Classification of Schizophrenia Versus Normal Subjects Using Deep Learning

Pinkal Patel
SBILab, Department of Electronics and Communication Engineering
Indraprastha Institute of Information Technology, Delhi (IIIT-D), India
pinkal08cece@gmail.com

Priya Aggarwal
SBILab, Department of Electronics and Communication Engineering
Indraprastha Institute of Information Technology, Delhi (IIIT-D), India
priyaa@iiitd.ac.in

Anubha Gupta
SBILab, Department of Electronics and Communication Engineering
Indraprastha Institute of Information Technology, Delhi (IIIT-D), India
anubha@iiitd.ac.in

ABSTRACT

Motivated by deep learning approaches to classify normal and neuro-diseased subjects in functional Magnetic Resonance Imaging (fMRI), we propose stacked autoencoder (SAE) based 2-stage architecture for disease diagnosis. In the proposed architecture, a separate 4-hidden layer autoencoder is trained in unsupervised manner for feature extraction corresponding to every brain region. Thereafter, these trained autoencoders are used to provide features on class-labeled input data for training a binary support vector machine (SVM) based classifier. In order to design a robust classifier, noisy or inactive gray matter voxels are filtered out using a proposed covariance based approach. We applied the proposed methodology on a public dataset, namely, 1000 Functional Connectomes Project Cobre dataset consisting of fMRI data of normal and Schizophrenia subjects. The proposed architecture is able to classify normal and Schizophrenia subjects with 10-fold cross-validation accuracy of 92% that is better compared to the existing methods used on the same dataset.

Keywords

Classification; Support Vector Machine; Stacked Autoencoder; Schizophrenia

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) has emerged as a powerful neuroimaging technique to study human brain function and dysfunction. It is one of the primary techniques to reveal how brain regions communicate with each other to accomplish specific tasks. It provides a promising way to study interaction between spatially remote distinct brain regions that are engaged simultaneously in a task. fMRI blood oxygenated level dependent (BOLD) signal is not random but temporally coherent between spatially remote functionally consistent regions. Functionally coupled regions together form functional brain networks [3, 6]. Identification of functional activation in fMRI data during task or task-free "resting state" is critical in neuroscience for pre-surgical planning and neuropsychiatric disease diagnosis such as Alzheimers dementia [11], autism [22], and schizophrenia [20].

A number of methods have been used for disease diagnosis or classification into normal and diseased subjects. In general, low SNR of fMRI data poses challenges in neuropsychiatric disorder diagnostics using this data. To the best of our knowledge, there is no standard automated fMRI tool in hospitals that can be used for disease diagnosis. This has motivated researchers to explore designing such systems that can aid doctors. For example, researchers have used stationary functional connectivity (FC) as a biomarker to classify patients of neurological and psychiatric diseases such as Alzheimer [10] or Schizophrenia from normal subjects [8] because FC has been wildly observed to be altered in several neuropsychological diseases [11, 16, 14, 15]. While these studies restricted the analysis to mean representative voxel time series, others have reported results on large scale networks.

In recent years, owing to the increasing success of deep learning methods in the areas of speech, signal, image, video and text mining, and recognition, these methods are being explored in neurodisorder disease diagnosis using structural MRI (sMRI) or functional MRI (fMRI) data [13, 23, 12]. In [13], deep belief network has been used to classify Attention Deficit Hyperactivity Disorder (ADHD) using ADHD200 dataset. However, decision is taken at the brain region level that is inappropriate because not every region may participate in a neuro-disorder and abnormality in one or more regions may not be the marker to classify either. In fact, this classification should have been done at the subject level. In [23], a hybrid architecture of DNN and HMM is used for Mild Cognitive Impairment (MCI) identification.

Recently, classification of Schizophrenia and normal subjects has been done on Cobre dataset [12, 1, 4, 5, 9, 21]. In [12], deep neural network with multiple hidden layers along with sparsity constraint is used to classify subjects. This deep architecture is able to achieve 85.8% accuracy. However, mean time series of all voxels of a region has been considered as input data for classification. This is a limitation because all voxels of a region may not be active at any time, hence, averaging across a whole region may suppress the signal of interest. At the same time, noise signal due to aliased components of undesired signals such as heart rhythm or breathing rhythms may get bolstered via averaging and may lead to spurious correlation values across regions. Moreover, in [12], functional connectivity matrix is computed using the Pearson correlation coefficient. Pearson correlation is computed via neglecting the effect of other regions. Hence, there may be a case when one region drives

two other regions, providing spurious correlation also called as triangular functional network.

In [9], resting-state data of 69 Schizophrenia and 72 healthy subjects has been acquired and correlation matrix is computed. An accuracy of 71.63% has been reported. This method also uses mean region representative time series and correlation metric that have limitations as noted above.

In [5], single layer feed forward network is used to extract features from correlation based functional connectivity matrix. 90% accuracy has been reported on Cobre dataset with 10 fold cross validation. In [4, 21], graph theoretic measures have been used as features to classify Schizophrenia and normal subjects. 80% accuracy has been reported on their own dataset. In [1], first whole brain is parcellated using independent component analysis (ICA). Next, various graph theoretic measures are computed to classify with 65% accuracy on the Cobre dataset. In [21], Amplitude of Low Frequency Fluctuations (ALFF), fractional Amplitude of Low Frequency Fluctuations (fALFF), Voxel-Mirrored Homotopic Connectivity (VMHC), and Regional Homogeneity (ReHo) are used as graph theoretic measures and average accuracy of 80% has been reported on the Cobre dataset.

In this paper, we propose a deep learning based architecture for the classification of normal and Schizophrenia subjects on publicly available Cobre dataset [18]. This paper has the following salient contributions:

1. We propose covariance based approach to filter out inactive gray matter voxels from the fMRI dataset before the data is fed to the first stage of the classifier.

2. We build a 2-stage classifier wherein the first stage is an autoencoder with 4 hidden layers in the encoder that is separately trained in unsupervised manner on every brain region (116 regions) without class-labeled input data. Encoder output is a 8-length feature vector for every region. Mean and standard deviation of every feature corresponding to all voxels of an individual brain region are used as representative regional feature vector. Feature vectors of all regions are applied to an SVM classifier with class-labeled data for the training purpose and later used for classification.

3. Instead of using the mean time series of all voxels of a region or the temporal mean of a voxel [23, 12, 1, 4, 5, 9, 21], we have used complete time series of all active voxels as input to SAEs. Thus, there is no information loss at the input end.

The rest of this paper is organized as follows. Section 2 presents briefly the theory of deep autoencoder and SVM classifier. Data description and pre-processing is described in Section 3. The proposed classifier methodology is presented in Section 4. In Section 4, we present results on 1000 Functional Connectome Cobre dataset. Finally, some concluding remarks are presented in Section 5.

2. BACKGROUND

Autoencoder (AE) [2, 19] is an unsupervised neural network that is composed of an encoder and a decoder with backpropagation algorithm to match the input and the output. A multiple hidden-layer autoencoder architecture is also called stacked autoencoder (SAE). The goal of autoencoder is to find representative features from the input data such that the data can be reconstructed from this feature vector. Hence, instead of searching for statistically significant features, one may use the output of encoder part of an AE as the feature vector. In the application of classification, this feature vector is used in a traditional classifier such as SVM, artificial neural network, etc. Since autoencoders have the ability to themselves extract statistically significant features, they have been successful in significantly improving the state-of-the-art classification accuracy.

Support Vector Machine (SVM) is a supervised learning based classifier [24, 7]. SVMs map the input feature data to higher dimensional feature space through some chosen non-linear Kernels such as Gaussian, linear, polynomial etc. Thus, in general, these classifiers can perform non-linear classification efficiently.

3. DATA DESCRIPTION AND PRE-PROCESSING

In this paper, we use the Schizophrenia COBRE dataset [18] for normal versus diseased classification using the proposed architecture. This dataset is a part of International Neuroimaging Data-Sharing Initiative under 1000 Functional Connectomes Project, comprised of fMRI data of 74 healthy subjects (controls) and 72 subjects of Schizophrenia with varying ages ranging from 18 to 65 years in both classes. Data has been collected with a repetition time TR=2s and echo time TE=29ms. The size of each slice is 64×64 with 32 slices in every brain volume, where voxel size is equal to $3 \times 3 \times 4mm^3$.

These fMRI data are pre-processed using SPM8 (Statistical Parametric Mapping; http://www.fil.ion.ucl.ac.uk/spm). First 5 volumes are discarded to allow magnetization to reach the steady state. Rest 145 functional volumes are slice time corrected using the middle slice as a reference followed by motion correction. Motion correction ensures head motion below 2 mm or voxel-to-voxel correspondence across time. Functional scans are spatially normalized onto Montreal Neurological Institute (MNI) space using DARTEL procedure so that results across subjects can be compared, resulting in functional images of dimension 53x63x52 (3-mm isotropic voxels). Further, data is smoothed with a Gaussian kernel (full width half maximum (FWHM)= 8mm).

The fMRI BOLD signal is active only at gray matter (GM) voxels. Hence, the normalized fMRI data was masked to filter out voxels of white matter (WM) and cerebro-spinal fluid (CSF). Next, we parcellated all gray matter voxels into 116 regions using the automated anatomical atlas (AAL).

4. PROPOSED WORK

In this section, we present the proposed two-stage SAE-based architecture for the classification of normal versus Schizophrenia subjects. The proposed strategy is as follows:

- Pre-stage: First, a covariance based strategy is proposed to filter out inactive or noisy GM voxels from the data.

- Stage-1: An autoencoder with 4-hidden layers in encoder and decoder each is trained in unsupervised manner on non-labeled data for every brain region to extract features.

- Stage-2: Learned features from all pre-trained SAEs are processed and fed to an SVM classifier that is trained on a class-labeled data for the binary (normal versus Schizophrenia subject) classification.

4.1 Pre-stage

Corresponding to any stimulus, either intrinsic or extrinsic, all of the gray matter voxels will not be activated at any particular time. Thus, it is important to filter out voxels that are *inactive* or *noisy* voxels. Although noise time series may be correlated, its covariance function given in (1) will exhibit a faster decay compared to the covariance function of the time-series of *active* or *non-noisy* voxels [17].

$$K_y[l] = \frac{E[(y[n] - \mu[n])(y[n-l] - \mu[n-l])]}{\sigma_y^2}, \qquad (1)$$

where n is the time index, $y[n]$ is the time series of a noisy voxel, $\mu[n]$ is the mean function, σ_y^2 is the variance of the time-series $y[n]$, and $K_y[l]$ is the normalized covariance function of $y[n]$ with l as the lag parameter.

In [17], authors employed K-mean based clustering to separate noisy and active voxels. The basic idea was that task-based data will induce periodicity in the covariance function of the active voxels' time-series besides having a slow decay.

However, K-means clustering based approach seemed to filter out some active voxels and leave some noisy voxels in the resting-state data. Deletion of active voxels from the useful data can lead to wrong results. Hence, we exploit the idea of difference in the decay rates of noisy and active voxels (refer to Fig.1). In order to filter out noisy voxels in a robust fashion, we propose a simple strategy.

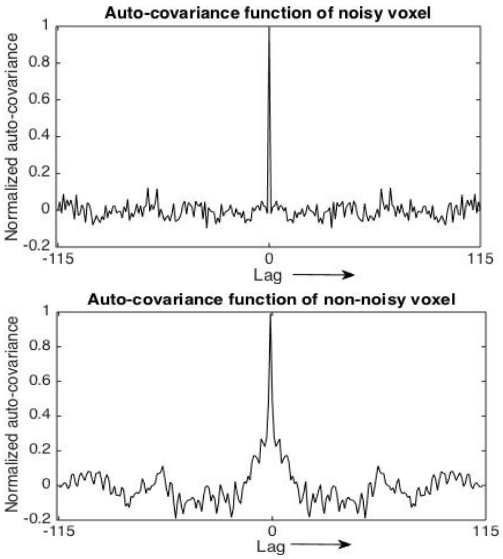

Figure 1: Covariance function of noisy and non-noisy voxels

We compute the area under the curve exhibited by $K_y[l]$ from $l = -35$ to $+35$. If this area is found to be greater than 2, we declare the voxel as noisy, else we declare it as active voxel. This is to note that, in general, hemodynamic response functions (HRFs) characterizing the system function at each voxel are found to exhibit a length of 30 to 40 taps. Hence, the correlation induced in the active voxel owing to system response function being active will roughly correspond to 40 taps or lags. Thus, we consider a lag of $l = \pm 35$ to compute area under the curve to distinguish slow decay covariance function from a fast decaying covariance function. Any value between 30 to 50 lags on each side provided us good results. We visually inspected the filtered voxels and noted that only noisy voxels got filtered out using this method.

4.2 Stage-1: Learning features via autoencoder

Figure 2 shows the architecture of stacked autoencoder (SAE) consisting of 4 hidden layers in the encoder and decoder each. Number of nodes in each of the hidden layers, i.e., $h^{(1)}$, $h^{(2)}$, $h^{(3)}$, $h^{(4)}$ are 200, 50, 100, and 8, respectively (Figure 2). In this work, we have considered 40 normal and 40 Schizophrenia subject data from the Cobre dataset, although there is no specific biasing in choosing these samples. Out of these, 10 normal and 10 Schizophrenia subject data are used for the training purposes and rest of the 30 normal and 30 Schizophrenia subjects' data are used for the testing purposes. In practice, over time, the trained classifier will be used on more and more test subjects. Hence, in order to ensure validity for realistic scenarios, we used more data for testing than training.

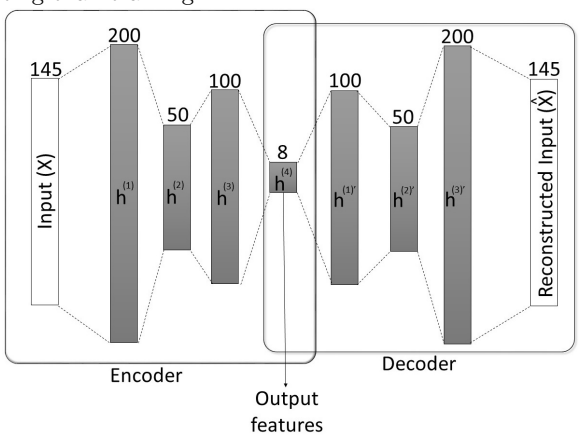

Figure 2: Architecture of Stacked Auto Encoder

In order to increase the size of data in the training phase, we have randomly chosen 50% voxels 30 times from each of the region for every subject. These provides us a dataset of 300 normal and 300 Schizophrenia subjects for the training phase.

Since we have 116 brain regions for each subject's data, we trained SAE for each region separately in an unsupervised manner with all 600 subjects' training data without class labels. The training parameters used for each of these SAE are listed in Table-1. Input to each SAE is the time series (of length 145) of active voxels corresponding to that region.

4.3 Stage-2: SVM Classifier

For training the SVM classifier, class-labeled input data is created from the pre-trained SAEs output. For every subject's input data to pre-trained SAEs, feature matrices are generated at the output. For example, let us assume that i^{th} region has N no. of active voxels. Hence, the i^{th} SAE will have an output feature matrix of size Nx8. It should be noted that in proposed architecture 145 length input feature vector is reduced to 8 length feature vector. This implies

Table 1: Parameters used in the training of proposed Stacked Sparse Autoenocder

Hidden Layer No.	l^2 weight	Sparsity regularizer	Sparsity proportion	Max epochs used
Layer 1	0.004	3	0.5	500
Layer 2	0.002	4	0.1	60
Layer 3	0.002	4	0.2	100
Layer 4	0.002	4	0.1	30

that the proposed architecture is learning informative feature vector corresponding to each voxel using the stacked autoencoder, instead of averaging over all voxels' time series that may dilute the variability within a region. In order to convert this Nx8 matrix to a vector of features for each subject and for each region, we compute the mean and standard deviation of every feature. This provides us a feature vector of length 16 for every region on every subject's data (Refer to Figure-3 for better clarity). In sum, corresponding to 116 regions, the proposed method provides a feature vector of length 16x116=1856 for every subject's data. This data is applied to a binary SVM classifier as class-labeled data during the training phase. This complete pipeline is shown in Figures 3 and 4. Radial basis function (RBF) was chosen as the Kernel type for training the SVM classifier.

5. RESULTS

In this section, we present experimental results. The model has been trained on a K40 GPU accelerator and CPU with model name: Intel(R) Xeon(R) CPU E5-2670 version 2 @ 2.5GHz with 10 cores having cache size 25600 KB with 100GB RAM.

The classifier's performance has been computed in terms of True Positive Rate (TPR) (2), False Discovery Rate (FDR) (3), Accuracy (4) and F-score (5) defined as below:

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

$$FDR = \frac{FP}{TP + FP} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{4}$$

$$TPR = \frac{2TP}{2TP + FP + FN} \tag{5}$$

where TP denotes true positives (no. of Schizophrenia classified as Schizophrenia), FN denotes false negatives (no. of Schizophrenia wrongly classified as normal), TN denotes true negatives (no. of normal subjects classified as normal), and FP denotes false positives (no. of normal subjects wrongly classified as Schizophrenia). In medical domain, both FN and FP are critical and should be as low as possible. Or, in other words, we should have high values for TPR, accuracy, and F-score, while FDR should be low.

Table-2 shows classification results with 10 fold cross- validation using the proposed method on the Cobre dataset. Results have also been tabulated using the proposed algorithm with SVM at the output of hidden layer vs softmax decision layer at the output of the hidden layer. From this table, it is observed that the proposed architecture provides 92% accuracy compared to 86.5% accuracy with single autoencoder.

In Table-3, accuracy results are presented on Cobre dataset as reported in the literature. From these results, it is noted that the proposed architecture has better accuracy compared to the existing methods. Table-2 presents averaged accuracy over all 10 folds, while Table-4 and Table-5 present accuracy results with the proposed method on each fold with SVM and softmax, respectively. It is observed that the proposed architecture achieves a maximum of 95% and 93.33% accuracy in one fold with SVM and Softmax, respectively that is much higher than existing methods tabulated in Table-3. The mimimum of accuracy over all individual folds with SVM, i.e., 90% is also higher than the maximum accuracy of 85.8% as achieved in the literature.

The better performance observed with the proposed methodology may be owing to the following reasons:

- Noisy or inactive gray matter voxels have been filtered out from each brain region before feeding data to SAEs.
- Instead of working with the mean time series and/or the temporal mean of the time series of a voxel, full 145 length time series of each of the active voxels has been considered. Hence, no information loss of data has happened, while the data is applied as input to the proposed architecture. Thus, variability within data of active voxels of a region is not lost with the proposed architecture.
- Features have been learned for every brain region via unsupervised training of 4-hidden layer SAEs on each region capturing the regional characteristics and considering all active gray matter voxels' data as input to each SAE. Later, statistical features have been captured from the regional voxels' feature vectors on each subject to better characterize the regional features that are then applied to train an SVM classifier.

6. CONCLUSION

In this paper, Stacked AutoEncoder (SAE) based two-stage architecture has been proposed for classification of normal versus Schizophrenia subjects from the functional MRI data of publicly available 1000 Functional Connectomes Project fMRI database. The proposed architecture works directly on active voxels' time series without converting them into region-wise mean time series. The proposed methodology provides a very good 10-fold cross-validation accuracy of 92% that is better than the existing methods used on the same dataset.

7. ACKNOWLEDGMENTS

Authors gratefully acknowledge the research funding support (Grant Number: 1(7)/2014-ME&HI) from the Ministry of Communication and IT, Govt. of India for this research work.

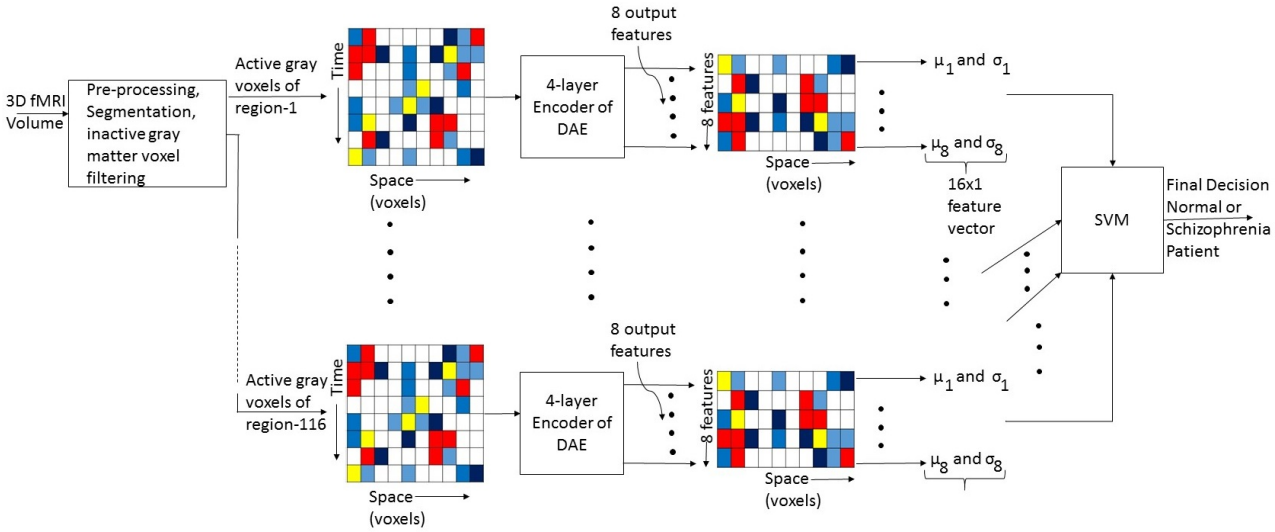

Figure 3: Proposed Two Stage Architecture

Table 2: Test results with 10 fold cross validation

Algorithm	TPR (%)	FDR (%)	Accuracy (%)	F-score (%)
Proposed algorithm (with multiple AEs and SVM)	88.67	4.83	92	92.27
Multiple AEs with Softmax	83.67	10.49	86.5	86.65

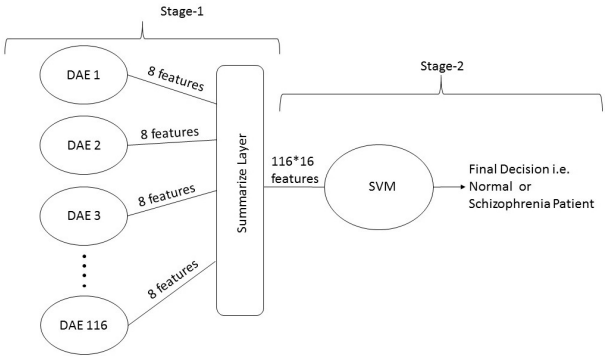

Figure 4: Proposed Two Stage Architecture

Table 3: Classification accuracy results on Cobre Dataset as reported in the literature

Sr.No.	References	Accuracy(%)
1	Proposed (with 10-fold cross validation)	92
2	Kim et al. (2016) [12]	85.8
3	Anderson and Cohen (2013) [1]	65
4	Cheng et al. (2015) [4]	80
5	Chyzhyk et al. (2015) [5]	90
6	Hsieh et al. (2014) [9]	71.6
7	Savio and Grana (2015) [21]	80

Table 4: Proposed algorithm accuracy (with SVM) on each of the individual 10 folds

Fold No.	Accuracy (%)
1	95
2	93.33
3	90
4	91.67
5	93.33
6	91.67
7	90
8	91
9	90
10	93.33

Table 5: Test accuracy using "Multiple AEs with Softmax" on each of the individual 10 folds

Fold No.	Accuracy (%)
1	91.67
2	81.67
3	85
4	93.33
5	86.67
6	91.67
7	80
8	78.33
9	93.33
10	83.33

8. REFERENCES

[1] A. Anderson and M. S. Cohen. Decreased small-world functional network connectivity and clustering across resting state networks in schizophrenia: an fMRI

classification tutorial, 2013.

[2] P. Baldi. Autoencoders, unsupervised learning, and deep architectures. *ICML unsupervised and transfer learning*, 27(37-50):1, 2012.

[3] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

[4] H. Cheng, S. Newman, J. Goñi, J. S. Kent, J. Howell, A. Bolbecker, A. Puce, B. F. O'Donnell, and W. P. Hetrick. Nodal centrality of functional network in the differentiation of Schizophrenia. *Schizophrenia research*, 168(1):345–352, 2015.

[5] D. Chyzhyk, A. Savio, and M. Graña. Computer aided diagnosis of schizophrenia on resting state fMRI data by ensembles of ELM. *Neural Networks*, 68:23–33, 2015.

[6] D. Cordes, V. M. Haughton, K. Arfanakis, G. J. Wendt, P. A. Turski, C. H. Moritz, M. A. Quigley, and M. E. Meyerand. Mapping functionally related regions of brain with functional connectivity MR imaging. *American Journal of Neuroradiology*, 21(9):1636–1644, 2000.

[7] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[8] M. D. Fox and M. Greicius. Clinical applications of resting state functional connectivity. *Frontiers in systems neuroscience*, 4:19, 2010.

[9] T.-H. Hsieh, M.-J. Sun, and S.-F. Liang. Diagnosis of Schizophrenia patients based on brain network complexity analysis of resting-state fMRI. In *The 15th International Conference on Biomedical Engineering*, pages 203–206. Springer, 2014.

[10] S. Huang, J. Li, L. Sun, J. Ye, A. Fleisher, T. Wu, K. Chen, E. Reiman, A. D. N. Initiative, et al. Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949, 2010.

[11] D. T. Jones, P. Vemuri, M. C. Murphy, J. L. Gunter, M. L. Senjem, M. M. Machulda, S. A. Przybelski, B. E. Gregg, K. Kantarci, D. S. Knopman, et al. Non-stationarity in the resting brain's modular architecture. *PloS one*, 7(6):e39731, 2012.

[12] J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *NeuroImage*, 124:127–146, 2016.

[13] D. Kuang, X. Guo, X. An, Y. Zhao, and L. He. Discrimination of ADHD based on fMRI data with Deep Belief Network. In *International Conference on Intelligent Computing*, pages 225–232. Springer, 2014.

[14] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J.-M. Annoni, M. Schluep, P. Vuilleumier, and D. Van De Ville. Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. *NeuroImage*, 83:937–950, 2013.

[15] N. Leonardi and D. Van De Ville. Identifying network correlates of brain states using tensor decompositions of whole-brain dynamic functional connectivity. In *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on*, pages 74–77. IEEE, 2013.

[16] X. Li, D. Zhu, X. Jiang, C. Jin, X. Zhang, L. Guo, J. Zhang, X. Hu, L. Li, and T. Liu. Dynamic functional connectomics signatures for characterization and differentiation of ptsd patients. *Human brain mapping*, 35(4):1761–1778, 2014.

[17] D. Liu, N. Zhong, and Y. Qin. Exploring functional connectivity networks in fMRI data using clustering analysis. In *International Conference on Brain Informatics*, pages 148–159. Springer, 2011.

[18] T. M. R. Network and the University of New Mexico. The center for biomedical research excellence (cobre), 2012.

[19] A. Ng. Sparse autoencoder. *CS294A Lecture notes*, 72:1–19, 2011.

[20] Ü. Sakoğlu, G. D. Pearlson, K. A. Kiehl, Y. M. Wang, A. M. Michael, and V. D. Calhoun. A method for evaluating dynamic functional network connectivity and task-modulation: application to Schizophrenia. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23(5-6):351–366, 2010.

[21] A. Savio and M. Graña. Local activity features for computer aided diagnosis of Schizophrenia on resting-state fMRI. *Neurocomputing*, 164:154–161, 2015.

[22] T. Starck, J. Nikkinen, J. Rahko, J. Remes, T. Hurtig, H. Haapsamo, K. Jussila, S. Kuusikko-Gauffin, M.-L. Mattila, E. Jansson-Verkasalo, D. L. Pauls, H. Ebeling, I. Moilanen, O. Tervonen, and V. J. Kiviniemi. Resting state fmri reveals a default mode dissociation between retrosplenial and medial prefrontal subnetworks in asd despite motion scrubbing. *Front Hum Neurosci*, 7:802, Nov 2013. 24319422[pmid].

[23] H.-I. Suk, S.-W. Lee, and D. Shen. A hybrid of deep network and hidden markov model for MCI identification with resting-state fMRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 573–580. Springer, 2015.

[24] V. Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.