). Abnormal regional spontaneous neural activity in first-episode, treatment-naive patients with late-life depression: a resting-state fMRI study. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 39, 326–331. doi: 10.1016/j.pnpbp.2012.07.004
- Lord, A., Horn, D., Breakpear, M., and Walter, M. (2012). Changes in community structure of resting state functional connectivity in unipolar depression. *PLoS ONE* 7:e41282. doi: 10.1371/journal.pone.0041282
- Lv, J., Jiang, X., Li, X., Zhu, D., Chen, H., Zhang, T., et al. (2015). Sparse representation of whole-brain fMRI signals for identification of functional networks. *Med. Image Anal.* 20, 112–134. 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
- Mäkinen, E. (1990). How to draw a hypergraph. *Int. J. Comput. Math.* 34, 177–185. doi: 10.1080/00207169008803875
- Marrelec, G., Horwitz, B., Kim, J., Pélégrini-Issac, M., Benali, H., and Doyon, J. (2007). Using partial correlation to enhance structural equation modeling of functional MRI data. *Magn. Reson. Imaging.* 25, 1181–1189. doi: 10.1016/j.mri.2007.02.012
- Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehericy, S., Doyon, J., et al. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *Neuroimage* 32, 228–237. doi: 10.1016/j.neuroimage.2005.12.057
- Meier, L., V., Sara, D. G., and Bühlmann, P. (2008). The group lasso for logistic regression. *J. R. Statist. Soc.* 70, 53–71. doi: 10.1111/j.1467-9868.2007.00627.x
- Montani, F. R., Ince, A. A., Senator, R., Arabzadeh, E., Diamond, M. E., and Panzeri, S. (2009). The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. *Philosoph. Trans. R. Soc. A Math. Phys. Eng. Sci.* 367, 3297–3310. doi: 10.1098/rsta.2009.0082
- Ogutu, J. O., Schulz-Streeck, T., and Piepho, H. P. (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proc.* 6(Suppl. 2):S10. doi: 10.1186/1753-6561-6-S2-S10
- Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466, 617–621. doi: 10.1038/nature09178
- Park, H. S., and Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Syst. Appl.* 36, 3336–3341. doi: 10.1016/j.eswa.2008.01.039
- Pievani, M., Agosta, F., Filippi, M., and Frisoni, G. (2011). Functional networks connectivity in patients with Alzheimer's disease and mild cognitive impairment. *Alzheimers Dementia* 7. doi: 10.1016/j.jalz.2011.05.614
- Qiao, L., Zhang, H., Kim, M., Teng, S., Zhang, L., and Shen, D. (2016). Estimating functional brain networks by incorporating a modularity prior. *Neuroimage* 141, 399–407. doi: 10.1016/j.neuroimage.2016.07.058
- Qiu, L., Huang, X., Zhang, J., Wang, Y., Kuang, W., Li, J., et al. (2014). Characterization of major depressive disorder using a multiparametric classification approach based on high resolution structural images. *J. Psychiatry Neurosci.* 39, 78–86. doi: 10.1503/jpn.130034
- Ren, P., Aleksić, T., Wilson, R. C., and Hancock, E. R. (2011). A polynomial characterization of hypergraphs using the Ihara zeta function. *Pattern Recognit.* 44, 1941–1957. doi: 10.1016/j.patcog.2010.06.011
- Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., and Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex* 15, 1332–1342. doi: 10.1093/cercor/bhi016
- Smith, S. M., Miller, K. L., Salimikhorsidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for FMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063
- Souly, N., and Shah, M. (2016). Visual saliency detection using group lasso regularization in videos of natural scenes. *Int. J. Comput. Vis.* 117, 93–110. doi: 10.1007/s11263-015-0853-6
- Sporns, O. (2010). *Networks of the Brain*. Cambridge, MA: MIT Press.
- Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125. doi: 10.1111/j.1749-6632.2010.05888.x
- Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nat. Neurosci.* 17, 652–660. doi: 10.1038/nn.3690
- Teipel, S. J., Grothe, M. J., Metzger, C. D., Grimmer, T., Sorg, C., Ewers, M., et al. (2017). Robust detection of impaired resting state functional connectivity networks in alzheimer's disease using elastic net regularized regression. *Front. Aging Neurosci.* 8:318. doi: 10.3389/fnagi.2016.00318
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Velez, D. R., White, B. C., Motsinger, A. A., Bush, W. S., Ritchie, M. D., Williams, S. M., et al. (2007). A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genet. Epidemiol.* 31, 306–315. doi: 10.1002/gepi.20211
- Wee, C. Y., Yap, P. T., Denny, K., Brownlyke, J. N., Potter, G. G., Welshboher, 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, 641–656. doi: 10.1007/s00429-013-0524-8
- Ye, M., Yang, T., Peng, Q., Xu, L., Jiang, Q., and Liu, G. (2015). Changes of functional brain networks in major depressive disorder: a graph theoretical analysis of resting-state fMRI. *PLoS ONE* 10:e0133775. doi: 10.1371/journal.pone.0133775
- Yu, J., Tao, D., and Wang, M. (2012). Adaptive hypergraph learning and its application in image classification. *IEEE Trans. Image Proc.* 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.* 31, 17514–17526. doi: 10.1523/JNEUROSCI.3127-11.2011
- Yu, S., Yoshimoto, J., Toki, S., Takamura, M., Yoshimura, S., Okamoto, Y., et al. (2015). Toward probabilistic diagnosis and understanding of depression based on functional MRI data analysis with logistic group LASSO. *PLoS ONE* 10:e0123524. doi: 10.1371/journal.pone.0123524
- 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
- Zeng, L. L., Shen, H., Liu, L., Wang, L., Li, B., Fang, P., et al. (2012). Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain* 135(Pt 5), 1498–1507. doi: 10.1093/brain/aww059
- Zhang, J., Cheng, W., Wang, Z. G., Zhang, Z. Q., Lu, W. L., Lu, G. M., et al. (2012). Pattern classification of large-scale functional brain networks: identification of informative neuroimaging markers for epilepsy. *PLoS ONE* 7:e36733. doi: 10.1371/journal.pone.0036733
- Zhou, L., Wang, L., and Ogunbona, P. (2014). “Discriminative sparse inverse covariance matrix: application in brain functional network classification,” in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on 2014* (Columbus, OH).
- Zhu, J. Y., Shen, X. Y., Qin, J. L., Wei, M. B., Yan, R., Chen, J. H., et al. (2016). “Altered anatomical modular organization of brain networks in patients with major depressive disorder,” in *The International Conference on Biological Sciences and Technology* (Amsterdam).
- Zou, H., and Trevor, H. (2005). Regularization and variable selection via the elastic net. *J. R. Statist. Soc.* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

**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 © 2018 Guo, Li, Xu, Jin, Xiang and Chen. 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) and the copyright owner 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.