

Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder

Chao-Zhe Zhu,^{a,b,*} Yu-Feng Zang,^{a,b} Qing-Jiu Cao,^c Chao-Gan Yan,^a Yong He,^d Tian-Zi Jiang,^b Man-Qiu Sui,^c and Yu-Feng Wang^c

^aState Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, PR China

^bNational Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, PR China

^cInstitute of Mental Health, Peking University, Beijing 100083, PR China

^dMcConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada H3A 2B4

Received 6 March 2007; revised 2 November 2007; accepted 14 November 2007

Available online 3 December 2007

In this study, a resting-state fMRI based classifier, for the first time, was proposed and applied to discriminate children with attention-deficit/hyperactivity disorder (ADHD) from normal controls. On the basis of regional homogeneity (ReHo), a mapping of brain function at resting state, PCA-based Fisher discriminative analysis (PC-FDA) was trained to build a linear classifier. Permutation test was then conducted to identify the brain areas with the most significant contribution to the final discrimination. Experimental results showed a correct classification rate of 85% using a leave-one-out cross-validation. Moreover, some highly discriminative brain regions, like the prefrontal cortex and anterior cingulate cortex, well confirmed the previous findings on ADHD. Interestingly, some important but less reported regions such as the thalamus were also identified. We conclude that the classifier, using resting-state brain function as classification feature, has potential ability to improve current diagnosis and treatment evaluation of ADHD.

© 2007 Published by Elsevier Inc.

Keywords: Fisher discriminative analysis; ADHD; ReHo; Resting state; fMRI

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is one of the most commonly diagnosed childhood behavioral disorders which affected approximately 5% of school-age children and characterized by the symptoms of inappropriate inattention, impulsivity, and hyperactivity. Children with ADHD have difficulties in controlling their behaviors or focusing their attentions which result in an

adverse effect on academic performance and social function. Moreover, 30%–60% of individuals diagnosed with ADHD in youth have symptoms that persist into adulthood (Biederman, 1998; Biederman et al., 2000). Currently available diagnosis and treatment evaluation of ADHD are mainly made according to the levels of the symptoms listed in the diagnostic criteria from DSM-IV (American Psychiatric Association, 1994). Ranking of the symptoms is usually conducted by the parents or teachers of the children, which is unfortunately subjective. Therefore more objective approaches are highly desired.

Structural and functional magnetic resonance imaging (MRI) techniques have been widely used in the quantitative analysis of the brain for ADHD, and various abnormalities have been reported as the objective evidences for some theoretical hypotheses on the disorder. Structural MRI studies have shown abnormalities of the whole brain and several specific brain areas, such as the frontal lobes, the basal ganglia, the parietal lobe, the occipital lobe, and the cerebellum in ADHD, in comparison to normal controls (Castellanos et al., 1996; Overmeyer et al., 2001; Sowell et al., 2003; Seidman et al., 2006). Using various experimental designs, task-related functional MRI (fMRI) studies found abnormal brain activation of ADHD in the dorsal anterior cingulate cortex (dACC), the ventrolateral prefrontal cortex (VLPFC), and the putamen (Bush et al., 1999; Durston et al., 2003; Teicher et al., 2000). Resting-state fMRI has also been used in the studies of ADHD and abnormalities were found in ACC, prefrontal cortex, putamen, temporal cortex, and cerebellum (Tian et al., 2006; Cao et al., 2006).

Although these studies have indicated that the pathophysiology of ADHD can be associated with the various brain regions, it has been argued that the analysis approaches based on the group-level statistics are less helpful to diagnosis (Seidman et al., 2004). Recently, increasing attention has been directed to the applications of pattern recognition techniques in brain image analysis. Compared with the traditional group-level analysis, such techniques can distinguish normal from abnormal at individual subject level. Hence they are potentially useful procedures for clinical diagnostic purposes. In such studies, various structural character-

* Corresponding author. State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, PR China. Fax: +86 10 58806154.

E-mail addresses: czzhu@bnu.edu.cn, chaozhe.zhu@yahoo.com.cn (C.-Z. Zhu).

Available online on ScienceDirect (www.sciencedirect.com).

istics or functional properties of the brain derived from neuroimaging data were used as the feature for classification. For structural MRI, shape of brain structures of interest, deformation field for registration, map of gray matter membership, map of cortical thickness, and so on have been employed to discriminate patients (e.g., schizophrenia and Alzheimer's disease) from healthy controls (Golland et al., 2002; Fan et al., 2007; Kawasaki et al., 2007; Yoon et al., 2007). For task-related fMRI, the original time series and activation maps have been used for discrimination of mental disorders (Kontos et al., 2004; Shinkareva et al., 2006).

Though the promising studies on some psychiatric disorders were reported using classification techniques, few were conducted on ADHD. In addition, though the task-induced brain activities have been used as the classification feature, the brain activities revealed by the resting-state fMRI have not been considered for. In the resting state, low-frequency (0.08 Hz) fluctuations (LFF) of the fMRI signal are considered to be related to spontaneous neuronal activity and the synchrony of LFF was first used to identify functional connectivity among motor cortices (Biswal et al., 1995) and then extended to other functional systems, e.g., between bilateral visual cortices, bilateral auditory cortices, bilateral amygdala, bilateral thalamus, and within the language system (Lowe et al., 1998; Cordes et al., 2000; Stein et al., 2000; Hampson et al., 2002). Abnormal LFF has been reported for ADHD (Tian et al., 2006; Cao et al., 2006) and other neuropsychiatric disorders (Li et al., 2002; Greicius et al., 2004; Liu et al., 2006). These findings inspired us to use the brain activity revealed by the resting-state fMRI as a classification feature to differentiate boys with ADHD from their normal controls in this study.

Besides classification features, learning algorithm is also an important aspect of a classification system and has a great impact on the performance of a classifier. Fisher discriminative analysis (FDA), with the advantages of simplicity, sound theoretical foundation, and ease of interpretation, has been widely used in the domain of pattern recognition (Duda et al., 2001). The traditional FDA, however, cannot be used directly when the within-scatter matrix is singular in the case of small sample size. In order to solve the problem, principal component analysis (PCA) is usually first employed to reduce the dimension of feature space and FDA was then performed in the transformed feature space. Such a PC-FDA approach was first proposed for face recognition (Swets and Weng, 1996) and extended into the field of task-related fMRI data analysis (Mørch et al., 1997). Recently, PC-FDA and its variants (e.g., canonical variate analysis, CVA) have been widely used in the analysis of task-related fMRI data (Carlson et al., 2003; LaConte et al., 2003; Strother et al., 2004; Mourao-Miranda et al., 2005). Carlson et al. (2003) used PC-FDA to investigate patterns of activity in the categorical representation of objects. LaConte et al. (2003) applied CVA on PCA basis to evaluate the impact of preprocessing choices and the number of principal components passed to the CVA on within-subject prediction and reliability. Strother et al. (2004) combined PCA and CVA to optimize preprocessing choices in fMRI data analysis. Mourao-Miranda et al., 2005 applied FDA and support vector machines (SVM) on PCA components for classification of different brain cognitive states using the whole-brain fMRI data. In this study, the PC-FDA was further extended into the resting-state fMRI analysis for the discrimination of mental disorders and applied to ADHD. Our initial trials were presented elsewhere (Zhu et al., 2005).

The rest of the article is organized as follows: the resting-state classification feature, learning algorithm, classifier performance,

and discriminative pattern are detailed in the Methodology section. Materials and experimental results are presented in the Materials and Results sections, respectively. The discussions are in the Discussion section followed by the Conclusion section.

Methodology

Regional homogeneity

As a mapping of brain function, regional homogeneity (ReHo) was originally proposed to measure the regional synchrony of brain activity recorded by fMRI (Zang et al., 2004). At a given voxel p , ReHo was defined as the Kendall's coefficient of concordance (KCC) of the time series of p with those of its $K-1$ nearest neighbors

$$W(p) = \frac{\sum(R_i)^2 - n(\bar{R})^2}{\frac{1}{12}K^2(n^3 - n)} \quad (1)$$

where W , ranging from 0 to 1, is the KCC within a cluster made up of voxel p and its $K-1$ neighbors; R_i is the sum rank of the i th time point; $\bar{R} = \frac{1}{2}K(n+1)$ is the mean of the R_i s; K is the number of time series within the measured cluster (one given voxel plus the number of its neighbors); n is the number of ranks. A larger ReHo value at a given voxel indicates higher regionally temporal synchronization within the cluster. According to Eq. (1), an individual W map (i.e., ReHo map) was obtained on a voxel by voxel basis for each subject.

In this study, ReHo map, derived from the resting-state fMRI, was used as our classification feature. To calculate the resting-state ReHo map, the preprocessed resting-state fMRI data were temporally band-pass filtered (0.01–0.08 Hz) to reduce the effect of low-frequency drifts and high-frequency noise (Biswal et al., 1995; Lowe et al., 1998) using AFNI (Cox, 1996). The number of neighboring voxels was set with 26 (i.e., $K=27$) as that in Zang et al. (2004). Regions outside the brain, contributing nothing to the discrimination, were masked and removed from the ReHo maps in order to reduce feature dimensionality. The structural images of an arbitrarily selected subject were shown with their corresponding map of gray matter membership and ReHo in Fig. 1.

PC-FDA

Suppose there are two classes of samples, each with a feature vector $x \in \mathfrak{R}^D$. FDA algorithm is trained with the samples to seek out the optimal projective direction, $\omega \in \mathfrak{R}^D$, along which the two classes of projected samples are separated with maximal ratio of between-class distance and within-class variability. A classifier can then be easily built up using the one-dimensional projected features. In this work, however, the total number of training samples, N , is far smaller than the feature dimension D , which is defined as the number of brain voxels in the ReHo map. Accordingly, PC-FDA approach was adopted to solve the ill-posed problem. PCA attempts to find linear combinations of the original features that explain most of the variance in these features using just a few components. A major concern is how many components should be reserved in PC-FDA to construct the dimension-reduced subspace? From the theory of linear algebra, the $N-2$ eigenvectors with the largest eigenvalues can be reserved to make FDA feasible in the subspace. Representing an original sample $x \in \mathfrak{R}^D$ in the subspace results in a low-dimension vector $y \in \mathfrak{R}^d$ ($d=N-2$) and the optimal projective direction, $\omega_d \in \mathfrak{R}^d$, can then be worked out directly. The mathematical

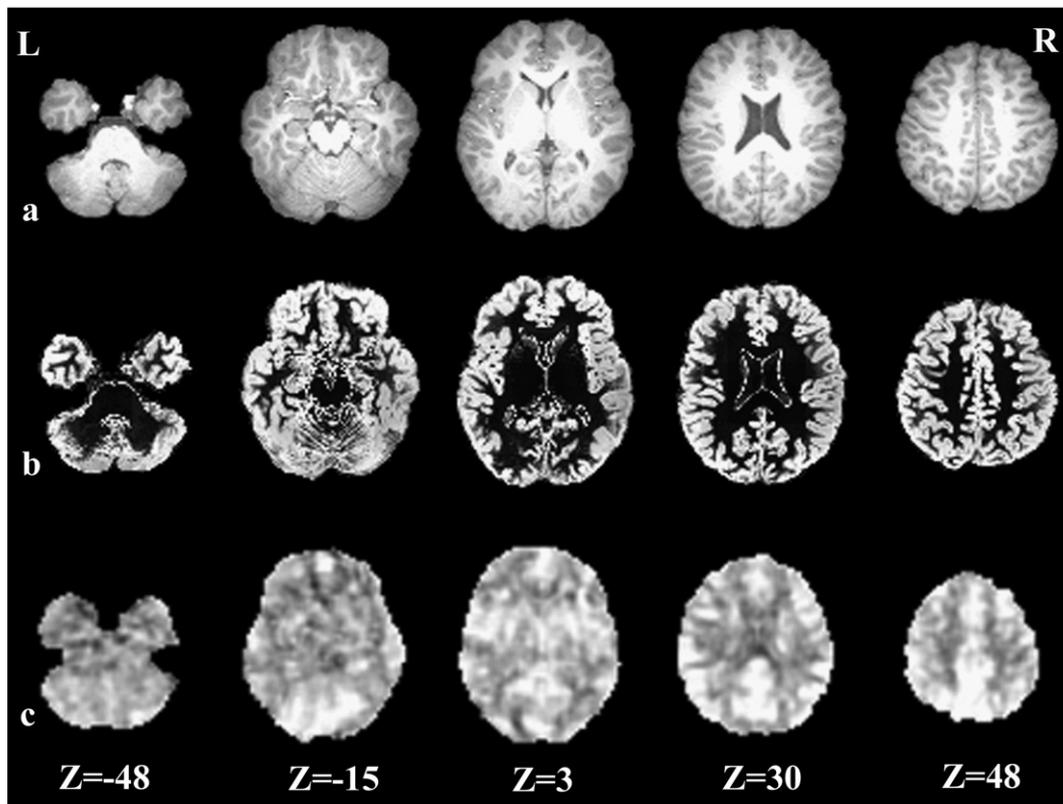


Fig. 1. Slice views of (a) the normalized structural images, (b) the corresponding membership map of gray matter, and (c) ReHo map, with z value indicating slice position; L and R indicating the left and right side of the brain.

description of PC-FDA was given in Appendix A. and the more theoretical foundation of PC-FDA was detailed elsewhere (Yang and Yang, 2003).

Classifier performance

Leave-one-out cross-validation was used in this work to estimate the performance of our classifier. The flow chart of leave-one-out cross-validation for our PC-FDA was shown in Fig. 2. Suppose there are N samples in total. The leave-one-out cross-validation trains classifier N times, each time leaving out one of the samples from training, but using the omitted one to compute classification error. Generalization rate, sensitivity, and specificity can be defined on the basis of results of leave-one-out cross-validation to quantify the performance of a classifier.

$$\text{Generalization Rate} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{TN} + \text{FP}) \quad (2)$$

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (4)$$

where TP is the number of patients correctly predicted; TN is the number of normal controls correctly predicted; FP is the number of normal controls classified as patients; FN is the number of patients classified as normal controls. We can see that the Sensitivity indicates the proportion of patients correctly predicted, while the Specificity indicates the proportion of normal controls correctly predicted. The Generalization Rate is the overall proportion of samples correctly predicted.

Discriminative pattern

If a classifier can predict new samples with a performance better than random accuracy, then we can believe that the two populations which the samples are drawn from are indeed different, and the classifier can capture the population differences (Golland and Fischl, 2003). When using FDA, it is straightforward to quantify and visualize the discriminative pattern by using the projective direction vector $\omega \in \mathcal{R}^D$. After the projective direction vector was reshaped back into a 3D matrix, the discriminative pattern was represented as a brain (Fisher brain). Thus the differences between the boys with ADHD and normal controls can be explored in a stereotaxic space, such as Talairach and Tournoux coordinates (Talairach and Tournoux, 1988). However, the optimal projective direction vector, $\omega_d \in \mathcal{R}^d$, cannot be reshaped directly into a brain space since it was obtained in the PCA dimension-reduced subspace. To solve the problem, a back projection procedure was carried out as in Eq. (A.12). The Fisher brain constructed from all 20 sample subjects was presented in Fig. 5(a).

As we know from the formula (A.3), the discriminating score is the weighted sum of ReHo values at each voxel and the weights were determined by the values at corresponding voxels in the Fisher brain. Thus, the larger the amplitude (positive or negative) of a voxel in the Fisher brain, the more the voxel contributes to the final discrimination. To determine the most important brain regions for discriminating ADHD and normal controls, a critical threshold is required, for each voxel, to set on the absolute amplitude of the Fisher brain. A statistically meaningful threshold can be derived, for each voxel, using a permutation test method. The permutation

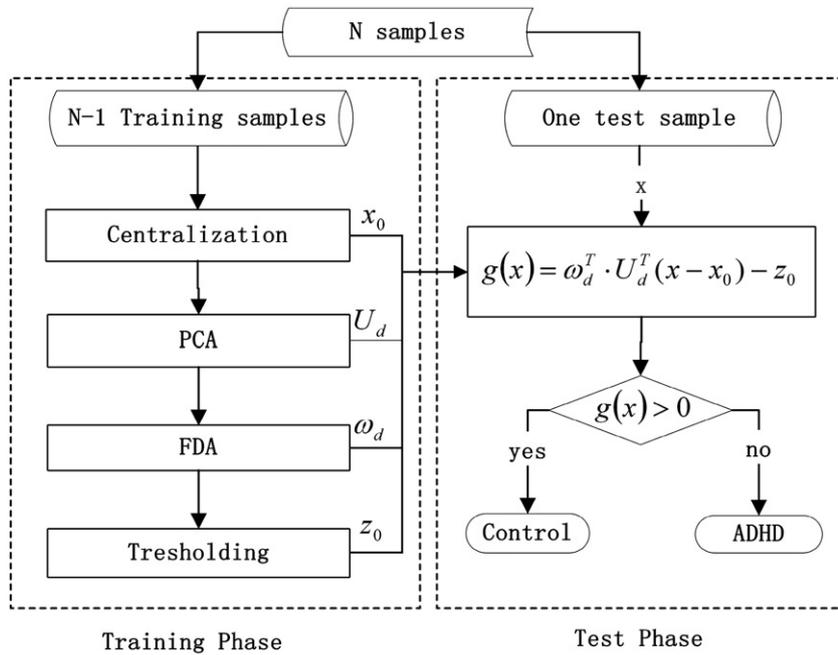


Fig. 2. Flow chart of leave-one-out cross-validation for PC-FDA. For detailed descriptions of the variables in the flow chart, refer to Appendix A..

test is one of nonparametric techniques which are used to estimate empirically the distribution of a statistic under a null hypothesis (Holmes et al., 1996; Frackowiak et al., 2004; Mourao-Miranda et al., 2005). A full permutation test, instead of an approximation permutation test, was applied to this study in order to produce a more powerful test (Frackowiak et al., 2004). In a full permutation test, all possible relabelings were used to compute the statistic to generate the full permutation distribution. Based on these null probability distributions and the observed statistic corresponding to the actual labeling, the P value, defined as the probability of a statistic value as or more extreme than the one observed, can be worked out. The smaller is the P value, the more reasonable to reject the null hypothesis. Usually a threshold of $P_0=0.05$ or $P_0=0.01$ is meaningful. In this study, we selected $P_0=0.05$.

In this work, the statistic is the absolute magnitude of each voxel of the Fisher brain and the null hypothesis is that there are no differences in the ReHo maps between ADHD patients and normal controls. The original class labels were permuted $C_N^{N_1}$ times for a full permutation, where N is the total number of samples and N_1 is the number of normal controls. For each relabeling, PC-FDA was trained and a Fisher brain was generated. After all possible relabelings were considered the full permutation distribution was estimated for each voxel under the null hypothesis. Based on the null probability distributions and the observed Fisher brain corresponding to the actual labeling, a P value was computed for each voxel and a P map was then constructed. Finally those voxels with P value smaller than the predefined threshold of P_0 were identified as significant for discrimination.

Materials

Participants

Participants included 12 boys with ADHD (age range: 11.00–16.50 years, mean±SD 13.34±1.44 years) and 12 age-matched

(within 0.5 year) control boys. All subjects are right-handed and have an intelligence quotient (IQ) >80. Written informed consent was obtained from parents or guardians of all participants. All children agreed to participate in this study. Three patients and one control were excluded from further analysis because of excessive head motion (translation greater than 1.2 mm or rotation greater than 1.2°). Hence there were 9 ADHD boys and 11 controls left for discriminative analysis.

MRI acquisition

The MR images were acquired on the SIEMENS TRIO 3-Tesla scanner in Institute of Biophysics, Chinese Academy of Sciences. For each subject we concerned the following two sets of imaging data: resting-state fMRI time courses and 3D structural MRI. Echo planar imaging (EPI) blood oxygenation level dependent (BOLD) images were acquired axially with the following parameters: 2000/30 ms (TR/TE), 30 slices, 4.5/0 mm (thickness/gap), 220×220 mm (FOV), 64×64 (resolution), 90° (flip angle), the whole session lasted for 480 s. 3D spoiled gradient-recalled whole-brain volume was acquired sagittally with the following parameters: 1700/3.92 ms (TR/TE), 192 slices, 1.0/0 mm (thickness/gap), 256×256 mm (FOV), 256×256 (resolution), 12° (flip angle).

Data preprocessing

The first 10 volumes of each functional time series were discarded for participant adaptation to the scanning. Preprocessing procedures for fMRI signals included motion correction, within-subject registration, time aligning across slices, time series linear detrending, voxels resampling to $3\times 3\times 3$ mm³, and spatial normalization. 3D structural MR images were also spatially normalized into the MNI space and non-brain tissue voxels were masked out. Among these processes, the linear detrending was

undertaken using the AFNI (Cox, 1996) and the others using the SPM2 (www.fil.ion.ucl.ac.uk/spm).

Results

As we described in Materials, there were totally 20 samples for discriminative analysis in this work, including 9 ADHD and 11 controls. First, the classifier was trained with all the 20 samples and then tested with the same 20 samples to indicate the separability of the classifier on the training set. Then a 20-round leave-one-out cross-validation, each with 19 training and 1 test samples, was conducted to estimate the prediction ability of the classifier. Classification results were listed in the top row of Table 1 from which we found that the training error was zero; Generalization Rate, Sensitivity, and Specificity of our resting-state fMRI based classifier computed from leave-one-out cross-validation were 85%, 78%, and 91% respectively.

In order to comprehensively investigate the discriminative ability of the proposed classification feature (ReHo map), structural brain information was also employed as a comparing feature. Map of gray matter (GM) membership has been used in voxel-based morphometry (VBM) analysis to identify and characterize abnormalities in GM density in ADHD (Overmeyer et al., 2001). The map presents, at each voxel, the membership of the local brain tissue belonging to GM and is usually derived from the algorithms of soft brain tissue segmentation. Multi-context fuzzy clustering (MCFC) was used here to generate the map because it is insensitive to the intensity inhomogeneities (Zhu and Jiang, 2003) and the resultant map of GM membership corresponding to the MR image in Fig. 1(a) was shown in Fig. 1(b). The classification results of the PC-FDA classifier based on the map of GM membership were depicted in the fourth row of Table 1. We found that the brain function at resting state (i.e., ReHo map) had a significantly higher Generalization Rate of 85% when compared with the structural feature of GM membership (55%).

The distributions of discriminative scores of both the training and test samples in each round of the cross-validation were shown in Figs. 3(a) and (b) for the ReHo map and the GM membership, respectively. From Fig. 3(a) we found that, for the ReHo map, only one testing control (round 5) and two testing patients (round 12 and 17) out of the 20-round leave-one-out cross-validation were placed on the wrong side of the boundaries by the classifier. Moreover, the within-class variations of discriminative scores of training samples were small relative to the between-class distance. For the GM membership, however, much more test samples were misclassified and a significantly larger within-class variance appeared as shown in Fig. 3(b).

For further validation, our PC-FDA classifier was compared with other two typical linear classifiers, Batch Perceptron (Duda et al., 2001) and linear support vector machine (SVM) (Vapnik,

1998). Batch Perceptron algorithm, for a linearly separable problem, can yield a separating hyperplane by minimizing a perceptron criterion function. Geometrically, the perceptron criterion function is proportional to the sum of the distances from misclassified samples to the separating hyperplane. In this study, the learning rate of Batch Perceptron was set to 1. Considering the stochastic property of the Batch Perceptron algorithm, we repeated the 20-round leave-one-out test for 10 times, each with a random initialization. Then the averaged results were used to quantify the final performance. SVM, based on the principle of structural risk minimization, is one of the most successful classification techniques over the last decade. For a linearly separable problem, linear SVM seeks a separating hyperplane maximizing the geometric margin, defined as the minimal Euclidean distance between any training example and the hyperplane. Intuitively, the margin measures how well the two classes were separated by a hyperplane. The classification results of the two alternative methods were listed in the second and third row in Table 1, respectively. From Table 1, we found that the Batch Perceptron approach yielded a very low Generalization Rate (55%). Both the linear SVM and the PC-FDA classifier performed much better than the Batch Perceptron. Moreover, the Generalization Rate of the PC-FDA classifier (85%) was higher than that of the linear SVM (75%) (see Discussion).

To give more insight into the Fisher brain, we compared it with the traditional mass-univariate analysis approach, i.e., *t*-test map (denoted as *T* brain) from the viewpoint of both mathematics and classification ability. The mathematical relationship between the *T* brain and the Fisher brain was detailed in Appendix B. In general, the *T* brain is a special case of the Fisher brain when voxel is assumed to be independent from each other and standard deviation at each voxel is introduced as a weighting factor. Thus the *T* brain can also be treated as a projective direction and we can evaluate their classification abilities by comparing the performance of the two projective directions in classifying the ReHo maps of ADHD and normal children. The Generation Rate of the classifier using the *T* brain as the projective direction was 55% which was significantly lower than that using the Fisher brain (85%). Moreover, as shown in Fig. 4, the discriminative scores derived from the *T* brain demonstrated a significantly larger within-class variation than that of the Fisher brain.

In addition to sensitively differentiate ADHD from normal controls, our classifier can also be used to characterize how the discriminative information is represented in the brain, which is important to assist basic and clinical neuroscience researchers to elucidate the pathophysiology. As shown in Fig. 5(b) the abnormal brain regions identified by the permutation test on the Fisher brains, like the prefrontal cortex and the anterior cingulate cortex, well confirmed previous findings on ADHD. Moreover, the thalamus was identified which was less reported in literatures.

Table 1
Classification results

Discriminative model		Classifier performance				
Classification feature	Classifier type	Training set correct rate	Leave-one-out cross-validation			
			Generalization Rate	Sensitivity	Specificity	
Function	ReHo Map	PC-FDA	100%	85%	78%	91%
		Linear SVM	100%	75%	56%	91%
		Batch Perceptron	100%	55%	44%	64%
Structure	GM membership	PC-FDA	100%	55%	56%	55%

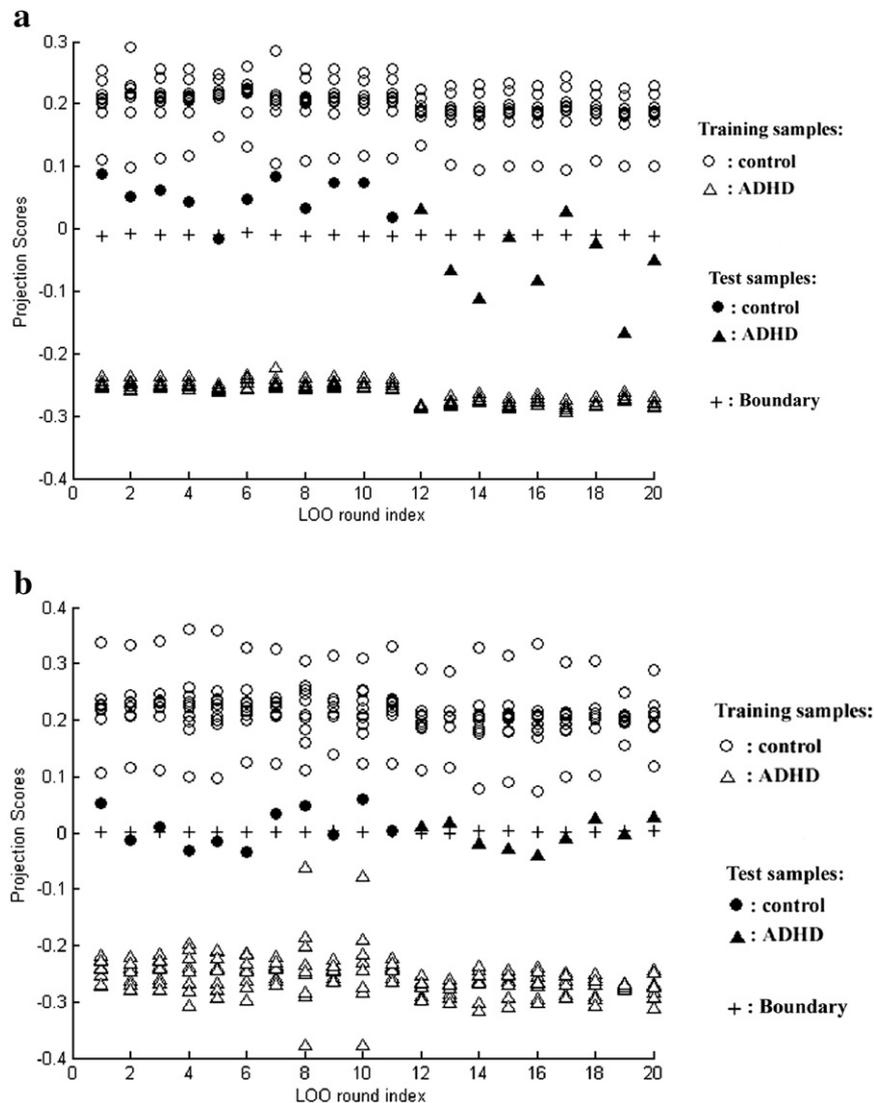


Fig. 3. Distributions of discriminative scores of 20-round leave-one-out (LOO) cross-validation of PC-FDA for (a) ReHo map and (b) GM membership.

We also compared the results of the Fisher brain and the T brain at the significant level ($P=0.05$). Generally speaking, the patterns of identified regions of the two approaches were quite similar. However, the Fisher brain demonstrated a larger size at most of the identified regions than that of the T brain.

Discussions

The resting-state fMRI currently received more and more interest since a baseline state is fundamental in understanding human brain functions (Raichle and Mintun, 2006) and has been intensively employed to study healthy (Biswal et al., 1995; Lowe et al., 1998; Cordes et al., 2000; Stein et al., 2000; Hampson et al., 2002) and abnormal (Li et al., 2002; Greicius et al., 2004; Tian et al., 2006; Cao et al., 2006; Liu et al., 2006) brains. Moreover, the resting-state fMRI, asking patients nothing but to remain still with eyes closed, is of more potential applications in clinical studies than the task-related fMRI where patients are required to follow relatively complicated cognitive tasks. Selection of classification feature is an important issue in pattern recognition and it greatly affects the classifier

performance. In this study, the resting-state fMRI was extended into the field of discriminative analysis of mental disorders. ReHo map, derived from resting-state fMRI data, has been used as the index of resting-state brain function to investigate regional spontaneous neural activity of patients with ADHD (Cao et al., 2006), which suggested that ReHo map could be potentially useful in revealing the pathophysiology of psychiatric disorders during resting state. Accordingly, the resting-state fMRI, reflecting spontaneous neural activities, is a promising feature for the classification of mental disorders and we used ReHo map in this study as the classification feature for ADHD.

As noted in the Introduction section, studies have also suggested various brain abnormalities in brain structures in ADHD population. For example, the map of GM membership was previously used to identify and characterize abnormalities in GM density in ADHD (Overmeyer et al., 2001). In order to validate the discriminative ability of the resting-state fMRI (i.e., ReHo map), structural MRI feature (i.e., GM density map) was used as a comparing classification feature. Though ADHD related abnormalities have been reported using ReHo (Cao et al., 2006) and GM

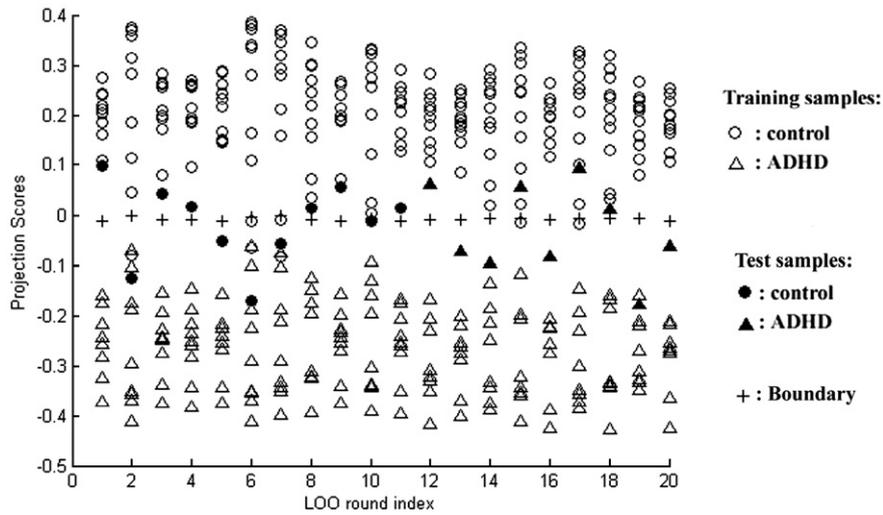


Fig. 4. Distribution of discriminative scores of 20-round leave-one-out (LOO) cross-validation when using T brain as projective direction.

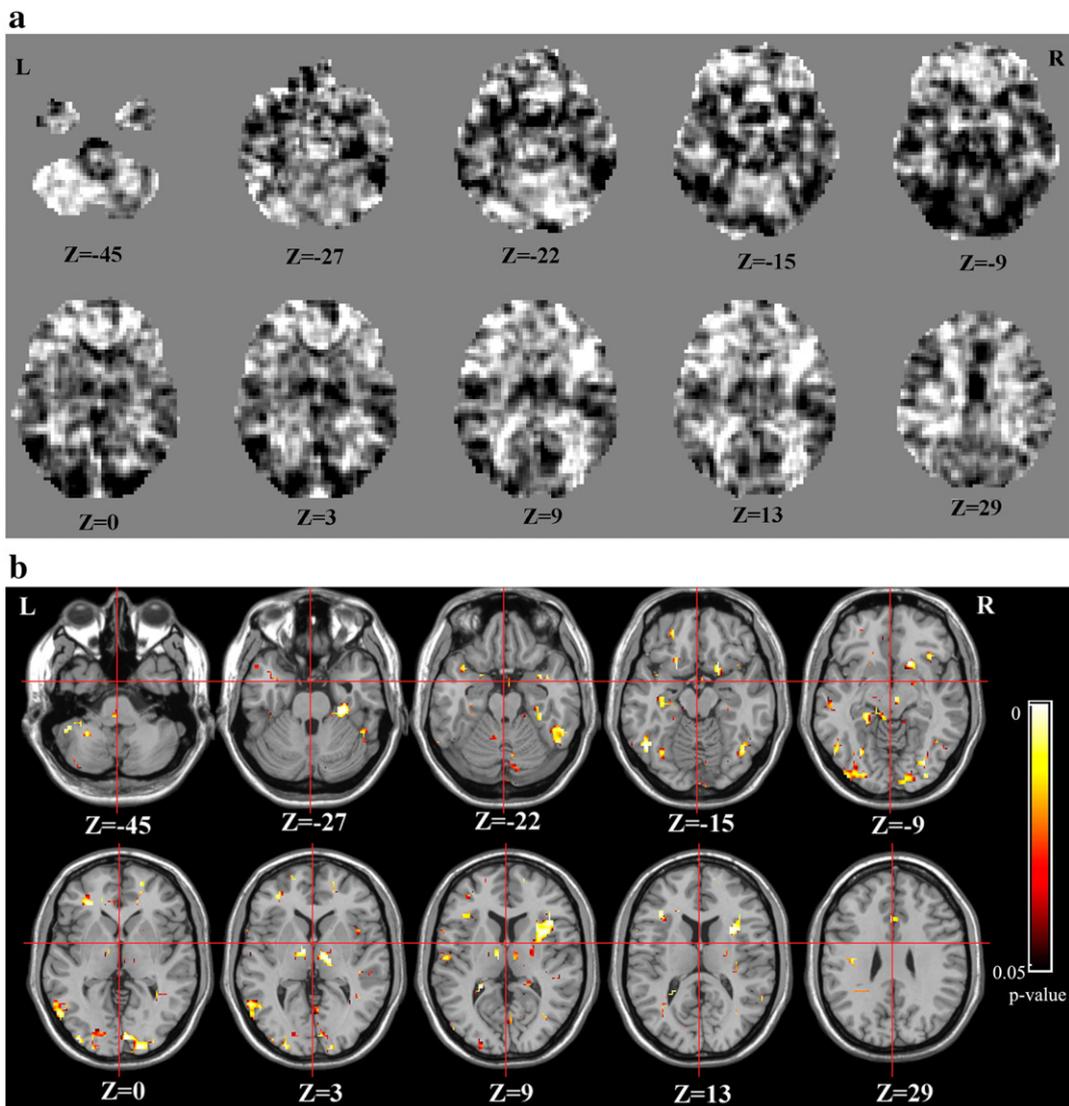


Fig. 5. Slice views of 3D images. (a) Fisher brain and (b) highly discriminative regions overlaid on structural images, with z value indicating slice position; L and R indicating the left and right side of the brain; color bar indicating the significant level (i.e., P value) of detected regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

membership (Overmeyer et al., 2001) independently, our experimental results in Table 1 clearly demonstrated that ReHo is more powerful (both in sensitivity and specificity) for discrimination of ADHD than GM membership. It implies that brain function may be more susceptible than brain structures for ADHD and the features derived from resting-state fMRI thus have more potential to assist in the diagnosis for ADHD.

Besides the classification feature, learning algorithm is another important aspect of pattern recognition. Although more complex nonlinear learning algorithms exist, linear learning algorithms were adopted in this study because they are more insensitive to overfitting problems than nonlinear ones, especially in the case of high feature dimension and small sample size (Mørch et al., 1997). In addition, the explanation of linear classifier is much more straightforward than that of nonlinear ones. In a linear classifier, the projective direction itself, like Fisher brain in FDA, is a natural way to clarify which voxels or brain regions are more important than others for the final discrimination. However, there is no uniform expression for the discriminative pattern between two populations due to the complicated separating boundaries of nonlinear classifiers, and thus it is difficult for a basic or clinical neuroscientist to understand the mechanism underpinning the classifier.

For further validation, the PC-FDA classifier was compared with two other typical linear classifiers, Batch Perceptron and linear SVM. Batch Perceptron algorithm, starting with a random initialization, may converge to any one of the possible solutions that correctly classifies all training samples (in the linear separable case). Therefore this algorithm does not care whether the solution has the best capacity of generalization or not. FDA and linear SVM, however, seek their own best one in all feasible solutions. FDA tries to find the solution to separate two classes' samples with maximal ratio of between-class distance and within-class variability. Linear SVM seeks out the solution to maximize the geometric margin, defined as the minimal Euclidean distance between any training example and the separating hyperplane. Therefore, the two linear classifiers should outperform Batch Perceptron and these are the theoretical reasons why Batch Perceptron algorithm gave the lowest Generation Rate of 55% among the three linear classifiers. Our experimental results further showed that the Generation Rate of PC-FDA (85%) is higher than that of linear SVM (75%). Though Mourao-Miranda and colleagues suggested, on the basis of their experimental results, that linear SVM had better generation performance than other linear classifiers (Mourao-Miranda et al., 2005). Some other studies, however, have found that SVM did not always outperform FDA (Yang et al., 2001; Tang et al., 2003). The theoretical comparison between FDA and linear SVM is not easy since they have different objective functions for classification and one possible reason to the difference in performance of the two classifiers may be the limited number of examples. Considering the small sample size (20 in total), the 10% difference in the Generation Rate between the two classifiers is, in practice, only two samples. After we carefully examined the distribution of the discriminative scores of the two samples additionally misclassified by linear SVM (LOO round 15 and 18), we found that both of them were located much near the separating boundary as shown in Fig. 3(a). Classification results are sensitive to such samples very near the separating boundary. If the decision on such samples were ignored the two classifiers then have no significant difference in classification performance.

As shown in Fig. 5(b), several localized areas were identified as highly discriminative regions, including the anterior cingulate gyrus, prefrontal cortex, putamen, occipital cortex, temporal cortex, cerebellum, and thalamus. Evidences from both structural and task fMRI studies have indicated that most of these regions may be involved in the pathology of ADHD (Castellanos et al., 1996; Durston et al., 2003; Bush et al., 1999; Sowell et al., 2003). The previous resting-state fMRI study also found the abnormality of prefrontal cortex, putamen, temporal cortex, and cerebellum in ADHD (Cao et al., 2006).

An involvement of the thalamus in the regulation of cortical arousal through thalamo-cortical connections has been reported earlier (Montaron and Buser, 1988). Then several functional neuroimaging studies found that the thalamus was activated during attentional tasks (Lawrence et al., 2003; Sturm et al., 2004; Fan et al., 2005) indicating the important role of the thalamus in attentional processes (Buchsbaum et al., 1990). Although the thalamus is presumed a very critical brain region subserving normal attentional processes, few studies from neuroimaging, to our knowledge, have found abnormality in thalamus in ADHD. In this study, however, the thalamus was identified in ADHD. The reason, as mentioned previously, may be that the traditional analysis approaches are essentially univariate. FDA is, in contrast, multivariate in nature. That is all voxels are explored simultaneously and the covariance structure is taken into account when quantifying group differences. Thanks to the intrinsically spatial correlations of brain activities, FDA should (at least, in theory) be more sensitive to the group difference than traditional analysis methods and, thus, the classifiers using Fisher brain as the projective direction be more powerful than that using voxel-wise *t*-test map (*T* brain). This has been supported by our experimental results. The Generation Rate of the *T* brain (55%) was much lower than that of the Fisher brain (85%). In addition, the discriminative scores derived from the *T* brain demonstrated a significantly larger within-class variation than that of the Fisher brain as shown in Fig. 4. In addition, we performed a statistical analysis on the *P* values over the identified bilateral thalamus for the Fisher brain and the *T* brain respectively and found that the distribution of significant value (i.e., *P* value) in the bilateral thalamus identified by the Fisher brain (mean±SD: 0.0188±0.0146) is obviously lower than that for the *T* brain (mean±SD: 0.0252±0.0128). This result indicated that the thalamus can still be identified in the Fisher brain, even with a more rigorous *P* threshold at which the thalamus cannot be identified by the *T* brain or other univariate approaches.

Though our classifier demonstrated a good performance from cross-validation, it is still a challenge to generalize our findings into clinical applications due to the limited size of samples with high dimensional feature and the impact of acquisition hardware and pulse sequence parameters for scanning. Accordingly it would be very important to evaluate our classifier with larger sample size and multi-center imaging data in the future. In addition, brain function features derived from task-related fMRI will be compared with the ReHo map in the future to further validate our proposed brain function feature derived from resting-state fMRI. Finally, future studies should be complemented to unravel the role of thalamus in the pathology of ADHD.

Conclusions

In this study, resting-state fMRI was proposed, for the first time, as classification feature and successfully applied into the

discriminative analysis for ADHD in the framework of PC-FDA. The discriminative model provided important evidence for potentially improving the diagnosis of ADHD. More importantly, other discriminative models could also be built in a similar way to identify subtypes of ADHD and to evaluate and predict treatment response of ADHD or extended even to other psychiatric disorders.

Acknowledgments

The authors thank Dr. Jun Wang and the two anonymous reviewers for constructive suggestions. This work was supported by the Natural Science Foundation of China, Grant No. 30500130, the National Key Basic Research and Development Program (973) Grant No. 2003CB716101, the Natural Science Foundation of China (NSFC, Chinese-Finnish NEURO program, No. 30470575), and Project of Science and Technology, Beijing (Y0204003040831).

Appendix A. PC-FDA

Mathematically, the objective function of FDA in the formula (A.1) is to be maximized.

$$J(\omega) = \frac{\omega^T S_b \omega}{\omega^T S_w \omega} \quad (\text{A.1})$$

where $S_w = \sum_{i=1}^{N_1} (x_i^{(1)} - \mu^{(1)})(x_i^{(1)} - \mu^{(1)})^T + \sum_{i=1}^{N_2} (x_i^{(2)} - \mu^{(2)})(x_i^{(2)} - \mu^{(2)})^T$ and $S_b = (\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})^T$ are the within-class and between-class scatter matrix respectively; $x_i^{(1)} (i=1, 2, \dots, N_1)$ and $x_i^{(2)} (i=1, 2, \dots, N_2)$ are the feature vectors, $\mu^{(1)}$ and $\mu^{(2)}$ the mean feature vector, and N_1 and N_2 the sample sizes, of the two classes, respectively. Theoretically, the optimal projective direction can be given by:

$$\omega = S_w^{-1} (\mu^{(1)} - \mu^{(2)}) \quad (\text{A.2})$$

Projecting each sample x onto ω results in a one-dimensional discriminating score, $z \in \mathfrak{R}^1$, by the inner product operation:

$$z = \langle x, \omega \rangle = \omega^T \cdot x \quad (\text{A.3})$$

Finally, the classification threshold, $z_0 \in \mathfrak{R}^1$, is determined by:

$$z_0 = (N_1 \bar{z}^{(1)} + N_2 \bar{z}^{(2)}) / N, \quad N = N_1 + N_2 \quad (\text{A.4})$$

where $\bar{z}^{(1)}$ and $\bar{z}^{(2)}$ are the centers of discriminating score of class 1 and 2 respectively. Finally a linear classifier is generated as

$$\begin{aligned} g(x) &= \omega^T \cdot x - z_0 > 0 & \forall x \in \text{class 1} \\ g(x) &= \omega^T \cdot x - z_0 < 0 & \forall x \in \text{class 2} \end{aligned} \quad (\text{A.5})$$

PCA is usually implemented by singular value decomposition (SVD). Suppose all training samples were contained in a data matrix $X \in \mathfrak{R}^{D \times N}$ with one column x for each subject and the sample center is $\mu \in \mathfrak{R}^D$.

$$X = [x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}, x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)}] \quad (\text{A.6})$$

$$\mu = \frac{1}{N} \left(\sum_{i=1}^{N_1} x_i^{(1)} + \sum_{i=1}^{N_2} x_i^{(2)} \right) \quad (\text{A.7})$$

Then the centered data matrix is

$$X_0 = [x_1^{(1)} - \mu, x_2^{(1)} - \mu, \dots, x_{N_1}^{(1)} - \mu, x_1^{(2)} - \mu, x_2^{(2)} - \mu, \dots, x_{N_2}^{(2)} - \mu] \quad (\text{A.8})$$

SVD on X_0 results in the following decomposition

$$X_{0D \times N} = U_{D \times N} S_{N \times N} T_{N \times N}^T \quad (\text{A.9})$$

S is a diagonal matrix with nonnegative diagonal elements s_i ($i=1, 2, \dots, N$) (i.e., singular values) in decreasing order. U contains the corresponding eigenvectors $\alpha_i \in \mathfrak{R}^D$ ($i=1, 2, \dots, N$).

Suppose d (less than D) components are reserved to generate the dimension-reduced space \mathfrak{R}^d . Then representing an original sample $x \in \mathfrak{R}^D$ in the subspace results in a low-dimension vector y

$$y = U_d^T (x - \mu), \quad U_d = [\alpha_1, \alpha_2, \dots, \alpha_d] \quad (\text{A.10})$$

Thus (A.2) can be directly used in \mathfrak{R}^d to find the optimal projective direction, $\omega_d \in \mathfrak{R}^d$, and the final classifier is then obtained as

$$\begin{aligned} g(x) &= \omega_d^T \cdot U_d^T (x - \mu) - z_0 > 0 & \forall x \in \text{class 1} \\ g(x) &= \omega_d^T \cdot U_d^T (x - \mu) - z_0 < 0 & \forall x \in \text{class 2} \end{aligned} \quad (\text{A.11})$$

Finally, ω_d in \mathfrak{R}^d can be back projected to the original feature space \mathfrak{R}^D according to the formula (A.10)

$$\omega = U_d \omega_d, \quad U_d = [\alpha_1, \alpha_2, \dots, \alpha_d] \quad (\text{A.12})$$

Appendix B. Relationship between Fisher brain and T brain

T brain is constructed by doing a t -test at each and every voxel between two groups:

$$T = [t_1, t_2, \dots, t_D]^T \quad (\text{B.1})$$

where $t_k = \frac{\mu_k^{(1)} - \mu_k^{(2)}}{\sigma_k \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$ is the t -value at voxel k ($k=1, 2, \dots, D$) and $\mu_k^{(1)}, \mu_k^{(2)}$ is mean observation of two classes; σ_k is the pooled standard deviation. T brain results essentially from mass-univariate statistical approach since each voxel was treated independently. Under the assumption of independency, within-class scatter matrix in (A.2) is degenerated into a diagonal matrix:

$$S_w = \begin{bmatrix} SS_1^{(1)} + SS_1^{(2)} & 0 & 0 & 0 \\ 0 & SS_2^{(1)} + SS_2^{(2)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & SS_D^{(1)} + SS_D^{(2)} \end{bmatrix} \quad (\text{B.2})$$

where $SS_k^{(1)}$ and $SS_k^{(2)}$ are the sum of squares of class 1 and 2 at voxel k respectively. The pooled standard deviation at voxel k is estimated by

$$\sigma_k = \sqrt{(SS_k^{(1)} + SS_k^{(2)}) / (N_1 + N_2 - 2)} \quad (\text{B.3})$$

Hence,

$$S_w = (N_1 + N_2 - 2) \cdot \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_D^2 \end{bmatrix} \quad (\text{B.4})$$

Therefore, the optimized projective direction vector is

$$\omega = S_w^{-1} \left(\mu^{(1)} - \mu^{(2)} \right) \\ = \frac{1}{N_1 + N_2 - 2} \left[\frac{\mu_1^{(1)} - \mu_1^{(2)}}{\sigma_1^2} \quad \frac{\mu_2^{(1)} - \mu_2^{(2)}}{\sigma_2^2} \quad \dots \quad \frac{\mu_D^{(1)} - \mu_D^{(2)}}{\sigma_D^2} \right]^T \quad (\text{B.5})$$

Replacing $\frac{\mu_k^{(1)} - \mu_k^{(2)}}{\sigma_k \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$ with t_k , we get

$$\omega = \frac{\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}{N_1 + N_2 - 2} \left[\frac{t_1}{\sigma_1} \quad \frac{t_2}{\sigma_2} \quad \dots \quad \frac{t_D}{\sigma_D} \right]^T \quad (\text{B.6})$$

which is very similar to the T brain in (B.1) when ignoring the scaling factor $\frac{\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}{N_1 + N_2 - 2}$. Comparing (B.1) with (B.6), we can see that classical T brain is just a Fisher brain when voxels are assumed to be independent from each other and standard deviations are introduced as weights for each voxel.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2007.11.029.

References

- American Psychiatric Association, 1994. Diagnostic and statistical manual of mental disorders, 4th ed. American Psychiatric Association, Washington, DC.
- Biederman, J., 1998. Attention deficit/hyperactivity disorder: a life-span perspective. *J. Clin. Psychiatry* 59, 4–16.
- Biederman, J., Mick, E., Faraone, S.V., 2000. Age-dependent decline of symptoms of attention deficit hyperactivity disorder: impact of remission definition and symptom type. *Am. J. Psychiatry* 157, 816–818.
- Biswal, B.B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541.
- Buchsbaum, M.S., Nuechterlein, K.H., Haier, R.J., Wu, J., Sicotte, N., Hazlett, E., et al., 1990. Glucose metabolic rate in normals and schizophrenics during the Continuous Performance Test assessed by positron emission tomography. *Br. J. Psychiatry* 156, 216–227.
- Bush, G., Frazier, J.A., Rauch, S.L., Seidman, L.J., Whalen, P.J., Jenike, M.A., et al., 1999. Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the Counting Stroop. *Biol. Psychiatry* 45, 1542–1552.
- Cao, Q.J., Zang, Y.F., Sun, L., Sui, M.Q., Long, X.Y., Zou, Q.H., Wang, Y.F., 2006. Abnormal neural activity in children with attention deficit hyperactivity disorder: a resting-state functional magnetic resonance imaging study. *NeuroReport* 17, 1033–1036.
- Carlson, T.A., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* 15, 704–717.
- Castellanos, F.X., Giedd, J.N., Marsh, W.L., Hamburger, S.D., Vaituzis, A.C., Dickstein, D.P., et al., 1996. Quantitative brain magnetic resonance imaging in attention deficit hyperactivity disorder. *Arch. Gen. Psychiatry* 53, 607–616.
- Cordes, D., Haughton, V.M., Arfanakis, K., et al., 2000. Mapping functionally related regions of brain with functional connectivity MR imaging. *Am. J. Neuroradiol.* 21, 1636–1644.
- Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.
- Duda, R., Hart, P., Stork, D., 2001. Pattern Classification. John Wiley & Sons, New York.
- Durston, S., Tottenham, N.T., Thomas, K.M., Davidson, M.C., Eigsti, I.M., Yang, Y., 2003. Differential patterns of striatal activation in young children with and without ADHD. *Biol. Psychiatry* 53, 871–878.
- Fan, J., McCandliss, B.D., Fossella, J., Flombaum, J.I., Posner, M.I., 2005. The activation of attentional networks. *NeuroImage* 26, 471–479.
- Fan, Y., Shen, D.G., Gur, R.C., Gur, R.E., Davatzikos, C., 2007. COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE Trans. Med. Imag.* 26 (1), 93–105.
- Frackowiak, R.S., et al., 2004. Human Brain Function (second edition). Elsevier Inc.
- Golland, P., Fischl, B., 2003. Permutation tests for classification: towards statistical significance in image-based studies. *IPMI'03. LNCS 2732*, 330–341.
- Golland, P., Fischl, B., Spiridon, M., Kanwisher, N., Buckner, R., Shenton, M., Kikinis, R., Dale, A., Grimson, W., 2002. Discriminative analysis for image-based studies, MICCAI 2002. *LNCS 2488*, 508–515.
- Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4637–4642.
- Hampson, M., Peterson, B.S., Skudlarski, P., Gatenby, J.C., Gore, J.C., 2002. Detection of functional connectivity using temporal correlations in MR images. *Hum. Brain Mapp.* 15, 247–262.
- Holmes, A.P., Blair, R.C., Watson, J.D.G., Ford, I., 1996. Nonparametric analysis of statistic images from functional mapping experiments. *J. Cereb. Blood Flow Metab.* 16, 7–22.
- Kawasaki, Y., Suzuki, M., Kherif, F., Takahashi, T., Zhou, S.Y., Nakamura, K., Matsui, M., Sumiyoshi, T., Seto, H., Kurachi, M., 2007. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *NeuroImage* 34, 235–242.
- Kontos, D., Megalooikonomou, V., Pokrajac, D., et al., 2004. Extraction of discriminative functional MRI activation patterns and an application to Alzheimer's disease, MICCAI 2004. *LNCS 3217*, 727–735.
- LaConte, S., Anderson, J., Muley, S., Ashe, J., Frutiger, S., Rehm, K., Hansen, L.K., Yacoub, E., Hu, X., Rottenberg, D., Strother, S., 2003. The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. *NeuroImage* 18, 10–27.
- Lawrence, N.S., Ross, T.J., Hoffmann, R., Garavan, H., Stein, E.A., 2003. Multiple neuronal networks mediate sustained attention. *J. Cogn. Neurosci.* 15, 1028–1038.
- Li, S.J., Li, Z., Wu, G.H., Zhang, M.J., Franczak, M., Antuono, P.G., 2002. Alzheimer disease: evaluation of a functional MR imaging index as a marker. *Radiology* 225, 253–259.
- Liu, H.H., Liu, Z.N., Liang, M., Hao, Y.H., Tan, L.H., Kuang, F., et al., 2006. Decreased regional homogeneity in schizophrenia: a resting-state functional magnetic resonance imaging study. *NeuroReport* 17, 19–22.
- Lowe, M.J., Mock, B.J., Sorenson, J.A., 1998. Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuation. *NeuroImage* 7, 119–132.
- Montaron, M.F., Buser, P., 1988. Relationships between nucleus medialis dorsalis, pericruciate cortex, ventral tegmental area and nucleus accumbens in cat: an electrophysiological study. *Exp. Brain Res.* 69, 59–66.
- Mørch, N., Hansen, L.K., Strother, S.C., Svarer, C., Rottenberg, D.A., Lautrup, B., Savoy, R., Paulson, O.B., 1997. Nonlinear versus linear models in functional neuroimaging: learning curves and generalization crossover. *Proceedings of the 15th International Conference on Information Processing in Medical Imaging*.
- Mourao-Miranda, J., Bokde, A.L., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *NeuroImage* 28, 980–995.
- Overmeyer, S., Bullmore, E.T., Suckling, J., Simmons, A., Williams, S.C., Santosh, P.J., Taylor, E., 2001. Distributed grey and white matter deficits

- in hyperkinetic disorder: MRI evidence for anatomical abnormality in an attentional network. *Psychol. Med.* 31, 1425–1435.
- Raichle, M.E., Mintun, M.A., 2006. Brain work and brain imaging. *Annu. Rev. Neurosci.* 29, 449–476.
- Seidman, L.J., Valera, E.M., Bush, G., 2004. Brain function and structure in adults with attention-deficit/hyperactivity disorder. *Psychiatr. Clin. North Am.* 27, 323–347.
- Seidman, L.J., Valera, E.M., Makris, N., Monuteaux, M.C., Boriel, D.L., Kelkar, K., Kennedy, D.N., Caviness, V.S., Bush, G., Aleardi, M., et al., 2006. Dorsolateral prefrontal and anterior cingulate cortex volumetric abnormalities in adults with attention-deficit/hyperactivity disorder identified by magnetic resonance imaging. *Biol. Psychiatry* 60, 1071–1080.
- Shinkareva, S.V., Ombao, H.C., Sutton, B.P., Mohanty, A., Miller, G.A., 2006. Classification of functional brain images with a spatio-temporal dissimilarity map. *NeuroImage* 33, 63–71.
- Sowell, E.R., Thompson, P.M., Welcome, S.E., Henkenius, A.L., Toga, A.W., 2003. Cortical abnormalities in children and adolescents with attention deficit hyperactivity disorder. *Lancet* 362, 1699–1707.
- Stein, T., Moritz, C., Quigley, M., Cordes, D., Haughton, V., Meyerand, E., 2000. Functional connectivity in the thalamus and hippocampus studied with functional MR imaging. *Am. J. Neuroradiol.* 21, 1397–1401.
- Strother, S., La Conte, S., Kai Hansen, L., Anderson, J., Zhang, J., Pulapura, S., Rottenberg, D., 2004. Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. *NeuroImage* 23 (Suppl. 1), S196–S207.
- Sturm, W., Longoni, F., Fimm, B., Dietrich, T., Weis, S., Kemna, S., Herzog, H., Willmes, K., 2004. Network for auditory intrinsic alertness: a PET study. *Neuropsychologia* 42, 563–568.
- Swets, D., Weng, J., 1996. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* 18, 831–836.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotaxic Atlas of the Human Brain*. Thieme, Stuttgart.
- Tang H.M., Lyu M.R. and King I., 2003, Face Recognition Committee Machine, International Conference on Acoustics, Speech, and Signal Processing (ICASSP2003), Hong Kong, April 6–10, 2003, pp.837–840.
- Teicher, M.H., Anderson, C.M., Polcari, A., Glod, C.A., Maas, L.C., Renshaw, P.F., 2000. Functional deficits in basal ganglia of children with attention-deficit/hyperactivity disorder shown with functional magnetic resonance imaging relaxometry. *Nat. Med.* 6, 470–473.
- Tian, L.X., Jiang, T.Z., Wang, Y.F., Zang, Y.F., He, Y., Liang, M., et al., 2006. Altered resting-state functional connectivity patterns of anterior cingulate cortex in adolescents with attention deficit hyperactivity disorder. *Neurosci. Lett.* 400, 39–43.
- Vapnik, V.N., 1998. *Statistical Learning Theory*. J. Wiley, New York.
- Yang, J., Yang, J.Y., 2003. Why can LDA be performed in PCA transformed space? *Pattern Recogn.* 36, 563–566.
- Yang, M.H., Kriegman, D., Ahuja, N., 2001. Face detection using multi-modal density models. *Comput. Vis. Image Underst.* 84, 264–284.
- Yoon, U., Lee, J.M., Im, K., Shin, Y.W., Cho, B.H., Kim, I.Y., Kwon, J.S., Kim, S.I., 2007. Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *NeuroImage* 34, 1405–1415.
- Zang, Y.F., Jiang, T.Z., et al., 2004. Regional homogeneity approach to fMRI data analysis. *NeuroImage* 22, 394–400.
- Zhu, C.Z., Jiang, T.Z., 2003. Multicontext fuzzy clustering for separation of brain tissues in magnetic resonance images. *NeuroImage* 18, 685–696.
- Zhu, C.Z., Zang, Y.F., Liang, M., Tian, L.X., He, Y., Li, X.B., Sui, M.Q., Wang, Y.F., Jiang, T., 2005. Discriminative analysis of brain function at resting-state for attention-deficit/hyperactivity disorder. *Proceedings of MICCAI: International Conference on Medical Image Computing and Computer Assisted Intervention (LNCS 3750: 468–475)*.