ISSN : 0974 - 7435 Volume 10 Issue 24

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(24), 2014 [15011-15017]

AD classification based on brain functional network using ADNI

Jie Xiang, Hong Liang, Cao Rui, Zheng Wu, Junjie Chen*
College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, 030024, (CHINA)
E-mail: chenjunjie@tyut.edu.cn

ABSTRACT

In order to assist the diagnosis of mild cognitive impairment (MCI) and provide a new diagnostic method, resting state brain functional networks of early mild cognitive impairment, late mild cognitive impairment and normal controls are constructed, node attributes of brain functional networks based on complex network are calculated and differences between groups are analyzed which served as the classification features. Then, the subjects are classified using support vector machine algorithm. The experimental result showed that this method can be used for diagnosis of MCI, having a certain application value.

KEYWORDS

Early mild cognitive impairment; Late mild cognitive impairment; Support vector machine; Classification; Alzheimer's disease neuroimaging initiative.

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disease with a high incidence rate, etiology and pathogenesis is unclear, the role of existing medications for AD is very limited. Mild Cognitive Impairment (MCI) is the prodromal stage of AD, this disease have a high conversion rate about 10%-15% to AD per year, while the conversion rate of the normal elderly is only 1%-2%. Therefore preclinical warning and early intervention is particularly important for patients with MCI [1, 2]. However, cognitive impairment of MCI patients is slighter than AD patients, present diagnosis methods of neuropsychological testing and neuroimaging studies can't make a definite diagnosis. How to identification diagnosis mark using for early find and prevent MCI converting to AD is an important issue.

With the development of magnetic resonance imaging, machine learning methods are gradually applied in the research of function magnetic resonance imaging. Studies have shown that the topology structure of brain network was different between MCI and AD, global and local attributes of MCI were different with normal control (NC), characteristic path length was greater [3] and node attributes of some brain regions in a single threshold was greater or less compared with normal controls [4]. Some researchers employed abnormal clustering coefficients and the value of functional connection as classification features and classified MCI and NC using support vector machine(SVM) algorithm[5,6].

Recent studies of the brain functional network of MCI concerning global attributes or local attributes in single threshold, there is no study concerning the brain functional network of MCI in continues threshold at present.

In this study, we analyzed the node attributes of brain networks in continues threshold about MCI and normal controls, node attributes with significant difference used as classification features to discriminate MCI patients from NCs using SVM algorithm to assist the diagnosis of MCI and provide a new angle of view to diagnosis of MCI.

THE THEORY OF SUPPORT VECTOR MACHINE

SVM is a type of machine learning method proposed by Vapnik based on statistics and structure risk minimization theory. SVM is one of the main trend machine learning methods, because it can solve problems of small samples, nonlinear and high-dimensional learning and have no problems of local extremum which neural network have and it showed successful performance in data analysis of fMRI [7].

The main idea of SVM is that the input vector is mapped to a high-dimensional feature space by kernel function and then to find the optimal hyperplane in the feature space, the hyperplane with maximum classification margin can distinguish two type samples accurately.

For linear problems, given a data set $\{x_i, y_i\} i = 1, 2, ...n$, where $x_i \in R^m$ are observations, and $y_i \in \{1, -1\}$ are corresponding labels, the linear discriminant function is $g(x) = (w \cdot x_i) + b$, the corresponding hyperplan is $w \cdot x + b = 0$, the optimal hyperplane that should meet the requirements as:

$$y_i[(w \cdot x_i) + b] - 1 \geq 0 \tag{2-1}$$

The class margin at this time is $\frac{1}{2}\|w\|^2$, the maximum margin is the minimum value of $\|w^2\|$.

For linear separable, according to the theory of optimum, that can be represented as a constraint optimization problem as the formula (2-2):

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ s.t.\ \ y_i(\langle w, x_i \rangle + b) \geq 1, \qquad i = 1, \cdots, n \end{cases} \tag{2-2}$$

To solve the above problem, we can transform it to the problem of lagrangian duality on the basis of optimization theory which can be expressed as:

$$\max(Q(\partial)) = \sum_{i=1}^{n} \partial_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \partial_i y_i \partial_j y_j (x_i^{\mathrm{T}} \cdot x_j) \tag{2-3}$$

Where $\sum_{i=1}^{n} \partial_i y_i = 0$ and $\partial_i \geq 0$ (i is lagrangian of each sample, i =1, 2…n). This function is a quadratic function optimization problem under the inequality constraints which have unique solution. The optimal classification function will be obtained when above problem was solved, it can be expressed as:

$$\mathrm{f}(x) = \mathrm{sgn}\left\{\sum_{i=1}^{n} \partial_i y_i (x_i \cdot x) + b\right\} \tag{2-4}$$

For nonlinear problems, a cost parameter c(c is a positive constant) and the slack variables ξi (ξi≥0) can be introduced to modify formula (2-1) to formula (2-5)

$$y_i \left((w \cdot x_i) + b\right) + \xi_i - 1 \geq 0 \tag{2-5}$$

To solve the function (2-6) in condition (2-5), we can transform (2-6) to the problem of lagrangian duality which is to solve (2-3) in condition $\sum_{i=1}^{n} \partial_i y_i = 0$ and $\mathrm{c} \geq \partial_i \geq 0$.

$$\min \varphi(w, \xi) = \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{n} \xi_i \tag{2-6}$$

We can construct optimal hyperplan in a high dimensional feature space transformed from input space by nonlinear transformation. Dual problem only involves inner product operation (xiT·xj) between samples, kernel function K(x,y) can take place of inner product operation, so optimal decision function of SVM can be expressed as:

$$\mathrm{f}(x) = \mathrm{sgn}\left\{\sum_{i=1}^{n} \partial_i y_i K(x_i, x) + b\right\} \tag{2-7}$$

Different kernel functions can construct different SVMs. There are four kind of kernel functions: linear kernel function, Radial Basis Function (RBF), polynomial kernel function and S kernel function.

MATERIALS AND METHODS

Subjects

Data used in this study were from the Alzheimer's disease Neuroimaging Initiative (ADNI) database, which was launched in 2003 by National Institute on Aging (NIA). The primary goal of ADNI is to aid researchers to research the mechanism of AD and develop new treatment, as well as lessen the time and cost. ADNI introduced the concept of Early MCI (EMCI) and Late MCI (LMCI) firstly, the distinction between them is the impairment level of delayed recall of logical memory. EMCI subjects were enrolled to narrow the gap between normal controls and LMCI subjects, so we can discover patient's condition early and offer effective treatment.

All participants are in the age range of 55-90 years, excluded Subjects of mental drug treatment has been ruled out。 The criterions of each group are described as follows:

1) normal controls: Mini-Mental State Examination (MMSE) scores between 24 and 30(inclusive), a Clinical Dementia Rating (CDR) score of 0, non-depressed and non-demented;

2) EMCI subjects: MMSE scores between 24 and 30(inclusive), a memory complaint by a study partner, objective memory loss measured by the Wechsler Memory Scale Logical Memory II(the maximum score is 25)(adjusted for age and education; >=16 years: 9-11; 8-15 years: 5-9; 0-7 years: 3-6), a CDR of 0.5, no significant impairment in other cognitive domains, preserved activities of daily living, and non-demented;

3) LMCI subjects: the criteria for LMCI subjects is the same as EMCI, except the scores of objective memory loss measured by delayed recall of Wechsler Memory Scale Logical Memory II (adjusted by age and education; >=16 years: <=8; 8-15 years:<=4; 0-7 years: <=2) [8,9].

Our study collected 70 subjects include 25 LMCI patients, 16EMCI patients and 29 NCs. The basic information of subjects was shown in Table 1.

Table 1 : Basic information of subjects used in the experiment

	LMCI	EMCI	NC	P
Gender(F/M)	9/16	8/8	15/14	0.47a
Age(year)	73.2±7.5	74.0±7.0	74.1±5.7	0.87b
MMSE	27.2±2.3	27.7±1.2	29.0±1.4	<0.01b
CDR	0.5±0.2	0.5±0.1	0.0±0.1	<0.01b
[a] Two-tailed chi-square test; [b] One-way ANOVA test.				

Data preprocessing

All the fMRI images and T1 structure images in this study were performed on Philips 3.0 Tesla MRI scanner. During the scan, subjects were instructed to have their eyes open and don't fall asleep, relax and try not to move. All the images we downloaded from the site are Analysis format to be preprocessing conveniently. The detailed scanning process and scanning parameters are available in the ADNI (http://www.loni.ucla.edu/ADNI).

All functional imaging data pre-processing was carried out using Data processing Assistant for Resting-State fMRI (DPARSF) [10]. The first 10 functional images were discarded, then the remaining 130 fMRI images were corrected for within-scan acquisition time differences between slices and realigned to correct for head motions to ensure that all levels of subjects were less than 3mm head movement and rotation of less than 3 degrees. Then the motion-corrected functional volumes were spatially normalized to standard brain. The images were performed with 12 dimension optimization affine transformation and normalized to 3 mm voxel of MNI standard space. Next the images were smoothed with a 4mm full-width at half maximum (FWHM) Gaussian kernel. Finally, low-frequency filter (0.01~0.08Hz) was performed to reduce the effects of low-frequency drifts and high-frequency noise.

Brain networks building and features extracting

After preprocessing, the fMRI data were segmented into 90 regions (45 for each hemisphere) using an anatomically labeled template (AAL) [11], each brain region was defined as a node in the network. Average time series of all voxels in a region was regarded as the time series of this region. To time sequence obtained by each brain region, partial correlation coefficients were calculated between each two and the correlation matrix of 90 × 90 can be obtained. Usually, the incidence matrix is transformed into binary adjacency matrix according to the setting threshold, that is when the partial correlation between node i and j greater than a certain threshold, we set aij to 1, while others to 0. But if we applied the same threshold to the matrixes of the all subjects, each resulting graph would have different number of edges, it would lead to the differences between each group of network properties and node properties. So most studies use sparsity to define whether edges will be established between two nodes. Sparsity is the ratio of the actual number of edges in the network and the maximum possible number of edges in the network. In this study, the range of sparsity is 8%-40%, an interval of 0.01. In this range of sparsity, the networks of each subject are connected [12, 13].

To each sparsity, we calculated node attributes of each subject include node degree, node centrality degree and node efficiency [14, 15]. The meaning of each attribute was described as follows, where N is the total number of nodes in the network G.

The degree of node i was defined as the number of edges connected to the node. The formula can be expressed as 3-1:

$$D_i = \sum_{j \in G} a_{ij} \tag{3-1}$$

Where aij is the number of connections between node i and node j in the network.

The centrality degree of node i was defined as the number of shortest paths through the node to all shortest paths among all other nodes. It reflects the importance of the node i in the network. The formula can be expressed as 3-2:

$$\mathrm{B}_i = \sum_{m \neq i \neq n} \frac{\sigma_{mn}(i)}{\sigma_{mn}} \tag{3-2}$$

Where σmn is the number of shortest paths between node m and node n, σmn(i) is the number of shortest paths from node m to n.

The node efficiency of node i was defined as the inverse of the mean distance between j and all other nodes. It reflects the degree of difficulty from the node to other nodes in the network, which can be expressed as the formula 3-3:

$$E_i = \frac{1}{N-1}\sum_{j\neq i}\frac{1}{d_{ij}} \tag{3-3}$$

Where dij is the shortest path length between nodes i and j, the shortest path length is the length of shortest path between two nodes.

In order to characterize the whole character in continuous sparsity, Area under the Curve (AUC) of each node attribute in selected sparsities are calculated. The AUC represents the area of a node attribute in a certain saprity, which can be expressed as the formula 3-4:

$$Y^{AUC} = \sum_{k=1}^{n-1}\left[Y\left(S_K\right)+Y\left(S_{K+1}\right)\right]\times\Delta S/2 \tag{3-4}$$

Where S1 and Sn separately mean the boundary of sparsity, ΔS is the interval of sparsity. In this study, S1 is 0.08, Sn is 0.40, ΔS is 0.01.Studies have shown that AUC was sensitive to the change of topological attributes in brain networks [16, 17].

CLASSIFICATION

Feature selection

As well known, the feature selection plays a crucial role in the classification. We want to select some significant features which can distinct EMCI, LMCI and NC. Some researchers used statistical test to screen significant features. In previous studies[18, 19], the paired T-test was used for the screening the brain areas with significant differences. However, this method is limited, because it should be conducted with the condition that all the samples are normally distributed. Since the fMRI signals cannot meet the Gaussian distribution, the paired T-test cannot be used. When nonparametric permutation test, such as Kolmogorov-Smirnov (K-S) test is performed, normal distribution of the samples is not required. Using the nonparametric permutation test [20], a single group can be tested whether the samples are distributed in accordance with a certain theory and, moreover, the two groups can be tested whether significant differences exist between them. It has been proven that the nonparametric permutation test is an effective and stable non-linear statistical test, primarily because of its strong anti-noise performance and high sensitivity to nonlinear signals [21]. In the present study, we used a nonparametric permutation test to investigate the statistical significance of two group differences in AUC values of the node attributes. We used this nonparametric test instead of parametric test, because it can avoid making any assumption about distribution of AUC values. Fig. (1) Shows the brain regions with significant difference of two groups. This figure has been drawn using BrainNet Viewer (http:// www.nitrc.org/projects/bnv). Table (2) shows the name of these brain regions, which were used as classification features in our classification.

Classification and result

Three classification experiments were performed on the subject set: EMCI and NC (EMCI/NC), LMCI and NC (LMCI/NC), EMCI and LMCI (EMCI/LMCI). In other words, in the EMCI/NC classification experiment, we try to train a classification model which can identify EMCI from NC, in the LMCI/NC classification experiment, we try to train a classification model which can identify LMCI from NC, and in the LMCI /NC classification experiment, a classification model were trained to identify LMCI and EMCI.

The procedure of training a classification model can be described as follows:

Firstly, all subjects were randomly divided into three parts, one parts as test set, the remaining subjects were as train set. Normalization was applied to the training set and testing set, the range of values is -1 to 1.

Secondly, RBF function was selected as kernel function. There is no unified criterion to optimize the parameters of kernel function, so we employed grid search method to optimize the parameters. Set a range to c and g, employed cross validation method to get classified accuracy of training set under the group c and g, the group c and g with highest classification accuracy as the best parameters. In this study, the range of c and g is $\left[2^{-10},2^{-9},...2^{9},2^{10}\right]$, 10-Fold cross validation was used.

Finally, classification model was obtained on train set used the best parameters c and g, test set was used to test this model. The accuracy of the model means the performance of SVM model. The result was shown as fig. (2):

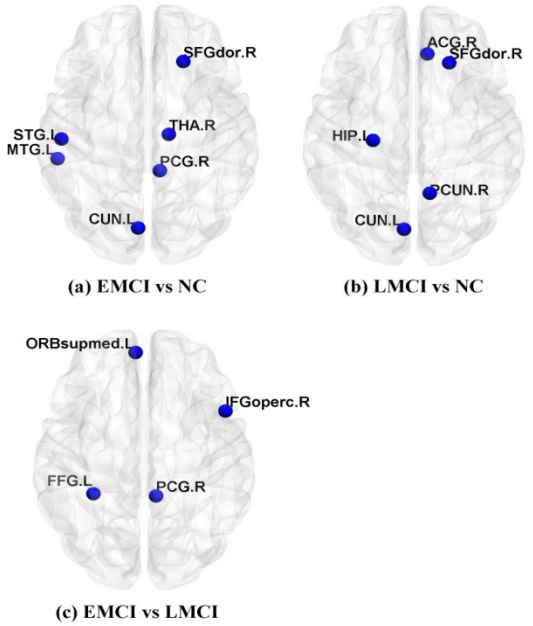

Fig. (1) : Brain regions with significant group differences

Table 2 : The names of brain regions

Regions	Full name of regions
SFGdor.R	Right Superior frontal gyrus, dorsolateral
IFGoperc.R	Right Inferior frontal gyrus, opercular part
ORBsupmed.L	Left Superior frontal gyrus, medial orbital
ACG.R	Right Anterior cingulate and paracingulate gyri
PCG.R	Right Posterior cingulate gyrus
HIP.L	Left Hippocampal
CUN.L	Left Cuneus
FFG.L	Left Fusiform gyrus
PCUN.R	Right Precuneus
THA.R	Right Thalamus
STG.L	Left Superior temporal gyrus
MTG.L	Left Middle temporal gyrus

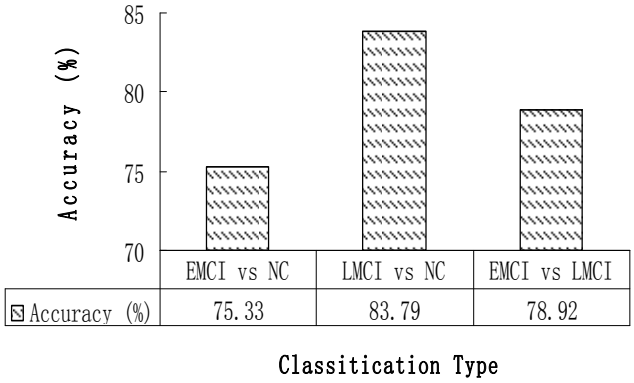

Fig. (2) : Classification accuracy

Figure 2 displays the classification accuracy of EMCI/MCI, LMCI/NC and EMCI/LMCI, from which we can see that the accuracy of EMCI/MCI, LMCI/NC and EMCI/LMCI is 75.33%、83.79%、78.92% respectively that is superior to chance level.

CONCLUSION

In this study, we constructed resting state brain functional network of EMCI patients, LMCI patients and normal controls based on complex network theory and calculated AUC values of node attributes in selected sparsities. AUC values which had significant difference were as classification features. SVM algorithm was applied to classify different two group subjects. Results have shown that AUC values of brain regions with significant difference include cingulate, temporal lobe frontal lobe and frontal lobe etc., These differences were consistent with the existing research results[3,4].These AUC values of brain regions as classification features can discriminate two kinds of MCI and normal controls. Therefore, this method can assist diagnosis of MCI to detect state of an illness early.

ACKNOWLEDGEMENT

This study was sponsored by the National Natural Science Foundation of China (Grant No. 61170136, 61373101), science and technology industrial project of Shanxi Province (No. 20140321002-01). We are also grateful to the data provided by ADNI.

REFERENCES

[1] 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 [J]. NeuroImage, 2009, 44(4): 1415-1422.

[2] Ron B, Elizabeth J, Kathryn Z G, et al. forecasting the global burden of Alzheimer's disease [J]. Alzheimer's and Dementia. 2007, 3(3): 186-191.

[3] Wang J H, Zuo X N, Dai Z J, et al. Disrupted Functional Brain Connectome in Individuals at Risk for Alzheimer's Disease [J]. Biological Psychiatry. 2013, 73(5): 472-481.

[4] Liu Zhenyu, Zhang Yumei, Yan Hao,et al. Altered topological patterns of brain networks in mild cognitive impairment and Alzheimer's disease: A resting-state fMRI study [J]. Psychiatry Research: Neuroimaging, 2012, 202(2): 118-125.

[5] Li Y, Wang Y, Wu G, et al. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's diseaseusing dynamic and network features [J]. Neurobiology of Aging. 2012, 33(2): 427.e15-427.e30.

[6] Chong-Yaw Wee, Pew-Thian Yap, Zhang Daoqiang, et al. Identification of MCI individuals using structural and functional connectivity networks [J]. NeuroImage, 2012, 59(3): 2045-2056.

[7] Kohannim Omid, Hua Xue, Hibar Derrek P, et al. Boosting power for clinical trials using classifiers based on multiple biomarkers [J]. Neurobiology of Aging. 2010, 31(8): 1429-1442.

[8] Aisen P S, Petersen R C, and Donohue M C, et al. Clinical Core of the Alzheimer's disease Neuroimaging Initiative: progress and plans [J]. Alzheimer's Dementia, 2010, 6: 239-246.

[9] Michael W. Weiner, Paul S. Aisen, Clifford R. Jack Jr, et al. The Alzheimer's disease neuroimaging initiative: progress report and future plans [J]. Alzheimer's Dementia, 2010, 6(3): 202-211(e7).

[10] Yan C G, Zang Y F. DPARSF: a MATLAB toolbox for "Pipeline" data analysis of resting-state fMRI [J]. Front Syst Neurosci. 2010, 4: 13.

[11] XUE Shaowei, TANG Yiyuan, LI Jian, et al. Method for constructing brain functional networks based on fMRI data [J]. APPLICATION RESEARCH OF COMPUTERS. 2010, 27(11): 4505-4507.

[12] Yao Zhijun, Zhang Yuanchao, Lei Lin,et al. Abnormal Cortical Networks in Mild Cognitive Impairment and Alzheimer's Disease [J]. PLoS Computational Biology, 2010, 6 (11): e1001006.

[13] He Y, Chen Z, Evans A. Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease [J]. The Journal of Neuroscience. 2008, 28(18): 4756-66.

[14] TIAN Lixia. Analysis of complex brain networks based on graph theory. Beijing Biomedical Engineering [J]. 2010, 29(1): 1-7.

[15] Sophie Achard, Ed Bullmore. Efficiency and cost of economical brain functional networks [J]. PLoS Computational Biology, 2007, 3(2): e17.

[16] Zhang Junran, Wang Jinhui, Wu Qizhu, et al. Disrupted Brain Connectivity Networks in Drug-Naive, First-Episode Major Depressive Disorder [J]. Biological Psychiatry. 2011, 70(4): 334–342.

[17] Wang Jinhui, Wang Liang, Zang Yufeng, et al. Parcellation-dependent small-world brain functional networks: A resting-state fMRI study [J]. Human Brain Mapping, 2009, 30(5): 1511-1523.

[18] D. Wang, D. Miao and G. Blohm, "Multi-class motor imagery eeg decoding for brain-computer interfaces", Front.Neurosci, vol. 6, Oct 2012.

[19] W. Deng, D.Q. Miao and G. Blohm, "A new method for eeg-based concealed information test", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, 2013, pp. 520-527.

[20] I. Alexei and G. Riccardi, "Kolmogorov-smirnov test for feature selection in emotion recognition fromspeech. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", IEEE: Kyoto, 2012, pp. 5125-5128.

[21] S.M. Hou, Y.R. Li and G.L. Liu, "A new method of detecting nonlinear for time series based on ks test", Journal of Electronics & Information Technology, vol. 29, pp. 808-810, Apr 2007.