HHS Public Access

Author manuscript

Med Image Anal. Author manuscript; available in PMC 2017 March 02.

Published in final edited form as:

Med Image Anal. 2016 August ; 32: 84–100. doi:10.1016/j.media.2016.03.003.

Hyper-connectivity of functional networks for brain disease diagnosis

Biao Jie[a,b], **Chong-Yaw Wee**[c], **Dinggang Shen**[d,e], and **Daoqiang Zhang**[a]

[a]Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

[b]Department of Computer Science and Technology, Anhui Normal University, Wuhu, 241000, China

[c]Department of Biomedical Engineering, National University of Singapore, 119077, Singapore

[d]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA

[e]Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea

Abstract

Exploring structural and functional interactions among various brain regions enables better understanding of pathological underpinnings of neurological disorders. Brain connectivity network, as a simplified representation of those structural and functional interactions, has been widely used for diagnosis and classification of neurodegenerative diseases, especially for Alzheimer's disease (AD) and its early stage - mild cognitive impairment (MCI). However, the conventional functional connectivity network is usually constructed based on the pairwise correlation among different brain regions and thus ignores their higher-order relationships. Such loss of high-order information could be important for disease diagnosis, since neurologically a brain region predominantly interacts with more than one other brain regions. Accordingly, in this paper, we propose a novel framework for estimating the hyper-connectivity network of brain functions and then use this hyper-network for brain disease diagnosis. Here, the functional connectivity hyper-network denotes a network where each of its edges representing the interactions among multiple brain regions (i.e., an edge can connect with more than two brain regions), which can be naturally represented by a hyper-graph. Specifically, we first construct connectivity hyper-networks from the resting-state fMRI (R-fMRI) time series by using sparse representation. Then, we extract three sets of brain-region specific features from the connectivity hyper-networks, and further exploit a manifold regularized multi-task feature selection method to jointly select the most discriminative features. Finally, we use multi-kernel support vector machine (SVM) for classification. The experimental results on both MCI dataset and attention deficit hyperactivity disorder (ADHD) dataset demonstrate that, compared with the conventional connectivity network-based methods, the proposed method can *not only* improve the classification performance, *but also* help discover disease-related biomarkers important for disease diagnosis.

Correspondence to: Dinggang Shen; Daoqiang Zhang.

Keywords

Functional MR imaging; Hyper-network; Classification; Alzheimer's disease

1. Introduction

As a neurodegenerative disorder, Alzheimer's disease (AD) is the most common form of dementia in people over 65 years old, which currently has no cure for AD. AD causes substantial, progressive neuron damage that is irreversible and eventually leads to death. The number of affected people is expected to double in the next 20 years, and 1 in every 85 people will be affected by 2050 (Brookmeyer et al., 2007). Mild cognitive impairment (MCI) as a prodromal stage of AD has gained a great deal of attention recently due to its high progression rate to AD. Existing studies have shown that MCI subjects progress to clinical AD with an annual rate of approximately 10% to 15%, while normal controls (NC) develop dementia with an annual rate of 1% to 2% (Petersen et al., 2001). A further study also showed that the cognitive impairment has a significant impact on life expectancy, similar to chronic conditions such as diabetes or chronic heart failure (Sachs et al., 2011). Thus, accurate diagnosis of MCI is important for possible early treatment and delay of the progression of AD.

Evidence from both anatomical and physiological studies suggests that cognitive processes depend on interactions among distributed brain regions (Sporns, 2014). In the past years, modern imaging techniques have provided efficient ways to explore the functional and structural interactions of the human brain regions, thus enabling better understanding of the pathological underpinnings of neurological disorders. These interaction patterns, which can be characterized via connectivity networks, have been applied recently to disease diagnosis and classification (Pievani et al., 2011; Stam et al., 2009; Wang et al., 2013). For example, based on the blood oxygen level-dependent (BOLD) signals (that can reflect the endogenous or spontaneous brain activity with both high spatial and temporal resolutions) can be extracted from resting -state functional magnetic resonance imaging (R-fMRI) images. Then, the inter-regional interactions of brain activities at rest can be characterized via functional connectivity networks derived from BOLD signals and used for classification of AD and MCI (Chen et al., 2011; Jie et al., 2014b; Richiardi et al., 2012; Wee et al., 2013a).

In the literature, many functional connectivity modeling methods have been proposed, including correlation-based methods, graphical models, partial-correlation-based methods, and sparse representation-based methods (Smith et al., 2011; Wee et al., 2014). Among them, most existing studies are based on the correlation-based methods (Bullmore and Sporns, 2009; Kaiser, 2011; Smith et al., 2013; Sporns, 2011; Xie and He, 2011). It was reported that the correlation-based methods can obtain relatively high sensitivity for detecting network connections, compared with other methods (Smith et al., 2011). Recently, correlation-based connectivity networks have been successfully used for classifying patients with AD/MCI from normal controls (Seeley et al., 2009; Shen et al., 2010; Wee et al., 2012). However, correlation-based methods can only capture pairwise information and thus cannot fully reflect the interactions among multiple brain regions (Huang et al., 2010; Wee et al.,

2014). In addition, correlation-based networks suffer from too many spurious connections due to arbitrarily thresholding of the correlation networks (Wee et al., 2014).

On the other hand, graphical models have been used to study brain connectivity, such as structural equation models (Bullmore et al., 2000; McIntosh et al., 1994) and dynamic causal models (Friston et al., 2003). However, most of these methods are confirmative, rather than exploratory, which makes them inadequate for studying AD/MCI brain connectivity since little prior knowledge (such as which brain regions should be involved and how they are connected) is available but is often required in those methods (Huang et al., 2010).

In addition, partial correlations, which correspond to the off-diagonal entries of inverse covariance matrix of the data, have been used for measuring the correlation between two brain regions by factoring out the influence of other brain regions. Estimation of partial correlation is normally achieved by using maximum-likelihood estimation (MLE) of the inverse covariance matrix. To reliably estimate the inverse covariance matrix with limited sample size, Huang et al. (2010) imposed the sparsity constraint via l_1-norm regularization on MLE, namely the sparse inverse covariance matrix (SICE), for learning brain connectivity of AD, MCI and NC from PET data. Although this method is effective for learning a sparse connectivity network, it is not suitable for estimating the magnitude of connectivity due to the shrinking effect, and also it has been reported that the SICE method is very sensitive to the regularization parameters (Smith et al., 2011).

Recent works (Supekar et al., 2008; Zanin et al., 2012) showed that the use of certain sparsity connectivity modeling can elucidate robust connections from a set of noisy connections and thus improves final performance for disease classification. In particular, sparse representation has been proposed for constructing functional connectivity network. For example, Lee et al. (2011) adopted a least absolute shrinkage and selection operator (LASSO) with a l_1 regularizer to construct the functional connectivity network from PET images for analysis of autism. Wee et al. (2014) adopted the Group LASSO with a $l_{2,1}$ regularizer to model the functional connectivity for classifying patients with MCI from NC. In the sparse representation based methods, a sparse linear regression model was adopted to enable the representation of a brain region by a linear combination of signals of other brain regions while, in the meantime, filtering out insignificant or spurious connections. This provides a new way on modeling how a brain region is interacted with the rest of the brain regions. However, in (Lee et al., 2011), it was performed independently across different subjects of same group (i.e., autism patients), thus not suitable for classification. On the other hand, in (Wee et al., 2014), sparse representation was applied across all subjects, including both patients and NC, to estimate the connectivity networks with identical topology but different connectivity strengths, ignoring the group-specific network topological patterns.

Functional connectivity networks constructed by measuring *pairwise* correlations can only reflect the second-order relationship among brain regions, ignoring high-order relationship among them (i.e., the interaction among more than two brain regions). Such loss of high-order information could be crucial for understanding the pathological underpinnings of the disease since neurological findings have demonstrated that a brain region predominantly

interacts directly with a few of other brain regions in neurological processes (Huang et al., 2010). Furthermore, recent studies in neuroscience have also identified significant high-order interactions in neuronal spiking, local field potentials, and cortical activities (Ganmor et al., 2011; Montani et al., 2009; Ohiorhenuan et al., 2010; Yu et al., 2011).

To address the above issues, in this paper, we proposed a novel functional connectivity network modeling method for the purpose of identifying patients with MCI from NC. Specifically, a hyper-network is constructed based on R-fMRI time series with each node on the network representing a brain region and each edge containing more than two nodes to denote interactions among multiple brain regions simultaneously, which can be naturally represented by using a hyper-graph. To the best of our knowledge, our work is among the first to use the hyper-graph in neuroimaging studies. Specifically, we first construct the connectivity hyper-networks using sparse representation (SR) approach (Wright et al., 2009). We then extract three types of brain-region specific features (i.e., clustering coefficients) from the constructed connectivity hyper-networks. Furthermore, we exploit a manifold regularized multi-task feature selection (M2TFS) method proposed in (Jie et al., 2014a) to jointly select the most discriminative features from those three sets of clustering coefficients. Finally, we use a multi-kernel support vector machine (SVM) technique (Zhang et al., 2011; Wee et al., 2012; Wee et al., 2013b) to fuse the selected clustering coefficients for classification.

The proposed method is first evaluated on a MCI dataset with promising results, compared to the conventional connectivity network-based methods. We also seek to explore the biological meaning of the hyper-network for the brain regions involved in classification. Moreover, we investigate the robustness of the proposed method with respect to parameters and further analyze the effects of various techniques (e.g., feature selection - M2TFS) for classification performance. To further evaluate the classification performance of our proposed method, we also apply it on a larger attention deficit hyperactivity disorder (ADHD) dataset. The obtained results further demonstrate the effectiveness of our proposed method.

2. Materials and methods

Fig. 1 illustrates a flowchart of the proposed classification framework, consisting of several major steps, i.e., image preprocessing, construction of connectivity hyper-network, feature extraction and selection, and classification. In this section, we will give the detailed descriptions for each of these steps.

2.1. Subjects

In this study, 12 amnestic MCI patients (6 males and 6 females) and 25 normal controls (9 males and 16 females) were recruited. Demographic information of the participants is shown in Table 1. All the recruited subjects were diagnosed by expert consensus panels. Data acquisition was performed using a 3.0-Tesla GE Signa EXCITE scanner. R-fMRI images of each participant were acquired with the following parameters: flip angle = 77°, TR/TE = 2000/32 ms, imaging matrix = 64 ×64, FOV = 256 ×256 mm^2, 34 slices, 150 volumes, and

voxel thickness = 4 mm. During the scanning, all subjects were instructed to keep their eyes open and stare at a fixation cross in the middle of the screen, which lasted for 5 min.

2.2. Data pre-processing

Data preprocessing is performed using Statistical Parametric Mapping software package (SPM8) (http://www.fil.ion.ucl.ac.uk/spm/software/SPM8/). Specifically, the first 10 acquired fMRI images of each subject are discarded to ensure magnetization equilibrium. The remaining 140 images are first corrected for the acquisition time delay among different slices before realigning them to the first volume of the remaining images for head motion correction. To further reduce the contributions of ventricles and WM regions as well as head motion, regression of nuisance signals including ventricle and WM signals as well as six head-motion profiles are performed. The first scan of remaining fMRI time series is co-registered to the T1-weighted MR image of the same subject. The estimated transformation is then applied to other fMRI scans of the same subject. The brain space of fMRI scans for each subject is further parcellated into 116 regions of interesting (ROIs) by warping the Automated Anatomical Labeling (AAL) (Tzourio-Mazoyer et al., 2002) template to the subject space using the deformation fields estimated via a deformable registration method called HAMMER (Shen and Davatzikos, 2002). For each subject, the mean time series of each individual ROI is computed by averaging the nuisance signals-regressed fMRI time series over all voxels in each particular ROI. In current study, the GM-masked mean time series of each region is *band-pass filtered* within frequency interval [$0.025 \leq f \leq 0.100$ Hz], since the fMRI dynamics of neuronal activities is the most salient within this frequency interval. Given the controversy of removing the global signal in the preprocessing of R-fMRI data (Fox et al., 2009; Murphy et al., 2009), we do not regress the global signal out (Achard et al., 2006; Lynall et al., 2010; Supekar et al., 2008). Notably, the head-motion profiles are matched between the MCI and NC groups ($p > 0.218$ in any direction).

2.3. Hyper-graph

It is well known that graph is a powerful tool for representing relationships among the objects of interest, where each node in the graph denotes one object and each edge links nodes with certain kind of relationship. In neuroimaging field, graph theory has been widely applied to the analysis of brain connectivity (Fornito et al., 2013; Kaiser, 2011; Sporns, 2012). In the conventional graph (i.e., simple graph), an edge connects only two related nodes. That is, the conventional graph only characterizes the pairwise relationships between paired nodes. Indeed, in addition to pairwise relationships, in many applications (e.g., functional interaction among brain regions), there may exist high-order relationships, which cannot be represented by the conventional graph. To overcome this limitation, hyper-graph has been proposed in this paper to characterize the high-order relationship among nodes. In general, a hyper-graph is an extended graph where an edge (called hyper-edge) can connect more than two nodes (Zhou et al., 2006).

By denoting a hyper-graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a node (vertex) set $\mathcal{V}$ and a hyper-edge set $\mathcal{E}$, we can then represent $\mathcal{G}$ using a $|\mathcal{V}| \times |\mathcal{E}|$ incidence matrix H with 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 $v\in\mathcal{V}$ is a node and $e\in\mathcal{E}$ is a hyper-edge of $\mathcal{G}$.

Base on H, the node degree of each vertex $v\in\mathcal{V}$ is

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

and the edge degree of hyper-edge $e\in\mathcal{E}$ is

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

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

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

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

It is worth noting that the conventional graph is a special kind of hyper-graph with each hyper-edge containing only two nodes. Fig. 2 illustrates an example of hyper-graph. In the literature, hyper-graph has been successfully applied to a variety of applications, such as image classification (Yu et al., 2012) and protein function prediction (Gallagher and Goldberg, 2013).

2.4. Construction of connectivity hyper-network

Inspired by recent works (Lee et al., 2011; Wee et al., 2014), in our study, we construct the connectivity hyper-networks from R-fMRI time series using sparse representation (Wright et al., 2009). Specifically, denote $X=[x_1,\ldots,x_m,\ldots,x_M]^T\in R^{M\times d}$ as a training subject with a total of M ROIs, where x_m represents the regional-mean time series of the m-th ROI, and d is the length of time series. Then, the regional mean time series of each ROI (i.e., x_m) can be regarded as a response vector, and can be estimated by using a linear combination of times series of other $M-1$ ROIs as follows:

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

where $A_m = [x_1,\ldots,x_{m-1},0,x_{m+1},\ldots x_M]$ denotes a data matrix including all time series except the m-th ROI (where we put a vector of all zeros in its location), α_m denotes the weight vector that quantifies the degree of influence of other ROIs to the m-th ROI, and $\tau_m \in R^d$ denotes a noise term. Note that a zero element in the weight vector implies that the corresponding ROIs are insignificant in estimating the time series.

A sparse learning is used to optimize the following objective function

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

This is a well-known NP problem due to the l_0-norm term and often been approximated by solving a standard l_1-norm regularized optimization problem with the following objective function (Chen et al., 1998):

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

where $\lambda > 0$ is a regularization parameter controlling the sparsity of the model. Different λ value corresponds to different sparsity solution, and a larger λ value indicates a sparser model, i.e., more elements in α_m are zero. Many sparse learning algorithms can be implemented to solve l_1-norm, such as least angle regress (LARS) (Efron et al., 2004). By using this sparse representation, we can obtain the interaction of one region with a few of other regions while simultaneously forcing the insignificant or spurious interactions to zero. That is, the regions with the corresponding zero elements in the weight vector α_m are considered redundant in estimating the time series of one region. This provides a way on modeling how a brain region is interacted with the rest of brain regions by filtering out the redundant interactions.

In our study, to characterize the interactions among different brain regions, for each subject, a hyper-network is constructed by performing sparse representation for each brain ROI used as a node, and a hyper-edge e_m includes a centroid ROI (i.e., m-th ROI) and other ROIs with the corresponding non-zero elements in the weight vector α_m computed in Eq. (7). To reflect multi-level interactions of information among brain regions, for each ROI (or node), we generate a group of hyper-edges, instead of generating a single hyper-edge, via varying the value of λ in a specified range. Here, multi-level means that different λ values determine different levels of interaction relationships among brain regions. In other words, in Eq. (7), the objective function with larger λ value yields a sparser solution and thus the hyper-edge contains fewer nodes. Specifically, in our experiment, for simplicity, we vary λ value from 0.1 to 0.9 with an incremental step of 0.1. It is worth noting that since ROIs with the same time series have the same values in the weight vector of Eq. (7), so all of them will be *either* included in the corresponding hyper-edge, *or* excluded jointly. In the experiment, we adopt the SLEP package (Liu et al., 2009) to resolve the optimization problem in Eq. (7).

2.5. Feature extraction

Feature extraction and feature selection are the two special forms of dimensionality reduction in machine learning and image processing fields. Their goal is to prevent the curse of dimensionality problem (Guyon and Elisseeff, 2003) and also identify the most relevant features that can lead to a better generalization performance of the learning models. In the current study, we adopt both approaches to the connectivity hyper-networks for improving the disease diagnosis performance and also identifying biomarkers that are relevant to disease pathology.

In the conventional connectivity network, the clustering coefficient is widely used to quantify the degree to which nodes in a network tend to cluster together (Rubinov and Sporns, 2010). Numerous studies have shown that the local clustering property of functional connectivity network has been disrupted in the AD and MCI patients at a group comparison level (Bai et al., 2009; Liu et al., 2012; Wang et al., 2012). Recently, the concept of clustering coefficient has been extended to the hyper-network domain. In this study, three different types of clustering coefficients defined in (Gallagher and Goldberg, 2013) are used to extract features from connectivity hyper-networks.

Given a connectivity hyper-network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $S(v) = \{e_i \in \mathcal{E} : v \in e_i\}$ represent a set of hyper-edges adjacent to the node v. Let $N(v) = \{u \in \mathcal{V} : \exists e \in \mathcal{E}, u, v \in e\}$ be the nodes that are neighbors of the node v. Then, three different types of clustering coefficients on the node v can be defined, respectively, as follows:

$$HCC^1(v) = \frac{2\sum_{u,t \in N(v)} I(u,t,\neg v)}{|N(v)|(|N(v)|-1)} \quad (8)$$

$$HCC^2(v) = \frac{2\sum_{u,t \in N(v)} I'(u,t,v)}{|N(v)|(|N(v)|-1)} \quad (9)$$

$$HCC^3(v) = \frac{2\sum_{e \in S(v)} (|e|-1) - |N(v)|}{|N(v)|(|S(v)|-1)} \quad (10)$$

where $I(u, t, \neg v) = 1$ if there exists $e_i \in \mathcal{E}$ such that $u, t \in e_i$ but $v \notin e_i$, and 0 otherwise. $I'(u, t, v) = 1$ if there exists $e_i \in \mathcal{E}$ such that $u, t, v \in e_i$, and 0 otherwise.

These three types of clustering coefficients reflect local clustering properties of hyper-network from different views. The HCC^1 computes the number of adjacent nodes that have connections not facilitated by node v, under the hypothesis that these connections are more robust because “independent” evidence is provided. Conversely, the HCC^2 calculates the number of adjacent nodes that have connections facilitated by node v, considering that those

nodes may share a function with each other and node v. The HCC^3 calculates the amount of overlap among adjacent hyper-edges of node v.

Finally, for each of these three definitions of clustering coefficients in Eqs. (8)–(10), we extract a set of clustering coefficients from the connectivity hyper-networks as features, thus producing three sets of features for each subject.

2.6. Feature selection

Features extracted from connectivity hyper-networks potentially include irrelevant or redundant features for subsequent MCI classification. On the other hand, three types of clustering coefficient features reflect the local clustering properties of the connectivity hyper-network in three different views. In order to select the intrinsic common subset of features (i.e., from the same brain regions) that are relevant to MCI pathology, we exploit the manifold regularized multi-task feature selection (M2TFS) method proposed in our previous work (Jie et al., 2014a) to jointly select the most discriminative features, where each task focuses on classification using one type of clustering coefficient features. Compared with the conventional single-task feature selection methods, multi-task feature selection can utilize related auxiliary information among tasks and hence often leads to better learning model (Argyriou et al., 2008; Obozinski et al., 2010).

Let $Z^c=[z_1^c,\ldots,z_n^c,\ldots,z_N^c]^T \in R^{N\times M}$ represent three sets of features obtained from totally N training subjects, each with M ROIs. Here, $z_n^c=[HCC^c(v_m)]_{m=1:M} \in R^M$ represents the vector of clustering coefficients from the n-th training subject according to the above definition of HCC^c. Let $Y=[y_1,\ldots,y_n,\ldots,y_N]^T \in R^N$ be the response vector for those N training subjects, where y_n is the corresponding class label (i.e., MCI patient or normal control) for the n-th training subject. Then, the M2TFS method optimizes the following objective function (Jie et al., 2014a):

$$\min_W \frac{1}{2}\sum_{c=1}^{C}\|Y-Z^c w^c\|_2^2+\beta\sum_{c=1}^{C}(Z^c w^c)^T L^c(Z^c w^c)+\gamma\|W\|_{2,1} \quad (11)$$

where $L^c = D^c - S^c$ represents a combinatorial Laplacian matrix on the c-th task. S^c denotes a similarity matrix that defines the similarity on task c across different training subjects, which can be defined as: $S^c(i,j) = 1$ if z_i^c and z_j^c have the same class label and 0 otherwise. D^c is the diagonal matrix defined as $D^c(i,i)=\sum_{j=1}^{N} S^c(i,j)$. In Eq. (11), $W=[w^1,w^2,\ldots,w^C] \in R^{M\times C}$ is the weight matrix, where C is the number of tasks (i.e., $C = 3$), and $W_{2,1}=\sum_{m=1}^{M} w_{m2}$ is the group sparsity regularizer encouraging features to be selected jointly. Here, w_j is the j-th row vector of W. The parameters β and γ are the corresponding regularization coefficients, which balance the contributions of three items in Eq. (11). The values of β and γ can be determined via inner cross-validation on training data.

In the M2TFS method, the group-sparsity regularizer (Ng and Abugharbieh, 2011; Yuan and Lin, 2006) ensures only a small number of ROI-specific features to be jointly selected across different tasks. The Laplacian regularization item preserves the discriminative information of the data from each type of clustering coefficient features by incorporating the label information of both classes, and thus can induce more discriminative features for classification.

2.7. Classification

Previous study (Zhang et al., 2011) demonstrated that multi-kernel SVM can effectively integrate features from different tasks or modalities, compared to single-kernel SVM. Here, we also adopt the multi-kernel SVM to fuse three types of clustering coefficient features for classification. Specifically, for each set of clustering coefficient features of training subjects, a linear kernel is first computed based on features selected by the M2TFS method, i.e., $\{ f_i^c$, $i = 1,2,\ldots,N\}$, $c = 1,\ldots,C$, where f_i^c denotes the selected c-th type of features from the original features z_i^c of the i-th subject. Then, we adopt the following multi-kernel technique to combine three types of selected clustering coefficient features:

$$k(f_i, f_j) = \sum_{c=1}^{C} \mu^c k^c \left(f_i^c, f_j^c\right) \quad (12)$$

where $k^c(f_i^c, f_j^c)$ denotes the kernel function (i.e., linear kernel used in our experiments) over the c-th type of selected clustering coefficient features between the i-th and j-th subjects, and μ^c is a no-negative weight parameter with $\sum_{c=1}^{C} \mu^c = 1$.

Following (Zhang et al., 2011), we use a coarse-grid search through cross-validation on the training subjects to find the optimal μ^c. Once obtaining the optimal μ^c, the standard SVM can be implemented for classification.

2.8. Implementation details

In our experiments, leave-one-out (LOO) cross-validation is used to evaluate the performance of the proposed method. Specifically, one subject is first left out for testing, and the remaining ones are used for training. The entire process is repeated for each subject. The linear SVM classifier is implemented using LIBSVM toolbox (Chang and Lin, 2001) with a default parameter value. The weights in the multi-kernel classification method are determined based on the training subjects through a grid search with the range from 0 to 1 at a step size of 0.1, via another LOO cross-validation. Also, for each type of clustering coefficient, a total of 116 features are extracted from the constructed connectivity hyper-network. For each extracted feature, we normalized it with its mean and standard deviation computed from all training subjects. These values of the mean and standard deviation will be also used to normalize the corresponding feature of each testing subject in the application stage. It is worth noting that the nested LOO cross-validation strategy is used to enhance the generalization power of the classifier. Specifically, the inner cross-validation loop on the

training data is used for determining certain parameters, while the outer cross-validation loop is performed to evaluate the generalizability of learning models for the unseen subjects.

3. Experiments and results

3.1. Classification performance

We evaluated the classification performance of a method by measuring the classification accuracy (i.e., the proportion of subjects that are correctly identified), sensitivity (i.e., the proportion of patients that are correctly identified), specificity (i.e., the proportion of NC that are correctly identified), and area under receiver operating characteristic (ROC) curve (AUC). Besides, to avoid inflated performance on imbalanced datasets, we also compute the balanced accuracy of classification (Velez et al., 2007), which can be defined as the arithmetic mean of sensitivity and specificity.

The proposed method was compared with the conventional (i.e., pairwise Pearson-correlation-based) connectivity network based classification method (denoted as CN-CC), where all negative correlation coefficients in the network were set to zero and also the weighted clustering coefficients (Rubinov and Sporns, 2010) were extracted as features. For extensive comparison, we also compare the results obtained using single type of clustering coefficients extracted from the connectivity hyper-networks (denoted as HN_HCC^1, HN_HCC^2 and HN_HCC^3, respectively). Note that, in those comparison methods (i.e., CN-CC, HN_HCC^1, HN_HCC^2 and HN_HCC$^{3)}$, LASSO-based method is used to perform feature selection and a linear SVM classifier is used for classification. In addition, for better comparison, we also concatenate all clustering coefficients extracted from our connectivity hyper-networks into a longer feature vector and perform feature selection with M2TFS for the case of using only one task (i.e., $C = 1$), followed by a linear SVM for classification. Classification results of all compared methods are summarized in Table 2. Fig. 3 provides the ROC curves of the methods.

As shown in both Table 2 and Fig. 3, the proposed method significantly outperforms the competing methods. Specifically, the proposed method yields a classification accuracy of 94.6% and a balanced accuracy of 93.9%, while the best accuracy is only 91.9% and the best balanced accuracy is 91.8% by other methods. A cross-validation estimation of the generalization performance shows an AUC of 0.96, indicating excellent diagnostic power of the proposed method. Also, hyper-network based methods, i.e., HN_HCC^1, HN_HCC^2, HN_HCC^3, and CONCAT with CN_CC, consistently outperform the conventional (correlation-based) network based method, implying the advantages of the hyper-network with high-order information over the conventional network, i.e., with the second-order information in characterizing brain functional connectivity. Moreover, the proposed method and CONCAT method, which use three types of clustering coefficient features, consistently outperform methods using only a single type of clustering coefficients, which implies that different clustering properties of the connectivity hyper-network convey the complementary information and should be integrated to further improve the classification performance.

To evaluate possible data overfitting by LOO cross-validation, we perform an additional experiment using 10-fold cross-validation. Specifically, the whole set of subjects is first

partitioned into 10 subsets (each with a roughly equal size). Then, a subset is selected as the testing data, and the remaining 9 subsets are combined as the training data. This process is repeated 10 times, where each time a different subset is used as the testing data. The proposed method achieves a classification accuracy of 93.8% and an AUC of 0.93, indicating robustness of the proposed method.

On the other hand, in Table 3, we also compare our proposed method with several other state-of-the-art methods for connectivity-network based MCI classification, as briefly decribed below. For example, (Wang et al., 2013) derived the brain network from wavelet-based correlations of both high- and low-resolution parcellation units and adopted graph-theory approaches to investigate topological organization of functional connectivity of 37 patients with MCI and 47 NC subjects; (Chen et al., 2011) extracted the time series of 116 ROIs from R-fMRI images and used the Pearson product moment correlation coefficients of pairwise ROIs for classification of AD and MCI. (Wee et al., 2013a) proposed a sparse multivariate autoregressive (MAR) modeling to infer effective connectivity networks and applied to MCI classification. Also, (Wee et al., 2014) adopted the Group LASSO, based on $l_{2,1}$-norm, to incorporate sparsity into connectivity modeling and applied for classifying patients with MCI from NC. Finally, (Jie et al., 2014) constructed pairwise Pearson-correlation-based connectivity networks and integrated multiple properties of brain network for MCI classification. As we can see from Table 3, our proposed method achieves the best classification accuracy and AUC value, which again validates the efficacy of our proposed method.

3.2. Brain regions involved in classification

In this subsection, we investigate the important features (corresponding to ROIs) selected by our method for MCI classification. Since the selected features are different for each LOO cross-validation fold, we choose features that are always selected in all folds as the most important features. For each selected important feature, the standard t-test is performed on all subjects to evaluate its discriminative power between MC and NC. For comparison, we also perform the same test on the features (i.e., clustering coefficients) from the conventional connectivity network (denoted as CC). Table 4 lists those important brain regions, and Fig. 4 shows those brain regions in the template space. These brain regions include *frontal gyrus, rectus gyrus, cingulate, parahippocampal gyrus, occipital gyrus, temporal gyrus, inferior temporal and temporal pole*, which are consistent with previous studies using group comparison. On the other hand, most of the selected features have p-values smaller than 0.05, indicating good discriminative power between patients and NC. It is worth noting that most of the selected features in the proposed method are more discriminative than the features computed from the conventional connectivity network. This partly explains why our method can achieve better performance when compared with the conventional connectivity-network based method.

3.3. Connectivity analysis

To analyze the interaction of selected brain regions and to graphically show differences on connectivity hyper-network between MCI patients and NC, we compute the average hyper-edges based on the selected ROIs in Table 4 for each group (i.e., MCI and NC). Specifically,

for each ROI listed in Table 4, we repeat the following steps to construct hyper-edges of each group. First, for each subject in each group, we first construct a hyper-edge using the Eq. (7) with a fixed λ value, and then calculate the number of occurrence for each ROI in the hyper-network. Next, we compute the average of degrees of the hyper-edges for all subjects, denoted as d. Finally, for each group, we select top d ROIs with the highest occurrence number to construct the corresponding average hyper-edge. Here, d denotes rounding d to the nearest integer greater than, or equal to, d. Fig. 5 graphically shows the average hyper-edges constructed on 8 selected ROIs with $\lambda = 0.3$. Here, each sub-figure in Fig. 5 denotes a hyper-edge, with the red node (i.e., the centroid node linked by other nodes) in each sub-figure representing the selected ROI, and the green nodes representing the selected regions from cerebellum that is placed between left cerebrum and right cerebrum.

As can be seen from Fig. 5, nearly all the hyper-edges of MCI group are obviously different from those in NC group. For example, as shown in Fig. 5(a), for the hyper-edge constructed based on left middle frontal gyrus (L.MFG), it is interacted with right middle frontal gyrus (R.MFG), right angular gyrus (R.ANG), and left crus I of cerebellar hemisphere (L.CICH) for MCI group, while it is interacted with left angular gyrus (L.ANG), left inferior frontal gyrus (opercular) (L.IFGoperc), and left orbitofrontal cortex (medial) (L.ORBmed) for NC group. Especially, a region from cerebellum (i.e., L.CICH) is also involved in the interaction for MCI group. Similarly, for the hyper-edge constructed based on left inferior temporal (L.IT), the interaction pattern of MCI patients is completely different from that of NC as shown in Fig. 5(h).

To graphically show the differences of connectivity hyper-network between patient and healthy groups, we induce *another connectivity graph* from the hyper-network with ROIs as nodes and the element of adjacency matrix of hyper-network (defined in Eq. (4)) as connectivity weight between nodes. It is worth noting that larger connectivity weight between a pair of nodes indicates more involvement of this pair of nodes in hyper-edges (i.e., interactions). For distinguishing this *new connectivity graph* from the above connectivity graph, we call it as *connection network*. Specifically, for each subject, we construct a connection network by the adjacency matrix of the connectivity hyper-network. Then, we evaluate the discriminative power of each connection between patients and NC using the standard t-test, and then select those connections with p-value less than 0.05 as the best discriminative connections. Besides, we choose the top 15 ROIs with the highest occurrence frequency in those selected connections as important brain regions. Fig. 6 graphically shows the obtained p-value on each connection, where Fig. 6(a) shows the p-values on all connections. Fig. 6(b) shows the thresholded p-value (i.e., p-value more than 0.05 is set to 1), while Fig. 6(c) shows only the threshoded p-values for the connections between selected 15 ROIs. Colors in Fig. 5 denote the corresponding p-value.

As we can see from Figs. 6(a) and(b), significant connections mainly exist between some specific brain regions (i.e., selected 15 ROIs), including *parahippocampus, amygdala, temporal pole, inferior temporal and orbitofrontal cortex*, which were found to be associated with MCI pathology. From Fig. 6(c), we can see that the connections between those selected ROIs are significantly different between MCI patients and NC (i.e., the corresponding p-

value is very small). It is worth noting that these regions are highly overlapped with the ROIs selected based on classification.

Moreover, for both groups (i.e., MCI patients and NC), we respectively compute the average connectivity weight among the selected 15 ROIs. Each element in the average connection sub-network represents the average weight of the corresponding edges across subjects within the same group. Figs. 7(a) and (b) respectively show the average connectivity weights for MCI and NC groups, while Fig. 7(c) shows the between-group difference of connectivity weights. As can be seen from Fig. 7, the connectivity weights of MCI are larger than those of NC, suggesting that the MCI patients may require more interactions among brain regions than NC for brain compensation of cognitive impairment.

To analyze the relationship between the topology of the proposed hyper-network and that of the traditional Pearson-correlation-based connectivity network, we define an coefficient R as follows:

$$R=P/Q \quad (13)$$

where

$$P=\frac{\text{The total number of hyper}-\text{edges passing highly correlated node pairs}}{\text{The number of highly correlated node pairs}}$$

and

$$Q=\frac{\text{The total number of hyper}-\text{edges passing all node pairs}}{\text{The number of all node pairs}}$$

Here, the highly correlated node pair means the corresponding edge in the traditional connectivity network with larger correlation coefficients. In our experiment, we set the corresponding values larger than 0.6 according to the previous studies (Luders et al., 2009). According to this definition, P represents the average number of hyper-edges passing all highly-correlated node pairs, and Q represents the average number of hyper-edges passing all node pairs. The larger R (i.e., $R > 1$) means that the number of hyper-edges passed highly-correlated node pairs is larger that of hyper-edges passing other node pairs. We compute and provide R values for every subject of our dataset in Fig. 8. The R values are consistently larger than 1 for all subjects, suggesting that the highly-correlated nodes are more likely to be included in the corresponding hyper-edges, compared with other nodes.

3.4. Results on attention deficit hyperactivity disorder (ADHD) classification

To further investigate the efficacy of our proposed method, we apply it on a larger dataset from New York University (NYU) site of ADHD-200 database (http://fcon_1000.projects.nitrc.org/indi/adhd200/), which includes 98 NC and 118 children with

attention deficit hyperactivity disorder (ADHD). For constructing the hyper-network, we used the time series from the Athena preprocessed data. The corresponding filtered time series files, ADHD200_AAL_TCs_filtfix.tar.gz, can be downloaded from the website with ADHD-200 preprocessed data. A detailed description of data acquisition and postprocessing can be found on the Athena website (http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline). In short, the data preprocessing pipeline is as follows: removing the first four echo-planar image (EPI) volumes; slice timing correction; deoblique of data; realignment for head motion correction; masking the volumes to exclude voxels at non-brain regions; averaging EPI volumes to obtain a mean image; co-registering the mean image into template space ($4 \times 4 \times 4$ $mm^{3)}$; extracting the fMRI time series from WM and CSF regions using masks obtained from segmenting the structural T1-weighted images; removing effects of WM, CSF, head motion and a low-order polynomial (detrending); temporal band-pass filtering ($0.009 < f < 0.08$ Hz); spatial smoothing the filtered data using a 6-mm full width at half maximum(FWHM) Gaussian filter. Here, the AAL (Tzourio-Mazoyer et al., 2002) template is used to parcellate the brain space into 116 ROIs, from which the regional mean time series are extracted.

In this experiment, we adopt the same setting as for the MCI experiment. Table 5 summarizes the classification performance, and Fig. 9 shows the ROC curves of all comparison methods. The proposed method consistently achieves better classification performance than the competing methods, which once again demonstrates the advantage of the connectivity hyper-network over the conventional network in characterizing brain functional connectivity.

4. Discussion

In this paper, we propose a new connectivity hyper-network based disease classification method. Different from the conventional connectivity-network based methods, which only measure pairwise relationship between paired brain regions, our proposed method can characterize high-order interaction information among different brain regions, which may contain useful information for identifying patients with MCI from normal controls.

4.1. Significance of results

In the network-based analysis, the construction of network is a very important step. In the literature, researchers have proposed many (functional) network models, as summarized in (Smith et al., 2011). But most existing network models are based on simple graph, which only reflects the interaction relationship between paired brain regions. In this paper we proposed to use hyper-graph to construct a connectivity hyper-network model for characterizing high-order interactions among multiple brain regions. Hyper-graph is an extension of conventional simple graph, which has been successfully applied to many problems (Huang et al., 2011; Yu et al., 2012). Inspired by recent works (Lee et al., 2011; Wee et al., 2014), we constructed the hyper-network using the sparse representation to characterize the high-order interactions among multiple brain regions. To the best of our knowledge, our work is among the first to use the hyper-graph in neuroimaging studies.

Connectivity-network based methods have been used for diagnosis and classification of neurodegenerative diseases, such as AD and MCI. However, high-order interaction relationship among different brain regions, which may contain useful information for identifying patients from NC, has often been ignored in the existing connectivity based methods. Our study demonstrated that, by exploring the high-order relationship information among brain regions, the proposed method can achieve the significantly improved performance in classification of MCI, when compared to the state-of-the-art connectivity-network based methods.

Besides the performance improvement in classification, we also found that the brain regions detected by our proposed method are relevant to MCI pathology. These detected brain regions include *frontal gyrus* (Bell-McGinty et al., 2005; Buckner et al., 2009; Wang et al., 2007), *rectus gyrus* (Fleisher et al., 2009), *cingulate* (Grady et al., 2003; Greicius et al., 2004; Han et al., 2011), *parahippocampal gyrus* (Grady et al., 2001; Van Hoesen et al., 2000), *occipital gyrus* (Nobili et al., 2010; Supekar et al., 2008), *temporal gyrus* (Fleisher et al., 2009; Smith et al., 2011; Wang et al., 2007), *inferior temporal* (De Santi et al., 2001; Hamalainen et al., 2007) and *temporal pole* (Davatzikos et al., 2011; Nobili et al., 2008; Wang et al., 2007). Moreover, two regions from the cerebellum were also selected. Recent study (Baldacara et al., 2011) suggested that, although the cerebellum might not be directly associated with the origin of AD, it may provide useful information for AD prognosis.

On the other hand, we analyzed the interaction of selected brain regions involved in our method, and found that the interaction patterns of MCI group among these regions are obviously different from those of NC group. Further analysis on the connectivity hyper-network shows that those significantly affected connectivities are mainly observed among specific brain regions which have been found associated with MCI pathology. For example, the abnormalities in connectivity within the temporal lobe, especially between the *hippocampus* and *parahippocampus*, have been reported in AD and MCI patients (Greicius et al., 2009). Also, the alteration of functional connectivity of *amygdala* in AD and MCI has been reported in a recent work (Yao et al., 2013). Moreover, we found that the connectivity weights in MCI group are usually larger than those in NC group in the selected important regions, suggesting that MCI may require more inter-region interactions compared to NC for compensating the loss of network efficiency, in line with existing studies. For instance, Wang et al. (2013) had reported an increase of characteristic path length and impaired functional connectivity between different functional modules in MCI patients. Some studies have also reported a loss in small-world characteristics (i.e., shorter path length and higher degree of clustering) in subjects with MCI and AD (Liu et al., 2012; Sanz-Arigita et al., 2010; Yao et al., 2010). These changes in interaction patterns indicate that some brain regions have been affected by the disease, consistent with the evidences of early functional abnormality in MCI patients (Dickerson and Sperling, 2008; Feng et al., 2012; Liu et al., 2014; Stam et al., 2007).

4.2. Effect of regularization parameter λ

In our method, the hyper-networks are constructed using sparse representation. In this method, $\lambda > 0$ (in Eq. (2)) is a regularization parameter that controls the sparsity of

representation. Instead of generating a single hyper-edge for each ROI (node), we generate a group of hyper-edges by varying the λ value in a specified range. To investigate the effect of different number of λ values on classification performance of proposed method, we tested 9 groups of λ values, i.e., {0. 1}, {0.1, 0.2}, {0.1, 0.2, 0.3}, …, {0.1, 0.2, …, 0.9}. Fig. 10 gives the classification result, which indicates that the classification accuracy can be improved with the increase of the number of λ values.

Furthermore, to evaluate the difference of network topology with different groups of λ values, we also perform the significant test on the clustering coefficients from network. Specifically, the standard t-test is performed on each kind of clustering coefficients (i.e., HCC^1, HCC^2 and HCC^3) from our constructed hyper-network with λ= 0.1 and other groups of λ values. Fig. 11 gives the obtained results. The results show that the topology of network with λ= 0.1 is significantly different from the topology of networks constructed by most of other groups of λ values (i.e., with the corresponding p-value < 0.05). Moreover, from Fig. 11, we can see that the p-values decreased dynamically with the increase of the number of λ values, indicating significant difference of network topology when using large λ values.

4.3. Effect of feature extraction and feature selection

To evaluate the effect of feature extraction and feature selection, we perform three additional experiments: 1) without feature selection step (**Exp1)**, 2) with a different feature selection method (**Exp2)**, and 3) extracting another set of different features (**Exp3)**. Table 6 summaries all experimental results.

- Specifically, in **Exp1** we perform our proposed classification framework without feature selection, i.e., we directly perform multi-kernel SVM technique on those extracted three types of clustering coefficient features. For comparison, we also perform the conventional connectivity network based method without feature selection (denoted as CN).
- **In Exp2**, we perform our proposed classification framework using t-test based feature selection, instead of using the M2TFS method. For comparison, we also perform the conventional connectivity network based method with the t-test based feature selection (still denoted as CN), instead of using its original LASSO-based feature selection.
- **In Exp3**, we extract the degree of each node (defined in Eq. 2) from the connectivity hyper-network, instead of three types of clustering coefficients, as features for our proposed method. For comparison, we also extract the degree of each node from the conventional connectivity network as feature for the conventional connectivity network based method (still denoted as CN), where LASSO-based feature selection and linear SVM based classification are still used.

As can be seen from Table 6, our proposed method still achieves better classification performance in all three experiments than the conventional connectivity network based method, which again shows the advantage of using connectivity hyper-network over the conventional network in characterizing brain interactions. Furthermore, from both Table 2

and Table 6, we can see that the classification performances of methods with feature selection are significantly better than methods without feature selection, indicating the importance of feature selection. Also, we can see that M2TFS achieves better performance than t-test method, indicating that M2TFS method can better characterize the complementary information of three types of clustering coefficient features and thus obtain more discriminative features for classification.

4.4. Effect of regularization parameters β and γ

Feature selection method (M2TFS) in our proposed classification framework includes two regularization items, i.e., a manifold regularization term and a group-sparsity regularizer. Two parameters β and γ balance the relative contributions of these two regularization terms. To investigate the effects of regularization parameters β and γ on classification performance of our proposed method, we test different values of β, i.e., β =[5, 10, 15, 20, 25, 30, 35, 40, 45, 50], and also test different values of γ, i.e., γ =[0, 2, 4, 6, 8, 10]. It is worth noting that, when $\gamma = 0$, no feature selection step is performed, i.e., all features extracted from connectivity hyper-network are used for classification. Fig. 12 shows the classification accuracies with respect to different combinations of β and γ values.

As we can see from Fig. 12 and Table 2, the classification accuracy of our proposed method with respect to the use of different combinations of β and γ values is consistently better than that of the conventional connectivity network based method, validating again the efficacy of our proposed method. On the other hand, Fig. 9 indicates that, with fixed γ, the classification accuracy changes smoothly with varied β, implying the robustness of our proposed method with respect to the parameter β. Also, Fig. 12 shows that, with fixed β, the classification performance is largely affected by the γ value, suggesting the importance of selecting the optimal γ value for final classification. Actually, this is reasonable since parameter γ controls the sparsity of M2TFS and hence determines the scale of optimal feature subset. Finally, Fig. 12 shows that the classification accuracy with feature selection (i.e., γ >0) is better than methods without feature selection (i.e., $\gamma = 0$), demonstrating again the importance of feature selection.

4.5. Comparison on different combination schemes

To investigate the contribution of each weight, i.e., μ^{HCC1}, μ^{HCC2} and μ^{HCC3}, on the classification performance of the proposed method, we test all their possible values, ranging from 0 to 1 at a step size of 0.1, with the constraint of $\mu^{HCC1} + \mu^{HCC2} + \mu^{HCC3} = 1$. Fig. 13 provides the classification performance, including classification accuracy, sensitivity, specificity and AUC value, with respect to different combination of coefficient weights. It is worth noting that, in each subplot, only the squares in the upper triangular shows valid values due to the constraint $\mu^{HCC1} + \mu^{HCC2} + \mu^{HCC3} = 1$. For each plot, the three vertices of the figure, i.e., the top left, top right and bottom left, denote results obtained when using only a single type of clustering-coefficients, i.e., $HCC^3(\mu^{HCC3} = 1)$, $HCC^2(\mu^{HCC2} = 1)$, or $HCC^1(\mu^{HCC1} = 1)$.

As we can see from Fig. 13, most inner squares in the upper triangle show larger values (i.e., better classification performance) than the three vertices, indicating the effectiveness of

combining three clustering coefficients for classification. Moreover, for most plots, there are a larger set of squares with higher classification accuracy, indicating both the robustness and consistency of our proposed method with combined weights.

4.6. Reliability and repeatability on the use of different brain atlases

It was reported that brain networks derived from different parcellation schemes or using different spatial scales may exhibit distinct topological architectures (Fornito et al., 2010; Hayasaka and Laurienti, 2010; Zalesky et al., 2010). To evaluate the reliability and repeatability of our results, we repeated the same experiments using the functional atlases proposed by Dosenbach (Dosenbach et al., 2010) and Craddock (Craddock et al., 2012), which partitions the human brain into 160 and 200 ROIs, respectively. The preprocessing steps prior to functional connectivity computation are the same as for AAL atlas. It is noting worth that each ROI in the Dosenbach atlas is defined as a 10 mm diameter square surrounding a selected seed point, and also the distance between all ROI centers is at least 10 mm with no spatial overlap, indicating that some brain areas are not covered by the set of these ROIs. Hence, we computed the regional mean time series based only on the regions covered by these ROIs. Table 7 shows the classification performance of all competing methods based on the functional atlases.

As we can see from Table 7, the proposed method consistently performed bettern than the competing methods, indicating both the reliability and robustness of our proposed method when using different parcellation schemes with varied spatial scales and numbers of ROIs.

Furthermore, we also evaluate the compulational cost of our proposed method using atlases with varied numbers of ROIs. Specifically, for each atlas, we calculate the average computational time of our proposed method for each step (i.e., hyper-network construction, learning model training, and prediction) separately. The experiments are carried out using an Intel (R) Core (TM) 2 Quad 2.83 GHZ processor and 4.00 G RAM. Fig. 14 shows the obtained results, indicating that only the computational time for hyper-network construction is increased dramatically with the number of ROIs. This is reasonable since hyper-edges are constructed for every ROI.

4.7. Reliability of features

To investigate the test-retest reliability of network topological features extracted from our constructed hyper-network, we calculate the mean intra-class correlation coefficient (ICC) of each type of clustering coefficients (i.e., HCC^1, HCC^2 and $HCC^{3)}$ when using the first and second half of the time series. The experimental results show that all three types of clustering coefficients can yield large ICC values (i.e., ICC = 0.67 ± 0.09, ICC = 0.68 ± 0.10, and ICC = 0.74 ± 0.04, respectively), indicating the reliability of these three types of features. Also, we compute the ICC values of these clustering coefficients from 9 hyper-networks built with 9 groups of λ values, i.e., {0.1}, {0.1, 0.2}, {0.1, 0.2, 0.3}, …, {0.1, 0.2, …, 0.9}, respectively, as shown in Fig. 15. The results in Fig. 15 show that 1) ICC values of three types of clustering coefficients increase with the increased number of λ values, and 2) these clustering coefficients become stable with the use of enough number of λ values.

4.8. Effect of feature normalization

Normalization is important to adjust features of different scales to a notionally common scale, thus enabling a more appropriate comparison among features during classification. To evaluate the effect of normalization on classification performance of our proposed method, we perform an additional experiment without feature normalization. The experiment results show that our proposed method without feature normalization can achieve a classification accuracy of 91.9% of and an AUC of 0.82, which again demonstrates the effectiveness of our proposed method. Meanwhile, these results are worse than the result obtained by the proposed method with feature normalization, indicating the importance of using a common scale for all features for improving the classification performance.

4.9. Limitation

In the current study, there are two major limitations. First, constructing the stable hyper-edge is a very important task for sparse-based hyper-network construction method. In the future work, we will explore some advanced techniques, such as robust LASSO (Xu et al., 2010) and group LASSO (Yuan and Lin, 2006), to address this limitation. Another limitation is the size of dataset, although we used a relatively large dataset for ADHD study. In the future work, we will evaluate the proposed method on dataset with larger sample sizes.

5. Conclusion

In summary, we have proposed a new (functional) connectivity hyper-network based classification method by utilizing high-order relationships among brain regions to facilitate disease classification. This method is completely different from conventional methods which often use only the pairwise relationships measured via Pearson correlation. Experimental results on both MCI and ADHD classifications indicate that our proposed method can not only improve brain disease classification, but also facilitate detection of disease-relevant structures.

Acknowledgments

The author acknowledges support from National Natural Science Foundation of China (Nos. 61573023, 61422204, 61473149, 61473190), Natural Science Foundation of Anhui Province (No. 1508085MF125), the Jiangsu Natural Science Foundation for Distinguished Young Scholar (No. BK20130034), the Open Projects Program of National Laboratory of Pattern Recognition (No. 201407361), the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20123218110009), the NUAA Fundamental Research Funds (No. NE2013105), and NIH grants (EB006733, EB008374, EB009634, MH100217, AG041721, AG049371, and AG042599).

References

Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J Neurosci. 2006; 26:63–72. [PubMed: 16399673]

Argyriou A, Evgeniou T, Pontil M. Convex multi-task feature learning. Mach Learn. 2008; 73:243–272.

Bai F, Zhang Z, Watson DR, Yu H, Shi Y, Yuan Y, et al. Abnormal functional connectivity of hippocampus during episodic memory retrieval processing network in amnestic mild cognitive impairment. Biol Psychiatry. 2009; 65:951–958. [PubMed: 19028382]

Baldacara L, Borgio JG, Moraes WA, Lacerda AL, Montano MB, Tufik S, et al. Cerebellar volume in patients with dementia. Rev Bras Psiquiatr. 2011; 33:122–129. [PubMed: 21829904]

Bell-McGinty S, Lopez OL, Meltzer CC, Scanlon JM, Whyte EM, Dekosky ST, et al. Differential cortical atrophy in subgroups of mild cognitive impairment. Arch Neurol. 2005; 62:1393–1397. [PubMed: 16157746]

Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement. 2007; 3:186–191. [PubMed: 19595937]

Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu HS, Hedden T, et al. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to alzheimer's disease. J Neurosci. 2009; 29:1860–1873. [PubMed: 19211893]

Bullmore E, Horwitz B, Honey G, Brammer M, Williams S, Sharma T. How good is good enough in path analysis of fMRI data? Neuroimage. 2000; 11:289–301. [PubMed: 10725185]

Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci. 2009; 10:186–198. [PubMed: 19190637]

Chang CC, Lin CJ. LIBSVM: a library for support vector machines. Acm Transactions on Intelligent Systems and Technology. 2011; 2:389–396.

Chen G, Ward BD, Xie C, Li W, Wu Z, Jones JL, et al. Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. Radiology. 2011; 259:213–221. [PubMed: 21248238]

Chen SSB, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. Siam J Sci Comput. 1998; 20:33–61.

Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum Brain Mapp. 2012; :33.doi: 10.1002/hbm.21333

Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging. 2011; 32:e2319–e2327. 2322.

De Santi S, de Leon MJ, Rusinek H, Convit A, Tarshish CY, Roche A, et al. Hippocampal formation glucose metabolism and volume losses in MCI and AD. Neurobiol Aging. 2001; 22:529–539. [PubMed: 11445252]

Dickerson BC, Sperling RA. Functional abnormalities of the medial temporal lobe memory system in mild cognitive impairment and Alzheimer's disease: insights from functional MRI studies. Neuropsychologia. 2008; 46:1624–1635. [PubMed: 18206188]

Dosenbach NUF, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, et al. Prediction of individual brain maturity using fMRI. Science. 2010; 329:1358–1361. [PubMed: 20829489]

Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat. 2004; 32:407–451.

Feng Y, Bai L, Ren Y, Chen S, Wang H, Zhang W, et al. FMRI connectivity analysis of acupuncture effects on the whole brain network in mild cognitive impairment patients. Magn Reson Imaging. 2012; 30:672–682. [PubMed: 22459434]

Fleisher AS, Sherzai A, Taylor C, Langbaum JB, Chen K, Buxton RB. Resting-state BOLD networks versus task-associated functional MRI for distinguishing Alzheimer's disease risk groups. Neuroimage. 2009; 47:1678–1690. [PubMed: 19539034]

Fornito A, Zalesky A, Breakspear M. Graph analysis of the human connectome: promise, progress, and pitfalls. Neuroimage. 2013; 80:426–444. [PubMed: 23643999]

Fornito A, Zalesky A, Bullmore ET. Network scaling effects in graph analytic studies of human resting-state fMRI data. Front Syst Neurosci. 2010; 4:22. [PubMed: 20592949]

Fox MD, Zhang D, Snyder AZ, Raichle ME. The global signal and observed anticorrelated resting state brain networks. J Neurophysiol. 2009; 101:3270–3283. [PubMed: 19339462]

Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage. 2003; 19:1273–1302. [PubMed: 12948688]

Gallagher, SR., Goldberg, DS. Clustering coefficients in protein interaction hyper-networks. In: Gao, J., editor. ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM BCB). 2013. p. 552-560.

Ganmor E, Segev R, Schneidman E. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. Proc Natl Acad Sci USA. 2011; 108:9679–9684. [PubMed: 21602497]

Grady CL, Furey ML, Pietrini P, Horwitz B, Rapoport SI. Altered brain functional connectivity and impaired short-term memory in Alzheimer's disease. Brain. 2001; 124:739–756. [PubMed: 11287374]

Grady CL, McIntosh AR, Beig S, Keightley ML, Burian H, Black SE. Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. J Neurosci. 2003; 23:986–993. [PubMed: 12574428]

Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc Natl Acad Sci U S A. 2004; 101:4637–4642. [PubMed: 15070770]

Greicius MD, Supekar K, Menon V, Dougherty RF. Resting-state functional connectivity reflects structural connectivity in the default mode network. Cereb Cortex. 2009; 19:72–78. [PubMed: 18403396]

Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003; 3:1157–1182.

Hamalainen A, Pihlajamaki M, Tanila H, Hanninen T, Niskanen E, Tervo S, et al. Increased fMRI responses during encoding in mild cognitive impairment. Neurobiol Aging. 2007; 28:1889–1903. [PubMed: 16997428]

Han Y, Wang J, Zhao Z, Min B, Lu J, Li K, He Y, Jia J. Frequency-dependent changes in the amplitude of low-frequency fluctuations in amnestic mild cognitive impairment: a resting-state fMRI study. Neuroimage. 2011; 55:287–295. [PubMed: 21118724]

Hayasaka S, Laurienti PJ. Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. Neuroimage. 2010; 50:499–508. [PubMed: 20026219]

Huang S, Li J, Sun L, Ye J, Fleisher A, Wu T, et al. Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. Neuroimage. 2010; 50:935–949. [PubMed: 20079441]

Huang Y, Liu Q, Lv F, Gong Y, Metaxas DN. Unsupervised image categorization by hypergraph partition. IEEE Trans Pattern Anal Mach Intell. 2011; 33:1266–1273. [PubMed: 21282850]

Jie B, Zhang D, Cheng B, Shen D. Manifold regularized multitask feature learning for multimodality disease classification. Hum Brain Mapp. 2014; doi: 10.1002/hbm.22642

Jie B, Zhang D, Gao W, Wang Q, Wee CY, Shen D. Integration of network topological and connectivity properties for neuroimaging classification. IEEE Trans Biomed Eng. 2014; 61:576–589. [PubMed: 24108708]

Kaiser M. A tutorial in connectome analysis: Topological and spatial features of brain networks. Neuroimage. 2011; 57:892–907. [PubMed: 21605688]

Lee H, Lee DS, Kang H, Kim BN, Chung MK. Sparse brain network recovery under compressed sensing. IEEE Trans Med Imaging. 2011; 30:1154–1165. [PubMed: 21478072]

Liu, J., Ji, S., Ye, J. SLEP: Sparse Learning with Efficient Projections. Arizona State University; Phoenix, Arizona: 2009.

Liu Y, Yu C, Zhang X, Liu J, Duan Y, Alexander-Bloch AF, et al. Impaired long distance functional connectivity and weighted network architecture in Alzheimer's disease. Cereb Cortex. 2014; 24:1422–1435. [PubMed: 23314940]

Liu Z, Zhang Y, Yan H, Bai L, Dai R, Wei W, et al. Altered topological patterns of brain networks in mild cognitive impairment and Alzheimer's disease: a resting-state fMRI study. Psychiatry Res. 2012; 202:118–125. [PubMed: 22695315]

Luders E, Narr KL, Thompson PM, Toga AW. Neuroanatomical correlates of intelligence. Intelligence. 2009; 37:156–163. [PubMed: 20160919]

Lynall ME, Bassett DS, Kerwin R, McKenna PJ, Kitzbichler M, Muller U, et al. Functional connectivity and brain networks in schizophrenia. J Neurosci. 2010; 30:9477–9487. [PubMed: 20631176]

McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B. Network analysis of cortical visual pathways mapped with PET. J Neurosci. 1994; 14:655–666. [PubMed: 8301356]

Montani F, Ince RA, Senatore R, Arabzadeh E, Diamond ME, Panzeri S. The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. Philos Trans A Math Phys Eng Sci. 2009; 367:3297–3310. [PubMed: 19620125]

Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? Neuroimage. 2009; 44:893–905. [PubMed: 18976716]

Ng B, Abugharbieh R. Generalized sparse regularization with application to fMRI brain decoding. Inf Process Med Imaging. 2011; 22:612–623. [PubMed: 21761690]

Nobili F, Mazzei D, Dessi B, Morbelli S, Brugnolo A, Barbieri P, et al. Unawareness of memory deficit in amnestic MCI: FDG-PET findings. J Alzheimers Dis. 2010; 22:993–1003. [PubMed: 20858977]

Nobili F, Salmaso D, Morbelli S, Girtler N, Piccardo A, Brugnolo A, et al. Principal component analysis of FDG PET in amnestic MCI. Eur J Nucl Med Mol Imaging. 2008; 35:2191–2202. [PubMed: 18648805]

Obozinski G, Taskar B, Jordan MI. Joint covariate selection and joint subspace selection for multiple classification problems. Stat Comput. 2010; 20:231–252.

Ohiorhenuan IE, Mechler F, Purpura KP, Schmid AM, Hu Q, Victor JD. Sparse coding and high-order correlations in fine-scale cortical networks. Nature. 2010; 466:617–621. [PubMed: 20601940]

Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, et al. Current concepts in mild cognitive impairment. Arch Neurol. 2001; 58:1985–1992. [PubMed: 11735772]

Pievani M, Agosta F, Galluzzi S, Filippi M, Frisoni GB. Functional networks connectivity in patients with Alzheimer's disease and mild cognitive impairment. J Neur. 2011; 258:170–170.

Richiardi J, Gschwind M, Simioni S, Annoni JM, Greco B, Hagmann P, et al. Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity. Neuroimage. 2012; 62:2021–2033. [PubMed: 22677149]

Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage. 2010; 52:1059–1069. [PubMed: 19819337]

Sachs GA, Carter R, Holtz LR, Smith F, Stump TE, Tu W, et al. Cognitive impairment: an independent predictor of excess mortality: a cohort study. Ann Intern Med. 2011; 155:300–308. [PubMed: 21893623]

Sanz-Arigita EJ, Schoonheim MM, Damoiseaux JS, Rombouts SA, Maris E, Barkhof F, et al. Loss of 'small-world' networks in Alzheimer's disease: graph analysis of FMRI resting-state functional connectivity. Plos One. 2010; 5:e13788. [PubMed: 21072180]

Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. Neuron. 2009; 62:42–52. [PubMed: 19376066]

Shen D, Davatzikos C. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans Med Imaging. 2002; 21:1421–1439. [PubMed: 12575879]

Shen H, Wang LB, Liu YD, Hu DW. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. Neuroimage. 2010; 49:3110–3121. [PubMed: 19931396]

Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, et al. Network modelling methods for FMRI. Neuroimage. 2011; 54:875–891. [PubMed: 20817103]

Smith SM, Vidaurre D, Beckmann CF, Glasser MF, Jenkinson M, Miller KL, et al. Functional connectomics from resting-state fMRI. Trends Cogn Sci. 2013; 17:666–682. [PubMed: 24238796]

Sporns O. The human connectome: a complex network. Ann N Y Acad Sci. 2011; 1224:109–125. [PubMed: 21251014]

Sporns O. From simple graphs to the connectome: networks in neuroimaging. Neuroimage. 2012; 62:881–886. [PubMed: 21964480]

Sporns O. Contributions and challenges for network models in cognitive neuroscience. Nat Neurosci. 2014; 17:652–660. [PubMed: 24686784]

Stam CJ, de Haan W, Daffertshofer A, Jones BF, Manshanden I, van Cappellen van Walsum AM, et al. Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. Brain. 2009; 132:213–224. [PubMed: 18952674]

Stam CJ, Jones BF, Nolte G, Breakspear M, Scheltens P. Small-world networks and functional connectivity in Alzheimer's disease. Cereb Cortex. 2007; 17:92–99. [PubMed: 16452642]

Supekar K, Menon V, Rubin D, Musen M, Greicius MD. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLOS Comput Biol. 2008; 4:e1000100. [PubMed: 18584043]

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002; 15:273–289. [PubMed: 11771995]

Van Hoesen GW, Augustinack JC, Dierking J, Redman SJ, Thangavel R. The parahippocampal gyrus in Alzheimer's disease. Clinical and preclinical neuroanatomical correlates. Ann N Y Acad Sci. 2000; 911:254–274. [PubMed: 10911879]

Velez DR, White BC, Motsinger AA, Bush WS, Ritchie MD, Williams SM, et al. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet Epidemiol. 2007; 31:306–315. [PubMed: 17323372]

Wang J, Zuo X, Dai Z, Xia M, Zhao Z, Zhao X, et al. Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. Biol Psychiatry. 2013; 73:472–481. [PubMed: 22537793]

Wang K, Liang M, Wang L, Tian L, Zhang X, Li K, et al. Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. Hum Brain Mapp. 2007; 28:967–978. [PubMed: 17133390]

Wang Z, Jia X, Liang P, Qi Z, Yang Y, Zhou W, et al. Changes in thalamus connectivity in mild cognitive impairment: evidence from resting state fMRI. Eur J Radiol. 2012; 81:277–285. [PubMed: 21273022]

Wee CY, Li Y, Jie B, Peng ZW, Shen D. Identification of MCI using optimal sparse MAR modeled effective connectivity networks. Med Image Comput Comput-Assis Interv. 2013a:8150.

Wee CY, Yap PT, Zhang D, Denny K, Browndyke JN, Potter GG, et al. Identification of MCI individuals using structural and functional connectivity networks. Neuroimage. 2012; 59:2045–2056. [PubMed: 22019883]

Wee CY, Yap PT, Shen DG. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Hum Brain Mapp. 2013b; 34:3411–3425. [PubMed: 22927119]

Wee CY, Yap PT, Zhang D, Wang L, Shen D. Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. Brain Struct Funct. 2014; 219:641–656. [PubMed: 23468090]

Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell. 2009; 31:210–227. [PubMed: 19110489]

Xie T, He Y. Mapping the Alzheimer's brain with connectomics. Front Psychiatry. 2011; 2:77. [PubMed: 22291664]

Xu HA, Caramanis C, Mannor S. Robust regression and Lasso. IEEE Trans Inf Theory. 2010; 56:3561–3574.

Yao H, Liu Y, Zhou B, Zhang Z, An N, Wang P, et al. Decreased functional connectivity of the amygdala in Alzheimer's disease revealed by resting-state fMRI. Eur J Radiol. 2013; 82:1531–1538. [PubMed: 23643516]

Yao Z, Zhang Y, Lin L, Zhou Y, Xu C, Jiang T. Abnormal cortical networks in mild cognitive impairment and Alzheimer's disease. Plos Comput Biol. 2010; 6:e1001006. [PubMed: 21124954]

Yu J, Tao D, Wang M. Adaptive hypergraph learning and its application in image classification. IEEE Trans Image Process. 2012; 21:3262–3272. [PubMed: 22410334]

Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc: Series B (Stat Methodol). 2006; 68:49–67.

Yu S, Yang HD, Nakahara H, Santos GS, Nikolic D, Plenz D. Higher-order interactions characterized in cortical activity. J Neurosci. 2011; 31:17514–17526. [PubMed: 22131413]

Zalesky A, Fornito A, Harding IH, Cocchi L, Yucel M, Pantelis C, et al. Whole-brain anatomical networks: does the choice of nodes matter? Neuroimage. 2010; 50:970–983. [PubMed: 20035887]

Zanin M, Sousa P, Papo D, Bajo R, Garcia-Prieto J, del Pozo F, et al. Optimizing functional network representation of multivariate time series. Sci Rep. 2012; 2:630. [PubMed: 22953051]

Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. 2011; 55:856–867. [PubMed: 21236349]

Zhou, D., Huang, J., Schölkopf, B. Advances in neural information processing systems (NIPS). MIT Press; Vancouver, British Columbia, Canada: 2006. Learning with hypergraphs: clustering, classification, and embedding; p. 1601-1608.

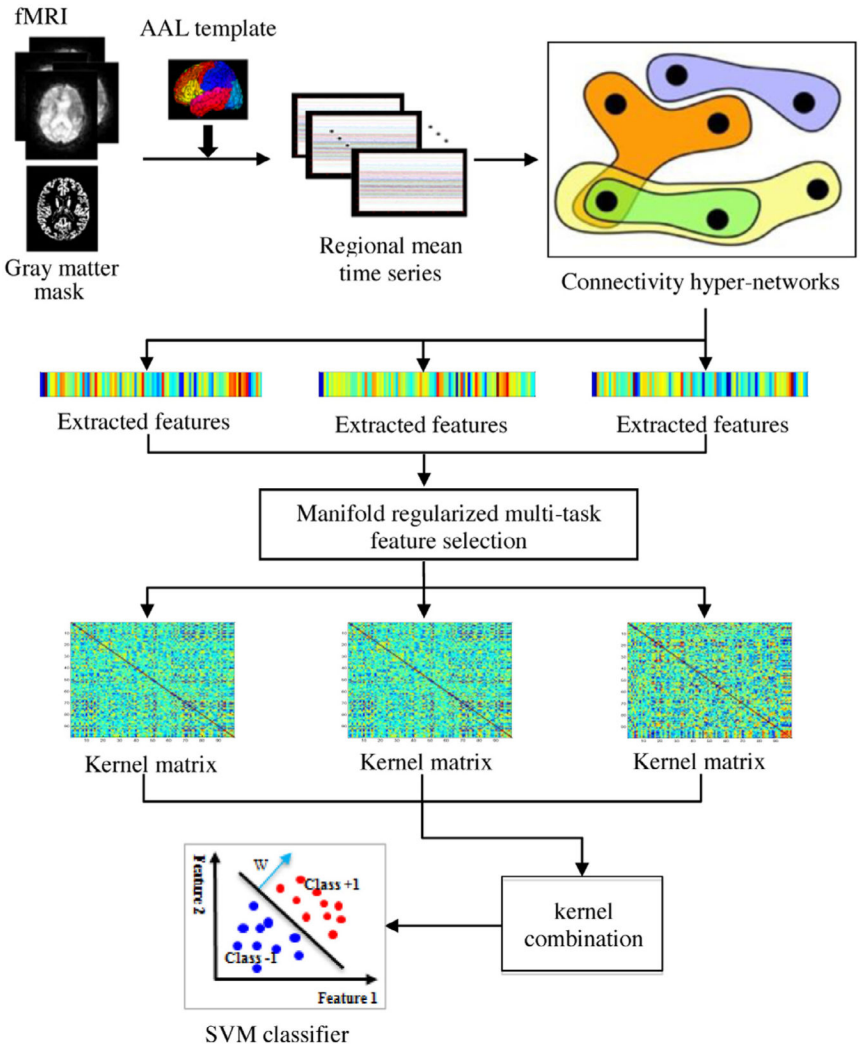
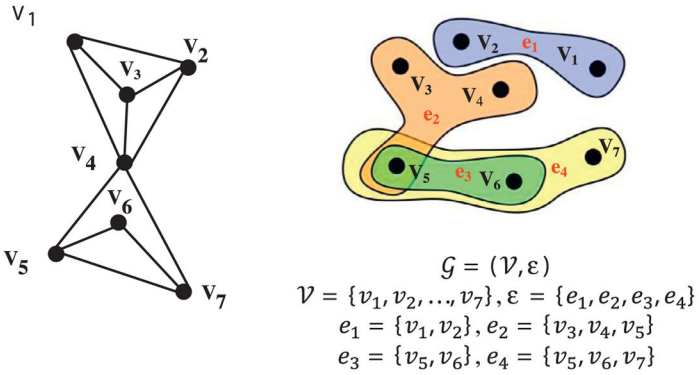


Fig. 1.
Flowchart of the proposed method.



	e_1	e_2	e_3	e_4
V_1	1	0	0	0
V_2	1	0	0	0
V_3	0	1	0	0
V_4	0	1	0	0
V_5	0	1	1	1
V_6	0	0	1	1
V_7	0	0	0	1

Fig. 2.

Hyper-graph vs. graph. Left: A conventional graph in which two nodes are connected together by an edge. Middle: a hyper-graph in which each hyper-edge can connect more than two nodes. Right: The incidence matrix for the hyper-graph in the middle.

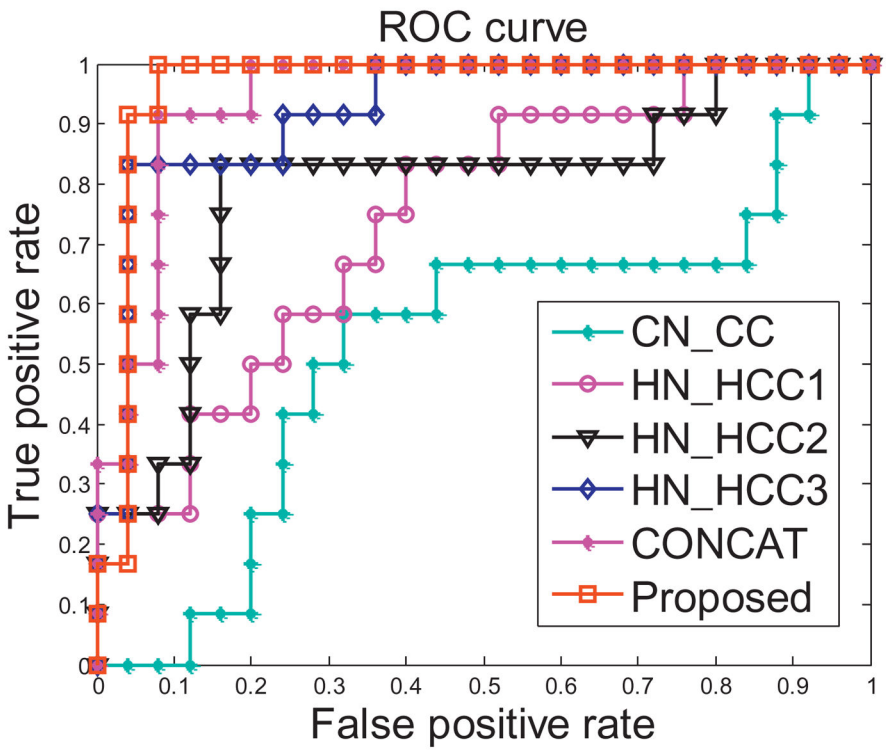


Fig. 3.
ROC curves of the compared methods for MCI classification.

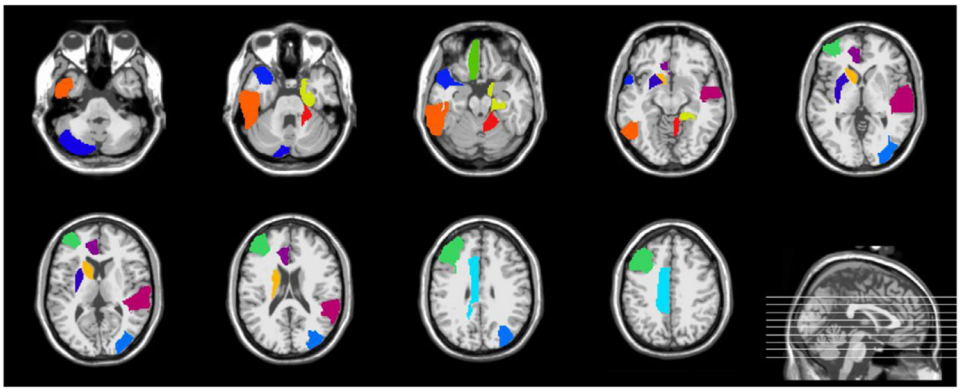


Fig. 4.
The important ROIs selected by the proposed method for MCI classification.

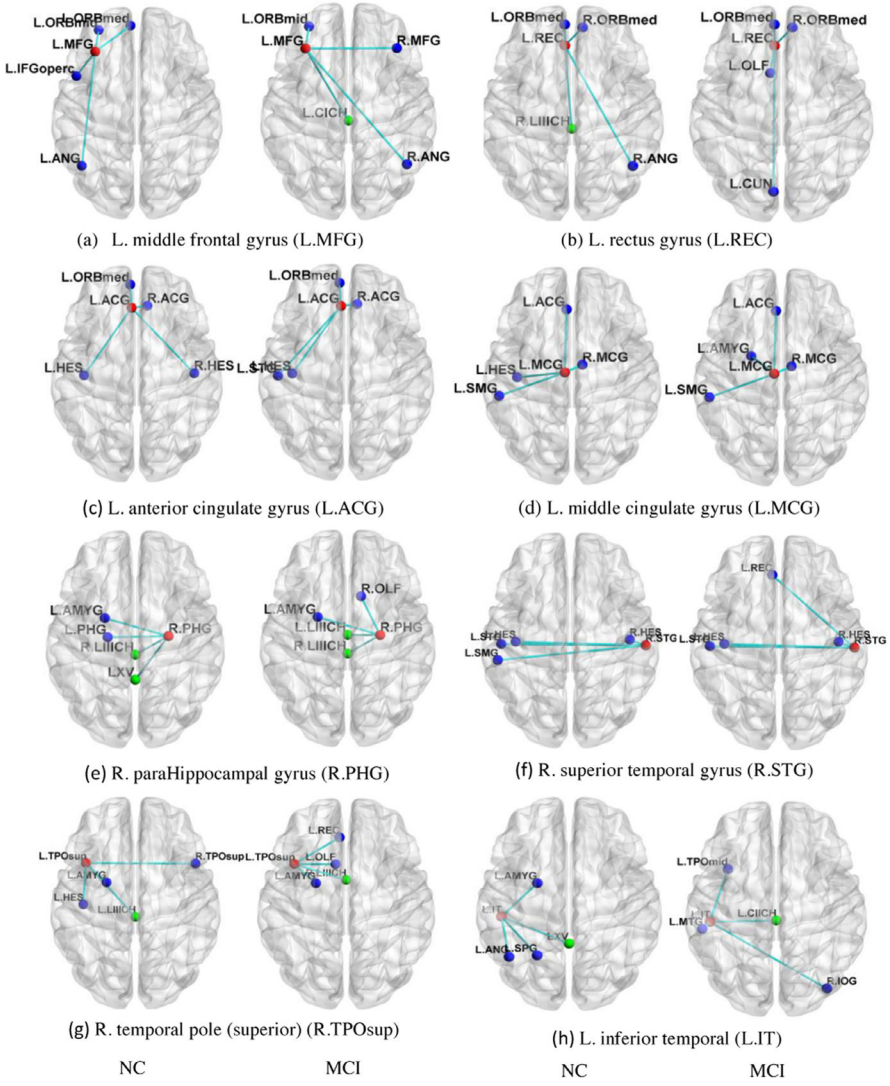
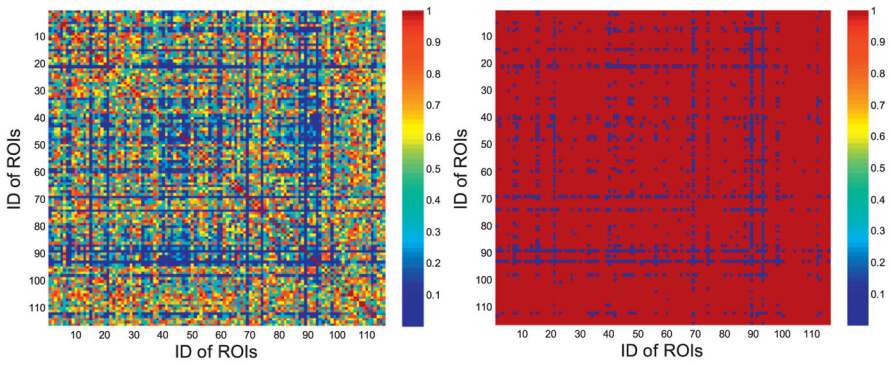


Fig. 5.

The average hyper-edges for NC (left) and MCI (right) groups based on 8 ROIs listed in Table 4 with λ=0.3. Here, each sub-figure denotes a hyper-edge constructed based on the corresponding ROI, where all nodes in each sub-figure form a hyper-edge, the red node (i.e., centroid node linked by other nodes) in each sub-figure represents the ROI used for constructing the hyper-edge, and the green nodes (i.e., nodes lying between left hemisphere and right hemisphere) represent the corresponding ROIs coming from cerebellum.



(a) The *p*-values on connection between all ROIs. (b) The thresholded *p*-values of connections between all ROIs.

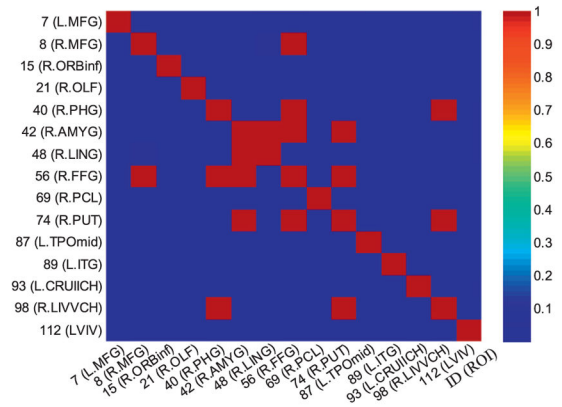
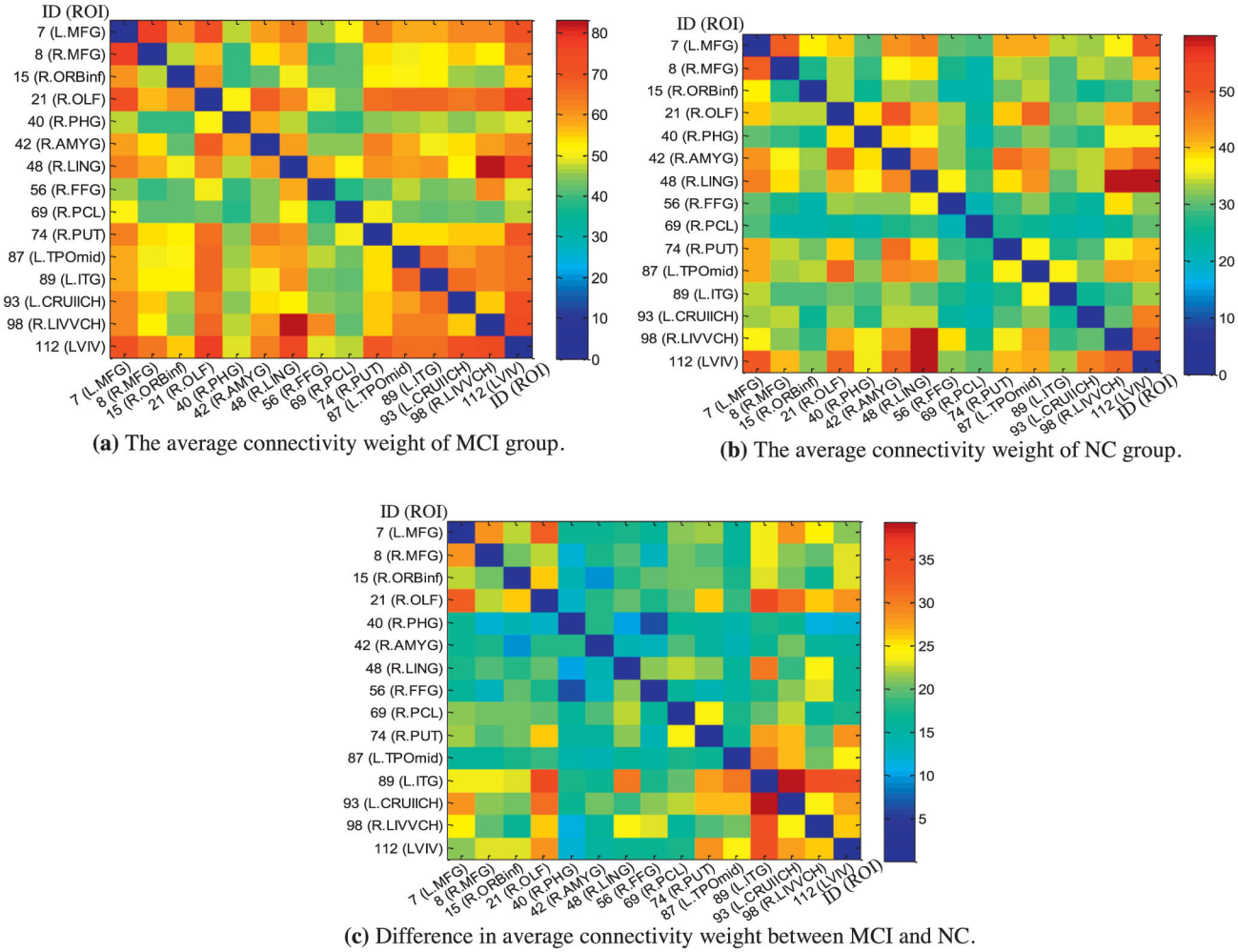


Fig. 6.

Visualization on p-values on connection between ROIs. Here, color denotes the corresponding *p*-value. (L.MFG = left Middle frontal gyrus, R.MFG = right Middle frontal gyrus, L.ORBinf = left Orbitofrontal cortex (inferior), L.OLF = left Olfactory, R.PHG = right ParaHippocampal gyrus, R.AMYG = right Amygdala, R.LING = right Lingual gyrus, R.FFG = right Fusiform gyrus, L.PCL = left Paracentral lobule, R.PUT = right Putamen, L.TPOmid = left Temporal pole (middle), L.ITG = left Inferior temporal, L.CIICH = Left crus II of cerebellar hemisphere, R.LIVVCH = Right lobule IV, V of cerebellar hemisphere, LVIV = Lobule VI of vermis.).



(a) The average connectivity weight of MCI group.

(b) The average connectivity weight of NC group.

(c) Difference in average connectivity weight between MCI and NC.

Fig. 7.

Average connectivity weights for MCI and NC groups and their differences. Colors in (a) and (b) represent the average connectivity weight of MCI and NC groups respectively, while colors in (c) represent difference of connectivity weights between MCI and NC.

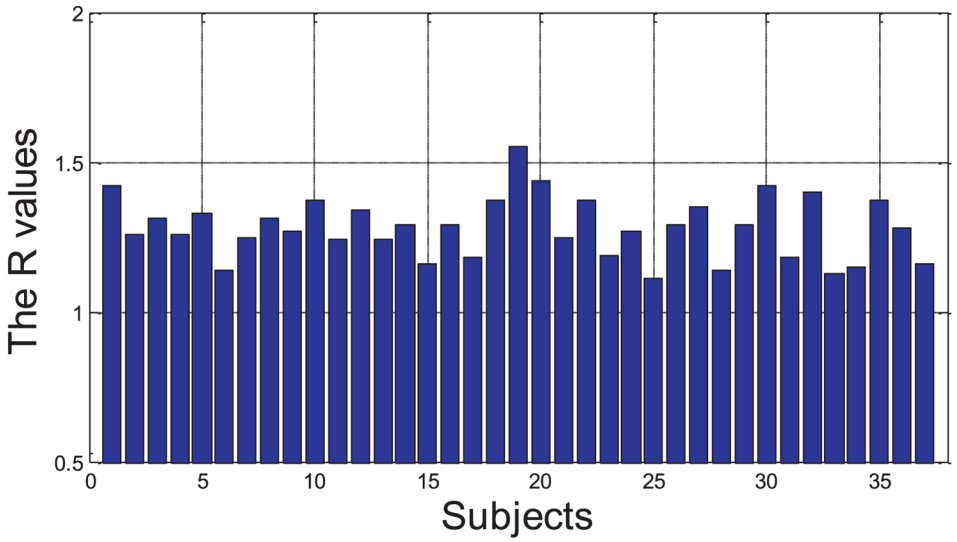


Fig. 8.
The R values for all subjects.

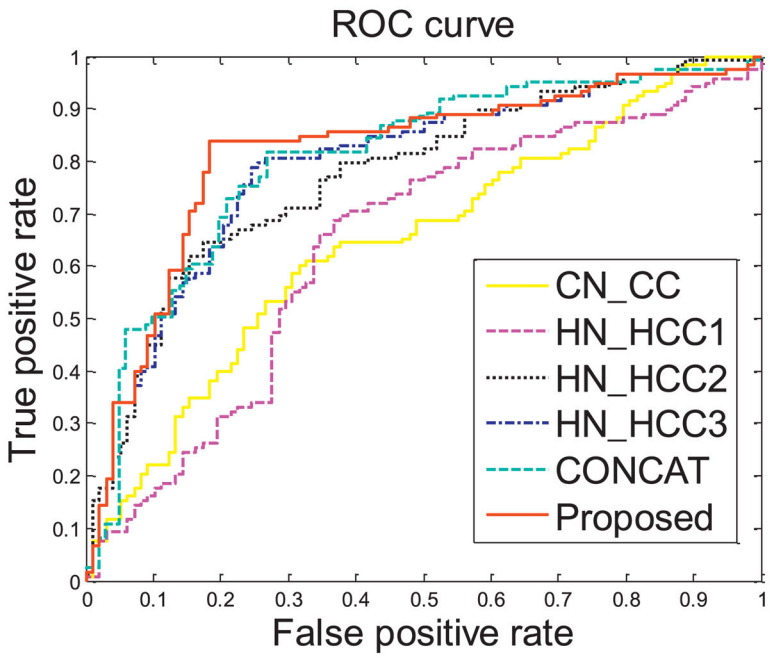


Fig. 9.
ROC curves of six different methods on ADHD classification.

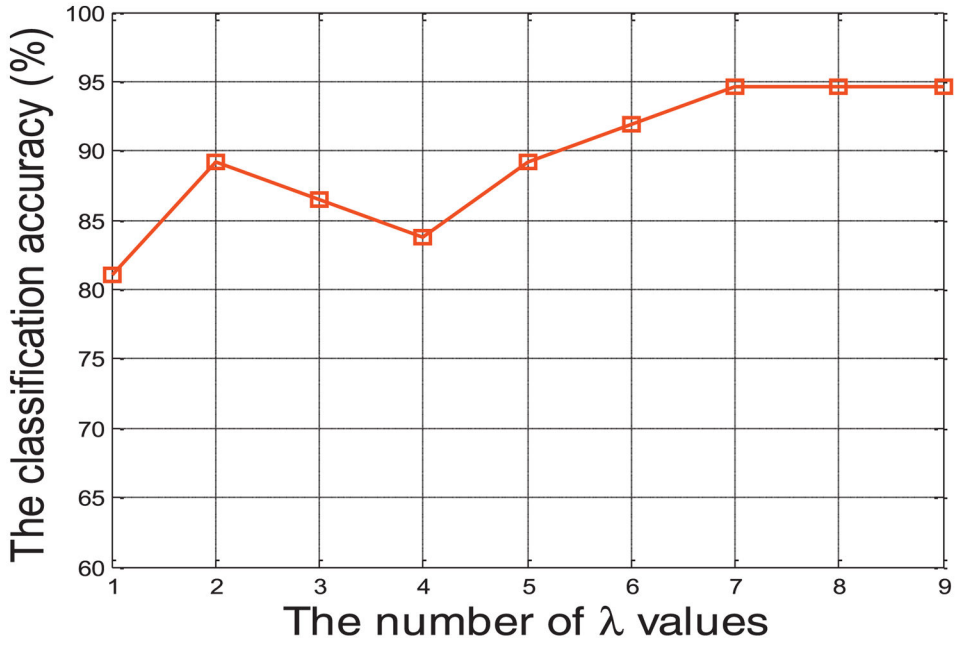


Fig. 10.
Classification accuracy w.r.t. the use of different number of λ values.

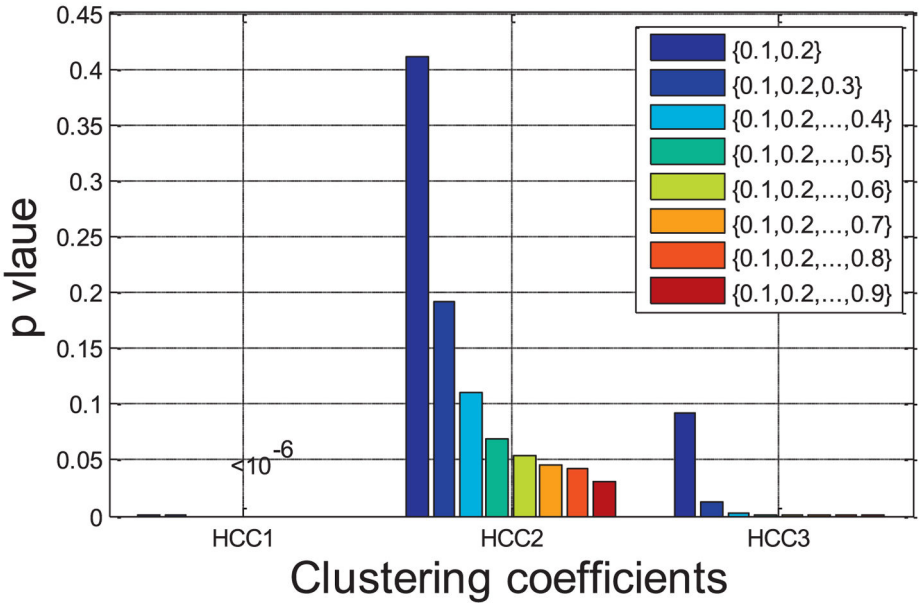


Fig. 11.

The p-value on three clustering coefficients from hyper-network with λ= 0.1 and other groups of λ values.

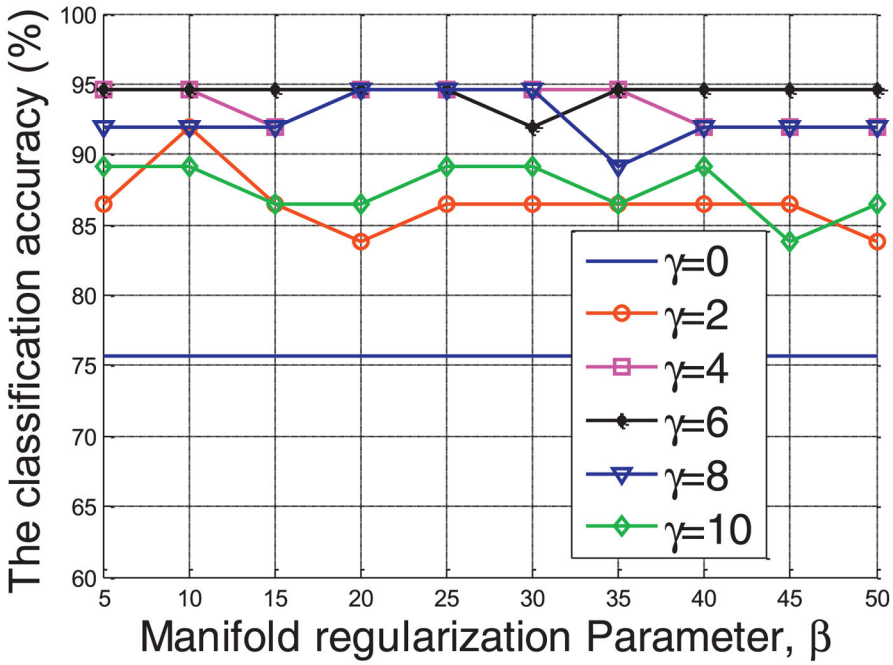


Fig. 12.
The classification accuracy w.r.t. the selections of β and γ values.

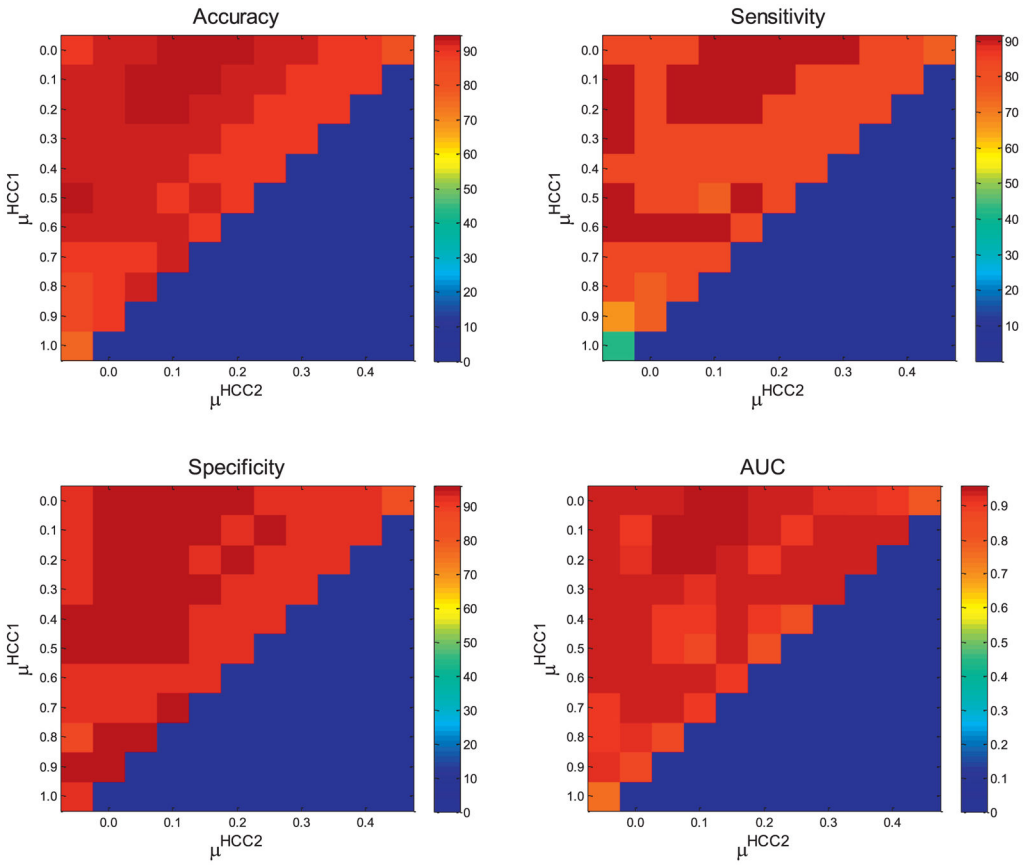


Fig. 13.

Classification results of proposed method when using different combinations of coefficient weights.

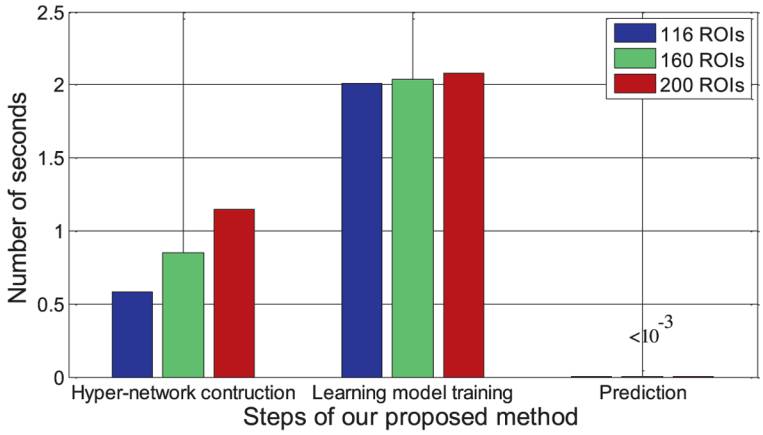


Fig. 14.

The computation cost of each step of our proposed method with different numbers of functional atlases.

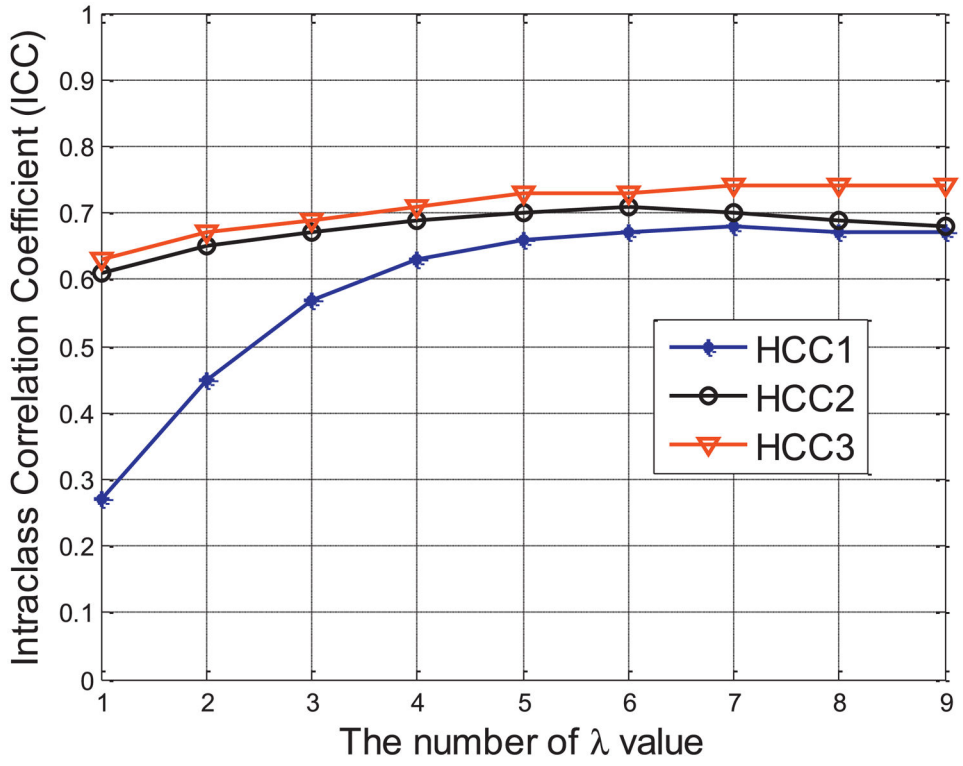


Fig. 15.

Changes of ICC values of three types of clustering coefficients w.r.t. the different number of λ values.

Table 1

Characteristics of the participants in this study.

Group	MCI	Normal
No. of subjects (male/female)	6/6	9/16
Age (mean ± SD)	75.0 ± 8.0	72.9 ± 7.9
Years of education (mean ± SD)	18.0 ± 4.1	15.8 ± 2.4
MMSE (mean ± SD)	28.5 ± 1.5	29.3 ± 1.1

MMSE: mini-mental state examination

Table 2

Classification performances of compared methods.

Method	Accuracy(%)	BAC(%)	Sensitivity(%)	Specificity(%)	AUC
CN-CC (*)	62.2	56.9	41.7	72.0	0.54
HN_HCC[1] (*)	75.7	66.9	41.7	92.0	0.75
HN_HCC[2]	81.1	79.5	75.0	84.0	0.80
HN_HCC[3]	89.2	87.7	83.3	92.0	0.93
CONCAT	91.9	91.8	91.7	92.0	0.94
Proposed	**94.6**	**93.9**	**91.7**	**96.0**	0.96

* indicates significant (i.e., p-value < 0.05) difference in terms of classification accuracy compared to the proposed method based on McNemar's test.

BAC: balanced accuracy; AUC: area under receiver operating characteristic (ROC) curve.

Table 3

Comparison on classification performance of state-of-the-art connectivity-network based methods.

Method	Subjects (MCI/NC)	Accuracy (%)	Sensitivity(%)	Specificity(%)	AUC
(Wang et al., 2013)	37/47	-	86.5	85.1	-
(Chen et al., 2011)	15/20	-	93	90	0.95
(Wee et al., 2013a)	12/25	91.9	-	-	0.90
(Wee et al., 2014)	25/25	84.0	84.0	84.0	0.87
(Jie et al., 2014b)	12/25	91.9	-	-	0.87
Proposed	12/25	**94.6**	**91.7**	**96.0**	**0.96**

Table 4

The important ROIs involved in classification.

ROI	*p*-value (HCC[1])	*p*-value (HCC[2])	*p*-value (HCC[3])	*p*-value (CC)
L. middle frontal gyrus	0.057	0.045	0.026	0.582
L. rectus gyrus	0.642	0.010	0.056	0.038
L. anterior cingulate gyrus	0.076	0.467	0.005	0.225
L. middle cingulate gyrus	0.869	0.521	0.044	0.283
R. paraHippocampal gyrus	0.018	0.078	0.014	0.380
R. middle occipital gyrus	0.025	0.001	0.010	0.454
L. caudate	0.002	0.002	0.001	0.925
L. putamen	0.128	0.055	0.017	0.409
R. superior temporal gyrus	0.056	0.037	0.439	0.645
R. temporal pole (superior)	0.022	0.011	0.002	0.909
L. inferior temporal	0.036	0.009	0.062	0.943
L. crus II of cerebellar hemisphere	0.295	0.033	0.071	0.911
R. lobule IV, V of cerebellar hemisphere	0.048	0.154	0.877	0.703

L. = left; R. = right.

Table 5

Performances on ADHD classification.

Method	Accuracy (%)	BAC (%)	Sensitivity (%)	Specificity (%)	AUC
CN-CC	61.1	61.1	61.0	61.2	0.64
HN_HCC[1]	65.7	65.7	66.1	65.3	0.64
HN_HCC[2]	69.0	68.8	71.2	66.3	0.77
CONCAT	77.5	77.0	81.8	72.3	0.80
Proposed	**82.9**	**82.8**	**83.9**	**81.6**	**0.82**

Table 6

Classification performance of different methods in three additional experiments.

Experiment	Accuracy		BAC		AUC	
	CN	Proposed	CN	Proposed	CN	Proposed
Exp1	54.1	**75.7**	46.5	**66.8**	0.51	**0.74**
Exp2	67.6	**83.8**	56.5	**79.3**	0.55	**0.88**
Exp3	59.5	**83.8**	54.8	**83.7**	0.65	**0.85**

BAC: balanced accuracy; AUC: area under receiver operating characteristic (ROC) curve.

Table 7

Classification performances of all methods with different functional atlases.

Method	Dosenbach atlas (160 ROIs)			Craddock atlas (200 ROIs)		
	Accuracy(%)	BAC(%)	AUC	Accuracy(%)	BAC(%)	AUC
CN-CC	73.0	69.2	0.68	59.5	46.2	0.63
HN_HCC[1]	75.7	64.7	0.60	81.1	75.2	0.72
HN_HCC[2]	89.2	89.8	0.92	83.8	83.7	0.82
HN_HCC[3]	91.9	89.7	0.96	81.1	77.3	0.82
CONCAT	83.8	88.0	0.91	83.8	85.8	**0.85**
Proposed	**97.3**	**95.8**	**0.98**	**89.2**	**87.7**	**0.85**