Tensor-Based Multi-index Representation Learning for Major Depression Disorder Detection with Resting-State fMRI

Dongren Yao[1,2], Erkun Yang[1], Hao Guan[1], Jing Sui[3], Zhizhong Zhang[4(✉)], and Mingxia Liu[1(✉)]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
mxliu@med.unc.edu

[2] Brainnetome Center & National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[3] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100678, China

[4] School of Computer Science and Technology, East China Normal University, Shanghai 200241, China
zzzhang@cs.ecnu.edu.cn

Abstract. Major depressive disorder (MDD) is a common and costly mental illness whose pathophysiology is difficult to clarify. Resting-state functional MRI (rs-fMRI) provides a non-invasive solution for the study of functional brain network abnormalities in MDD patients. Existing studies have shown that multiple indexes derived from rs-fMRI, such as fractional amplitude of low-frequency fluctuations (fALFF), voxel-mirrored homotopic connectivity (VMHC), and degree centrality (DC) help depict functional mechanisms of brain disorders from different perspectives. However, previous methods generally treat these indexes independently, without considering their potentially complementary relationship. Moreover, it is usually very challenging to effectively fuse multi-index representations for disease analysis, due to the significant heterogeneity among indexes in the feature distribution. In this paper, we propose a tensor-based multi-index representation learning (TMRL) framework for fMRI-based MDD detection. In TMRL, we first generate multi-index representations (i.e., fALFF, VMHC and DC) for each subject, followed by patch selection via group comparison for each index. We further develop a tensor-based multi-task learning model (with a tensor-based regularizer) to align multi-index representations into a common latent space, followed by MDD prediction. Experimental results on 533 subjects with rs-fMRI data demonstrate that the TMRL outperforms several state-of-the-art methods in MDD identification.

Keywords: Major depressive disorder · rs-fMRI · Diagnosis

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-87240-3_17) contains supplementary material, which is available to authorized users.

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12905, pp. 174–184, 2021.
https://doi.org/10.1007/978-3-030-87240-3_17

1 Introduction

Major depression disorder (MDD) is one of the most prevalent disabling disorder, characterized by depressed mood, loss of interest or pleasure in nearly all activities. This mental illness has a high mortality rate due to the suicidal behavior of MDD patients, while the high cost of treatment troubles patients, their family members, and society [1,2]. Even though many efforts have been made in clinical neuroscience and psychiatric research, the unknown etiology and pathological mechanism still prevent us from fully understanding the disease.

Resting-state functional MRI (rs-fMRI) has become an essential non-invasive tool for assessing the brain substrates underlying mental disorders. [3–5]. Recent studies report that MDD is not only related to regional deficits, but also related to abnormal functional integration of distributed brain regions [6–8]. To measure spontaneous neural activities in rs-fMRI, multiple indexes have been designed, such as fractional amplitude of low-frequency fluctuation (fALFF) [9], voxel-mirrored homotopic connectivity (VMHC) [10], and degree centrality (DC) [11]. These indexes have been shown to help reveal the functional mechanisms of brain disorders [12–14]. However, existing methods generally treat multiple indexes independently, without considering their potentially complementary relationship. Also, it is usually very challenging to effectively fuse multi-index representations because of the significant between-index heterogeneity in feature distribution.

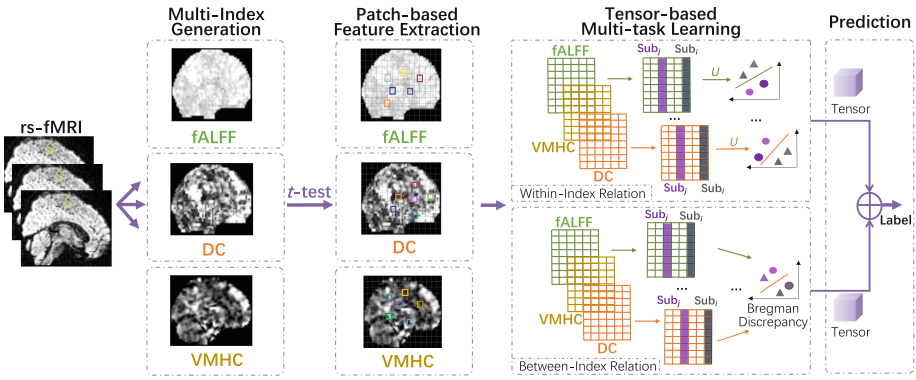

Fig. 1. Framework of the tensor-based multi-index representation learning (TMRL) for MDD identification with rs-fMRI, including (1) multi-index generation, (2) patch-based feature extraction, (3) tensor-based multi-task learning, and (4) prediction.

To address these issues, we propose a tensor-based multi-index representation learning (**TMRL**) framework for MDD detection with rs-fMRI. As illustrated in Fig. 1, we first generate three indexes (i.e., fALFF, VMHC, and DC) from preprocessed fMRI. For each index, we perform patch-wise group comparison (via

Table 1. Demographic information of studied subjects. M: Male; F: Female; Y: Yes; N: No; D: Lack of record; Mean ± Standard.

Category	Gender	Age	Education year	First period	On medication	Illness time
MDD	99M/183F	38.7 ± 13.6	10.8 ± 3.6	209(Y)/49(N) 24(D)	124(Y)/125(N) 33(D)	50.0 ± 65.9 35(D)
HC	87M/164F	39.6 ± 15.8	13.0 ± 3.9	–	–	–

t-test) between MDD and healthy controls (HCs) to select the most discriminative patches based on training images. We hypothesize that *patches at the same location for the same subject from different indexes contain similar neural activity*. Accordingly, we propose a tensor-regularized multi-task learning model (with a tensor-based regularizer) to align multi-index features into a common latent space. This helps mitigate distribution differences and capture potential relationships between indexes. We finally perform prediction using the learned new features. Experiments on 533 subjects with rs-fMRI suggest the efficacy of our method in MDD detection compared with previous state-of-the-arts.

2 Materials and Method

2.1 Subjects and Image Pre-processing

The public rs-fMRI dataset consists of 533 subjects[1], including 282 MDD and 251 age-matched healthy controls (HCs). The rs-fMRI scans were acquired by using a Siemens scanner with an echo-planar imaging sequence at the Southwest University (Table 1). The following lists the scanning parameters: repetition time (TR) = 2,000 ms, echo time (TE) = 30 ms, flip angle = 90°, thickness/gap = 3.0/1.0 mm, time points = 242 s, field of view (FOV) = 220 mm × 220 mm, voxel size = 3.44 × 3.44 × 4.00, and matrix size = 61 × 73 × 61.

Each fMRI scan was basically pre-processed by using the Data Processing Assistant for Resting-State fMRI (DPARSF). In this pipeline, we first discard the first 10 time points, followed by slice timing correction, head motion correction, regression of nuisance co-variants of head motion parameters, white matter, and cerebrospinal fluid. Images are then normalized with an EPI template in the MNI space, resampling to $3 \times 3 \times 3\,\text{mm}^3$ resolution, and spatial smoothing using a 6 mm full-width at half-maximum (FWHM) Gaussian kernel. This pipeline is slightly changed to generate multi-index representations (see below).

2.2 Methodology

Multi-index Generation. For each subject with rs-fMRI scan, we first extract multi-index representations, i.e., fALFF, VMHC, and DC (size: 61 × 73 × 61). (1) The fALFF index is employed to depict the relationship between field potential activity and cognitive-emotional processing [9]. Fast Fourier Transform is

[1] http://rfmri.org/REST-meta-MDD.

used to transform the time series of each voxel to the frequency domain. And the mean square root is obtained across $0.01 - 0.1Hz$ band for each voxel. The fALFF as a ratio is standardized by the mean fALFF of the global brain for all voxels. No temporal bandpass filtering $(0.01 - 0.1Hz)$ is required to extract the fALFF index. (2) The VMHC index is used to measure the functional connection between two voxels across the contralateral hemispheres [10]. In specific, it represents the Pearson's correlation coefficient between the time series of each voxel and that of its counterpart voxel at the same location in the opposite hemisphere. (3) The DC index is a graph-based measurement of brain networks, which captures the number of instantaneous functional connections between a voxel and the rest voxels in the entire brain network. We calculate the weighted sum of positive correlations by keeping inter-voxel connectivities with Pearson's correlation coefficient greater than a threshold $r = 0.25$, leading to the DC matrix.

Patch-Based Feature Extraction. Three activation maps (i.e., indexes) are calculated at the voxel level, but some voxels may not provide informative information for MDD diagnosis. To this end, we employ a patch-based feature extraction strategy by selecting the most discriminative patches in each index. We first partition each index into multiple non-overlapping $3\times3\times3$ patches. We then perform *patch-wise group comparison* between MDD and HC populations to select the most discriminative patches in one index based on training subjects. For each location in a specific index, we vectorize the corresponding patches of MDD and HC subjects, yielding two 27-dimensional feature vectors. Each element in the feature vector denotes the intensity or degree value of a specific voxel within a patch. We then leverage the t-test algorithm to compare these two feature vectors. If a patch contains more than 9 elements whose $p \leq 0.05$, we treat it as a discriminative patch. To make features between different indexes consistent and comparable, we perform a union operation on patches selected within three indexes and obtain a total of 308 patches. The feature vector of each of three indexes is $8,316$-dimensional to represent each subject.

Tensor-Regularized Multi-task Learning. To alleviate the inter-index heterogeneity and capture the underlying relationship among indexes, we develop a tensor-regularized multi-task learning model to fuse multi-index representations. Let $\mathbf{X}^{(v)} = [\boldsymbol{x}_1^{(v)}, \boldsymbol{x}_2^{(v)}, \cdots, \boldsymbol{x}_n^{(v)}] \in \mathbb{R}^{d\times n}$ denotes the v-th $(v = 1, \cdots, V)$ index feature matrix, where each column is a d-dimensional feature vector corresponding to the v-th index and n is the number of training samples. To explore the consistent cues from various functional characteristics, we further assume that there exist V projection matrices $\{\mathbf{U}^{(v)}\}_{v=1}^{V}$ to transform multi-index representations $\{\mathbf{X}^{(v)}\}_{v=1}^{V}$ into a latent space. Therefore, an ideal latent space is expected in which discrepancy and redundancy of multi-index representations is eliminated, and the useful complementary information is easily captured.

To this end, we formulate multi-index representation fusion as a tensor-regularized multi-task learning problem, with each task corresponding to a spe-

cific index. Denote C ($C = 2$ in this work) as the number of categories and $\mathbf{Y}^{(v)} \in \mathbb{R}^{C\times n}$ as the label matrix for training samples in the v-th index space. Let $\mathbf{U}^{(v)} \in \mathbb{R}^{d\times C}$ denotes the v-th projection matrix corresponding to the v-th index/task. Denote $\|\cdot\|_F$ as the Frobenius norm of a matrix. Our tensor-regularized multi-task learning model is formulated as:

$$\min_{\{\mathbf{U}^{(v)}\}_{v=1}^{V}} \sum_{v=1}^{V} ||\mathbf{U}^{(v)\top}\mathbf{X}^{(v)} - \mathbf{Y}^{(v)}||_F^2 + \alpha\|\boldsymbol{\mathcal{U}}\|_{\circledast} + \beta\sum_{1\le i<j\le V}\|\mathbf{U}^{(i)} - \mathbf{U}^{(j)}\|_F^2, \quad (1)$$

where the 1^{st} term is the empirical loss on training samples, the 2^{nd} term is a tensor-based regularizer, while the last term is the Bregman divergence [15].

In Eq. (1), $\boldsymbol{\mathcal{U}} = \Phi(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(V)}) \in \mathbb{R}^{d\times C\times V}$ is a tensor by merging different $\mathbf{U}^{(v)}$ to a 3-order tensor along the third dimension, where each frontal slice of $\boldsymbol{\mathcal{U}}$ is our task-specific projection matrix (i.e., $\boldsymbol{\mathcal{U}}(:,:,v) = \mathbf{U}^{(v)}$). The high-order tensor low-rank norm $\|\boldsymbol{\mathcal{U}}\|_{\circledast}$ measures the rank of a block circulant matrix constructed by all projection matrices, where $\|\cdot\|_{\circledast}$ is the tensor nuclear norm [16]. By comparing each column and each row of frontal slices (i.e., $\mathbf{U}^{(v)}$) via a low-rank constraint, this tensor-based regularizer helps explore shared and complementary information among multi-index representations. Also, all task-specific classifiers are correlated via this regularizer, so that knowledge can be flexibly transferred between tasks/indexes. In addition, the Bregman divergence is used here to encourage that the discrepancy between multiple transformations should be small, which helps reduce the heterogeneity among different indexes. That is, with the last two terms in Eq. (1), the proposed model helps mitigate distribution differences and capture the underlying relationship between indexes.

Due to the use of tensor nuclear norm and Bregman divergence regularizers, the objective function defined in Eq. (1) is difficult to optimize. We develop a unique optimization algorithm to solve this problem. To facilitate optimization, we first construct two block matrices as follows:

$$\tilde{\mathbf{U}} = [\tilde{\mathbf{U}}^{(1)}; \tilde{\mathbf{U}}^{(2)}; \cdots; \tilde{\mathbf{U}}^{(V)}],\ \tilde{\mathbf{Y}} = [\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \cdots, \mathbf{Y}^{(V)}]. \quad (2)$$

Denote $\tilde{\mathbf{X}}$ as a diagonal matrix, where its v-th diagonal element is $\mathbf{X}^{(v)}$ and others are 0, and $\mathbf{I} \in \mathbb{R}^{d\times d}$ as an identity matrix. We also define a matrix $\mathbf{M}$ as:

$$\mathbf{M} = \begin{bmatrix} (V-1)\mathbf{I} & -\mathbf{I} & \cdots & -\mathbf{I} \\ -\mathbf{I} & (V-1)\mathbf{I} & \cdots & -\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{I} & -\mathbf{I} & \cdots & (V-1)\mathbf{I} \end{bmatrix}. \quad (3)$$

Then, Eq. (1) can be rewritten as follows:

$$\min_{\tilde{\mathbf{U}}} ||\tilde{\mathbf{U}}^{\top}\tilde{\mathbf{X}} - \tilde{\mathbf{Y}}||_F^2 + \alpha\|\boldsymbol{\mathcal{U}}\|_{\circledast} + \beta tr(\tilde{\mathbf{U}}^{\top}\mathbf{M}\tilde{\mathbf{U}}), \quad (4)$$

which can be solved by using the Augmented Lagrange Multiplier (ALM) [17]. Specifically, by introducing an auxiliary tensor variable $\boldsymbol{\mathcal{G}}$, the problem in Eq. (3) can be solved by minimizing the following problem:

$$\min_{\tilde{\mathbf{U}},\boldsymbol{\mathcal{G}}} ||\tilde{\mathbf{U}}^{\top}\tilde{\mathbf{X}} - \tilde{\mathbf{Y}}||_F^2 + \alpha\|\boldsymbol{\mathcal{G}}\|_{\circledast} + \beta tr(\tilde{\mathbf{U}}^{\top}\mathbf{M}\tilde{\mathbf{U}}) + \langle\boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{U}} - \boldsymbol{\mathcal{G}}\rangle + \frac{\rho}{2}||\boldsymbol{\mathcal{U}} - \boldsymbol{\mathcal{G}}||_F^2, \quad (5)$$

where the tensor $\mathcal{W}$ is the Lagrange multiplier and ρ is the penalty parameter. We employ an alternating scheme to optimize Eq. (5), with details given below.

(1) $\tilde{\mathbf{U}}$-subproblem: When the tensor $\mathcal{G}$ is fixed, since $\mathbf{G}^{(v)} = \Phi_v^{-1}(\mathcal{G})$ and $\mathbf{W}^{(v)} = \Phi_v^{-1}(\mathcal{W})$, where Φ_v^{-1} is the inverse operation w.r.t Φ by clipping v-th frontal slice of the tensor, the optimization task can be rewritten as:

$$\min_{\tilde{\mathbf{U}}} ||\tilde{\mathbf{U}}^\top \tilde{\mathbf{X}} - \tilde{\mathbf{Y}}||_F^2 + \beta tr(\tilde{\mathbf{U}}^\top \mathbf{M} \tilde{\mathbf{U}}) + < \tilde{\mathbf{W}}, \tilde{\mathbf{U}} - \tilde{\mathbf{G}} > + \frac{\rho}{2} ||\tilde{\mathbf{U}} - \tilde{\mathbf{G}}||_F^2, \tag{6}$$

where $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{G}}$ are the block matrices constructed by $[\mathbf{W}^{(1)}; \mathbf{W}^{(2)}; \cdots; \mathbf{W}^{(V)}]$ and $[\mathbf{G}^{(1)}; \mathbf{G}^{(2)}; \cdots; \mathbf{G}^{(V)}]$, respectively.

(2) $\mathcal{G}$-subproblem: When $\tilde{\mathbf{U}}$ is fixed, the objective function in Eq. (5) can be solved through Theorem 1 in the *Supplementary Materials*.

The above two steps are alternately repeated until the convergence condition is satisfied. Based on [18], we prove the convergence of our optimization algorithm through Theorem 2 in the *Supplementary Materials*. The code will be freely released to the public via GitHub.

Prediction with Metric Learning. With the learned $\mathbf{U}$, we can calculate the distance between two subjects $\mathbf{x}_i$ and $\mathbf{x}_j$ with their new representations as:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)\mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)^\top} = \left\| \mathbf{U}^\top \mathbf{x}_i - \mathbf{U}^\top \mathbf{x}_j \right\|_2, \tag{7}$$

where $\mathbf{M} = \mathbf{U}^\top \mathbf{U}$. Given a new unseen test subject $\mathbf{z}$ with rs-fMRI, we first extract its multi-index representations and generate the new feature vector via $\hat{\mathbf{z}} = \mathbf{U}^\top \mathbf{z}$. A metric learning method [19] is used for prediction at the test stage. We randomly select m MDD and m HC subjects from training data and calculate the average distance from the test subject to the m sample within each group (i.e., MDD or HC). We set $m = 5$ to avoid the bias caused by random selection. The class label of the group with the smaller distance will be assigned to $\mathbf{z}$.

3 Experiment and Results

Experimental Settings. We evaluate the effectiveness of the proposed model using a 5-fold cross-validation (CV) strategy. To avoid bias introduced by a random partition, this 5-fold CV procedure is repeated 5 times. The performance of MDD identification from age-matched HCs is measured by four metrics, including accuracy (ACC), sensitivity (SEN), specificity (SPE), and F1-score (F1).

Competing Methods. We compare the proposed TMRL with two traditional methods, i.e., (1) **Baseline** that concatenates all features from different views into one vector, (2) **MKLpy** [20] with a multi-kernel learning technique, as well as three state-of-the-art multi-view learning methods, i.e., (3) **SNMF** [21] that uses a shallow non-negative matrix factorization method, (4) **DMF** [22] that employs a deep matrix factorization technique, and (5) **McDR** [23] that

uses a multi-view feature reduction method. The parameters α and β in our TMRL model are chosen from $\{0.01, 0.02, \cdots, 1\}$ via an inner cross-validation strategy based on only the training data. The parameters for five competing methods are set according to the original papers, and they use a linear support vector machine (SVM) as the classifier to detect MDD patients from HC subjects. In these methods, the parameter C for SVM is selected from the range of $\{0.01, 0.05, 0.1, 0.15, \cdots, 10\}$ via inner cross-validation on the training samples.

Table 2. Classification results of different methods in MDD detection.

Method	ACC	SEN	SPE	F1
SVM	0.561 ± 0.012	0.599 ± 0.018	0.518 ± 0.021	0.591 ± 0.014
MKLpy	0.585 ± 0.026	0.589 ± 0.023	0.579 ± 0.019	0.606 ± 0.029
NMF	0.579 ± 0.006	0.596 ± 0.014	0.559 ± 0.016	0.602 ± 0.011
DMF	0.588 ± 0.033	0.614 ± 0.027	0.562 ± 0.039	0.614 ± 0.020
McDR	0.594 ± 0.021	0.621 ± 0.015	0.577 ± 0.022	0.626 ± 0.019
TMRL (Ours)	**0.642 ± 0.027**	**0.643 ± 0.013**	**0.639 ± 0.016**	**0.654 ± 0.028**

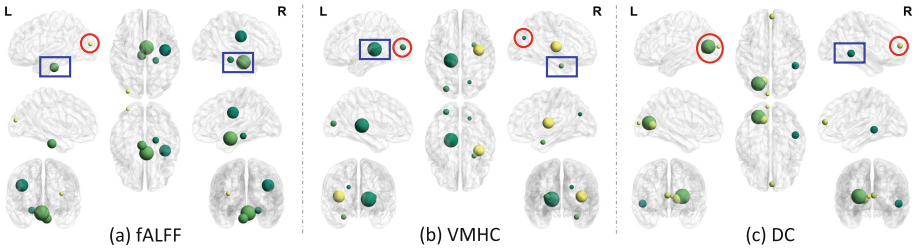

Fig. 2. Top 5 discriminative patches identified by our TMRL in three indexes. (Color figure online)

Results of MDD Detection. We first report the results of six different methods in MDD vs. HC classification in Table 2. It can be seen from this table that the proposed TMRL produces the best performance in MDD detection among the six methods. Compared with five competing methods, the TMRL achieves at least 4% improvement in terms of ACC and SPE values and 2% improvement in terms of SEN and F1 metrics. This implies that our method is able to learn effective representations from three indexes and boost the detection performance. Besides, compared with the simple feature concatenation method (i.e., SVM), the multi-kernel method (i.e., MKLpy) and the multi-view learning methods (i.e., NMF, DMF, and McDR) generally yield better results. This further validates the necessity of modeling the potential relationships of multiple indexes (as we do in this work) in order to improve the detection performance of MDD.

Visualization of Discriminating Regions. To display the most discriminative features identified by the proposed TMRL, we visually show the top 5 clusters of our selected patches within each of three indexes in Fig. 2. Since many patches (size: $3 \times 3 \times 3$ without overlap) are very close, we show their clusters here to facilitate visualization. Specifically, if the distance between two patches along a single axis is less than 9, we treat them as the same cluster.

As can be observed from Fig. 2, the temporal regions (marked by blue rectangles) are simultaneously identified by TMRL in the three indexes. Similarly, the occipital lobe (marked by red circles) also appears in three indexes and is regarded as the discriminative region by the proposed TMRL. This is consistent with previous studies that both temporal and occipital lobe regions are associated with MDD [24–26]. These visual results validate the reliability of TMRL in identifying MDD-affected brain regions with multi-index representations.

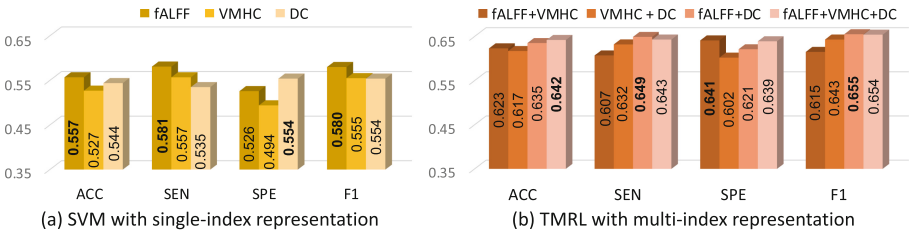

Fig. 3. Results of SVM with single-index representation and our TMRL that uses different combinations of three indexes in MDD vs. HC classification.

Table 3. Comparison with state-of-the-art methods for rs-fMRI based MDD detection.

Method	Model	Index	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Gu *et al.* [27]	Kendall+SVM	DC	0.533	0.578	0.482	0.567
Li *et al.* [28]	SVM	DC	0.544	0.535	0.554	0.554
Jie *et al.* [29]	SVM-FoBa	fALFF	0.561	0.622	0.491	0.605
Guo *et al.* [30]	sROC	VMHC	0.509	0.553	0.462	0.544
Ours	TMRL	DC+fALLF+VMHC	**0.642**	**0.643**	**0.639**	**0.654**

Multi-Index Vs. Single-Index Representation. We further compare the results achieved by SVM with single-index representation and our TMRL with different combinations of three indexes, with results shown in Fig. 3. Figure 3 (a) suggests that compared with the VMHC and DC indexes, the fALFF with SVM shows the best performance. It can be seen from Fig. 3 (b) that our TMRL using multi-index representation (regardless of two or three indexes) consistently outperforms SVM that uses only one index in terms of four metrics. This implies that multi-index representation may provide complementary information for MDD detection, and using only one index cannot produce good performance.

Comparison with State-of-the-Arts. Even though many studies propose to use different indexes to study the brain changes associated with MDD [27–32], only a few studies directly employ these indexes to distinguish MDD from HCs. In Table 3, we briefly summarize several state-of-the-art (SOTA) studies for MDD vs. HC classification based on three indexes derived from rs-fMRI data. For a fair comparison, we reproduced their algorithms, performed experiments using the same data set as this work, and reported the classification results.

As can be seen from Table 3, the proposed TMRL yields the best performance in terms of four evaluation metrics, compared with four SOTA methods. This may be due to the use of multi-index representation in TMRL, while four SOTA methods use only a single index. On the other hand, the SVM-FoBa model proposed in [29] with the fALFF index produces the overall better performance (e.g., with higher ACC, SEN and F1 values), compared with the other three SOTA methods that use the DC or VMHC index. This is consistent with the results reported in Fig. 3, that is, the fALFF may be more powerful in capturing brain changes associated with MDD when compared with VMHC and DC.

4 Conclusion

In this paper, we proposed a tensor-based multi-index representation learning (TMRL) framework for automated MDD detection based on multiple indexes derived from rs-fMRI data. Experimental results on 533 subjects with rs-fMRI demonstrate that our method is superior to several state-of-the-art methods in MDD detection. In the current work, we only employ a multi-index representation derived from rs-fMRI for MDD detection. Considering the potential complementary property of different imaging techniques, we think it is interesting to explore multi-modality data such as DTI and structural MRI to further study the neurobiological mechanisms of MDD, which will be our future work.

Acknowledgements. This work was finished when D. Yao was visiting the University of North Carolina at Chapel Hill. D. Yao and M. Liu was partly supported by NIH grant (No. AG041721). Z. Zhang was partly supported by the National Key Research and Development Program of China (No. 2016YFD0700100).

References

1. Otte, C., et al.: Major depressive disorder. Nat. Rev. Dis. Primers **2**(1), 1–20 (2016)
2. Gray, J.P., Müller, V.I., Eickhoff, S.B., Fox, P.T.: Multimodal abnormalities of brain structure and function in major depressive disorder: a meta-analysis of neuroimaging studies. Am. J. Psychiatry **177**(5), 422–434 (2020)
3. Yao, D., et al.: A mutual multi-scale triplet graph convolutional network for classification of brain disorders using functional or structural connectivity. IEEE Trans. Med. Imaging **40**(4), 1279–1289 (2021)
4. Wang, M., Lian, C., Yao, D., Zhang, D., Liu, M., Shen, D.: Spatial-temporal dependency modeling and network hub detection for functional MRI analysis via convolutional-recurrent network. IEEE Trans. Biomed. Eng. **67**(8), 2241–2252 (2019)

5. Guo, X., et al.: Shared and distinct resting functional connectivity in children and adults with attention-deficit/hyperactivity disorder. Transl. Psychiatry **10**(1), 1–12 (2020)
6. Zhi, D., et al.: Aberrant dynamic functional network connectivity and graph properties in major depressive disorder. Front. Psychiatry **9**, 339 (2018)
7. Wang, M., Zhang, D., Huang, J., Yap, P.T., Shen, D., Liu, M.: Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation. IEEE Trans. Med. Imaging **39**(3), 644–655 (2019)
8. Yao, D., Sui, J., Yang, E., Yap, P.-T., Shen, D., Liu, M.: Temporal-adaptive graph convolutional network for automated identification of major depressive disorder using resting-state fMRI. In: Liu, M., Yan, P., Lian, C., Cao, X. (eds.) MLMI 2020. LNCS, vol. 12436, pp. 1–10. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59861-7_1
9. Zou, Q.H., et al.: An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. J. Neurosci. Methods **172**(1), 137–141 (2008)
10. Zuo, X.N., et al.: Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy. J. Neurosci. **30**(45), 15034–15043 (2010)
11. Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. J. Math. Soc. **2**(1), 113–120 (1972)
12. Wei, J., et al.: Voxel-mirrored homotopic connectivity of resting-state functional magnetic resonance imaging in blepharospasm. Front. Psychol. **9**, 1620 (2018)
13. Sun, H., et al.: Regional homogeneity and functional connectivity patterns in major depressive disorder, cognitive vulnerability to depression and healthy subjects. J. Affect. Disord. **235**, 229–235 (2018)
14. Liu, W., et al.: Abnormal degree centrality of functional hubs associated with negative coping in older Chinese adults who lost their only child. Biol. Psychol. **112**, 46–55 (2015)
15. Banerjee, A., Merugu, S., Dhillon, I.S., Ghosh, J., Lafferty, J.: Clustering with Bregman divergences. J. Mach. Learn. Res. **6**(10), 1705–1749 (2005)
16. Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., Yan, S.: Tensor robust principal component analysis with a new tensor nuclear norm. IEEE Trans. Pattern Anal. Mach. Intell. **42**(4), 925–938 (2019)
17. Lin, Z., Chen, M., Ma, Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055 (2010)
18. Boyd, S., Parikh, N., Chu, E.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Now Publishers Inc, Boston (2011)
19. Yang, L., Jin, R.: Distance metric learning: a comprehensive survey. Mich. State Univ. **2**(2), 4 (2006)
20. Lauriola, I., Aiolli, F.: MKLpy: A python-based framework for multiple kernel learning. arXiv preprint arXiv:2007.09982 (2020)
21. Févotte, C., Idier, J.: Algorithms for nonnegative matrix factorization with the β-divergence. Neural Comput. **23**(9), 2421–2456 (2011)
22. Zhao, H., Ding, Z., Fu, Y.: Multi-view clustering via deep matrix factorization. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (2017)
23. Zhang, C., Fu, H., Hu, Q., Zhu, P., Cao, X.: Flexible multi-view dimensionality co-reduction. IEEE Trans. Image Process. **26**(2), 648–659 (2016)

24. Beneyto, M., Kristiansen, L.V., Oni-Orisan, A., McCullumsmith, R.E., Meador-Woodruff, J.H.: Abnormal glutamate receptor expression in the medial temporal lobe in schizophrenia and mood disorders. Neuropsychopharmacology **32**(9), 1888–1902 (2007)
25. Caetano, S.C., et al.: Medial temporal lobe abnormalities in pediatric unipolar depression. Neurosci. Lett. **427**(3), 142–147 (2007)
26. Liao, Y., et al.: Is depression a disconnection syndrome? Meta-analysis of diffusion tensor imaging studies in patients with MDD. J. Psychiatry Neurosci. JPN **38**(1), 49 (2013)
27. Gu, L., Huang, L., Yin, F., Cheng, Y.: Classification of depressive disorder based on rs-fMRI using multivariate pattern analysis with multiple features. In: 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), pp. 61–66. IEEE (2017)
28. Li, M., et al.: Clinical utility of a short resting-state MRI scan in differentiating bipolar from unipolar depression. Acta Psychiatrica Scandinavica **136**(3), 288–299 (2017)
29. Jie, N.F., et al.: Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. IEEE Trans. Auton. Ment. Dev. **7**(4), 320–331 (2015)
30. Guo, W., et al.: Decreased interhemispheric resting-state functional connectivity in first-episode, drug-naive major depressive disorder. Prog. Neuropsychopharmacol. Biol. Psychiatry **41**, 24–29 (2013)
31. Zhou, M., et al.: Intrinsic cerebral activity at resting state in adults with major depressive disorder: a meta-analysis. Prog. Neuropsychopharmacol. Biol. Psychiatry **75**, 157–164 (2017)
32. Fan, H., Yang, X., Zhang, J., Chen, Y., Li, T., Ma, X.: Analysis of voxel-mirrored homotopic connectivity in medication-free, current major depressive disorder. J. Affect. Disord. **240**, 171–176 (2018)