Exploring Common and Distinct Structural Connectivity Patterns Between Schizophrenia and Major Depression via Cluster-driven Nonnegative Matrix Factorization

Junming Shao*, Zhongjing Yu*, Peiyan Li*, Wei Han*, Christian Sorg†, and Qinli Yang*

*Big Data Research Center, School of Computer Science and Engineering
University of Electronic Science and Technology of China, Chengdu 611731, China
Email: {junmshao@uestc.edu.cn, qinli.yang@uestc.edu.cn}
†Technische Universität München, Ismaninger Strasse 22, 81675 Munich, Germany

Abstract—In this paper, we introduce a novel method to discover common and distinct structural connectivity patterns between SZP and MDD via a Cluster-Driven Nonnegative Matrix Factorization (called CD-NMF). Specifically, CD-NMF is applied to decompose the joint structural connectivity map into common and distinct parts, and each part is further factorized into two sub-matrices (i.e. common/distinct basis matrix and common/distinct encoding matrix) correspondingly. By imposing the clustering constraints on common and distinct encoding matrices, the discriminative patterns as well as the common patterns between the two disorders are extracted simultaneously. Experimental results demonstrate that CD-NMF allows finding the common and distinct structural patterns effectively. More importantly, the derived distinct patterns, show powerful ability to discriminate the patients of schizophrenia and major depressive disorder.

Keywords-nonnegative matrix factorization; structural connectivity; biomarker;

I. INTRODUCTION

Schizophrenia (SZP) and major depressive disorder (MDD) are frequent psychiatric disorders with lifetime prevalence of about 0.3-0.7%, and 3-17%, respectively. While SZP and MDD differ in many symptomatic aspects, they also show relevant overlaps in several dimensions and levels. For example, clinical symptoms of patients with SZP and MDD often transcend diagnostic categories, e.g. loss of pleasure and interest can be found in patients with schizophrenia and major depression, respectively [1], [2]. The genetic correlation for single nucleotide polymorphism is about 0.43 between SZP and MDD, suggesting shared genetic etiology and common pathophysiological pathways [3]. It is also estimated that approximately half of patients with schizophrenia have comorbid with depression [4]. Such overlaps are not limited to schizophrenia and major depression, but a very general and fundamental aspect of psychiatric-diagnostic categories, resulting in significant doubts about sharp diagnostic borders between distinct psychiatric disorders in general [5].

Nonnegative matrix factorization (NMF), has been shown to be a powerful tool to handle high-dimensional data in many applications. Considering the unique property of neuroimaging data (i.e. small number of samples yet with high dimensionality), in this paper, we propose a novel cluster-driven nonnegative matrix factorization method, called CD-NMF, which aims at discovering the common and distinct brain features between major depression and schizophrenia simultaneously.

A. Basic Idea

Our cluster-driven NMF, works on a joint data matrix (i.e. including the data of two groups), and decomposes it into two parts: common part and distinct part directly. For both parts, they are further factorized into two matrices like traditional nonnegative matrix factorization, where one matrix corresponds to the basis matrix (e.g. connectivity-pattern matrix in this study) and the other characterizes the encoding matrix (e.g. the subject-pattern matrix). To ensure the two factorized basis matrices allow capturing the common and distinct patterns simultaneously, we leverage the available label information to impose the clustering constraints on the two encoding matrices. Specifically, for the distinct part, if we want to extract distinct patterns between two groups, we may expect that the corresponding encodings for each group are as similar as possible, and meanwhile, the corresponding encodings of subjects between two groups are as different as possible. Similarly, for the common part, we may expect that the corresponding encodings of all subjects for both groups are as similar as possible. By adding the clustering constraints on encoding matrices, the two basic matrices characterize the common and distinct patterns between two groups intuitively.

To illustrate the cluster-driven nonnegative matrix factorization, Fig. 1 gives an illustrative example. Given a joint matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, it is factorized into two parts: common connectivity matrix $\mathbf{A}_c$ and distinct connectivity matrix $\mathbf{A}_d$, respectively (see Figure 2). The common connectivity matrix $\mathbf{A}_c$ is further factorized into two matrices $\mathbf{W}_c \in \mathbb{R}^{m \times k_c}$ and $\mathbf{H}_c \in \mathbb{R}^{n \times k_c}$, where $\mathbf{W}_c$ represents the common structural patterns shared by both disorders while $\mathbf{H}_c$ indicates its corresponding weights for each pattern, and k_c is number of common structural patterns. Similarly,
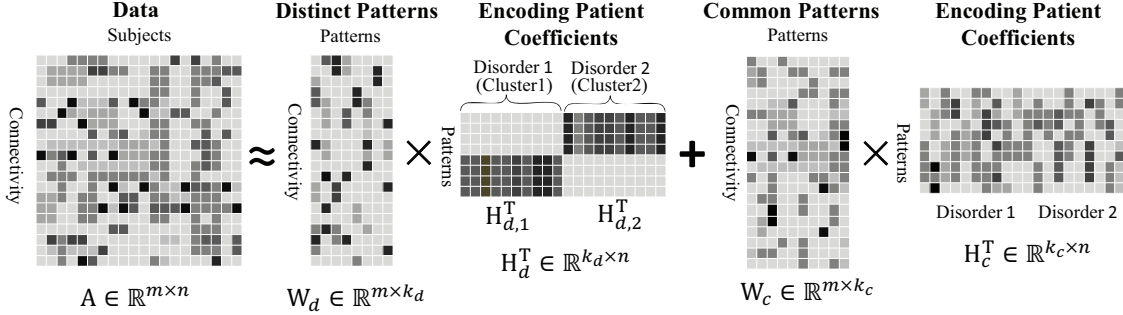
Figure 1: The illustration of our cluster-driven NMF for common and distinct structural pattern discovery. Given a joint subject-connectivity map **A**, it is decomposed into common and distinct parts by imposing the clustering constraints on the encoding coefficients, where the encoding coefficients of subjects in $\mathbf{H}_{d,1}$, $\mathbf{H}_{d,2}$ or $\mathbf{H}_c$ are as similar as possible, while the encoding coefficients of subjects between $\mathbf{H}_{d,1}$ and $\mathbf{H}_{d,2}$ are as dissimilar as possible.

the distinct connectivity matrix $\mathbf{A}_d$ is factorized into two matrices $\mathbf{W}_d \in \mathbb{R}^{m \times k_d}$ and $\mathbf{H}_d \in \mathbb{R}^{n \times k_d}$, where $\mathbf{W}_d$ represents the distinct structural patterns for two disorders, $\mathbf{H}_d$ indicates its corresponding weights for each pattern and k_d is number of distinct structural patterns. To identify the common and distinct connectivity patterns, we add the constraints on the encodings $\mathbf{H}_d$ and $\mathbf{H}_c$, respectively (Fig. 1). For $\mathbf{H}_d$, the encodings of subjects in one disorder $\mathbf{H}_{d,1}$ are similar and the encodings of subjects in another disorder $\mathbf{H}_{d,2}$ are similar. However, the $\mathbf{H}_{d,1}$ and $\mathbf{H}_{d,2}$ are dissimilar. For $\mathbf{H}_c$, the encodings of all subjects for both disorders are similar. With this strategy, the common and distinct patterns are intuitively obtained simultaneously.

II. RELATED WORK

During the past decade, many approaches have been proposed to find the biomarker for MDD and SZP. In recent years, an interesting line of biomarker research has focused on the in-vivo brain neuroimaging data. Many studies have demonstrated that the reduced regional gray matter volumes [6], and aberrant white matter structural connectivity [7], are frequently found in the cortical parts of intrinsic triple networks (i.e. default mode network, central executive network and salience network [8]) in patients of MDD and SZP [9].

On the methodology side, during past decades, many group-based NMF variants have been proposed, such as [10], [11], [12]. For instance, Zafeiriou et al.[10] have developed the discriminant NMF (DNMF), which extracts features that enforce not only the spatial locality, but also the separability between classes in a discriminant manner. Similarly, Potluru et al.[13] have introduced a contrast term to NMF (called co-NMF), to identify distinctive features between two groups. Lee et al.[11] have proposed a semi-supervised of NMF (SSNMF) by jointly exploiting both limited labeled and plenty of unlabeled data to extract more discriminative features. For most existing group-based NMF approaches, they mainly focus on the discriminative pattern

discovery. Recently, Kim et al. [12] introduce a new joint matrix factorization approach to simultaneously discovers common as well as discriminative topics given multiple document sets in an unsupervised way.

III. DATA ACQUISITION AND PREPROCESSING

Subjects and Data acquisition. 25 patients with major depression, 21 patients with schizophrenia, and 25 healthy controls participated in the study. All participants provided informed consent in accordance with the Human Research Committee guidelines of the Klinikum Rechts der Isar, TechnischeUniversität, München. All subjects underwent T1-weighted, and diffusion-weighted imaging in a 3T Philips Achieva using an eight-channel phased-array head coil. We instructed the participants to keep their eyes closed and not to fall asleep during the 10-minute scan. T1-weighted structural data were obtained using a magnetization-prepared rapid acquisition gradient echo sequence, and Diffusion weighted MRI was based on a pulsed gradient spin-echo echo planar imaging sequence.

Construction of brain structural connectivity network. To construct brain structural network, selected nodes of default mode, salience network and executive network were pre-defined by analysis of rs-fMRI data of an independent sample of healthy controls. In brief, 25 healthy controls were scanned on the same MRI scanner by the same rs-fMRI sequence. Data were preprocessed and analyzed in the same way as described in [9], [14]. Preprocessing includes motion correction, smoothing, and normalization. High-model-order group independent component analysis was performed to identify components reflecting intrinsic networks. Using spatial regression and spatial templates of default mode, salience network and executive network based on Uddin et al.[15], components-of-interest were selected. Nodes of default mode, salience network and central executive network were defined as spherical regions-of-interest (ROI) of $3mm$ radius and local peaks of networks. In total, 105 ROIs were
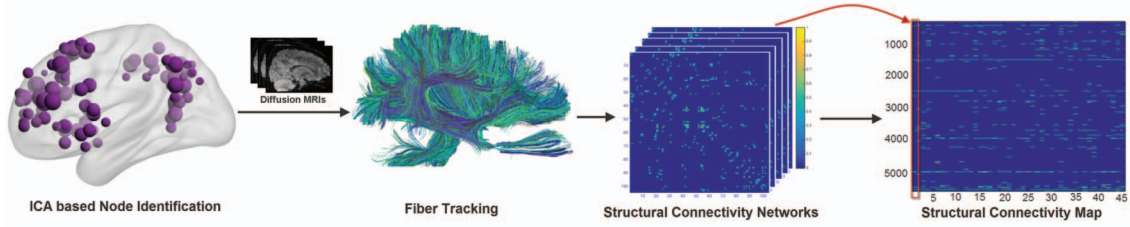
Figure 2: The chartflow of structural brain network construction.

generated to represent three networks of interest: 30 ROIs for the default mode network, and 37 ROIs for the salience network and 38 ROIs for executive network. Building upon these nodes, the flowchart of constructing the brain structural connectivity network is illustrated in Figure 2. For more information, please refer to the work [16].

IV. MINING COMMON AND DISTINCT STRUCTURAL CONNECTIVITY PATTERNS

Building upon the derived joint structural connectivity map, in this section we will introduce our cluster-driven nonnegative matrix factorization to discover the common and distinct structural connectivity patterns between major depressive disorder and schizophrenia.

A. Cluster-driven Nonnegative Matrix Factorization

Formally, suppose that there are two patient groups with distinct psychiatric disorders (e.g. major depression and schizophrenia), the objective function is defined as follows.

$$\min_{\mathbf{W}_c,\mathbf{H}_c,\mathbf{W}_d,\mathbf{H}_d} f(\mathbf{W}_c,\mathbf{H}_c,\mathbf{W}_d,\mathbf{H}_d) = ||\mathbf{A}-\mathbf{W}_c\mathbf{H}_c^T-\mathbf{W}_d\mathbf{H}_d^T||_F^2$$
$$+f_d(\mathbf{H}_d) + f_c(\mathbf{H}_c) \quad (1)$$

$$\text{s.t. } \mathbf{W}_c \geq 0, \mathbf{W}_d \geq 0 \text{ AND } \mathbf{H}_c \geq 0, \mathbf{H}_d \geq 0$$

$$||\mathbf{H}_d(\cdot,l)||_2 = 1, \text{for } l = 1,\cdots,k_d$$

$$||\mathbf{H}_c(\cdot,l)||_2 = 1, \text{for } l = 1,\cdots,k_c$$

where $f_c(\cdot)$ and $f_d(\cdot)$ are the two penalty functions on the encoding matrices. k_d and k_c are the number of distinct and common structural connectivity patterns, respectively.

To identify the common and distinct connectivity patterns between two disorders, we introduce a clustering-style constraints on the encoding matrices on both $\mathbf{H}_c$ and $\mathbf{H}_d$. As stated in Section I-A, as the encoding matrices represent the weight for each connectivity pattern, if we want to extract the distinct patterns, the corresponding weights for different disorders should be as different as possible, and meanwhile, the weights for subjects in the same disorder are as similar as possible. Similarly, for extracting common patterns, the corresponding weights for all subjects in both

disorders should be similar. Therefore, we define our penalty functions of f_d and f_c as follows.

$$f_d(\mathbf{H}_d) = \alpha_1\left(||\mathbf{H}_{d,1} - \mathbf{M}_{d,1}||_F^2 + ||\mathbf{H}_{d,2} - \mathbf{M}_{d,2}||_F^2\right)$$
$$+\alpha_2||\mathbf{H}_{d,1} \odot \mathbf{H}_{d,2}||_1^2 \quad (2)$$

$$f_c(\mathbf{H}_c) = \alpha_1||\mathbf{H}_c - \mathbf{M}_c||_F^2 \quad (3)$$

where α_1 and α_2 are the scalars controlling the relative contribution of the corresponding terms. $\mathbf{H}_{d,1} \in \mathbb{R}^{n_1 \times k_d}$ and $\mathbf{H}_{d,2} \in \mathbb{R}^{n_2 \times k_d}$ are the encoding matrices for two groups, respectively. $\mathbf{H}_d = [\mathbf{H}_{d,1}; \mathbf{H}_{d,2}]$. $\mathbf{M}_{d,1}$ is a $n_1 \times k_d$ matrix where each row is a same vector of the average encoding coefficients for major depression (i.e. $\frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{H}_d(i,:)$). Similarly, $\mathbf{M}_{d,2}$ is a $n_2 \times k_d$ matrix where each row is a same vector of average encoding coefficients for schizophrenia (i.e. $\frac{1}{n_2}\sum_{i=1}^{n_2}\mathbf{H}_d(n_1 + i,:)$). $\mathbf{M}_c \in n \times k_c$ contains n number of rows with the same vector of the encoding coefficients for the whole $\mathbf{H}_c$ ($\frac{1}{n}\sum_{i=1}^{n}\mathbf{H}_c(i,:)$). The two terms: $||\mathbf{H}_{d,1}-\mathbf{M}_{d,1}||_F^2+||\mathbf{H}_{d,2}-\mathbf{M}_{d,2}||_F^2$ and $||\mathbf{H}_{d,1}\odot\mathbf{H}_{d,2}||_1^2$ are thus used to construct two clusters, where the corresponding encoding coefficients $\mathbf{H}_{d,1}$ are a cluster and the $\mathbf{H}_{d,2}$ is another cluster. The (i,j)-th component of $\mathbf{H}_{d,1} \odot \mathbf{H}_{d,2}$ corresponds to the inner product between $\mathbf{H}_{d,1}^i$, the i-th row encoding coefficient of $\mathbf{H}_{d,1}$, and $\mathbf{H}_{d,2}^j$, the j-th row encoding coefficient of $\mathbf{H}_{d,2}^j$. In addition, the constraints $||\mathbf{H}_d(l,\cdot)||_2 = 1$ and the minimization of $||\mathbf{H}_{d,1} \odot \mathbf{H}_{d,2}||_1^2$ are finally to encourage the sparsity of $\mathbf{H}_d$ so that some of them become exactly zero.

To solve this objective function, the block-coordinate descent (BCD) framework is applied which guarantees every limit point is a stationary point [17], and we obtain the updating as follows.

$$\mathbf{w}_c^{(l)} \leftarrow \mathbf{w}_c^{(l)} + \left[\frac{(\mathbf{A}\mathbf{H}_c)_{\cdot l} - \mathbf{W}_d(\mathbf{H}_d^T\mathbf{H}_d)_{\cdot l} - \mathbf{W}_c(\mathbf{H}_c^T\mathbf{H}_c)_{\cdot l}}{(\mathbf{H}_c^T\mathbf{H}_c)_{ll}}\right]_+ \quad (4)$$

$$\mathbf{w}_d^{(l)} \leftarrow \mathbf{w}_d^{(l)} + \left[\frac{(\mathbf{A}\mathbf{H}_d)_{\cdot l} - \mathbf{W}_c(\mathbf{H}_c^T\mathbf{H}_c)_{\cdot l} - \mathbf{W}_d(\mathbf{H}_d^T\mathbf{H}_d)_{\cdot l}}{(\mathbf{H}_d^T\mathbf{H}_d)_{ll}}\right]_+ \quad (5)$$

$$\mathbf{h}_c^{(l)} \leftarrow \left[\frac{(\mathbf{A}^T\mathbf{W}_c)_{\cdot l} - \mathbf{H}_d(\mathbf{W}_d^T\mathbf{W}_c)_{\cdot l} - \mathbf{H}_c(\mathbf{W}_c^T\mathbf{W}_c)_{\cdot l}}{(\mathbf{W}_c^T\mathbf{W}_c)_{ll} + \alpha_1}\right.$$

Algorithm 1 The Pseudocode of CD-NMF algorithm.

1: **Input:** the joint matrix A, integers k_c and k_d, and parameters α_1, α_2

2: Initialize $\mathbf{W}_c$, $\mathbf{W}_d$, $\mathbf{H}_c$ and $\mathbf{H}_d$
3: //Update each column of $\mathbf{W}_c$ and $\mathbf{H}_c$
4: **repeat**
5:     **for** i = 1 to k_c **do**
6:         Update $\mathbf{W}_c$ using Eq. (4);
7:         Update $\mathbf{H}_c$ using Eq. (6);
8:     **end for**
9:     **for** i = 1 to k_d **do**
10:         Update $\mathbf{W}_d$ using Eq. (5);
11:         Update $\mathbf{H}_d$ using Eqs. (7) and (8) ;
12:     **end for**
13: **until** stopping criteria is satisfied.

14: **Output:** $\mathbf{W}_c$, $\mathbf{W}_d$, $\mathbf{H}_c$ and $\mathbf{H}_d$;

$$
\left. +\frac{\mathbf{H}_c^{(l)}(\mathbf{W}_c^T\mathbf{W}_c)_{ll}+\alpha_1\times\mathbf{m}_c^{(l)}}{(\mathbf{W}_c^T\mathbf{W}_c)_{ll}+\alpha_1} \right]_+ \tag{6}
$$

$$
\mathbf{h}_{d,1}^{(l)} \leftarrow \left[\frac{(\mathbf{A}_1^T\mathbf{W}_d)_{\cdot l}-\mathbf{H}_{d,1}(\mathbf{W}_d^T\mathbf{W}_d)_{\cdot l}-\mathbf{H}_{c,1}(\mathbf{W}_c^T\mathbf{W}_d)_{\cdot l}}{(\mathbf{W}_d^T\mathbf{W}_d)_{ll}+\alpha_1} \right.
$$

$$
\left. +\frac{\mathbf{h}_{d,1}^{(l)}(\mathbf{W}_d{}^T\mathbf{W}_d)_{ll}+\alpha_1\mathbf{m}_{d,1}^{(l)}-\frac{\alpha_2}{2}\mathbf{1}_{n_1\times 1}\cdot\mathbf{h}_{d,2}^{(l)}{}^T\cdot\mathbf{1}_{n_2\times 1}}{(\mathbf{W}_d^T\mathbf{W}_d)_{ll}+\alpha_1} \right]_+ \tag{7}
$$

$$
\mathbf{h}_{d,2}^{(l)} \leftarrow \left[\frac{(\mathbf{A}_2^T\mathbf{W}_d)_{\cdot l}-\mathbf{H}_{d,2}(\mathbf{W}_d^T\mathbf{W}_d)_{\cdot l}-\mathbf{H}_{c,2}(\mathbf{W}_c^T\mathbf{W}_d)_{\cdot l}}{(\mathbf{W}_d^T\mathbf{W}_d)_{ll}+\alpha_1} \right.
$$

$$
\left. +\frac{\mathbf{h}_{d,2}^{(l)}(\mathbf{W}_d{}^T\mathbf{W}_d)_{ll}+\alpha_1\mathbf{m}_{d,2}^{(l)}-\frac{\alpha_2}{2}\mathbf{1}_{n_2\times 1}\cdot\mathbf{h}_{d,1}^{(l)}{}^T\cdot\mathbf{1}_{n_1\times 1}}{(\mathbf{W}_d^T\mathbf{W}_d)_{ll}+\alpha_1} \right]_+ \tag{8}
$$

Initialization. Here, we first partition the whole matrix $\mathbf{A}$ into $k = k_c + k_d$ clusters via K-Means algorithm (For stable, the initialization of K-Means is fixed as the first k objects to avoid randomness). Afterwards, the k_d number of the most different centroids is selected as the initial distinct patterns and the remaining centroids are regarded as the initial common patterns. For $\mathbf{H}_c$, it is initialized as: $\mathbf{H}_c = \frac{ones(n,k_c)}{n}$. Finally, $\mathbf{H}_d$ is obtained by using $\mathbf{H}_d^T = \mathbf{W}_d^{-1}(\mathbf{A} - \mathbf{W}_c\mathbf{H}_c^T)$.

B. CD-NMF Based Classification

Building upon the cluster-driven nonnegative matrix factorization, the classification procedure mainly involves the following steps:

1) Using any cross-validation strategy (e.g. 10-fold cross validation) to partition the joint structural connectivity map $\mathbf{A}$ into $\mathbf{A}_{\text{TRAIN}}$ and $\mathbf{A}_{\text{TEST}}$ for each fold.

2) For each fold, decompose $\mathbf{A}_{\text{TRAIN}}$ using CD-NMF, and obtain the distinct structural connectivity patterns $\mathbf{W}_d$.
3) Building upon $\mathbf{W}_d$, generate the new data $\mathbf{S}_{\text{TRAIN}} = \mathbf{W}_d^T\mathbf{A}_{\text{TRAIN}}$ and $\mathbf{S}_{\text{TEST}} = \mathbf{W}_d^T\mathbf{A}_{\text{TEST}}$.
4) Finally, $\mathbf{S}_{\text{TRAIN}}$ and $\mathbf{S}_{\text{TEST}}$ are used to train a support vector machine classifier. The prediction performance is finally averaged by all folds.

V. EXPERIMENT

A. Synthetic Data

We start the evaluation with synthetic data to prove the concept of CD-NMF. To test whether CD-NMF can extract the common and distinct patterns between two groups, here a synthetic data matrix with size of 20×40 is generated to contain two groups, where each column corresponds to one sample and each group has 20 samples (see Fig. 3). Five common patterns, across subjects of both groups, are created (corresponding to the rows of 4,8,12,16,20). The six distinct patterns for the first group (corresponding to the rows of 2,3,5,6,9,11) and the second group (corresponding to the rows of 1,7,10,13,14,17) are further generated in the data set. Fig. 3 (a) and (d) show the extracted common and distinct patterns detected by CD-NMF by setting different values of k_d and k_c. It is interesting to notice that CD-NMF allows finding all generated common and distinct patterns effectively. More importantly, with different k_d and k_c, the obtained common and distinct patterns are almost the same. The only difference among these factorizations is that common and distinct patterns are split into several parts with the increase of k_c and k_d. Beyond, the number of distinct patterns for each group, is automatically obtained. Namely, the number distinct patterns for each group are naturally obtained during the factorization process due to the clustering constraint.

B. Evaluation on Real-world Data

Following the brain connectomes stated in Section III, the joint structural connectivity map of major depression and schizophrenia is obtained. Building upon the matrix, we applied the CD-NMF algorithm to factorize it into four sub-matrices with $k_d = 5$ and $k_c = 10$. We set $\alpha_1 = 16$ and $\alpha_2 = 2$ in this study. Fig. 5 plots the five distinct connectivity patterns, where two connectivity patterns belong to major depression and the remaining connectivity patterns characterize the schizophrenia. Similarly, we plot five out of ten common structural connectivity patterns shared by the two disorders (see Fig. 4). It is interesting note that although MDD and SZP shares many common ROIs, they have their specific altered structural connectivity (see Fig. 5). These common and specific changed structural connectivity, which are highly consistent with the previous studies, such as [18].
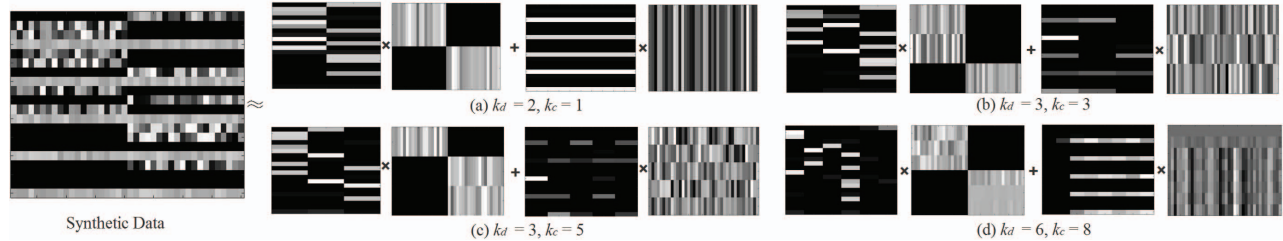
Figure 3: The common and distinct pattern discovery on a toy example. By specifying the parameters α_1 and α_2 with suitable values, CD-NMF allows extracting all common and distinct patterns with different k_d and k_c.
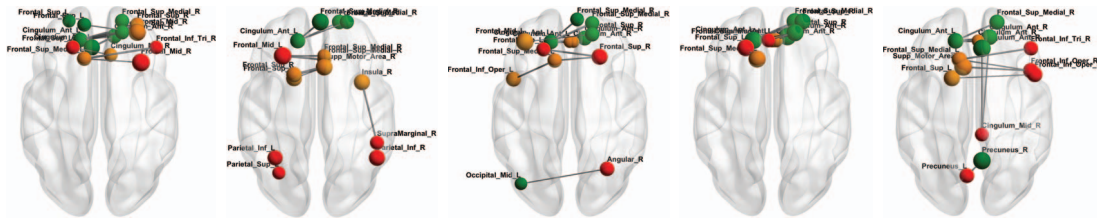

Figure 4: Common structural connectivity patterns shared by major depression and schizophrenia. Due to space limitation, only five of ten common patterns are displayed.
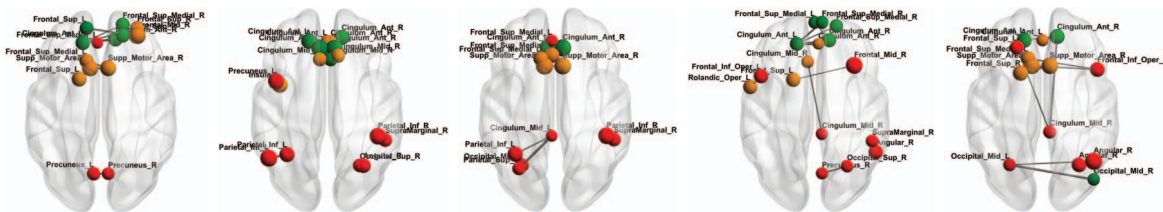

Figure 5: Distinct structural connectivity patterns between major depression and schizophrenia. Here, the first two plots visualize the specific changed patterns in MDD and the rest plots correspond to the specific changes for SZP.

C. Prediction performance with distinct connectivity

To further evaluate whether the extract distinct connectivity patterns really characterize the discriminative information of the two disorders, the prediction performance is reported and also compared its performance to several representatives of discriminative feature representation techniques: (1) the baseline support vector machine (SVM) with original features; (2) the feature selection strategy (Information Gain) [19]; (3) the wide-spread discriminant NMF: DNMF [10]; (4) a local NMF representation with Fisher linear discriminant analysis : FNMF [20]; (5) the popular semi-supervised NMF: SSNMF [11]. The last three comparing algorithms are also combined with SVM for prediction. For all comparing algorithms, the best prediction performance is reported. For FNMF, DNMF and SSNMF, due to the random initialization (All comparing algorithms cannot produce their best results with our initialization strategy, and thus the random initialization is used.), all reported results are the averaged classification accuracies over ten times. Fig. 6 plots the classification accuracies of different algorithms for predicting the two disorders. From the plot, we can see that CD-

NMF shows its superiority over all comparing algorithms, with more than 16% gained classification accuracy.

VI. CONCLUSION

In this paper, we propose a novel cluster-driven nonnegative matrix factorization approach to discover the common and distinct connectivity patterns between major depression and schizophrenia. In contrast to existing ROI-based investigation in traditional neuroimaging data analysis, our new approach, provides a new way to bring deep insight into the two disorders at the network level. The high discriminative power of derived distinct patterns, further demonstrates the effectiveness of CD-NMF.
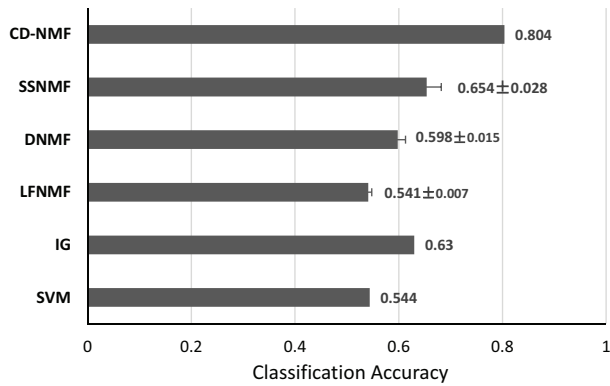
VII. ACKNOWLEDGEMENT

Figure 6: The best prediction performance with different algorithms. SVM: LibSVM with default parameters; IG: Feature selection with top 3000 features based on information gain; FNMF: $k = 20$ and $\alpha = 0.01$; SSNMF: $k = 38$ and DNMF: $k = 20$, $\delta = 0.0001$, and $\gamma = 0.1$; CD-NMF: $k_c = 5$, $k_d = 10$, $\alpha_1 = 16$ and $\alpha_2 = 2$. IG, FNMF, SSNMF and DNMF are all combined with LibSVM for Classification.

REFERENCES

[1] M. A. Marissen, N. Arnold, and I. H. Franken, "Anhedonia in borderline personality disorder and its relation to symptoms of impulsivity," *Psychopathology*, vol. 45, no. 3, pp. 179–184, 2012.

[2] D. A. Pizzagalli, "Depression, stress, and anhedonia: toward a synthesis and integrated model," *Annual review of clinical psychology*, vol. 10, pp. 393–423, 2014.

[3] C.-D. G. of the Psychiatric Genomics Consortium *et al.*, "Genetic relationship between five psychiatric disorders estimated from genome-wide snps," *Nature Genetics*, vol. 45, no. 9, pp. 984–994, 2013.

[4] P. F. Buckley, B. J. Miller, D. S. Lehrer, and D. J. Castle, "Psychiatric comorbidities and schizophrenia," *Schizophrenia Bulletin*, vol. 35, no. 2, pp. 383–402, 2009.

[5] T. Insel, B. Cuthbert, M. Garvey, R. Heinssen, D. Pine, K. Quinn, C. Sanislow, and P. Wang, "Research domain criteria (rdoc): toward a new classification framework for research on mental disorders." *The American journal of psychiatry*, vol. 167, no. 7, pp. 748–751, 2010.

[6] I. Ellison-Wright and E. Bullmore, "Meta-analysis of diffusion tensor imaging studies in schizophrenia," *Schizophrenia Research*, vol. 108, no. 1, pp. 3–10, 2009.

[7] M. Quan, S.-H. Lee, M. Kubicki, Z. Kikinis, Y. Rathi, L. J. Seidman, R. I. Mesholam-Gately, J. M. Goldstein, R. W. McCarley, M. E. Shenton *et al.*, "White matter tract abnormalities between rostral middle frontal gyrus, inferior frontal gyrus and striatum in first-episode schizophrenia," *Schizophrenia Research*, vol. 145, no. 1, pp. 1–10, 2013.

[8] V. Menon, "Large-scale brain networks and psychopathology: a unifying triple network model," *Trends in Cognitive Sciences*, vol. 15, no. 10, pp. 483–506, 2011.

[9] A. Manoliu, V. Riedl, A. Zherdin, M. Mühlau, D. Schwerthöffer, M. Scherr, H. Peters, C. Zimmer, H. Förstl, J. Bäuml *et al.*, "Aberrant dependence of default mode/central executive network interactions on anterior insular salience network activity in schizophrenia," *Schizophrenia Bulletin*, p. sbt037, 2013.

[10] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 683–695, 2006.

[11] H. Lee, J. Yoo, and S. Choi, "Semi-supervised nonnegative matrix factorization," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 4–7, 2010.

[12] H. Kim, J. Choo, J. Kim, C. K. Reddy, and H. Park, "Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 567–576.

[13] V. K. Potluru and V. D. Calhoun, "Group learning using contrast nmf: Application to functional and structural mri of schizophrenia," in *IEEE International Symposium on Circuits and Systems*, 2008, pp. 1336–1339.

[14] J. Shao, Q. Yang, A. Wohlschlaeger, and C. Sorg, "Discovering aberrant patterns of human connectome in alzheimer's disease via subgraph mining," in *IEEE 12th International Conference on Data Mining Workshops*, 2012, pp. 86–93.

[15] L. Q. Uddin, K. S. Supekar, S. Ryali, and V. Menon, "Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development," *Journal of Neuroscience*, vol. 31, no. 50, pp. 18 578–18 589, 2011.

[16] J. Shao, N. Myers, Q. Yang, J. Feng, C. Plant, C. Böhm, H. Förstl, A. Kurz, C. Zimmer, C. Meng *et al.*, "Prediction of alzheimer's disease using individual structural connectivity networks," *Neurobiology of Aging*, vol. 33, no. 12, pp. 2756–2765, 2012.

[17] J. Kim, Y. He, and H. Park, "Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework," *Journal of Global Optimization*, vol. 58, no. 2, pp. 285–319, 2014.

[18] S. Kühn and J. Gallinat, "Resting-state brain activity in schizophrenia and major depression: a quantitative meta-analysis," *Schizophrenia Bulletin*, vol. 39, no. 2, pp. 358–365, 2013.

[19] G. Forman, "An extensive empirical study of feature selection metrics for text classification," *Journal of Machine Learning Research*, vol. 3, pp. 1289–1305, 2003.

[20] Y. W. Y. Jia and C. H. M. Turk, "Fisher non-negative matrix factorization for learning local features," in *Asian Conference on Computer Vision*, 2004, pp. 27–30.