Hierarchical High-Order Functional Connectivity Networks and Selective Feature Fusion for MCI Classification

Xiaobo Chen[a], **Han Zhang**[a], **Dinggang Shen**[a,b,*], and **the Alzheimer's Disease Neuroimaging Initiative**[1]

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

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

Abstract

Conventional Functional connectivity (FC) analysis focuses on characterizing the correlation between two brain regions, whereas the *high-order FC* can model the correlation between two brain region pairs. To reduce the number of brain region pairs, clustering is applied to group all the brain region pairs into a small number of clusters. Then, a high-order FC network can be constructed based on the clustering result. By varying the number of clusters, multiple high-order FC networks can be generated and the one with the best overall performance can be finally selected. However, the important information contained in other networks may be simply discarded. To address this issue, in this paper, we propose to make full use of the information contained in all high-order FC networks. *First*, an agglomerative hierarchical clustering technique is applied such that the clustering result in one layer always depends on the previous layer, thus making the high-order FC networks in the two consecutive layers highly correlated. As a result, the features extracted from high-order FC network in each layer can be decomposed into two parts (blocks), i.e., one is redundant while the other might be informative or complementary, with respect to its previous layer. *Then*, a selective feature fusion method, which combines sequential forward selection and sparse regression, is developed to select a feature set from those informative

feature blocks for classification. Experimental results confirm that our novel method outperforms the best single high-order FC network in diagnosis of mild cognitive impairment (MCI) subjects.

Keywords

Functional connectivity; high-order network; hierarchical clustering; brain disease diagnosis

Introduction

Alzheimer's disease (AD) is the most prevalent dementia, accounting for about 60–80% of dementia cases among the elderly population worldwide. It has been reported that the incidence of AD will double every five years after the age of 65 (Bain, et al., 2008), and 1 in every 85 people will be affected by the year 2050 (Brookmeyer, et al., 2007). As a typical neurological disorder disease, AD is commonly characterized by some predominant clinical symptoms, including progressive memory loss and cognitive deficits, which severely interfere with patients' daily lives. This disease is incurable and worsens over time due to the degeneration of specific nerve cells, presence of neuritic plaques, and neurofibrillary tangles (McKhann, et al., 1984), eventually causing death. Thus, it is important for AD to be diagnosed as early and accurately as possible, so the effective pharmacological and behavioral treatments can be provided to potentially delay the progress of AD.

Some works (Johnson, et al., 2006; Whitwell, et al., 2007) suggest that the pathological manifestation of AD begins many years before it can be diagnosed. Mild cognitive impairment (MCI), as a prodromal stage of AD, has attracted much attention because MCI subjects tend to convert to clinical AD with an average conversion rate of 10% to 15% per year (Misra, et al., 2009), and more than 50% within 5 years (Gauthier, et al., 2006; Petersen, et al., 2001). Therefore, earlier diagnosis of MCI is very important for treatment and possible delaying the progression of MCI to moderate and severe stages. However, identifying MCI subjects from subjects undergoing normal aging is much more difficult because of the subtlety of involved cognitive impairment. Through analyzing neuroimaging data with machine learning algorithms (Mitchell, 1997), it is possible to identify subtle diagnostic biomarkers that can effectively distinguish MCI from normal controls (NC). Some successfully applied machine learning algorithms include support vector machine and multiple kernel learning (Cortes and Vapnik, 1995; Jie, et al., 2014b; Zhang, et al., 2011), multi-task and sparse learning (Friedman, et al., 2008; Huang, et al., 2010; Suk, et al., 2014a), and the recently emerged deep neural networks (Liu, et al., 2015; Suk, et al., 2015; Suk, et al., 2013; Suk, et al., 2014b) amongst many others.

Resting-state functional magnetic resonance imaging (RS-fMRI), a cutting-edge technology at disposal of cognitive neuroscience, can measure the blood oxygenation level dependent (BOLD) signal. This signal reflects low-frequency spontaneous fluctuations in the resting brain, which are related to intrinsic neural activity within the brain (Fox and Raichle, 2007). From RS-fMRI data, it is possible to infer functional connectivity (FC) between structurally separated brain regions. Here, the FC is defined as the temporal correlation of BOLD signals measured in different brain areas (Friston, et al., 1993). It has been proven (Greicius, 2008)

that FC is a useful tool for understanding the pathological underpinnings of MCI at the whole-brain level (Rombouts, et al., 2005; Sorg, et al., 2007; Wang, et al., 2007) and has great potential in providing a significant biomarker for diagnosis of MCI (Wee, et al., 2015; Wee, et al., 2012; Wee, et al., 2014). This is because the topological structure and strength of FC are disrupted due to the pathological attack of AD (Anderson and Cohen, 2013; dos Santos Siqueira, et al., 2014; Fekete, et al., 2013; Wang, et al., 2007). In most studies, FC is represented by a graph (Brier, et al., 2014; Toussaint, et al., 2014) in which a node corresponds to a brain region and an edge characterizes the FC strength between different brain regions. Therefore, constructing FC network based on RS-fMRI holds great promise for distinguishing MCI from normal aging (Stam, et al., 2009; Stam, et al., 2007). Most studies have focused on estimating FC from BOLD signals by using different statistical methods. For instance, Pearson's correlation (Jie, et al., 2014b; Wee, et al., 2012) can measure the pairwise similarity of BOLD signals. But this method handles each pair of brain regions independently, without taking into account the effects of other brain regions. Partial correlation characterizes the pairwise correlation and, at the same time, factors out the effects of other brain regions. Sparse representation (Yu, et al., 2016; Suk, et al., 2014c; Wee, et al., 2015; Wee, et al., 2014; Wright, et al., 2009) is able to approximate the BOLD signals of one brain region by linearly combining a smallest set of the signals from the rest brain regions, thus resulting in a sparse FC network.

Traditionally, the FC is estimated based on the entire length of BOLD time series (Jie, et al., 2014a; Jie, et al., 2014b; Wee, et al., 2014). It assumes the FC is stationary, by ignoring the complexity and dynamic property of the brain activities. To deal with this problem, a sliding window approach has been utilized in (Wee, et al., 2013; Wee, et al., 2015) to partition the entire BOLD signals (derived from each brain region) into several overlapping segments. For each segment, a FC network is constructed to characterize the time-varying FC between two brain regions during a certain time interval. Consequently, a set of FC networks can be generated and fused for MCI classification. Nevertheless, as in most traditional methods, the involved correlation (and thus the FC network) is still low-order in the sense that *1)* it is calculated based on the raw BOLD signals; *2)* it only describes how two brain regions are functionally interacted with each other; *3)* it cannot be directly used because of the possible phase mismatch across different subjects. In our previous study (Chen, et al., 2016), we regarded these low-order FC networks as a set of *FC time series*, each of which is associated with a specific pair of brain regions and characterizes the variation of their FC over time. Thus, by computing the correlation between two FC time series (involving up to four brain regions), a novel high-order FC network can be constructed for each subject. It should be emphasized that, different from the low-order case above, the correlation (and thus the FC network) obtained in such a manner is high-order because *1)* it is computed based on FC time series and is thus a high-level feature; *2)* it characterizes how different brain region pairs functionally interact with each other; *3)* it owns a merit of invariance, avoiding the adverse impact of the possible phase mismatch across different subjects. Moreover, the proposed high-order FC network is different from *either* static *or* dynamic FC. First, the high-order FC network uses time-varying FC obtained by the sliding window, which is different from the procedure of computing the static FC. Second, by calculating a second round of correlation, our method aggregates the time-varying FCs into a single high-order

functional network (*i.e.*, computing the correlation's correlation), thus very different from those traditional dynamic (low-order) FC computational approaches.

Note that the scale of high-order FC networks is much larger than that of low-order FC networks, because the quantity of brain region pairs is much larger than that of brain regions (e.g., 116 brain regions yield about 116^2 brain region pairs). As a result, the feature vector extracted from the high-order FC networks is of high-dimensionality, which increases the model complexity as well as the risk of over-fitting. To overcome these problems, a clustering algorithm (Chen, et al., 2014; Ward Jr, 1963) is first used to group the FC time series (associated with each pair of brain regions) into U different clusters, where U is a predefined number of clusters. Then, the mean of the FC time series within each cluster is used to compute high-order FC and further construct high-order FC networks. This strategy succeeds in reducing the scale of high-order FC networks as well as the number of features, thus improving computation efficiency and alleviating over-fitting. However, we found that the selected value of U will severely influence the discriminative ability of the resulting high-order FC networks in diagnosis (Chen, et al., 2016), since different values of U lead to different high-order FC networks. To find an optimal high-order FC network, a candidate set of high-order FC networks with different discriminative ability is generated by varying U (Chen, et al., 2016). Then, each high-order FC network is evaluated independently on the validation data and the one with the best performance is finally selected, while others are simply discarded. Although this method achieves promising results, there is still room for improvement.

In order to boost the diagnosis performance of high-order FC networks, we hypothesize in this paper that those discarded high-order FC networks in (Chen, et al., 2016), although inferior to the optimal network in overall discriminative performance, may contain some essential and complementary information for classification. In this sense, choosing a single optimal high-order FC network while discarding others, may cause the loss of useful information for diagnosis. Therefore, inspired by the above insight, we propose a novel hierarchical high-order FC network and feature fusion framework to take advantage of all high-order FC networks and further improve the diagnosis performance. Notably, by selection and integration of the information contained in the candidate high-order FC networks in an appropriate way, this method takes full advantage of all the available discriminative information and also avoids over-fitting. This study is featured by the two following aspects:

First, in order to exploit multiple high-order FC networks while avoiding large quantity of features, we put forward to construct these high-order FC networks in a hierarchical way, where the network in one layer is closely dependent on the network in the previous layer. In view of this inherent relationship between two consecutive layers, the features extracted from the network in each layer can be decomposed into two feature blocks, depending on the correlation with the features extracted from the network in the previous layer. Those highly-correlated features with respect to the previous layer are redundant and thus can be eliminated before feature fusion, without losing vital information. In contrast, those less-correlated features may contain complementary information and should be taken into account in the next stage (i.e., feature fusion). By constructing the hierarchical high-order

FC networks, we can *not only* produce useful and complementary features, *but also* reduce the redundancy significantly.

Second, considering the layer-wise structure of features, we propose a novel feature fusion method to utilize the information contained in all candidate high-order FC networks. This method elaborately combines the sequential forward selection (Jain and Zongker, 1997) and sparse regression (Tibshirani, 1996) under the classic wrapper-based feature selection framework (Kohavi and John, 1997). First, sequential forward selection operates on feature block from each layer and adds one block each time. Then, sparse regression operates on the combined feature blocks and attempts to select only a small number of individual features. Finally, support vector machine (Chen, et al., 2011a; Chen, et al., 2011b; Cortes and Vapnik, 1995) with simple linear kernel is trained based on the selected features, and the classification performance on validation data is employed to guide the selection procedure above. This feature fusion method further removes redundant and uncorrelated features, with respect to the classification, thus mitigating the influence of over-fitting.

It is worth noting the existence of the previous work (Zhou, et al., 2011) on modelling the hierarchical structure of brain anatomical network. However, there are two main differences from our proposed work. First, their brain anatomical network focuses on multi-scale structural relationship generated based on T1-weighted MRI data, while our high-order FC networks model multi-scale functional relationship derived from RS-fMRI data. Second, their brain anatomical network is still a low-order network since it considers only the direct pairwise interaction between brain regions. Our proposed hierarchical high-order FC network reveals high-order correlations, reflecting how different *pairs* of brain regions (involving at least four brain regions) interact with each other.

Moreover, compared to the previous work (Chen, et al., 2016), this study can be viewed as an important extension with the following novelties: *1)* a hierarchical strategy is developed to generate multiple related high-order FC networks for containing as more information as possible; *2)* due to the use of hierarchical structure, a feature decomposition technique is further developed based on correlation analysis of hierarchical high-order FC networks from consecutive layers in order to reduce feature redundancy; *3)* due to the complexity of those obtained features, a feature fusion method (by integration of sequential forward selection and sparse regression) is developed to select a few discriminative features for classification.

Data and Methodology

Data preparation

The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset is used in this study. ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and nonprofit organizations. The goal of ADNI is to validate the use of various biomarkers, including MRI, PET imaging and related neuropsychological assessments for AD clinical trials and diagnosis.

In this work, 50 MCI subjects and 49 normal controls (NCs) are randomly selected from ADNIGo and ADNI2 dataset. (For more details about imaging parameters, please see the ADNI protocols at adni.loni.ucla.edu.) The patient group consists of 27 early MCI subjects and 23 late MCI subjects. Subjects from both groups were age-matched and scanned using 3.0T Philips scanner. SPM8 software package (http://www.fil.ion.ucl.uk/spm/software/spm8) was used to preprocess the acquired RS-fMRI data. The scanning time for each subject is 7 min (*i.e.*, 140 volumes), and the subjects with more than 2.5 min of large frame-wise displacement (FD > 0.5) were not included in this study (*i.e.*, excluded from data inclusion). For magnetization equilibrium, the first 3 volumes of each subject were discarded before preprocessing, and then the remaining 137 volumes were used for subsequent analysis. A rigid-body transformation was used to correct head motion during the scan; and the subject with head motion larger than 2 mm or 2° were not included in this study. The fMRI images were registered to the Montreal Neurological Institute (MNI) space and spatially smoothed using a Gaussian kernel with full width at half maximum (FWHM) = $6 \times 6 \times 6$ mm^3. We did not perform data scrubbing (*i.e.*, removing volumes with FD > 0.5), as it would introduce additional artifacts to the subsequent dynamic FC analysis. The RS-fMRI images were parcellated into 116 regions according to the Automated Anatomical Labeling (AAL) template. The mean RS-fMRI time series of each brain region was band-pass filtered (0.015–0.15 Hz). Head motion parameters (*Friston24*), mean BOLD signal of white matter, and mean BOLD signal of cerebrospinal fluid were all regressed out from the RS-fMRI data to further reduce artifacts.

Framework

The main steps of hierarchical high-order FC networks construction are shown in Fig. 1 where four brain regions are denoted by A, B, C, and D. Generally speaking, this method consists of the following steps: *1)* A sliding window with length *N* and step size *s* is applied to partition the entire BOLD signal into multiple overlapping segments, each of which characterizes the neural activity of brain region in a relatively small time interval. *2)* For each subject, a set of low-order FC networks is constructed, each of which is based on a BOLD signal segment. By doing so, we actually obtain a set of *FC time series*, each describing the temporal variation of correlation between two brain regions. *3)* All subjects' FC time series associated with the same brain region pair are concatenated together to form a long FC time series which is represented by a point in high-dimensional space. *4)* In the high-dimensional space, the long FC time series from all brain region pairs are grouped into *U* clusters by the clustering algorithm, thus, yielding consistent clustering results across different subjects. *5)* For each subject, the mean of the FC time series within the same cluster is computed and then a high-order FC network (HON) is constructed based on the correlation between the mean FC time series of different clusters. *6)* High-order feature vector (e.g., weighted local clustering coefficients in this paper) is extracted from the constructed high-order network. *7)* Repeating steps 4–6 multiple times with different *U*s generates multiple high-order FC networks, each of which characterizes the high-order correlation at different scales, and also multiple high-order feature vectors for each subject. *8)* The feature vectors extracted from all high-order FC networks are analyzed based on correlation and then a feature subset is selected by a feature selection that combines the

sequential forward selection and sparse regression. *9)* Support vector machine (SVM) is trained with the selected features to classify MCI and NC subjects.

We can observe that the resulting high-order FC network from Steps 4–5 closely depends on the number of clusters *U*. By varying *U*, we can obtain a candidate set of high-order FC networks, from which their corresponding features can be extracted. Then, we can ensemble these networks in a principled way to make full use of all available features. On one hand, more networks will produce more features, which may bring benefits to MCI diagnosis. On the other hand, more networks also make the relationship among the features more complex and difficult to analyze. Also, due to the limited sample size in practice, this will probably cause the curse of dimensionality and over-fitting. As a result, taking advantage of the information brought by all high-order FC networks and also avoiding over-fitting become a crucial problem for the multiple high-order FC networks ensemble.

To deal with the problems above, we propose hierarchical high-order FC networks, as well as an effective feature fusion method, for MCI classification. *First*, the agglomerative hierarchical clustering algorithm is used to produce multiple clustering results with different *Us*. Then, each clustering result is utilized to generate a high-order FC network in the corresponding layer. *Second*, features are extracted from all high-order FC networks, and the correlation analysis is performed to reveal their inherent relationship, based on which features in each layer can be decomposed into the redundant part and the informative part, respectively. *Third*, a sequential forward selection procedure is employed to combine the informative features from different high-order FC networks, followed by a sparse regression to select the individual features beneficial for classification. The detailed procedure is explained as follows.

Hierarchical clustering and feature decomposition

Suppose the FC time series in high-dimensional space are progressively grouped into u_i and u_{i+1} clusters, respectively, in the layer i and layer $i + 1$, where $u_i > u_{i+1}$. Due to the nature of agglomerative hierarchical clustering, some clusters in the layer i, which are highly similar to each other, will be merged to form new clusters in the layer $i + 1$, while the remaining clusters are retained without change. An illustration of the agglomerative hierarchical clustering procedure is shown in the panel of Step 4 and its upper panel in Fig. 1. Notice that the FC time series in Fig. 1 are generated by the sliding window approach in Step 1. Due to the hierarchical clustering in Fig. 1, we can observe that, from the layer i to the layer $i + 1$, the clusters in the blue and purple ellipses are merged to form new clusters as shown in red ellipse, while all other clusters are retained, thus reducing the total number of clusters in the layer $i + 1$. Based on the clustering results in the layer i and the layer $i + 1$, the high-order FC networks HON_i and HON_{i+1} are constructed, respectively, where one vertex corresponds to one cluster in the corresponding layer, as shown by the high-order FC network in the panel of Step 5 and its upper panel in Fig. 1. Subsequently, the feature vectors $Fea_i \in R^{u_i}$ and $Fea_{i+1} \in R^{u_{i+1}}$ (i.e., weighted local clustering coefficients (Rubinov and Sporns, 2010; Watts and Strogatz, 1998) in this paper) can be extracted from the high-order FC networks HON_i and HON_{i+1}, respectively. For this type of features, each entry in Fea_i and Fea_{i+1} corresponds to a vertex in HON_i and HON_{i+1}, respectively, thus also corresponding to a

cluster in the layer i and the layer $i + 1$. Due to the high overlapping between the clustering results in the layer $i + 1$ and the layer i, Fea_{i+1} can be decomposed as $Fea_{i+1} = [D_{i+1}, S_{i+1}]$, where D and S respectively indicate the corresponding clusters that are newly formed from and already exist in the previous layer. As a result, S_{i+1} may be highly correlated with some features in Fea_i. This implies that only D_{i+1} in the layer $i + 1$ may contain useful information, whereas S_{i+1} is redundant with respect to the previous layer. This observation can be generalized to the case of multiple layers. Suppose hierarchical high-order FC networks are generated from the layer 1 to the layer L. Following the above analysis, the feature vector Fea_i extracted from the layer i can be decomposed into two parts $Fea_i = [D_i, S_i]$ by comparing it with the previous layer. Note that, to decompose the features in the layer 1, an extra layer 0 is used. For the layer i (2), only D_i may contain useful information while S_i is redundant with respect to the previous layer. But for the layer 1, both S_1 and D_1 should be reserved to guarantee the completeness of information contained in all networks. Therefore, for each subject, the features extracted from all hierarchical high-order FC networks can be condensed and expressed as $Fea = [D_0, D_1, D_2, …, D_L]$ where $D_0 = S_1$. In this paper, each D_i is called a feature block since it is associated with a specific layer.

Selective feature fusion

Note that although applying the above agglomerative hierarchical clustering and correlation analysis can reduce the dimensionality of features to a large extent, the redundancy between different layers may still exist, especially when taking into account multiple layers. In addition, not all of the features in Fea are discriminative in terms of MCI classification. To maximally benefit from the information contained in Fea and also further reduce redundancy simultaneously, we propose a feature fusion method, which combines sequential forward selection and sparse regression (Liu, et al., 2009), under the framework of wrapper-based feature selection (Kohavi and John, 1997). Here, sequential forward selection can find feature block progressively, while sparse regression can select individual features that are predictive for classification. Specifically, given a current set A of feature blocks, a new feature block D_i from $Fea - A$ can be selected and combined with A. Then, sparse regression (Zhang et al., 2016, Zhang et al., 2017) is performed on all training samples with this enlarged feature subset to find a small subset C, which is beneficial for classification. Next, the selected features of all training subjects are used to train a linear SVM model, and the classification accuracy on the validation subjects is used to guide the selection of D_i, which means that the one yielding the optimal accuracy is finally selected. The procedure above is repeated until a required number of feature blocks is reached. This procedure is illustrated in the following Table 1.

Evaluation protocol

In this study, because of limited samples, the leave-one-out cross validation (LOOCV) is adopted to evaluate the performance of different methods. Specifically, given a total of N subjects, $N - 1$ subjects are used for training a SVM classifier and the rest one is used for testing. The classification results on the testing subject are recorded. The above procedure is repeated N times, each time leaving out a different subject for testing. Finally, the average of the N evaluation results is computed to compare the generalization performance of different methods. For the hyper-parameter in each method, we tune its value on the training subjects

by using the nested LOOCV. That is, for each fold of the above LOOCV, we have $N-1$ training subjects. Then, for the $N-1$ training subjects, one is left out for testing and the remaining $N-2$ subjects are for training, using each specific hyper-parameter. This procedure will repeat $N-1$ times, each time leaving out a different subject from the $N-1$ training subjects. Then, the average of $N-1$ evaluation results is computed and the hyper-parameter that gives rise to the best accuracy is finally selected for this fold.

Experimental Analysis and Results

Feature correlation analysis and feature decomposition

In this study, all methods were implemented in MATLAB 2012b environment. To generate high-order FC networks, we use a sliding window with $s=1$ and $N=50$. To generate multiple layers, we start from one layer with a relatively large number of clusters ($U=220$), for retaining sufficient information. Then, the sequent layers are added by gradually reducing U by 30 until the optimal performance is achieved. In such a way, it eventually generates 4 high-order FC networks from the layer 1 to the layer 4: HON_1, HON_2, HON_3, and HON_4, where the number of clusters U equals 220, 190, 160, and 130, respectively. The feature vectors (*i.e.*, weighted local clustering coefficients) $Fea_1 \in R^{220}, Fea_2 \in R^{190}, Fea_3 \in R^{160}$, and $Fea_4 \in R^{130}$ are extracted from HON_1, HON_2, HON_3, and HON_4, respectively. Since our method is a feature fusion method, the correlation between features from different high-order FC networks provides important prior information. To verify the analysis in Section 2.3, we show the correlation between Fea_i and Fea_{i+1} ($i=1,2,3$) in Fig. 2. As shown by the red lines in Fig. 2, most features in Fea_{i+1} are actually highly correlated with features in Fea_i, implying that most features in the current layer are redundant *with respect to* those in the previous layer and thus should be eliminated before feature fusion. Based on this correlation with the previous layer, each feature vector Fea_i ($i=1,2,3,4$) can be decomposed into two feature blocks, as shown in Fig. 3. As we can see, only about 30 features of each layer are less correlated with the previous layer. To guarantee the inclusion of sufficient information and also reduce the redundancy, five feature blocks $S_1 \in R^{191}$, $D_1 \in R^{29}$, $D_2 \in R^{30}$, $D_3 \in R^{29}$, and $D_4 \in R^{30}$ are engaged in the subsequent feature fusion, while others are eliminated. In short, the total number of features decreases from 700 to 309 by this unsupervised correlation analysis.

Classification accuracy

For sparse regression, the SLEP toolbox (Liu, et al., 2009) is utilized. The hyper-parameter involved in sparse regression is determined by the nested LOOCV described in Section 2.5. To construct SVM classifier, the well-known LIBSVM toolbox (Chang and Lin, 2001) is applied with default hyper-parameter. The proposed sequential forward selection and sparse regression based *hierarchical high-order FC networks* feature fusion (HHON-SFS) is compared with some closely related methods, including 1) a simple feature fusion method (HHON-CON), which directly concatenates all features extracted from four high-order FC networks, 2) four individual high-order FC networks (HON_1, HON_2, HON_3, and HON_4), 3) two low-order FC networks based on partial correlation (LON-PAC) and Pearson's correlation (LON-PEC), and 4) sparse representation (SR) and weighted SR (WSR) based FC networks (Yu, et al., 2016), respectively. To measure the performance of different

methods, we use the following 7 indices: accuracy (ACC), area under ROC curve (AUC), sensitivity (SEN), specificity (SPE), Youden's Index (YI), F_score, and balanced accuracy (BAC). These statistical measures are defined as (Sokolova, et al., 2006)

$$ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

$$SEN = \frac{TP}{TP+FN}$$

$$SPE = \frac{TN}{TN+FP}$$

$$YI = SEN + SPE - 1$$

$$F\_soore = 2 \times \frac{precision \times recall}{precision + recall}$$

$$BAC = \frac{1}{2} \times (SEN + SPE)$$

where $precision = \frac{TP}{TP+FP}$, $recall = \frac{TP}{TP+FN}$, and TP, TN, FP, and FN denote the true positive, true negative, false positive and false negative, respectively.

The experimental results are shown in Table 2. As we can see, the high-order FC networks achieve better accuracy than the two low-order FC networks. This is consistent with our previous research (Chen, et al., 2016), indicating that high-order FC networks provide more discriminative biomarkers for MCI identification. Two sparse representation based networks, *i.e.*, SR and WSR, remarkably improve correlation based networks; however, they are still inferior to the proposed hierarchical high-order FC networks. Comparing the four individual high-order FC networks, we can observe its sensitive performance to the number of clusters U. For example, too large or too small U will adversely affect the performance. This can be understood as follows: large U will cause too many redundant features, while small U will lead to significant information loss. We observe that the high-order FC network HON_2 ($U = 190$) achieves better performance than other individual high-order FC networks (HON_1, HON_3, and HON_4). Nevertheless, it does not mean the information contained in other suboptimal networks is completely useless, as they can also distinguish MCI and NC subjects to a certain extent, although their overall accuracies are lower than HON_2. As an attempt to make use of all information, HHON-CON directly concatenates Fea_1, Fea_2, Fea_3,

and Fea_4 together to form a combined feature vector of length 700. Although this method involves all features, the accuracy falls just between the best one and the worst one, as indicated by Table 2. This may be due to too many redundant features that make the relationship between features more complex, thus causing difficulty in individual feature selection and also potential over-fitting. In contrast, the proposed feature fusion method, HHON-SFS, achieves the best performance among all competitive methods. On one hand, we attribute this improvement to the feature correlation analysis and also the resulting feature decomposition, which eliminates many redundant features. On the other hand, the combination of sequential forward selection and sparse regression makes it possible to evaluate the importance of feature blocks and individual features progressively. As a result, the crucial and complementary features have more probability to be selected and fused for classification.

Sequential forward selection and individual feature selection

For the sequential forward selection procedure, we vary the number of feature blocks from one to five and show the corresponding classification accuracies in Fig. 4(a). We can see that two feature blocks yield preferable performance. Moreover, it is necessary to choose some feature blocks according to the specific training subjects, rather than simply using all feature blocks. The proposed method can automatically determine which feature blocks should be selected based on the training subjects. Due to the use of LOOCV, different training subjects may cause differences in selection results of feature blocks. If one feature block is selected due to the nested LOOCV on the training subjects, but no individual feature is selected from this block based on the whole training subjects by sparse regression, this feature block will be eventually discarded. Therefore, we show the total number of occurrences of each feature block in Fig. 4(b). As we can see, S_1 is selected in each fold of LOOCV, which means it contains many features beneficial for disease classification. In fact, we can observe from Fig. 3 that, without S_1, the rest feature blocks cannot form a complete high-order FC network and much information will be lost. S_1, D_2 and D_4 are also selected many times, while D_2 is never chosen, which indicates that the information contained in D_2 cannot bring remarkable guidance for classification. For the individual feature selection, the experimental result is shown in Fig. 4(c). This is consistent with Fig. 4(b), which further confirms that most features are located in S_1, but D_2 and D_4 also provide some complementary features for improving classification. As a whole, we can see from these results that, despite the overall performances of individual high-order FC networks are different, they do contain complementary information and thus integrating them can boost the overall performance.

Hierarchical high-order FC networks

The averaged low-order FC networks built by LON-PEC and LON-PAC on MCI and NC subjects are shown in Figs. 5–6, respectively. Similarly, the illustrations of averaged sparse representation and weighted sparse representation based FC networks are shown in Figs. 7–8, respectively. As we can see, the difference between MCI and NC is imperceptible for low-order FC networks, which implies the inferior classification performance. In contrast, the high-order FC networks in the four layers, as shown in Figs. 9–12, have more remarkable differences. Specifically, the number of strong high-order correlations in MCI subjects is generally more than that in NC subjects, which, to a certain sense, reflects the disorder

characteristic of neural degenerative disease. This difference also explains why the high-order FC networks can better distinguish MCI and NC subjects than the traditional low-order FC networks. This observation is also consistent with our previous research. Moreover, due to the agglomerative hierarchical clustering, the high-order FC networks actually have much more overlap between two consecutive layers.

Limitations

One important issue of the current study is related to the limited number of subjects used, which may lead to less statistical power. In addition, the leave-one-out cross validation approach was used to evaluate the classification performance of different approaches, which often leads to optimistic performance. This issue can be better overcome by other approaches such as bootstrapping. In the future, our proposed method will be further evaluated by the bootstrapping technique, along larger dataset. Besides, the number of clusters could be also determined by the bootstrapping technique.

Another limitation is related to the use of Pearson's correlation which requires the samples to be independent to each other, although it has been used extensively in RS-fMRI and brain network studies. In our case, the dynamic FC time series generated by using the sliding window might violate this requirement, since neighboring windows are highly overlapped. Thus, the estimated correlation coefficient could be biased from the "true" correlation and less robust. But, in this study, we just simply calculate a "second level" of correlation based on such dynamic FC time series because of its simplicity and popularity in brain network construction (*e.g.*, most of the previous FC analysis studies computed temporal correlation using the RS-fMRI signals of each pair of brain regions, without considering the sample dependency problem). In the future work, how to robustly measure high-order relationship between dynamic FC time series using other statistical techniques will be investigated.

Conclusion

In this paper, we propose integrating the information contained in multiple high-order FC networks for MCI classification. Hierarchical clustering technique is utilized to generate multiple high-order FC networks, each of which is located in one layer. Due to such hierarchy structure, the features extracted from the network in each layer can be simplified, and only the informative feature block is taken into account. By combining sequential forward selection and sparse regression, a novel feature fusion method is further developed. This method is able to selectively fuse informative feature blocks from different layers and further determine a small number of distinctive features for early diagnosis. Finally, support vector machine with simple linear kernel is used for MCI classification. The experimental results demonstrate the capability of the proposed approach in making full use of information contained in multiple high-order FC networks. It also shows that the appropriate combination of multiple high-order FC networks can yield much better classification performance than any single optimal high-order FC network.

Acknowledgments

References

Anderson A, Cohen MS. Decreased small-world functional network connectivity and clustering across resting state networks in schizophrenia: an fMRI classification tutorial. Frontiers in human neuroscience. 2013; 7:520.doi: 10.3389/fnhum [PubMed: 24032010]

Bain LJ, Jedrziewski K, Morrison-Bogorad M, Albert M, Cotman C, Hendrie H, Trojanowski JQ. Healthy brain aging: A meeting report from the Sylvan M. Cohen Annual Retreat of the University of Pennsylvania Institute on Aging. Alzheimer's & dementia: the journal of the Alzheimer's Association. 2008; 4:443.

Brier MR, Thomas JB, Fagan AM, Hassenstab J, Holtzman DM, Benzinger TL, Morris JC, Ances BM. Functional connectivity and graph theory in preclinical Alzheimer's disease. Neurobiology of aging. 2014; 35:757–768. [PubMed: 24216223]

Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimer's & dementia. 2007; 3:186–191.

Chang C, Lin C. LIBSVM: a library for support vector machines. Citeseer. 2001

Chen X, Xiao Y, Cai Y, Chen L. Structural max-margin discriminant analysis for feature extraction. Knowledge-Based Systems. 2014; 70:154–166.

Chen X, Yang J, Ye Q, Liang J. Recursive projection twin support vector machine via within-class variance minimization. Pattern Recognition. 2011a; 44(10–11):2643–2655.

Chen X, Yang J, Liang J. Optimal Locality Regularized Least Squares Support Vector Machine via Alternating Optimization. Neural Processing Letters. 2011b; 33:301–315.

Chen X, Zhang H, Gao Y, Wee CY, Li G, Shen D. High-Order Resting-State Functional Connectivity Network for MCI Classification. Human brain mapping. 2016; 37(9):3282–96. [PubMed: 27144538]

Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995; 20:273–297.

dos Santos Siqueira A, Biazoli Junior CE, Comfort WE, Rohde LA, Sato JR. Abnormal functional resting-state networks in ADHD: graph theory and pattern recognition analysis of fMRI data. BioMed research international. 2014

Fekete T, Wilf M, Rubin D, Edelman S, Malach R, Mujica-Parodi LR. Combining classification with fMRI-derived complex network measures for potential neurodiagnostics. PLoS one. 2013; 8(5):e62867. [PubMed: 23671641]

Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Reviews Neuroscience. 2007; 8:700–711. [PubMed: 17704812]

Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008; 9:432–441. [PubMed: 18079126]

Friston K, Frith C, Liddle P, Frackowiak R. Functional connectivity: the principal-component analysis of large (PET) data sets. Journal of cerebral blood flow and metabolism. 1993; 13:5–5. [PubMed: 8417010]

Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H. Mild cognitive impairment. The Lancet. 2006; 367:1262–1270.

Greicius M. Resting-state functional connectivity in neuropsychiatric disorders. Current opinion in neurology. 2008; 21:424–430. [PubMed: 18607202]

Huang S, Li J, Sun L, Ye J, Fleisher A, Wu T, Chen K, Reiman E, Initiative A.s.D.N. Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. NeuroImage. 2010; 50:935–949. [PubMed: 20079441]

Jain A, Zongker D. Feature selection: Evaluation, application, and small sample performance. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1997; 19:153–158.

Jie, B., Shen, D., Zhang, D. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014. Springer; 2014a. Brain Connectivity Hyper-Network for MCI Classification; p. 724-732.

Jie B, Zhang D, Gao W, Wang Q, Wee CY, Shen D. Integration of network topological and connectivity properties for neuroimaging classification. Biomedical Engineering, IEEE Transactions on. 2014b; 61:576–589.

Johnson S, Schmitz T, Moritz C, Meyerand M, Rowley H, Alexander A, Hansen K, Gleason C, Carlsson C, Ries M. Activation of brain regions vulnerable to Alzheimer's disease: the effect of mild cognitive impairment. Neurobiology of aging. 2006; 27:1604–1612. [PubMed: 16226349]

Kohavi R, John GH. Wrappers for feature subset selection. Artificial intelligence. 1997; 97:273–324.

Liu J, Ji S, Ye J. SLEP: Sparse learning with efficient projections. Arizona State University. 2009; 6:491.

Liu S, Liu S, Cai W, Che H, Pujol S, Kikinis R, Feng D, Fulham MJ. Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease. Biomedical Engineering, IEEE Transactions on. 2015; 62:1132–1140.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease Report of the NINCDS - ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984; 34:939–939. [PubMed: 6610841]

Misra C, Fan Y, Davatzikos C. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage. 2009; 44:1415–1421. [PubMed: 19027862]

Mitchell, TM. Machine learning. McGraw-Hill; New York: 1997.

Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, Ritchie K, Rossor M, Thal L, Winblad B. Current concepts in mild cognitive impairment. Archives of neurology. 2001; 58:1985–1992. [PubMed: 11735772]

Rombouts SA, Barkhof F, Goekoop R, Stam CJ, Scheltens P. Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study. Human brain mapping. 2005; 26:231–239. [PubMed: 15954139]

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

Sokolova, M., Japkowicz, N., Szpakowicz, S. AI 2006: Advances in Artificial Intelligence. Springer; 2006. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation; p. 1015-1021.

Sorg C, Riedl V, Mühlau M, Calhoun VD, Eichele T, Läer L, Drzezga A, Förstl H, Kurz A, Zimmer C. Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. Proceedings of the National Academy of Sciences. 2007; 104:18760–18765.

Stam C, De Haan W, Daffertshofer A, Jones B, Manshanden I, Van Walsum AVC, Montez T, Verbunt J, De Munck J, Van Dijk B. Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. Brain. 2009; 132:213–224. [PubMed: 18952674]

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

Suk, HI., Lee, SW., Shen, D. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Springer; 2015. A Hybrid of Deep Network and Hidden Markov Model for MCI Identification with Resting-State fMRI; p. 573-580.

Suk HI, Lee SW, Shen D, Initiative ADN. Subclass-based multi-task learning for Alzheimer's disease diagnosis. Frontiers in aging neuroscience. 2014a; 6

Suk HI, Lee SW, Shen D, Initiative A.s.D.N. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Structure and Function. 2013; 220:841–859. [PubMed: 24363140]

Suk HI, Lee SW, Shen D, Initiative A.s.D.N. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage. 2014b; 101:569–582. [PubMed: 25042445]

Suk HI, Wee CY, Lee SW, Shen D. Supervised Discriminative Group Sparse Representation for Mild Cognitive Impairment Diagnosis. Neuroinformatics. 2014c:1–19. [PubMed: 24174261]

Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996:267–288.

Toussaint PJ, Maiz S, Coynel D, Doyon J, Messé A, de Souza LC, Sarazin M, Perlbarg V, Habert MO, Benali H. Characteristics of the default mode functional connectivity in normal ageing and Alzheimer's disease using resting state fMRI with a combined approach of entropy-based and graph theoretical measurements. NeuroImage. 2014; 101:778–786. [PubMed: 25111470]

Wang K, Liang M, Wang L, Tian L, Zhang X, Li K, Jiang T. Altered functional connectivity in early Alzheimer's disease: A resting - state fMRI study. Human brain mapping. 2007; 28:967–978. [PubMed: 17133390]

Ward JH Jr. Hierarchical grouping to optimize an objective function. Journal of the American statistical association. 1963; 58:236–244.

Watts DJ, Strogatz SH. Collective dynamics of 'small-world'networks. nature. 1998; 393:440–442. [PubMed: 9623998]

Wee, CY., Yang, S., Yap, PT., Shen, D. Machine Learning in Medical Imaging. Springer; 2013. Temporally dynamic resting-state functional connectivity networks for early MCI identification; p. 139-146.

Wee CY, Yang S, Yap PT, Shen D, Initiative, A.s.D.N. Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. Brain Imaging and Behavior. 2015:1–15. [PubMed: 25724689]

Wee CY, Yap PT, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D. Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. PloS one. 2012; 7:e37828. [PubMed: 22666397]

Wee CY, Yap PT, Zhang D, Wang L, Shen D. Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. Brain Structure and Function. 2014; 219(2):641–656. [PubMed: 23468090]

Whitwell JL, Przybelski SA, Weigand SD, Knopman DS, Boeve BF, Petersen RC, Jack CR. 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. Brain. 2007; 130:1777–1786. [PubMed: 17533169]

Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2009; 31:210–227.

Yu, R., Zhang, H., An, L., Chen, X., Wei, Z., Shen, D. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016. Springer; Correlation-Weighted Sparse Group Representation for Brain Network Construction in MCI Classification; p. 37-45.

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

Zhang Y, Zhou G, Jin J, Zhao Q, Wang X, Cichocki A. Sparse Bayesian Classification of EEG for Brain-Computer Interface. IEEE Transactions on Neural Networks and Learning Systems. 2016; 27(11):2256–2267. [PubMed: 26415189]

Zhang Y, Wang Y, Jin J, Wang X. Sparse Bayesian Learning for obtaining sparsity of EEG frequency bands based feature vectors in Motor Imagery Classification. International Journal of Neural Systems. 2017; 27:1650032. [PubMed: 27377661]

Zhou L, Wang Y, Li Y, Yap P, Shen D. Hierarchical Anatomical Brain Networks for MCI Prediction: Revisiting Volumetric Measures. PLoS ONE. 2011; 6(7):e21935. [PubMed: 21818280]
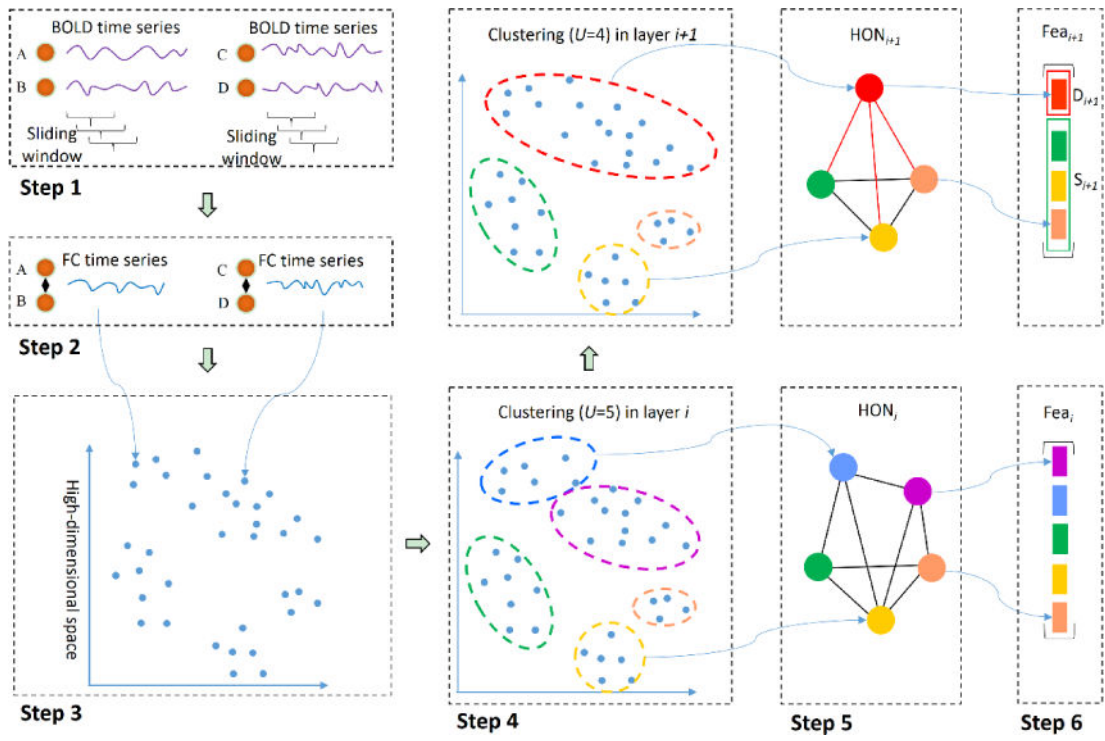
Fig. 1.
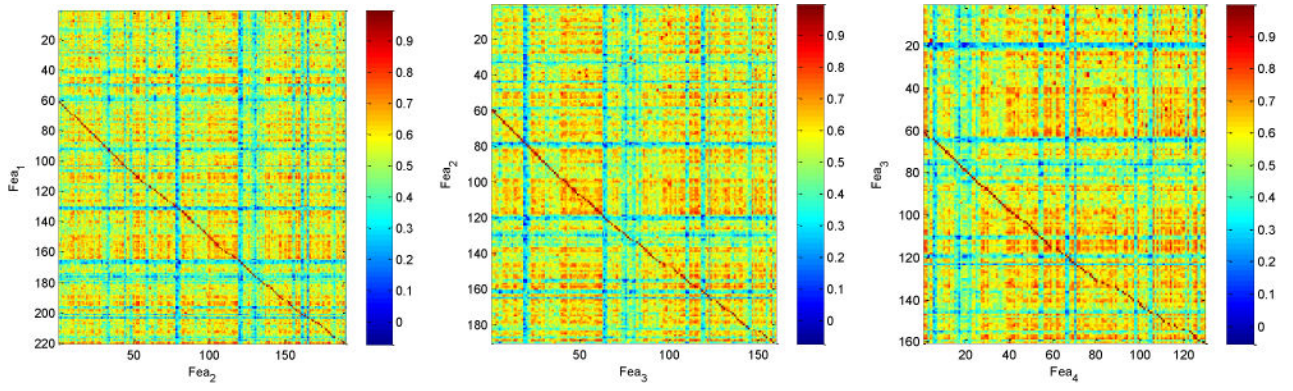Framework of the proposed hierarchical high-order FC networks construction.

Fig. 2.
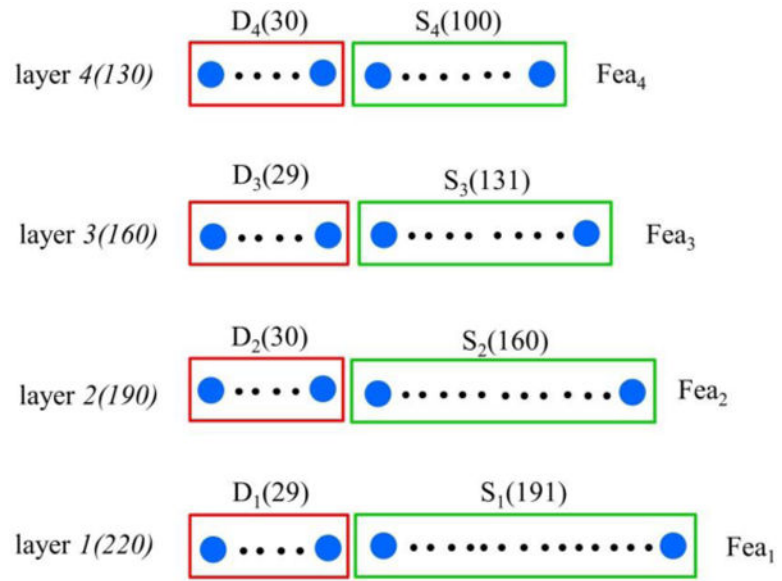Correlation between features from neighboring high-order FC networks.

Fig. 3.
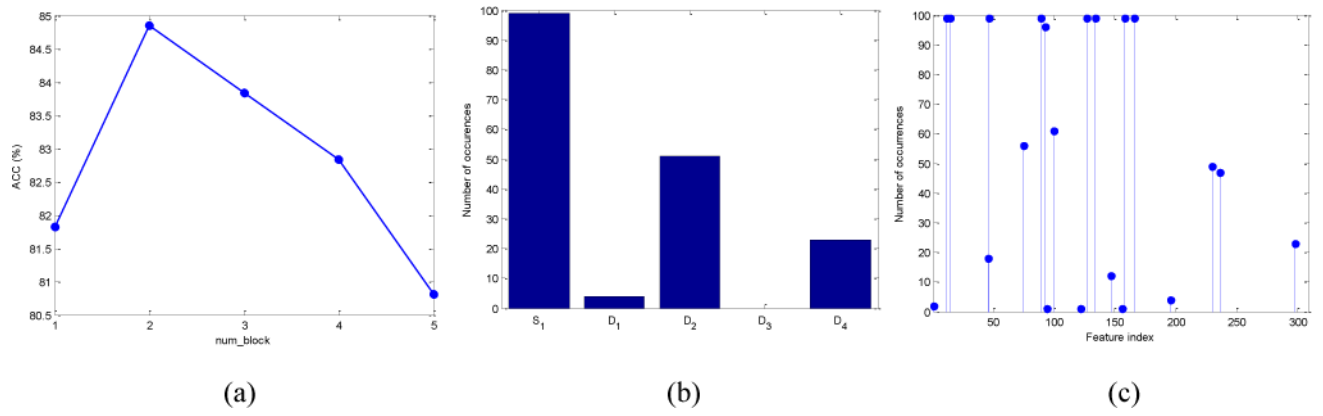Illustration of feature decomposition in each layer.

Fig. 4.
(a) Variation of accuracy versus the number of feature blocks; (b) Number of occurrences of each feature block; (c) Number of occurrences of each individual feature.
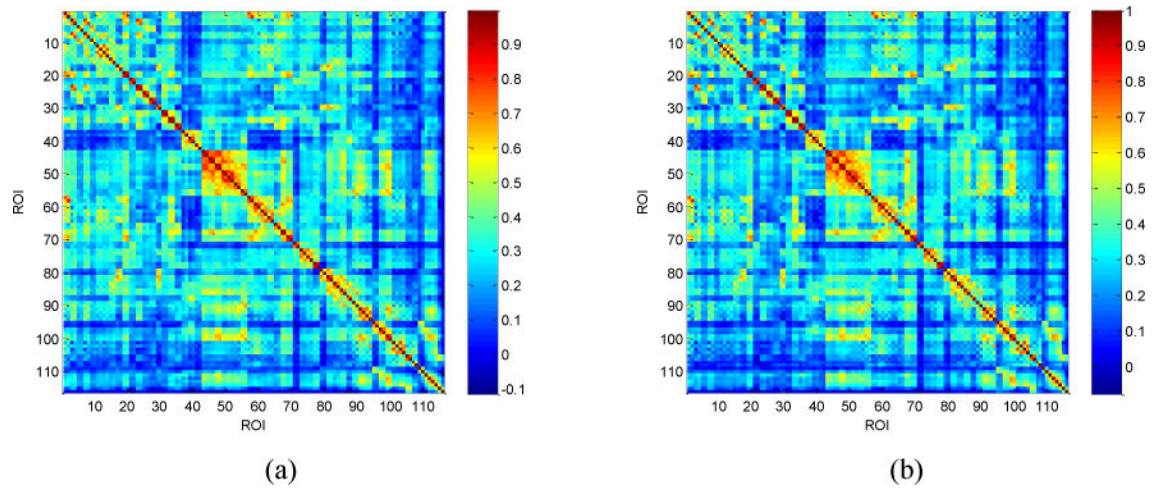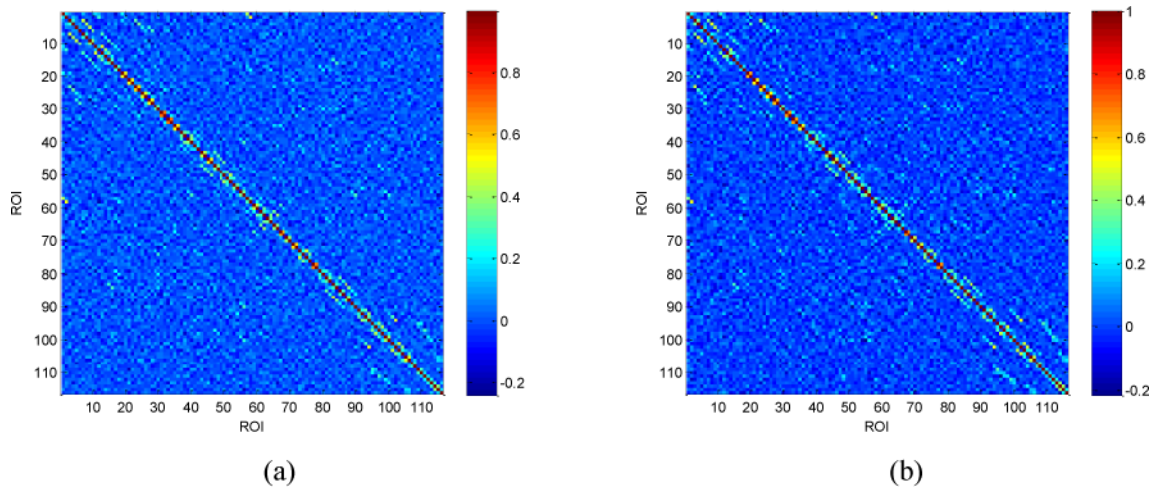
Fig. 5.
Averaged low-order FC networks built by LON-PEC for (a) MCI and (b) NC subjects.

Fig. 6.
Averaged low-order FC networks built by LON-PAC for (a) MCI and (b) NC subjects.
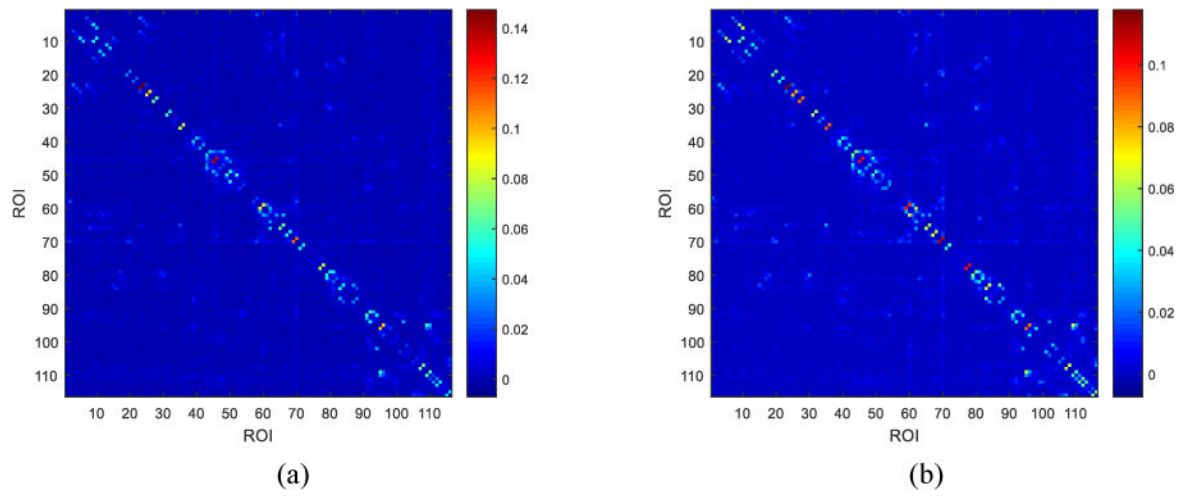
Fig. 7.
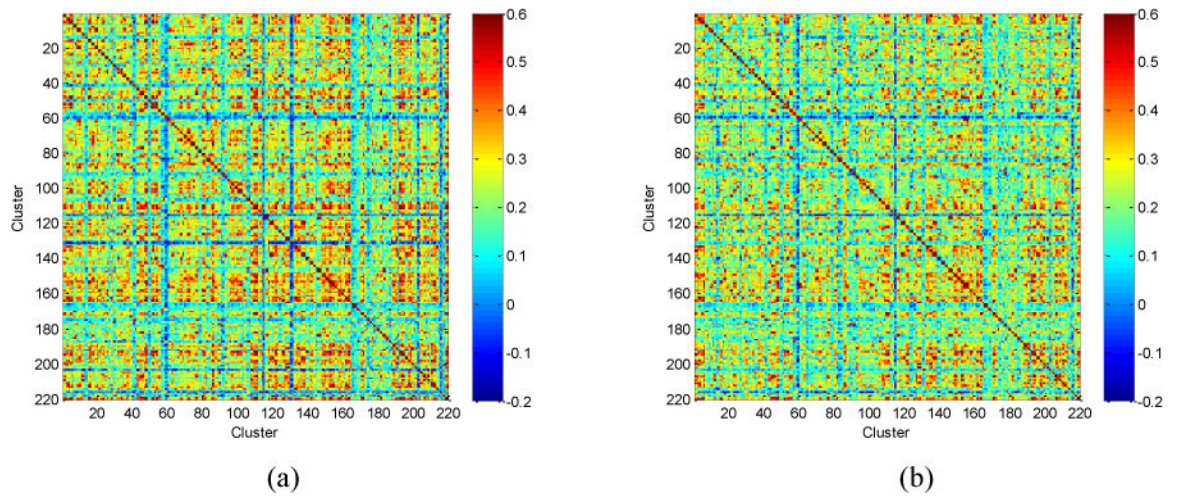Averaged FC networks built by SR for (a) MCI and (b) NC subjects.
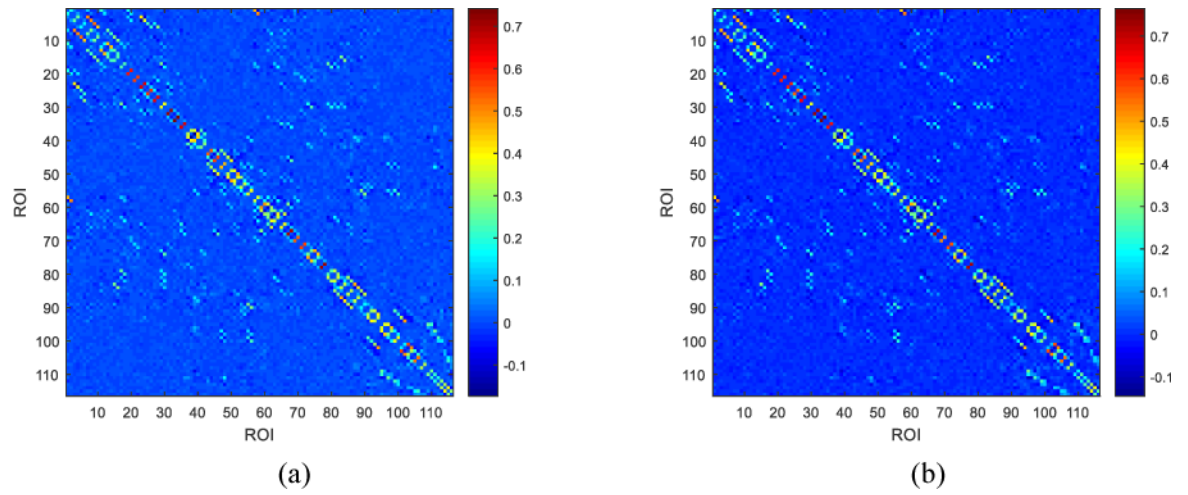
Fig. 8.
Averaged FC networks built by WSR for (a) MCI and (b) NC subjects.

Fig. 9.
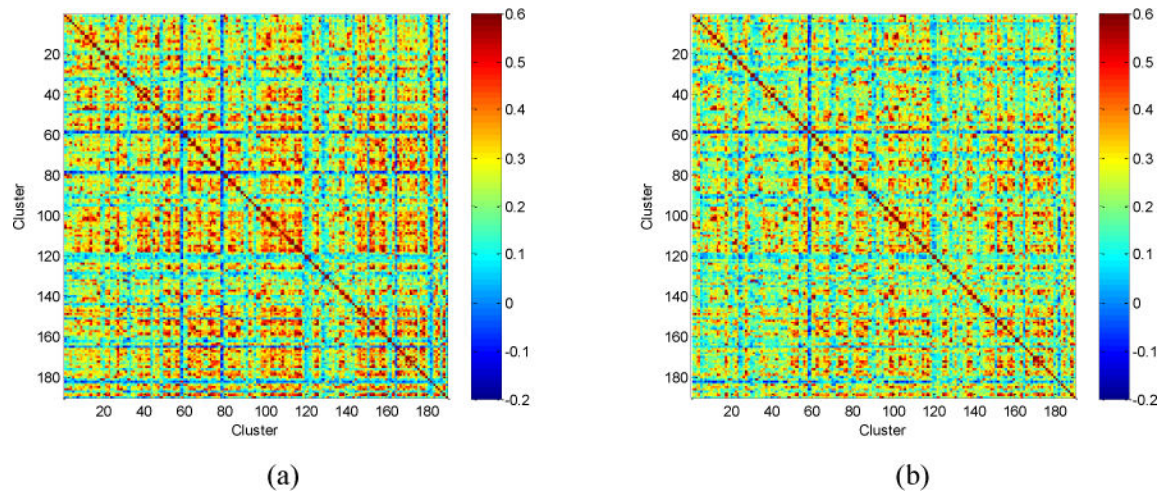Averaged high-order FC networks for (a) MCI and (b) NC subjects in the layer 1.

Fig. 10.
Averaged high-order FC networks for (a) MCI and (b) NC subjects in the layer 2.
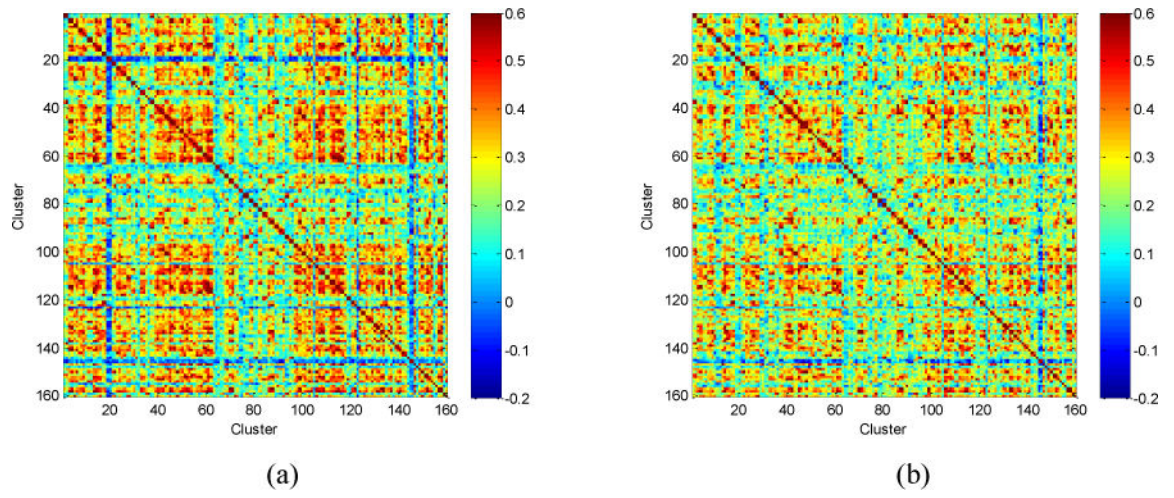
Fig. 11.
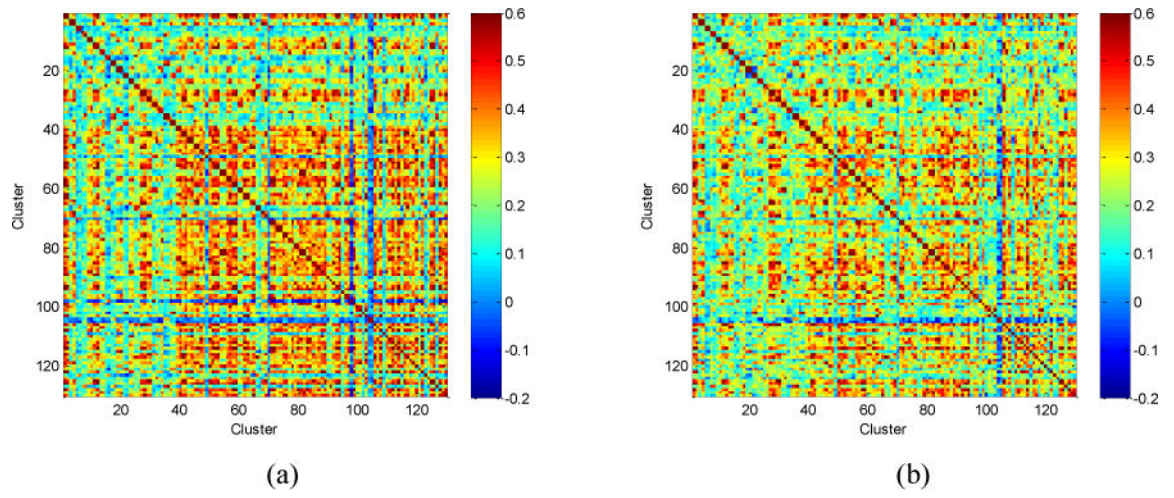Averaged high-order FC networks for (a) MCI and (b) NC subjects in the layer 3.

Fig. 12.
Averaged high-order FC networks for (a) MCI and (b) NC subjects in the layer 4.

Table 1

Sequential forward selection and sparse regression

Initialize $A = \Phi$, and the number of feature blocks *num_block*

For t=1: *num_block*

   $acc_{max} = 0$

   For each D_i in *Fea − A*

     $B = A \cup D_i$

     Perform sparse regression on the training subjects with features in B to select a subset C;

     Train a linear SVM model on the training subjects with feature subset C;

     Obtain accuracy acc of the SVM model on validation data;

     If acc > acc_{max}

       $acc_{max} = acc$; $i_{max} = i$

     End If

   End For

$$A = A \cup D_{i_{\max}}$$

End For

Table 2

Performance comparison of different methods in MCI classification.

Method	ACC	AUC	SEN	SPE	YI	F-score	BAC
LON-PEC	57.58	0.6008	58.00	57.14	15.14	58.00	57.57
LON-PAC	60.61	0.6249	58.00	63.27	21.27	59.79	60.63
SR	64.65	0.6155	68.00	61.22	29.22	66.02	64.61
WSR	72.73	0.7318	72.00	73.47	45.47	72.73	72.73
HON_1	80.81	0.8567	86.00	75.51	61.51	81.90	80.76
HON_2	81.82	0.8743	82.00	81.63	63.63	82.00	81.82
HON_3	68.69	0.8094	74.00	63.27	37.27	70.48	68.63
HON_4	72.73	0.7857	74.00	71.43	45.43	73.27	72.71
HHON-CON	78.79	0.8702	80.00	77.55	57.55	79.21	78.78
HHON-SFS	**84.85**	**0.9057**	**88.00**	**81.63**	**69.63**	**85.44**	**84.82**