

DISCRIMINATING BIPOLAR DISORDER FROM MAJOR DEPRESSION BASED ON KERNEL SVM USING FUNCTIONAL INDEPENDENT COMPONENTS

*Shuang Gao^{1, 2}, Elizabeth A. Osuch³, Michael Wammes³, Jean Théberge³, Tian-Zi Jiang^{1, 2, 4},
Vince D. Calhoun^{5, 6, *}, Jing Sui^{1, 2, 4, 5, *}*

1 Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China, 100190

2 University of Chinese Academy of Sciences, China

3 Lawson Health Research Institute, London Health Sciences Centre; 860 Richmond Street; London, Ontario; Canada N6A 3H8

4 CAS Centre for Excellence in Brain Science and Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, China

5 The Mind Research Network, 1101 Yale Blvd NE, Albuquerque, NM 87131, USA

6 Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM 87131, USA

* Correspondence to: vcalhoun@mrn.org, jing.sui@nlpr.ia.ac.cn

ABSTRACT

Bipolar disorder (BD) and major depressive disorder (MDD) both share depressive symptoms, so how to discriminate them in early depressive episodes is a major clinical challenge. Independent components (ICs) extracted from fMRI data have been proved to carry distinguishing information and can be used for classification. Here we extend a previous method that makes use of multiple fMRI ICs to build linear subspaces for each individual, which is further used as input for classifiers. The similarity matrix between different subjects is first calculated using distance metric of principal angle, which is then projected into kernel space for support vector machine (SVM) classification among 37 BDs and 36 MDDs. In practice, we adopt forward selection technique on 20 ICs and nested 10-fold cross validation to select the most discriminative IC combinations of fMRI and determine the final diagnosis by majority voting mechanism. The results on human data demonstrate that the proposed method achieves much better performance than its initial version [8] (93% vs. 75%), and identifies 5 discriminative fMRI components for distinguishing BD and MDD patients, which are mainly located in prefrontal cortex, default mode network and thalamus etc. This work provides a new framework for helping diagnose the new patients with overlapped symptoms between BD and MDD, which not only adds to our understanding of functional deficits in mood disorders, but also may serve as potential biomarkers for their differential diagnosis.

Index Terms—independent component analysis, linear subspace, kernel SVM, bipolar disorder, major depression disorder

1. INTRODUCTION

The diagnosis of bipolar disorder (BD) and major depression disorder (MDD) is an important research challenge in the most costly brain diseases. The reason is depressive symptoms are more frequent than manic symptoms during the course of bipolar disorder, and moreover the manic symptoms are often hypomanic ones, or even subthreshold. [1] Thus, bipolar disorder is misdiagnosed as major depression disorder with high probability in clinical treatment. This misdiagnosis has many undesirable consequences, leading to futile treatment, inappropriate medication and risk of deterioration. [2] Differentiating BD from MDD at an early stage is in eager need. Specifically, making a high-precision diagnosis in young people before the episode of manic symptoms is of great benefit to patients who experience ambiguous hypomania and high risk group who have a family history of BD. Identifying discriminative biomarkers between BD and MDD could help provide more efficient and correct treatment.

During the last decade, an increasing amount of researchers explore the value of neuroimaging investigated the pathophysiology difference between both mood disorders. Especially, resting-state function magnetic resonance imaging (fMRI) plays an important role in studies of

functional connectivity of brain. Independent component analysis (ICA) is a typical method for brain functional connectivity research [3], [4], and is a data driven technique compared to regional correlation analysis approaches which are hypothesis based methods. The latter one usually needs priori knowledge to find regions of interests (ROIs) for the calculation of correlation. On the contrary, ICA method requires almost no efforts of human intervention. Spatial ICA discovers spatial independent components encoding temporally coherent brain regions, which is the most commonly used [4].

Some methods have been proposed combining spatial ICA and classification in brain disorders, for example Calhoun et al. made value of temporal lobe and default mode component to distinguish subjects with bipolar disorder, chronic schizophrenia, and healthy control [5] and Demirci et al. chose temporal lobe and lateral frontal parietal mode components to achieve better classification performance in schizophrenia study [6]. Support vector machines (SVM) [7] classifier is a supervised binary classification algorithms for linear classification and has been developed into kernel SVM for nonlinear classification. SVM is a widely used machine learning classification algorithm and has achieved good performance in a number of neuroimaging studies. As for functional brain networks, it's born to be a nonlinear structure which can be analyzed on manifold structure and projected into kernel space for SVM classification.

In this paper, the proposed method is an extended version of a discriminant analysis algorithm proposed by Fan et al. [8], capable of automatically selecting discriminative spatial components and combining them for classification. We change the strategy to extract the kernel classification features and make use of nested 10-fold cross validation combined with particular kernel matrix partition of different subjects instead of leave-one-out cross validation. Main steps are as follows. We obtain a bunch of independent components (ICs) of each subjects by applying ICA on their fMRI data firstly, whose combinations span different linear subspaces. Then we calculate the subspace distance between different subjects using principle angles, and define the corresponding similarity metric. After projecting the similarity matrix into kernel space, we partition it into several different symmetric matrices for further training stage and selection of the optimal combination of ICs. Finally, we adopt nested 10-fold cross validation strategy in conjunction of SVM classifier to facilitate the optimization process and make advantage of votes for the final diagnosis of uncertain patients. This improved algorithm is applied to a BD and MDD study. Specifically, we classify BD and MDD patients and make diagnosis by their resting-state fMRI data. We also analyze the corresponding functional brain networks of selected discriminative components in consideration of prior knowledge. The improved method in this study has brought about a vast increase in the classification accuracy and significantly reduces the time complexity.

2. METHODS

2.1. Individual-specific spatial component extraction

Each subject's resting-state fMRI data can be represented as a $(t \times v)$ matrix denoted by \mathbf{X} , where t is the number of time points and v is the number of brain voxels. The ICA model of fMRI is $\mathbf{X} = \mathbf{A} \cdot \mathbf{S}$, where \mathbf{A} is the $(t \times c)$ mixing matrix and \mathbf{S} is the $(c \times v)$ source signal matrix each row of which is an independent component corresponding to a specific functional brain network. The component number c is a free parameter which can be estimated using information theoretic approaches [9], or determined using prior knowledge about functional brain networks.

To make independent components of different subjects comparable, we adopt group information guided ICA (GIGICA) [10], which is a group ICA method [4] to guarantee the same component number and correspondence across different subjects. In GIGICA, the fMRI data of each subject are firstly temporally concatenated after two stages of principal component analysis (PCA) reduction; then the standard ICA model above is conducted on the concatenated matrix, yielding group independent components (GICs). To get individual-specific independent components, a back-reconstruction step is implemented by GIGICA with group independent components as references.

As Fig.1a shows the schematic flowchart of GIGICA, group-specific independent components are built from BD, MDD and healthy control (HC), while new data are not involved in the computation of GICs to avoid potential bias. We also want to find the biomarkers between patients and NC, which is an ongoing work. The GICs built on BD, MDD and HC turn to be standard references in this classification, and make it available to any other new arrivals.

2.2. Subspace similarity between different subjects

A linear subspace spanned by independent components is adopted to represent fMRI data. From a multivariate data representation perspective, the independent components are basis vectors which span a subspace for each individual. Different combinations of ICs can represent regions of distinct interests. The subspace distance between two subspaces is measured by projection metric, defined as

$$d_p = \left(\sum_{i=1}^k \sin^2 \theta_i \right)^{\frac{1}{2}} = \left(k - \sum_{i=1}^k \cos^2 \theta_i \right)^{\frac{1}{2}}, \quad (1)$$

where p is short for projection, k is the number of subspace dimensions and θ_i is the principle angle of two k -dimensional subspaces. Given two subspaces $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$ and $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k\}$, where $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$ and $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k\}$ are orthonormal basis vectors for subspaces \mathbf{A} and \mathbf{B} , the principle angles

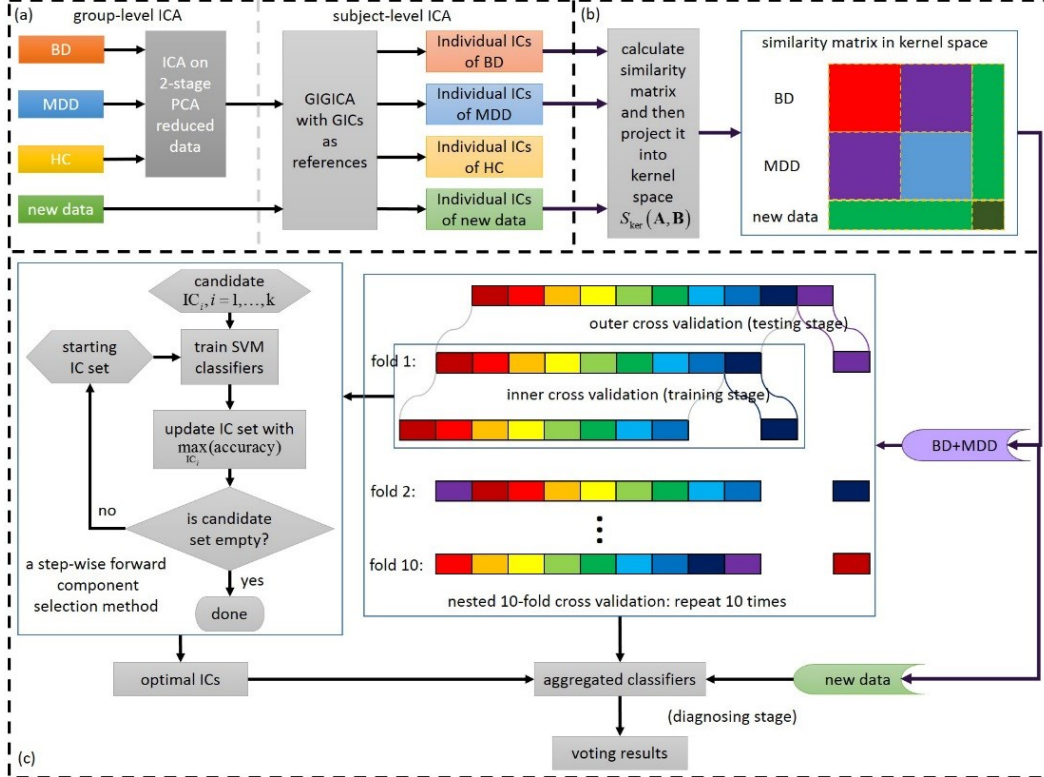


Fig.1. Flowchart of method. (a) The group-level ICA is computed on temporally concatenated fMRI data of BD, MDD patients and HCs after 2-stage PCA dimensionality reduction, and the subject-level ICA is computed by GIGICA [10] with GICs as references, which is applicable to subjects involved (BD, MDD and HC) and not involved (new data) in the computation of group-level ICA. (b) Different combinations of ICs are adopted to calculate subspace similarity matrix of BD, MDD and new data, then the matrix is projected into kernel space. (c) A step-wise forward selection method of linear subspaces spanned by individual ICs is adopted to find the optimal IC combination with best performance, while 10 times nested 10-fold cross validation is applied to build the aggregated classifiers. Making diagnosis of new data based on the majority voting mechanism uses the ensemble classifiers from nested cross validation.

$0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_k \leq \frac{\pi}{2}$ between two subspaces \mathbf{A} and \mathbf{B} are defined as

$$\cos \theta_i = \max_{\mathbf{a}_i \in \mathbf{A}} \max_{\mathbf{b}_i \in \mathbf{B}} \mathbf{a}_i' \mathbf{b}_i = \max_{\mathbf{b}_i \in \mathbf{B}} \max_{\mathbf{a}_i \in \mathbf{A}} \mathbf{b}_i' \mathbf{a}_i, \quad (2)$$

$$\text{subject to } \mathbf{a}_i' \mathbf{a}_i = \mathbf{b}_i' \mathbf{b}_i = 1, \mathbf{a}_i' \mathbf{a}_j = \mathbf{b}_i' \mathbf{b}_j = 0, (i \neq j).$$

Principle angles can be computed from singular value decomposition (SVD) of the dot multiplication of two subspaces. Cosines of principle angles numerically correspond to ordered singular values of $\mathbf{A}'\mathbf{B}$, and thus $\sum_{i=1}^k \cos^2 \theta_i = \sum_{i=1}^k s_i$, where $s = \text{svd}(\mathbf{A}'\mathbf{B})$.

The subspace similarity between different subjects can be defined correspondingly as

$$S_p = \left(\frac{1}{k} \sum_{i=1}^k \cos^2 \theta_i \right)^{\frac{1}{2}}, \quad (3)$$

where p is short for projection and k is the number of subspace dimensions. Suppose the number of BD and MDD patients were m and n respectively, we can construct a $(m+n)$ -dimensional symmetric matrix based on the similarity metric between different subjects as in Eq.(3).

2.3. Partition of similarity matrix in kernel space

The $(m+n)$ -dimensional matrix above is mapped into a high dimensional feature space via sigmoid kernel function for applying SVM algorithm [8] resulting in $K(\mathbf{A}, \mathbf{B}) = \tanh(\gamma S(\mathbf{A}, \mathbf{B}))$, where $S(\mathbf{A}, \mathbf{B})$ is the similarity in Eq.(3), and γ is the kernel parameter. With the kernel function, we can partition the $(m+n)$ -dimensional symmetric matrix in kernel space to build SVM classifier.

The whole classification process included three stages, the training and testing stage using known BD and MDD patients in nested 10-fold cross validation (Fig.1c, 9/10 data run inner cross validation to choose the best combination of ICs, then test 1/10 left-out data, repeated 10 times), and the final diagnosing stage (Fig.1c) for any possible new data. It's obvious that we have to extract smaller matrices from the $(m+n)$ -dimensional symmetric matrix for three different classification stages. The details are shown in Fig.2. We select the columns and rows from the whole matrix of BD, MDD according to group assignment to form the symmetric matrices for training and testing stages, and combine with

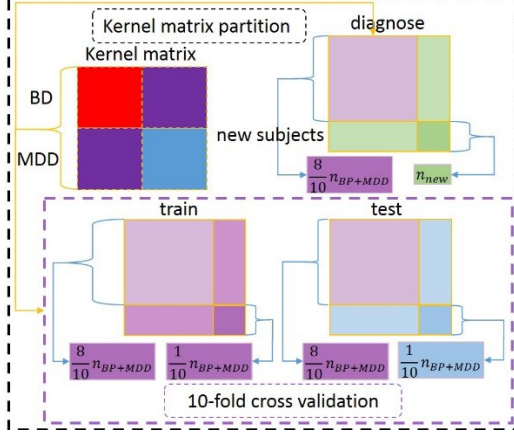


Fig.2. Kernel matrix partition. The whole matrix of BD and MDD is partitioned into different symmetric matrices for SVM, according to different cross validation group assignment. The training sets of train, test and diagnosis stage are all the same shown in pink square, while the testing sets are different corresponding to Fig.1c.

new data if there is any to form the symmetric matrices for diagnosing stage.

2.4. Component selection and cross validation via SVM

We adopt step-wise forward selection method [11] and nested 10-fold cross validation. The component selection procedure is performed in inner cross validation to find optimal independent components. In particular, the initial optimizing component set is empty, and all ICs compose the candidate component set. At each iteration, one component in candidate set is added to the optimizing set iteratively and build a new component set with other components which already exist in it. Then, we perform the propose kernel SVM classification on new sets, and select the component with the highest accuracy to add into the optimization set. This process continues until the candidate set is empty. The flowchart of component selection is shown in Fig.1c.

In nested 10-fold cross validation, the outer cross validation is the testing stage and the inner cross validation is the training stage (component selection procedure) on known BD and MDD. We obtain 1 optimal IC combination and 9 SVM models in one inner cross validation (total 10 optimal combinations in nested validation), so the aggregated classifiers result in 9 votes for testing data. Nested cross validation repeats 10 times. The optimal ICs and aggregated classifiers are obtained after nested validation (Fig.1c). In diagnosing stage, we can predict labels for any new subjects with 10×10 IC combination and $9 \times 10 \times 10$ SVM models (leading to 900 classification results for voting). Note that new data information was untouched in both training and testing stages, ensuring unbiased prediction.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed method was applied to the classification of BD and MDD patients on resting-state fMRI data. We also compared our method with the initial method. SVM classifier

TABLE I
Demographic and Clinical Data of Participants

Group	BD	MDD	NC
Number	37	36	33
Age (years, mean \pm SD)	21 \pm 3	20 \pm 3	20 \pm 2
Gender (M:F)	18:19	10:26	13:20

was implemented by a library for support vector machines (LIBSVM [12], <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>), and GIGICA was computed using group ICA of fMRI toolbox (GIFT, <http://mialab.mrn.org/software/>). The kernel parameter γ was set to 1, and the trade-off parameter C of all SVM classifiers was set to 10. They were all fixed during the calculation. Total number of independent components was set to 20 [8] in consideration of reducing computational complexity.

3.1. Demographic information, imaging data, and image processing

The data collection was approved by the University of Western Ontario Research Ethics Board. All the participants were provided with a written description of the study and had the opportunity to ask questions about the procedure. Written, informed consent was then obtained from willing participants.

A total of 37 participants with BD and 36 age matched participants with MDD were recruited from the First Episode Mood and Anxiety Program in London, Ontario, Canada. Each participant met diagnostic criteria for BD, type I, or MDD using the Structured Clinical Interview for DSM disorders-IV, research version (SCID-IV) or the Diagnostic Interview for Genetic Studies (DIGS). In addition, agreement between the clinical chart diagnosis and the SCID/DIGS diagnosis of MDD or BD-type I was required for the patient groups. Medications were unchanged for three weeks prior to scanning.

Patients were excluded if they had a history of head injury leading to unconsciousness for longer than a few seconds, or significant non-psychiatric medical illness. Individuals with an active substance use disorder (except possibly caffeine and/or nicotine abuse/dependence), posttraumatic stress disorder or obsessive-compulsive disorder were also excluded. Youth were excluded if they were in imminent danger to themselves or others, if they were actively psychotic, or if they had exclusions for MRI scanning. Any family history of BD in the MDD group was exclusionary. For comparison purposes, 33 healthy participants were also included as the control group (NC). A summary of demographic information is presented in Table I.

All MRI imaging was collected using a 3.0 T MRI scanner (MAGNETOM Verio, Siemens, Erlangen, Germany) at the Lawson Health Research Institute, and a 32-channel phased-array head coil (Siemens). A T1-weighted, 3D magnetization-prepared rapid gradient echo sequence was used to collect anatomical images. Acquisition parameters were as follows: TR=3000ms, TE=2.98ms, flip angle=9°,

FOV=256mm×256mm, matrix size=256×256, 176sagittal slices, voxel size=1mm×1mm×1mm. Functional scans consisted of gradient-echo, echo-planar scans with repetition time (TR)=2000ms, echo time (TE)=30ms, flip angle=90°, field of view (FOV)=240mm×240mm, matrix size=80×80, 40 axial slices and thickness=3mm, with no parallel acceleration. Scans covered the whole brain with an isotropic spatial resolution of 3 mm for a total time of approximately 8 min (164 brain volumes). No participant reported falling asleep during the scan when asked immediately after scanning.

The fMRI images were preprocessed using statistical parametric mapping software (SPM12, <http://www.fil.ion.ucl.ac.uk/spm/>). To allow for magnetization equilibrium, the first ten images were discarded. The remaining 154 images were first corrected for the acquisition time delay among different slices, and then were realigned to the first volume for head-motion correction. The fMRI images were further spatially normalized to the Montreal Neurological Institute (MNI) space by diffeomorphic anatomical registration using exponentiated Lie algebra (DARTEL) with T1-weighted anatomical images and spatially smoothed with a 6 mm FWHM Gaussian kernel.

3.2. Comparison and discriminative performance

We investigated the performance of our method with different component numbers compared with the initial method, as shown in Fig.3a. It was demonstrated that our method could achieve classification rates over 90% as a function of component number in training stage. Our method achieved $92\% \pm 0.01$ classification accuracy, while the original method only got $70\% \pm 0.03$ averagely. The decision values of testing subjects yielded from SVM classifiers. Positive (negative) decision values corresponded to data labels predicted as BD (MDD). The voting value was the aggregated value of 90 decision values yielding from their corresponding optimal component combination SVM classifiers. We had total 73 BD and MDD patients. Fig.3b showed the final classification results of them. The proposed method achieved 93.1% classification accuracy. The associated sensitivity and specificity were 92.0% and 94.4%. All made great improvement compared with original method. Besides, we effectively reduced time complexity. It only took 39.8 hours for one 10-fold cross validation in our method, while the initial method needed about 143.1 hours on a computer cluster with 19 servers, 188 Intel Xeon CPUs and CentOS linux operating system. Although both took a rather long time, it was very helpful to discover the potential discriminative functional brain networks of BD and MDD patients. The step-wise forward selection method was born to be with high time complexity, let alone combined with cross validation method.

3.3. Functional brain networks corresponding to the optimal components

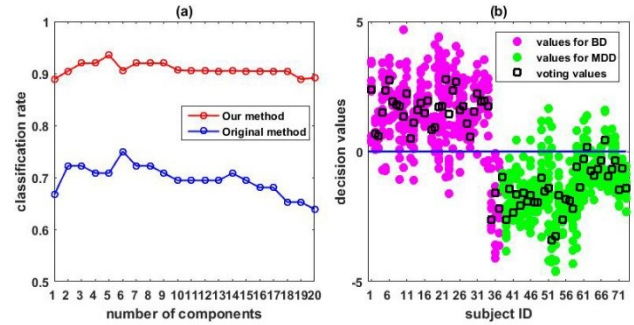


Fig.3. (a) Classification rates as a function of component number. Both of our method and original one were added up in the testing stage of cross validation. (b) Decision values of all testing 37 BDs and 36 MDDs. The voting values were the aggregated values of 90 SVM results based on optimal ICs of 10 times nested validation.

We tracked the selection order of independent components in every SVM classifiers until all of them reached the best performance, and the corresponding accuracies of inner cross validation and details were illustrated in Fig.4b. There were 100 combinations of independent components yielding from 10 times of nested 10-fold cross validation. It was obvious that recurring combinations appeared several times. The frequency of all 20 component appearing in best combination was concluded in Fig.4a. It was worth noting that the selection order was consistent with the frequency. Furthermore, as shown in Fig.3a five components achieved the best performance in testing stage. Thus, we chose the first 5 independent components as the most discriminative functional brain networks according to the frequency figure. The spatial maps of component 2, 16, 6, 1, 11 were shown in Fig.5. Component 2 and 1 were subcortical regions including amygdala, caudate, thalamus, insula, hippocampus etc. Component 16 and 11 covered part of frontal lobe. Component 6 was default mode network (DMN). All of them were reported to demonstrate discriminative power in previous studies [13], [14]. According to our results, amygdala, inferior frontal gyrus and DMN seemed to have more prominent performance comparatively.

4. CONCLUSION

This study presented an extended version of a pattern classification method, capable of identifying discriminative functional connectivity patterns for distinguishing BD and MDD, and making reliable diagnosis of completely new subjects. Experimental results demonstrated that our method achieved high classification accuracy and performed much better than the initial method. Discriminative functional brain networks was identified in training stage and the combinations of corresponding independent components helped make diagnosis of new subjects. Future work is devoted to reduce the time complexity and combine the construction of kernel similarity matrix with clustering algorithm or deep learning for more complex new data. We also take other multi-classification methods into

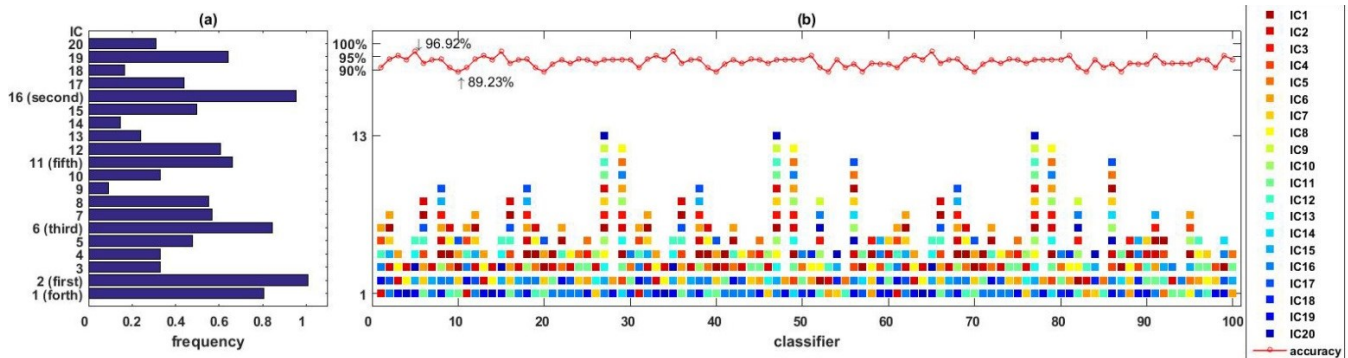


Fig.4. (a) Frequency of components in 100 combinations. Frequency was added up from all optimal combinations of (b). Each color in (b) corresponded to the bar of same component number in (a). (b) Accuracy and selected optimal component combination. The selection order of 20 ICs in 100 combinations of ICs yielding from 10 times nested 10-fold cross validation was shown in irregularly arranged little squares with different colors corresponding to different ICs. The top line is classification accuracy marked with the highest and lowest values in training stage of cross validation.

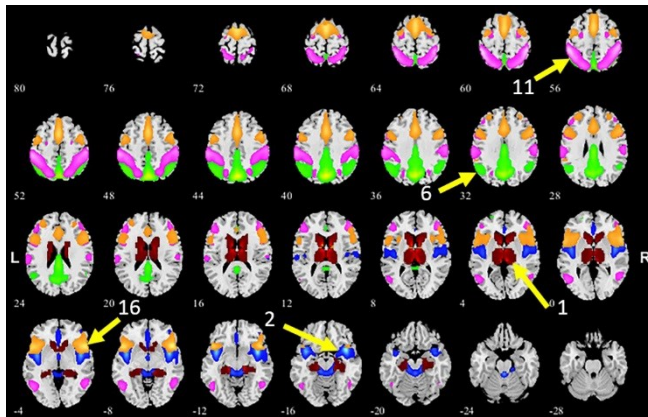


Fig.5. Spatial maps of optimal components. The frequency order of this 5 top selected ICs is 2, 16, 6, 1, and 11, which were shown in different colors as blue, orange, green, red and purple. This 5 ICs included subcortical regions, prefrontal cortex, DMN etc.

consideration as the multi-classification SVM is of large time consumption.

5. ACKNOWLEDGEMENT

This project was supported by the China National High-Tech Development Plan (863 plan) No. 2015AA020513 and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB02060005); The Chinese National Natural Science Foundation (81471367); and the National Institutes of Health grants P20GM103472, 1R01EB006841 and R01EB005846; the Lawson Health Research Institute, grant No. LHR D1374; and a Pfizer Independent Investigator Award, grant No. WS2249136.

6. REFERENCES

[1] J. R. C. de Almeida, M. L. Phillips, "Distinguishing between unipolar depression and bipolar depression: current and future clinical and neuroimaging perspectives," *Biological Psychiatry*, vol. 73, no. 2, pp. 111–118, Jan, 2013.

[2] D. Grotegerd, T. Suslow, J. Bauer et al., "Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: a pilot study," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 263, no. 2, pp. 119–131, Mar, 2013.

[3] V. D. Calhoun, T. Eichele, G. Pearlson, "Functional brain networks in schizophrenia: a review," *Front. Hum. Neurosci.*, 3, 17, Aug, 2009.

[4] V. D. Calhoun, J. Liu, T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *Neuroimage*, vol. 45, no. 1, pp. S163–S172, Mar, 2009.

[5] V. D. Calhoun, P. Maciejewski, G. Pearlson et al., "Temporal lobe and 'default' hemodynamic brain modes discriminate between schizophrenia and bipolar disorder," *Human Brain Mapping*, vol. 29, no. 11, pp. 1265–1275, Sep, 2008.

[6] O. Demirci, V. P. Clark, V. D. Calhoun, "A projection pursuit algorithm to classify individuals using fMRI data: application to schizophrenia," *Neuroimage*, vol. 39, no. 4, pp. 1774–1782, Feb, 2008.

[7] C. Cortes, V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep, 1995.

[8] Y. Fan, Y. Liu, H. Wu et al., "Discriminant analysis of functional connectivity patterns on Grassmann manifold," *Neuroimage*, vol. 56, no. 4, pp. 2058–2067, Jun, 2011.

[9] Y.-O. Li, T. Adali, V. D. Calhoun, "Estimating the number of independent components for functional magnetic resonance imaging data," *Human Brain Mapping*, vol. 28, no. 11, pp. 1251–1266, Feb, 2007.

[10] Y. Du, Y. Fan, "Group information guided ICA for fMRI data analysis," *Neuroimage*, vol. 69, pp. 157–197, Apr, 2013.

[11] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection," *Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar, 2003.

[12] C.-C. Chang, C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, Apr, 2011.

[13] C. Gentili, E. Ricciardi, M. I. Gobbini et al., "Beyond amygdala: default mode network activity differs between patients with social phobia and healthy controls," *Brain research bulletin*, vol. 79, no. 6, pp. 409–413, Aug, 2009.

[14] D. L. Greenberg, H. J. Rice, J. J. Cooper et al., "Co-activation of the amygdala, hippocampus and inferior frontal gyrus during autobiographical memory retrieval," *Neuropsychologia*, vol. 43, no. 5, pp. 659–674, 2005.