2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)

A Weighted Discriminative Dictionary Learning Method for Depression Disorder Classification using fMRI Data

Xin Wang*, Yanshuang Ren†, Yehui Yang*, Wensheng Zhang*§ and Neal N. Xiong‡*

*Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

†Department of Radiology, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100053, China

‡Department of Business and Computer Science, Southwestern Oklahoma State University, OK, 73096, USA

Email: wangxin2012@ia.ac.cn, yanshr2006@163.com, yehui.yang@ia.ac.cn, wensheng.zhang@ia.ac.cn, neal.xiong@swosu.edu

§Corresponding author

***Abstract*—In this paper, we present a novel depression disorder classification algorithm, named weighted discriminative dictionary learning (WDDL), based on functional magnetic resonance imaging (fMRI) data. The underlying relationship between samples and dictionary atoms is exploited by introducing an adaptive weighting scheme. Tested on fMRI data of 29 patients with depression and 29 healthy controls, our algorithm outperforms all other classification methods compared in this work. Furthermore, we detect the discriminative brain regions of patients which can reveal the pathogenesis of depression disorder.**

1. Introduction

Functional magnetic resonance imaging (fMRI), which is based on blood-oxygen-level-dependent (BOLD) techniques, has become an effective approach to study the functional activities of the brain [21], [34]. There are evidences for altered fMRI activation patterns in patients with depression disorder [6], [9], [24]. Meanwhile, the diagnosis of depression disorder is mainly dependent on clinical signs and symptoms, which involves risks for misdiagnose [6], [7], [14]. Therefore, developing automatic depression disorder classification methods of fMRI data is of great importance for diagnosis and treatment of depression disorders [2], [4], [11].

In the past few years, sparse representation has emerged as a powerful tool in image classification. Wright *et al.* propose sparse representation based classification (SRC) to deal with face recognition problem [30]. They first take all the training samples of each class as dictionaries to represent testing samples and then classify them to the class with the minimal representation error. The dictionaries for each class can represent the samples of an individual class well in SRC, but they are pre-defined rather than learnt from a training set. Recent work has shown that methods with data adaptive and learned dictionaries outperform ones with pre-defined dictionaries [10]. In particular, several class-specific dictionary learning methods have been proposed to further enhance classification performance [28], [29], [32]. For instance, Yang *et al.* propose Fisher discrimination dictionary learning (FDDL) method, which utilizes representation coefficients and representation errors for image classification [32]. Recently, a simple and effective image classification method named discriminative feature-oriented dictionary learning (DFDL) has been proposed [28]. This method requires intra-class similarities and emphasizes the inter-class differences.

Inspired by the success of dictionary learning and sparse representation (DLSR) in image classification, some researchers apply DLSR to neuroimaging data classification [18], [24], [26], [31], [35]. Liu *et al.* introduce SRC into neuroimaging study and demonstrate its effectiveness for discriminating Alzheimer's disease or mild cognitive impairment from healthy control with magnetic resonance imaging (MRI) data [15]. FDDL is used to characterize the brain's functional status into task-free or task-performance states [33]. Ramezani *et al.* utilize sparse representation of brain cognitive patterns to classify individuals based on fMRI data [22]. A method with two-stage sparse representations is proposed to differentiate task-based and resting state fMRI signals [34].

Although the above researches have obtained some achievements of DLSR for neuroimaging classification, they still have the following issues: these dictionary learning methods treat all the samples indiscriminately and ignore the valuable relationship between the samples and dictionary atoms. To address the issue, we propose an automatic depression disorder classification method based on fMRI data, the weighted discriminative dictionary learning (WDDL) method. WDDL codes each test sample using two class-specific dictionaries respectively and classifies it to the class with the smaller representation error. Compared with the above DLAR classification methods, the proposed algorithm has the following advantages: we introduce a weighting scheme based on the similarity between the samples and dictionary atoms, which make the representation model more discriminative.

The contributions of this paper are as follows: (1) An adaptive weighting scheme is introduced to improve the classification performance. (2) Our method is interpretable, as it discovers the discriminative information in the brain of patients with depression disorders. The revealed information can provide guidance for diagnosis.

2. Materials and methods

2.1. Depression database

2.1.1. Participants. This study is approved by Department of Radiology, Guang An Men Hospital of China Academy of Traditional Chinese Medicine, Beijing, China. Written informed consents are obtained from all participants. 29 patients with depression (14 females, 15 males) and 29 age-, sex- and education-matched healthy controls (15 females, 14 males) are analyzed in this study. All subjects are right-handed native Chinese speakers. They are recruited at the Department of Radiology, Guang An Men Hospital of China Academy of Traditional Chinese Medicine. The diagnosis of depression disorder is made according to Structured Clinical Interview for the DSM-IV, patient version (SCIDI/P) [8] by experienced psychiatrists. All the patients have no history of other neurological illness or head injury. None of the patients have received treatment within at least 6 months prior to screening. Healthy controls are interviewed using the Structured Clinical Interview for DSM-IV, nonpatient edition (SCIDI/NP). They have no current or history of depression disorder or other psychiatric disorders.

2.1.2. Data acquisition. Experiments are performed on a General Electric (GE) signa 1.5T echo speed superconducting MRI scanner. Subjects are asked to relax themselves with their eyes closed, not to think of anything, but not to fall asleep. Foam padding is used to minimize head motion and reduce scanner noise. Functional images are acquired using an echo-planar imaging (EPI) sequence with whole-brain coverage (repetition time (TR) = 2000ms, echo time (TE) = 30ms, flip angle = 90°, field of view (FOV) = 24cm, matrix = 64×64, thickness = 3mm, slices= 41).

2.2. Data preprocessing

We use Statistical Parametric Mapping(SPM8, http://www.fil.ion.ucl.ac.uk/spm/software/spm8/), Resting-State fMRI Data Analysis Toolkit (REST, http://restfmri.net/forum/index.php/) and Data Processing Assistant for Resting-State fMRI (DPARSF, http://www.rest fmri.net/forum/taxonomy/term/36) for fMRI data preprocessing. In consideration of subject's adaptation to the scanning and the scanner calibration, the first 10 time points are discarded. Further preprocessing procedures include slice timing, realignment for head motion correction. No subject of Depression database is discarded as all the subjects have no excessive head movement (translation$<$ 2.0mm or rotation$<$ 2.0°). Next, the functional images are spatially normalized to the standard EPI template in SPM8 and resampled to a voxel size of $3 \times 3 \times 3$ mm^3. After smoothing with an isotropic Gaussian kernel (FWHW=4mm), temporal band-pass filtering (0.01HZ-0.08HZ) is performed to remove physiological high-frequency noises and low-frequency drifts. What's more, 6 head motion parameters, global mean signal, white matter signal and cerebrospinal fluid signal are regressed out as covariates. Finally, the fMRI time series are segmented into 90 brain regions using the Automated Anatomical Labeling (AAL) atlas [27]. We take the mean of all voxels in a region as its fMRI time series.

2.3. Proposed weighted discriminative dictionary learning

2.3.1. Objective function in WDDL. A subject can be represented as a matrix $\mathbf{Y}^t \in \mathbb{R}^{n\times m}$, where n is the number of brain regions and m is the number of time series. By collecting all the samples from each class, we build the sample matrix of healthy class (HC), $\mathbf{Y} \in \mathbb{R}^{n\times mp}$, and the sample matrix of patient class (PC), $\tilde{\mathbf{Y}} \in \mathbb{R}^{n\times mq}$, where p and q are the numbers of subjects corresponding to each class, respectively. The aim of training stage is to learn dictionary $\mathbf{D} \in \mathbb{R}^{n\times r}$ of HC and dictionary $\tilde{\mathbf{D}} \in \mathbb{R}^{n\times r}$ of PC, where r is the number of atoms in each dictionary. For simplicity, we only describe the model corresponding to healthy class below. The model of patient class can be solved in a similar way. The objective function related to the healthy class of our WDDL method is formulated as:

$$\begin{aligned} \mathrm{J} = \arg \min_{\mathbf{D},\mathbf{S},\tilde{\mathbf{S}}} & \Big(\frac{1}{K}\sum_{i=1}^{K} \mathbf{w}_i \|\mathbf{y}_i - \mathbf{D}\mathbf{s}_i\|_2^2 \\ & - \frac{\rho}{\tilde{K}}\sum_{j=1}^{\tilde{K}} \tilde{\mathbf{w}}_j \|\tilde{\mathbf{y}}_j - \mathbf{D}\tilde{\mathbf{s}}_j\|_2^2\Big) \\ & s.t. \|\mathbf{d}_k\|_2^2 = 1, \|\mathbf{S}\|_0 < \epsilon_1, \|\tilde{\mathbf{S}}\|_0 < \epsilon_2, k = 1, ..., r, \end{aligned} \quad (1)$$

where $K = mp$ is the number of columns of HC matrix $\mathbf{Y}$, $\mathbf{y}_i$ is the i-th column of $\mathbf{Y}$, $\mathbf{s}_i$ and $\mathbf{w}_i$ are the coding coefficient and weight of $\mathbf{y}_i$, $\tilde{K} = mq$ is the number of columns of PC matrix $\tilde{\mathbf{Y}}$, $\tilde{\mathbf{y}}_j$ is the j-th column of $\tilde{\mathbf{Y}}$, $\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{w}}_j$ are the coding coefficient and weight of $\tilde{\mathbf{y}}_j$. ρ is a positive regularization parameter and k is the index of dictionary atoms. The first term requires intra-class differences to be small, while the second term emphasizes inter-class differences. We set the weights to be inversely proportional to the distances between the training samples and the mean of dictionary atoms. The weight $\mathbf{w}_i$ is defined as

$$\mathbf{w}_i = \frac{1}{Z} exp(-\|\mathbf{y}_i - \overline{\mathbf{d}}\|_2^2), \quad (2)$$

where Z is the normalization constants, $\overline{\mathbf{d}}$ is the mean vector of dictionary atoms in $\mathbf{D}$. The weight $\tilde{\mathbf{w}}_j$ can be solved similarly. Distance between a sample and the mean of dictionary atoms can measure the similarity between the sample and the dictionary atoms. If the distance between a sample and dictionary atoms is small, the weight would be large, making the representation error small enough to satisfy the convergence condition. In other words, the aim of adding the weighting scheme is to learn a dictionary which can better represent the samples similar to it. The better representation can further enhance the classification performance.

2.3.2. Optimization strategy. The optimization procedure of WDDL can be divided into two sub-problems: updating $\mathbf{S}$ and $\tilde{\mathbf{S}}$ with $\mathbf{D}$ fixed, updating $\mathbf{D}$ with $\mathbf{S}$ and $\tilde{\mathbf{S}}$ fixed.

Suppose that dictionary $\mathbf{D}$ is fixed, the objective function in Eq. (1) is reduced to the solver of coding coefficients. Since the samples from the two classes share the same dictionary, they can be coded in the same way as follows:

$$\mathbf{s}_i^* = \arg\min_{\mathbf{s}_i}(\mathbf{w}_i\|\mathbf{y}_i - \mathbf{D}\mathbf{s}_i\|_2^2 + \lambda\|\mathbf{s}_i\|_0), \tag{3}$$

where λ is the regularization parameter. The coding coefficient $\tilde{\mathbf{s}}_j$ can also be solved by the above equation.

When coding coefficients $\mathbf{S}$ and $\tilde{\mathbf{S}}$ are fixed, dictionary $\mathbf{D}$ can be updated class by class. The objective function in Eq. (1) is reduced to:

$$\begin{aligned}\mathbf{D}^* =& \arg\min_{\mathbf{D}}(\frac{1}{K}\sum_{i=1}^{K}\mathbf{w}_i\|\mathbf{y}_i - \mathbf{D}\mathbf{s}_i\|_2^2 \\ &- \frac{\rho}{\tilde{K}}\sum_{j=1}^{\tilde{K}}\tilde{\mathbf{w}}_j\|\tilde{\mathbf{y}}_j - \mathbf{D}\tilde{\mathbf{s}}_j\|_2^2) \\ & s.t.\|\mathbf{d}_k\|_2^2 = 1, k = 1, ..., r.\end{aligned} \tag{4}$$

The above objective function can also be written as:

$$\begin{aligned}\mathbf{D}^* =& \arg\min_{\mathbf{D}}(\frac{1}{K}\|diag(\sqrt{\mathbf{W}})(\mathbf{Y} - \mathbf{D}\mathbf{S})\|_F^2 \\ &- \frac{\rho}{\tilde{K}}\|diag(\sqrt{\tilde{\mathbf{W}}})(\tilde{\mathbf{Y}} - \mathbf{D}\tilde{\mathbf{S}})\|_F^2) \\ =& \arg\min_{\mathbf{D}}(-2trace(\mathbf{G}\mathbf{D}^T) + trace(\mathbf{D}\mathbf{H}\mathbf{D}^T)).\end{aligned} \tag{5}$$

By using the equation $\|\mathbf{R}\|_F^2 = trace(\mathbf{R}\mathbf{R}^T)$ for any matrix $\mathbf{R}$, we can derive the first equation of Eq. (5) to the second and denote:

$$\begin{aligned}\mathbf{G} &= \frac{diag(\mathbf{W})}{K}\mathbf{Y}\mathbf{S}^T - \frac{\rho diag(\tilde{\mathbf{W}})}{\tilde{K}}\tilde{\mathbf{Y}}\tilde{\mathbf{S}}^T \\ \mathbf{H} &= \frac{diag(\mathbf{W})}{K}\mathbf{S}\mathbf{S}^T - \frac{\rho diag(\tilde{\mathbf{W}})}{\tilde{\mathbf{K}}}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T.\end{aligned} \tag{6}$$

The object function in Eq. (5) is convex if and only if $\mathbf{H}$ is positive semidefinite. To guarantee the matrix $\mathbf{H}$ in Eq. (5) is positive semidefinite, we replace $\mathbf{H}$ with $\hat{\mathbf{H}} = \mathbf{H} - \lambda_{min}(\mathbf{H})\mathbf{I}_k$, where $\lambda_{min}(\mathbf{H})$ is the minimum eigenvalue of $\mathbf{H}$ and $\mathbf{I}_k$ is the identity matrix. In addition, the replacement will not change the optimal solution to Eq. (5) according to [28]. The object function in Eq. (5) is equivalent to:

$$\begin{aligned}&\mathbf{D}^* = \arg\min_{\mathbf{D}}(-2trace(\mathbf{G}\mathbf{D}^T) + trace(\mathbf{D}\hat{\mathbf{H}}\mathbf{D}^T)) \\ &s.t.\|\mathbf{d}_k\|_2^2 = 1, k = 1, ..., r.\end{aligned} \tag{7}$$

The dictionary atoms $\mathbf{d}_k$ can be updated one after another by solving Eq. (7) via the algorithm like [28] or [19]:

$$\begin{aligned}\mathbf{u}_k &\leftarrow \frac{1}{\hat{H}_{k,k}}(\mathbf{g}_k - \mathbf{D}\hat{\mathbf{h}}_k) + \mathbf{d}_k \\ \mathbf{d}_k &\leftarrow \frac{1}{\|\mathbf{u}_k\|_2}\mathbf{u}_k,\end{aligned} \tag{8}$$

where $\hat{H}_{k,k}$ is the value of $\hat{\mathbf{H}}$ at (k,k), $\hat{\mathbf{h}}_k$ is the k-th column of $\hat{\mathbf{H}}$ and $\mathbf{g}_k$ is the k-th column of $\mathbf{G}$. The algorithm of WDDL is summarized as follows. The algorithm of WDDL is summarized as follows.

Algorithm 1 Weighted Discriminative Dictionary Learning

Input:
$\mathbf{Y}$, $\tilde{\mathbf{Y}}$, dictionary size r, regularization parameter λ and ρ.
Output:
Dictionary $\mathbf{D}$.
Initialize $\mathbf{D}$ by randomly picking r columns of $\mathbf{Y}$.
while not converged **do**
  Update the weight $\mathbf{W}$ and $\tilde{\mathbf{W}}$ by solving Eq. (2).
  Fix $\mathbf{D}$, update $\mathbf{S}$ and $\tilde{\mathbf{S}}$ by solving Eq. (3).
  Fix $\mathbf{S}$ and $\tilde{\mathbf{S}}$, update $\mathbf{D}$ by solving Eq. (8).
end while
Return $\mathbf{D}$.

2.4. Patient classification

For the purpose of verifying the proposed method, we perform classification on the testing samples. Classification is based on the representation error of each class. First, a testing sample $\mathbf{Y}^t$ is represented by the two class-specific dictionaries learned in training step as $\mathbf{S}^t$ and $\tilde{\mathbf{S}}^t$, respectively. Secondly, the testing sample is assigned to the class with the smaller representation error: $\arg\min_{c\in 1,2} r_c(\mathbf{Y}^t)$ where

$$\begin{aligned}\mathrm{r}_1(\mathbf{Y}^t) &= \|\mathbf{Y}^t - \mathbf{D}\mathbf{S}^t\|_2 \\ \mathrm{r}_2(\mathbf{Y}^t) &= \|\mathbf{Y}^t - \tilde{\mathbf{D}}\tilde{\mathbf{S}}^t\|_2.\end{aligned} \tag{9}$$

The overall classification procedure of our algorithm is shown in Fig. 1.

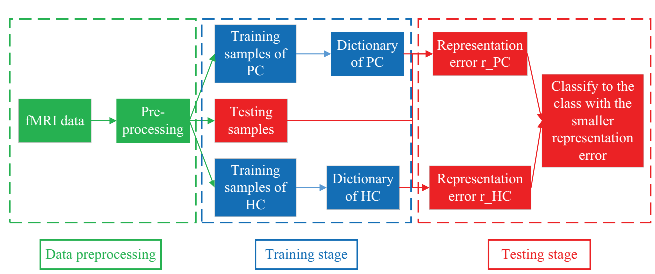

Fig. 1. The schematic diagram of WDDL method used for depression disorder classification of fMRI data.

2.5. Identification of the most discriminative brain regions

To identify the discriminative brain regions of depression disorder, we perform a two-sample t-test (p-value < 0.05) on the two class-specific dictionaries. We compare each row vector related to each brain region between the dictionary $\mathbf{D}$ and $\tilde{\mathbf{D}}$. The number of significant differences of each related brain region is counted in leave-one-out cross-validation (LOOCV). The more times a brain region is significantly different, the more discriminative it is.

3. Results

In this section, we conduct detailed experiments to demonstrate the effectiveness of our method on Depression database.

3.1. Classification performance

Due to the limited number of subjects in Depression database, we employ LOOCV to classifiers. Specifically, each of the subjects is treated as testing sample in turn, and the rest of subjects are treated as training samples. To assess the classification performance of our method, we measure accuracy, sensitivity and specificity which are commonly used in patient classification problems. Table 1 summarizes the classification performance of WDDL and the compared methods. These methods include support vector machine (SVM) [1], [20], naive Bayes classifier (NB) [23], single dictionary based classification (SDC) [5], SRC [30] and DFDL [28]. The parameters of these models are all optimized by LOOCV. Experimental results show that WDDL outperforms the compared methods, achieving 79.31% accuracy, 75.86% sensitivity and 82.76% specificity.

TABLE 1. Classification performance of WDDL and the compared methods on Depression database.

Algorithm	Accuracy(%)	Sensitivity(%)	Specificity(%)
SVM	58.62	65.52	51.72
NB	56.90	48.28	65.52
SDC	58.62	68.97	48.28
SRC	63.79	55.17	72.41
DFDL	72.41	**75.86**	68.97
WDDL	**79.31**	**75.86**	**82.76**

3.2. Discriminative brain regions of patients with depression

The ten most discriminative brain regions between patients with depression and healthy controls are shown in Fig. 2.

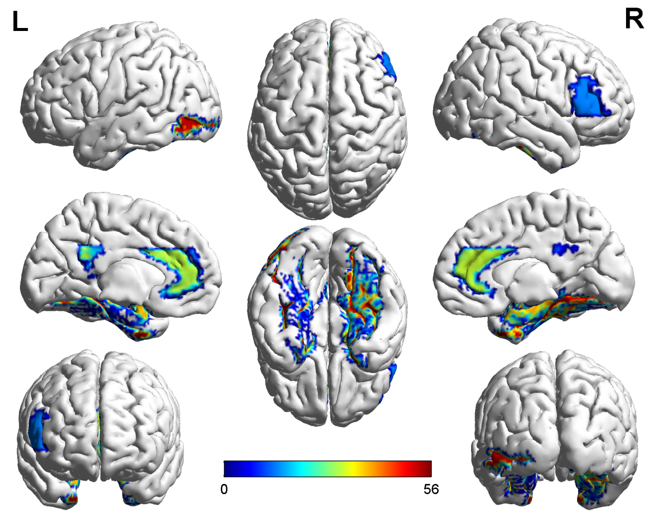

Fig. 2. The ten most discriminative brain regions in patients with depression compared with healthy controls. The color bar indicates the index of brain regions shown in this figure.

4. Discussion

In this paper, we develop a new method named WDDL to classify fMRI data of patients with depression and healthy controls. Experimental results in Table 1 show that our method outperforms compared methods on Depression database. The results of WDDL compared with SVM and NB indicate that our algorithm performs better than traditional classifiers. Classification based on one dictionary is used to compare the effect of one dictionary and two class-specialfic dictionaries. The comparisons with SRC and DFDL indicate that our algorithm outperforms the state-of-the-art class-specific dictionary learning methods. The better performance of WDDL over these methods verifies that the weighting scheme do make a positive contribution to the improvement of WDDL.

The discriminative brain regions of depression disorder detected by our method, shown in Fig. 2, are consistent with previous studies. We find that the discriminative brain regions are mainly involved in the limbic-cortical networks, such as hippocampus, parahippocampal gyrus, anterior cingulate cortex, posterior cingulate cortex, middle frontal gyrus and amygdala. The present study adds an additional literature to the key role of limbic-cortical networks in the pathogenesis of depression disorder [12]. Among these brain regions, anterior cingulate cortex plays a key role in the cognition and emotional regulation [17]. The fusiform and inferior occipital gyrus, which are related to visual recognition, are also discriminative between patients with depression and healthy controls [13]. As all the participants are scanned with eyes closed, this may imply the aberrant visual recognition processing in depression disorder. In addition, the brain regions of default mode network (DMN) [3], [16], [21], [36], such as posterior cingulate cortex, amygdala are also significantly different in patients with depression, which has been detected in previous works [25].

5. Conclusions

In this paper, we proposed a depression disorder classification method named WDDL, which provided better classification performance than the compared methods. More specifically, weighting scheme was introduced into the representation model to improve classification performance. In this method, two class-specific dictionaries were learned and fMRI data of each sample was represented by the two dictionaries, respectively. One sample was classified to the class with the smaller representation error. Experimental results demonstrated the effectiveness and improved classification performance of WDDL.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants 61305018, 61432008, 61472423 and 61532006. Wensheng Zhang is the corresponding author of this paper.

References

[1] Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. Data Min. knowl. Discov. 2, 121-167.

[2] Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., Cercignani, M., 2015. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. NeuroImage 112, 232-243.

[3] Damoiseaux, J., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C., Smith, S.M., Beckmann, C., 2006. Consistent resting-state networks across healthy subjects. Proc. Nati. Acad. Sci. 103, 13848-13853.

[4] Delavarian, M., Towhidkhah, F., Gharibzadeh, S., Dibajnia, P., 2011. Automatic classification of hyperactive children: Comparing multiple artificial intelligence approaches. Neurosci. Lett. 498, 190-193.

[5] Deshpande, H., Maurel, P., Barillot, C., 2015. Classification of multiple sclerosis lesions using adaptive dictionary learning. Comput. Med. Imag. Graph. 46, 2-10.

[6] Du, J., Wang, L., Jie, B., Zhang, D., 2016. Network-based classification of ADHD patients using discriminative subnetwork selection and graph kernel PCA. Comput. Med. Imag. Graph. 52, 82-88.

[7] Ferrari, A.J., Charlson, F.J., Norman, R.E., Patten, S.B., Freedman, G., Murray, C.J., Vos, T., Whiteford, H.A., 2013. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 10, e1001547.

[8] First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B., 1995. Structured clinical interview for DSM-IV axis I disorders. New York: New York State Psychiatric Institute.

[9] Ford, J., Farid, H., Makedon, F., Flashman, L.A., McAllister, T.W., Megalooikonomou, V., Saykin, A.J., 2003. Patient classification of fMRI activation maps. MICCAI: Int. Conf. Med. Image Computing Computer Assist. Intervent. Springer, pp. 58-65.

[10] Gangeh, M.J., Farahat, A.K., Ghodsi, A., Kamel, M.S., 2015. Supervised dictionary learning and sparse representation-a review. arXiv preprint arXiv:1502.05928.

[11] Garg, G., Prasad, G., Coyle, D., 2013. Gaussian Mixture Model-based noise reduction in resting state fMRI data. J. Neurosci. Methods 215, 71-77.

[12] Guo, W.-B., Liu, F., Xue, Z.-M., Yu, Y., Ma, C.-Q., Tan, C.-L., Sun, X.-L., Chen, J.-D., Liu, Z.-N., Xiao, C.-Q., 2011. Abnormal neural activities in first-episode, treatment-naive, short-illness-duration, and treatment-response patients with major depressive disorder: a resting-state fMRI study. J. Affect. Disord. 135, 326-331.

[13] Guo, W.-B., Liu, F., Xue, Z.-M., Xu, X.-J., Wu, R.-R., Ma, C.-Q., Wooderson, S.C., Tan, C.-L., Sun, X.-L., Chen, J.-D., 2012. Alterations of the amplitude of low-frequency fluctuations in treatment-resistant and treatment-response depression: a resting-state fMRI study. Prog. Neuropsychopharmacol. Biol. Psychiatry 37, 153-160.

[14] Liu, F., Guo, W., Liu, L., Long, Z., Ma, C., Xue, Z., Wang, Y., Li, J., Hu, M., Zhang, J., 2013. Abnormal amplitude low-frequency oscillations in medication-naive, first-episode patients with major depressive disorder: a resting-state fMRI study. J. Affect. Disord. 146, 401-406.

[15] Liu, M., Zhang, D., Shen, D., Initiative AsDN., 2012. Ensemble sparse classification of Alzheimer's disease. NeuroImage 60, 1106-1116.

[16] Long, X.-Y., Zuo, X.-N., Kiviniemi, V., Yang, Y., Zou, Q.-H., Zhu, C.-Z., Jiang, T.-Z., Yang, H., Gong, Q.-Y., Wang, L., 2008. Default mode network as revealed with multiple methods for resting-state functional MRI analysis. J. Neurosci. Methods 171, 349-355.

[17] Lu, Q., Wang, Y., Luo, G., Li, H., Yao, Z., 2013. Dynamic connectivity laterality of the amygdala under negative stimulus in depression: A MEG study. Neurosci. Lett. 547, 42-47.

[18] Lv, J., Li, X., Zhu, D., Jiang, X., Zhang, X., Hu, X., Zhang, T., Guo, L., Liu, T., 2013. Sparse representation of group-wise fMRI signals. MICCAI: Int. Conf. Med. Image Computing Computer Assist. Intervent. Springer, pp. 608-616.

[19] Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2010. Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. 11, 19-60.

[20] Misaki, M., Luh, W.-M., Bandettini, P.A., 2013. The effect of spatial smoothing on fMRI decoding of columnar-level organization with linear support vector machine. J. Neurosci. Methods 212, 355-361.

[21] Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Nat. Acad. Sci. 98, 676-682.

[22] Ramezani, M., Abolmaesumi, P., Marble, K., MacDonald, H., Johnsrude, I., 2012. Classification of individuals based on sparse representation of brain cognitive patterns: a functional MRI study. Proc. 34th IEEE Int. Eng. Med. Biol.Conf. pp. 2688-2691.

[23] Rish, I., 2001. An empirical study of the naive Bayes classifier. Proc. IJCAI-01 on Empirical Methods in AI Workshop pp. 41-46.

[24] Rosa, M.J., Portugal, L., Hahn, T., Fallgatter, A.J., Garrido, M.I., Shawe-Taylor, J., Mourao-Miranda, J., 2015. Sparse network-based models for patient classification using fMRI. Neuroimage 105, 493-506.

[25] Sheline, Y.I., Barch, D.M., Price, J.L., Rundle, M.M., Vaishnavi, S.N., Snyder, A.Z., Mintun, M.A., Wang, S., Coalson, R.S., Raichle, M.E., 2009. The default mode network and self-referential processes in depression. Pro. Nati. Acad. Sci. 106, 1942-1947.

[26] Shi, J., Liu, X., Li, Y., Zhang, Q., Li, Y., Ying, S., 2015. Multi-channel EEG-based sleep stage classification with joint collaborative representation and multiple kernel learning. J. Neurosci. Methods 254, 94-101.

[27] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273-289.

[28] Vu, T.H., Mousavi, H.S., Monga, V., Rao, G., Rao, U.A., 2016. Histopathological image classification using discriminative feature-oriented dictionary learning. IEEE Trans. Med. Imag. 35, 738-751.

[29] Wang, H., Yuan, C., Hu, W., Sun, C., 2012. Supervised class-specific dictionary learning for sparse modeling in action recognition. Pattern Recogn. 45, 3902-3911.

[30] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., 2009. Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Int. 31, 210-227.

[31] Xiong, N., Vasilakos, A.V., Yang, L.T., Song, L., Pan, Y., Kannan, R., Li, Y., 2009. Comparative analysis of quality of service and memory usage for adaptive failure detectors in healthcare systems. IEEE Journal on Selected Areas in Communications 27, 495-509.

[32] Yang, M., Zhang, L., Feng, X., Zhang, D., 2014. Sparse representation based fisher discrimination dictionary learning for image classification. Int. J. Comput. Vis. 109, 209-232.

[33] Zhang, X., Guo, L., Li, X., Zhu, D., Li, K., Sun, Z., Jin, C., Hu, X., Han, J., Zhao, Q., 2012. Characterization of task-free/task-performance brain states. MICCAI: Int. Conf. Med. Image Computing Computer Assist. Intervent. Springer, pp. 237-245.

[34] Zhang, S., Li, X., Lv, J., Jiang, X., Guo, L., Liu, T., 2016. Characterizing and differentiating task-based and resting state fMRI signals via two-stage sparse representations. Brain Imag. Behav. 10, 21-32.

[35] Zhang, C., Song, S., Wen, X., Yao, L., Long, Z., 2015. Improved sparse decomposition based on a smoothed L0 norm using a Laplacian kernel to select features from fMRI data. J. Neurosci. Methods 245, 15-24.

[36] Zhu, X., Wang, X., Xiao, J., Liao, J., Zhong, M., Wang, W., Yao, S., 2012. Evidence of a dissociation pattern in resting-state default mode network connectivity in first-episode, treatment-naive major depression patients. Biol. Psychiatry 71, 611-617.