Automatic identification of ADHD and Autsim based on ICA and SVM using resting state fMRI

Jinze Li[1], Gang Yu[1], Pan Lin[2], Yanhua Gao[3], and Kai Ai[1]

[1]*School of Geosciences and Info-Physics, Central South University, Changsha, China,* [2]*Institute of Biomedical Engineering, Xi'an Jiaotong University, Xi'an, China,* [3]*Department of B Ultrasound, Shaanxi Provincial People's hospital, Xi'an, China*

Introduction

Psychiatric disorders are harmful to children and adolescents. And it's a hard work to distinguish the corresponding patients from the healthy in early diagnosis. Previous studies have proved that the brain functional networks show abnormal pattern in children and adolescents who suffer from mental diseases such as attention deficit hyperactivity disorder (ADHD) and autism disorder [1]. This paper presents a combined method based on independent components analysis (ICA) and support vector machine (SVM) to classify ADHD, Autism and control group automatically. Based on the combined method, more psychiatric disorders of children and adolescents are expected to be automatically distinguished in the future.

Materials and Methods

Image acquisition All the resting state fMRI data were obtained from the 1000 Functional Connectomes Project [2] that were collected from a 3T Siemens scanner. The data includes three groups: the ADHD group (37 male and 23 female, age 10.07±0.77 years), the autism group (28 male and 4 female, age 19.33±5.09 years) and the normal controls (27 male and 22 female, age 10.17±0.95 years). All the groups show no significant differences in both age and gender.

Data Preprocessing As the fMRI data is a high dimensional signal, it is necessary to reduce the dimension in a manner that preserves its temporal structure. We applied ICA by using tools in FSL [3] to reduce spatial dimension. After that, a set of ICA spatial activation maps and the corresponding time series are obtained.

Methods The proposed method can be implemented as follows: First of all, we use an automatic independent component labeling method [4] to reduce data artifacts in ICA components. Secondly, after de-noising we remain relevant ICA components. Based on the cross-correlation function [1], we compute distance matrix by using corresponding time series of those ICA components. The distance matrices of three certain subjects (ADHD, Autism and Control subject, respectively) are shown in Fig.1. Thirdly, because the above distance matrices only describe straight-line Euclidean distance of independent components, we still need to calculate the geodesic distance which can reflect real low-dimensional geometry relationship of manifolds (non-Euclidean space). According to previous study [5], we warp the distance matrices by using the embedding technique of isomap [5] to fit the manifolds. Within each distance matrix, the isomap procedure will gain a new matrix which includes the recalculated shortest path (i.e. geodesic distance). Then the new created matrix is exhibited in "graph" format shown as Fig.2 (the same subjects are used as above). Fourthly, by using the graphs derived from isomap processes, we define a feature matrix that contains measures of the directed graph properties including average path length, clique number, graph density, edge connectivity, median closeness, median degree, maximum degree, vertex count, edge count and transitivity. We apply these properties as the feature vector to train SVM classifier subsequently. After training, it is proved that average path length and vertex count for ADHD or transitivity and edge count for Autism are most important features to get best classification accuracy respectively (Fig.3).

Result

We utilize SVM to classify three types of data (ADHD, Autism and Control respectively). Due to the small sample size, we performed cross-validation on testing data to ensure that the accuracy remains unbiased. The SVM based classification results on testing datasets are summarized in Table1. For each method, overall classification accuracy, sensitivity value and specificity value are provided. We can find that the classification accuracy of ADHD is slightly higher than autism. Generally, the classification accuracy of ADHD and Autism is satisfactory on our dataset.

Conclusion

The proposed method can validly capture the differences of brain connectivity between discords patients and healthy controls. And the results not only show that the geometrical temporal structure (Fig.2) of ADHD and autism can reveals the complex brain functional network efficiently, but also show that SVM can be applied to extracted features for discrimination. Current result has demonstrated the accuracy and feasibility of our presented classification method. In the future, we will improve the accuracy of the classification and try to classify more psychiatric disorders of children and adolescents.

Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities of Central South University (2013zzts251), by Doctoral Fund of Ministry of Education of China (20120201120071) and by the Fundamental Research Funds for the Central Universities of China.

Reference

[1] Ariana Anderson, et al. *NeuroImage* 2010, vol. 49, p. 2509–2519. [2] Milham, M.P. et al. 2009, www.nitrc.org/projects/fcon_1000/. [3] Smith, S.M., et al. *NeuroImage* 2004, vol.23, p. 208–219. [4] Jussi Tohka, et al. NeuroImage 2008, vol.39, p.1227–1245. [5] Joshua B. Tenenbaum, et al. *Science* 2000, vol.290, p.22.

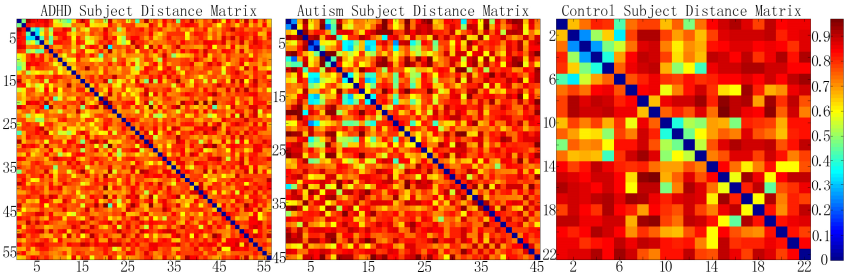

Figure.1 Subject distance matrices showing unequal number of ICA components. Hot color represents longer distance (i.e. weaker correlation) between two certain components. Cool color represents shorter distance (i.e. stronger correlation) between two certain components.

	SVM		
	Overall Acc. (%)	**Sens. (%)**	**Spec. (%)**
ADHD/Control	87.6	88.0	87.8
Autism/Control	78.6	75.6	79.6

Table.1 Accuracy of the learning machine

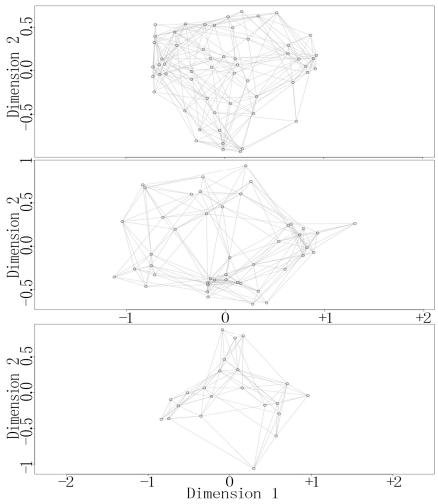

Figure.2 Isomap of different kind of subjects. From top to bottom: ADHD, Autism and control.

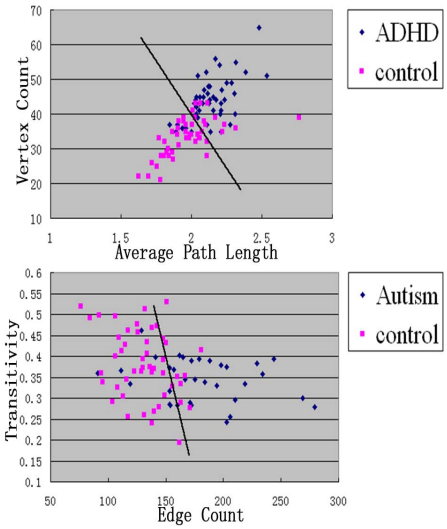

Figure.3 The scatter plot of different feather based subjects' distribution. The black line represents the ideal surface for classification.