HHS Public Access

Author manuscript

IEEE Trans Cybern. Author manuscript; available in PMC 2019 August 01.

Published in final edited form as:

IEEE Trans Cybern. 2019 August ; 49(8): 3141–3154. doi:10.1109/TCYB.2018.2839693.

Sparse Multiview Task-Centralized Ensemble Learning for ASD Diagnosis Based on Age- and Sex-Related Functional Connectivity Patterns

Jun Wang [Member, IEEE],
Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, also with the School of Digital Media, Jiangnan University, Wuxi 214122, China, and also with the Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China (wangjun_sytu@hotmail.com)

Qian Wang [Member, IEEE],
Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China (wang.qian@sjtu.edu.cn).

Han Zhang [Senior Member, IEEE],
Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (hanzhang@med.unc.edu).

Jiawei Chen [Member, IEEE],
Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (jawaechan@gmail.com).

Shitong Wang, and
School of Digital Media, Jiangnan University, Wuxi 214122, China, and also with the Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China (wxwangst@aliyun.com)

Dinggang Shen [Fellow, IEEE]
Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, and also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (dgshen@med.unc.edu)

Abstract

Autism spectrum disorder (ASD) is an age- and sex-related neurodevelopmental disorder that alters the brain's functional connectivity (FC). The changes caused by ASD are associated with different age- and sex-related patterns in neuroimaging data. However, most contemporary computer-assisted ASD diagnosis methods ignore the aforementioned age-/sex-related patterns. In this paper, we propose a novel sparse multiview task-centralized (Sparse-MVTC) ensemble classification method for image-based ASD diagnosis. Specifically, with the age and sex information of each subject, we formulate the classification as a multitask learning problem, where

Correspondence to: Qian Wang; Dinggang Shen.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

each task corresponds to learning upon a specific age/sex group. We also extract multiview features per subject to better reveal the FC changes. Then, in Sparse-MVTC learning, we select a certain central task and treat the rest as auxiliary tasks. By considering both task-task and view-view relationships between the central task and each auxiliary task, we can learn better upon the entire dataset. Finally, by selecting the central task, in turn, we are able to derive multiple classifiers for each task/group. An ensemble strategy is further adopted, such that the final diagnosis can be integrated for each subject. Our comprehensive experiments on the ABIDE database demonstrate that our proposed Sparse-MVTC ensemble learning can significantly outperform the state-of-the-art classification methods for ASD diagnosis.

Keywords

ABIDE; autism spectrum disorder (ASD); diagnosis; high-order functional connectivity (FC); machine learning; multiview multitask (MVMT) learning; sparse multiview task-centralized (Sparse-MVTC) learning

I. INTRODUCTION

A. Background

AUTISM spectrum disorder (ASD) is a mental disorder characterized by the impairments including social and communication difficulties, restricted interest, and repetitive behavior. According to the report released by the centers for disease control and prevention, one out of 68 American children suffer from ASD, which makes the disease an important public health issue and also financial burdens for both the family and the society [1].

In recent years, resting-state functional magnetic resonance imaging (rs-fMRI) has become a pivotal tool in understanding the mechanism of ASD, as the imaging technique is capable of identifying biomarkers potentially for ASD diagnosis and follow-ups [2]–[8]. It reveals the interactions between activities of individual brain regions when a subject is placed at the resting state. The brain activities are observed through the changes in the cerebral blood supply, which are referred to as the blood-oxygen-level dependent (BOLD) signals. Since the BOLD signals are sensitive to spontaneous and intrinsic neural activities within the brain, they are used for noninvasive and effective observations toward ASD at the whole-brain level [9]. Specifically, functional connectivity (FC), which is defined as the correlation of temporal BOLD signals in different brain regions, exhibits the functional association of the structurally segregated brain regions [10]. In the literature, FC is often modeled as a network using the graph theory. A vertex in the network/graph is assigned to a specific brain region, and an edge between two vertices characterizes the FC between the corresponding brain regions. Since the pathology may disrupt the FC network, we are able to extract important biomarker information (i.e., features) from the FC network for the sake of understanding and diagnosing ASD.

Recent studies have discovered that ASD is an age- and sex-related neurodevelopmental disorder [11]–[14]. Alaerts *et al.* [15] investigated age-related FC changes in functional circuitry of posterior superior temporal sulcus (pSTS), and reported the complexity of

developmental patterns of the pSTS circuitry. Besides, Wiggins *et al.* [16], [17] observed that ASD patients differ from normal controls regarding their age-related changes in the FC of the default-mode networks. Moreover, ASD is far more prevalent in males compared to females, with the ratio from 3:1 to 4:1 [18]. Alaerts *et al.* [19] also found that males and females have different neural correlates with ASD, characterized by predominant hypo-connectivity patterns in males but hyper-connectivity patterns in females.

The aforementioned studies are based on the conventional FC analysis, in which pairwise Pearson's correlation coefficient is often calculated as the FC between brain regions. Although the Pearson's correlation is easy to compute, it utilizes the mean regional BOLD time series to derive only a single scalar for measuring the FC, which ignores the complex and high-order interactions of the brain regions. In order to better reveal the hidden interactions within the BOLD signals, Zhang *et al.* [20] proposed high-order FC (HOFC), in which the connectivity of the brain regions is defined by the *correlation of correlations.* Specifically, given a region under consideration, its *correlation feature vector* is first calculated by following the conventional FC analysis method to measure its correlations with all other regions in the brain. After that, a high-level correlation is calculated between the *correlation feature vectors* (i.e., FC profiles) of a pair of regions, thus obtaining their HOFC. Conclusively, FC can directly reflect the temporal correlation of the BOLD signals, whereas the HOFC does not. On the other hand, HOFC integrates knowledge from more brain regions and thus provides higher-level functional interaction knowledge compared to the FC. In the context of neuroscience, HOFC focuses on the correlation of the topographical connectivity properties and has several advantages, i.e., more sensitive to group difference, able to better capture individual variability, and able to show more prominent modular structures [20]. Therefore, FC and HOFC have different statistical properties and can be complementary to each other in neural disease diagnosis.

B. Motivation

In order to account for the relationship of ASD with age and sex, one can partition subjects into different age/sex groups. Specifically, each group acts as a *task*, while computer-assisted ASD diagnosis upon all groups can be solved by multitask learning (MTL). The rationale is that, by learning upon multiple closely related tasks jointly, we are able to utilize task-task relationship and attain more precise classification for all groups [21]–[26]. Meanwhile, since FC and HOFC are complementary to each other for ASD diagnosis, they are often perceived as two different views that should be utilized simultaneously in the multiview learning (MVL) [27]–[29]. In this way, the multiview multitask (MVMT) learning scheme is derived for more accurate ASD diagnosis, such that individual learning tasks with coupled views of features can be jointly considered.

In the literature, several solutions to MVMT learning have been presented. He and Lawrence [30] proposed a graph-based iterative algorithm *IteM*^{2}, which fully utilized information among the multiple tasks and multiple views for MVMT problems; Zhang and Huan [31] proposed *regMVMT* for MTL on the multiview data, in which both view-view co-regularization and task-task similarity are considered simultaneously; Jin *et al.* [32] proposed CSL-MTMV for MTL on multiview data by utilizing shared structure learning.

Although these methods integrate MTL and MVL together, they still have several shortcomings that make them inapplicable to our problem. First, the time complex of *IteM*^{2} is cubic to the number of tasks whereas *regMVMT* requires a large matrix to store the data, which makes both methods insufficient to be applied to large number of tasks. Second, both methods do not integrate feature selection in the learning criteria and, therefore, they cannot identify disease-related features in the neuroimaging applications. Third, *regMVMT* assumes that all tasks should be similar and CSL-MTMV ignores intertask relationship between task pairs, which is not appropriate in this paper.

The aim of this paper is to develop an efficient MVMT ensemble classification method for individualized ASD diagnosis. We group subjects into T individual groups/tasks according to their age/sex information, and utilize both views of FC and HOFC for the classification. In order to address the computation as well as joint feature selection issues in MVMT, we propose a novel sparse multiview task-centralized ensemble classification method (Sparse-MVTC-E) in this paper. Sparse-MVTC-E utilizes a novel learning mechanism, namely, task-centralized learning. Specially, at each turn, we first pick up one task as a *central* task and treat others as the *auxiliary* tasks. We then learn T classifiers, corresponding to the selected central task and the $T-1$ auxiliary tasks, to classify subjects in all groups. The sparse learning-based classification utilizes the task-task and view-view relationships between the central task and each auxiliary task for better diagnosis. As each task can be regarded as a central task in turn, we can perform the above procedure for Z ($Z \leq T$) times nonrepeatedly. In this way, we acquire Z classifiers for each group in the training stage. In the testing stage, we associate the subject with a specific group according to the age/sex information, and obtain the classification results from the Z classifiers corresponding to different groups/tasks. Then, a selective ensemble strategy is utilized to derive the final diagnosis of the testing subject. Note that, in our method, the complex relationships between all tasks are decoupled by using the task-centralized learning mechanism; thus, the computational efficiency of the proposed method can be greatly improved. From the perspective of MTL, the central task and auxiliary tasks in Sparse-MVTC share common features, which can be considered as an inductive bias. Besides, the constraint upon the central-auxiliary relation in (4) also produces a specific inductive bias. Both biases act as regularizers for the optimization procedure [33]. The proposed Sparse-MVTC is therefore superior to the traditional single-task learning methods according to [33].

The rest of this paper is organized as follows. In Section II, we propose the Sparse-MVTC-E learning method for ASD classification. In Section III, we present the experimental results, as well as the discriminant FC and HOFC features, for ASD diagnosis. In Section IV, we discuss some related works with our studies. In Section V, we draw the conclusion of this paper.

II. METHOD

The notations used in this paper are summarized as follows. We denote matrices with boldface uppercase letters, vectors with boldface lowercase letters, and scalars with normal italic letters, respectively. Specifically, we denote the identity matrix as **I**, and its ith column vector as $\mathbf{e}_i$ where the ith element is one and others are all zeros. Obviously, we can obtain

the ith row of a matrix $\mathbf{X}$ by the operation $\acute{e}_i \cdot \mathbf{X}$, and the jth column of $\mathbf{X}$ by the operation $\mathbf{X} \cdot \mathbf{e}_j$. We further denote the transpose operator, the trace operator, and the inverse of a matrix $\mathbf{X}$ as $\mathbf{X}'$, tr($\mathbf{X}$), and $\mathbf{X}^{-1}$, respectively. We also use vec($\mathbf{X}$) to vectorize the matrix $\mathbf{X}$ to a single column vector.

Fig. 1 provides an overview of our proposed Sparse-MVTC-E classification framework. There are four main steps in the pipeline: 1) extraction of the FC and HOFC features; 2) subject grouping or task partitioning according to the age and sex information; 3) Sparse-MVTC learning upon all tasks; and 4) an ensemble strategy to make the final diagnostic decision.

A. Sparse Multiview Task-Centralized Model

Assume that there are T supervised learning tasks for the data with V views. In the tth task ($t = 1, 2, \ldots, T$), we have N_t training subjects $\{\mathbf{x}_i, y_i\}$, $i = 1, \ldots N_t$. We define $\mathbf{x}_i = \left(\left(\mathbf{x}_i^1\right)', \ldots, \left(\mathbf{x}_i^V\right)'\right)' \in \mathbb{R}^D$ to be the vector to include all features of the V views for the ith subject. Each view has D_v features $\left(D = \sum_{v=1}^{V} D_v\right)$. $y_i \in \{-1, 1\}$ is the class label of $\mathbf{x}_i$ (e.g., "−1" for healthy controls, and "1" for patients). Let $X_t = \left[\mathbf{x}_1, \ldots, \mathbf{x}_{N_t}\right]' \in \mathbb{R}^{N_t \times D}$ and $y_t = \left(y_1, \ldots, y_{N_t}\right)'$ be the data matrix and the training label vector for the tth task, respectively. The vector $\mathbf{w}_t^v \in \mathbb{R}^{D_v}$ indicates the weights of all features in the vth view to linearly regress the labels in the tth task. We further define $\mathbf{w}_t = \left(\left(\mathbf{w}_t^1\right)', \ldots, \left(\mathbf{w}_t^V\right)'\right)' \in \mathbb{R}^D$ and $\mathbf{W} = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_T\right] \in \mathbb{R}^{D \times T}$.

In our proposed task-centralized strategy, we select a certain central task from all tasks, and treat the others as the auxiliary tasks. Specifically, suppose that the tth ($t = 1, 2, \ldots, T$) task is treated as the central task. The proposed task-centralized learning model can be formulated as the following optimization problem:

$$\min_{\mathbf{w}} \left(L_t\left(\mathbf{x}_t, \mathbf{w}_t\right) + \sum_{\substack{s=1 \\ s \neq t}}^{T} L_s\left(\mathbf{x}_s, \mathbf{w}_s\right) + R\left(\mathbf{w}_t; \left\{\mathbf{w}_s\right\}_{s=1, s \neq t}^{T}\right) \right) \quad (1)$$

where $L_t(\cdot)$ and $L_s(\cdot)$ are the loss functions for the central and the auxiliary tasks, respectively. $R\left(\mathbf{w}_t; \left\{\mathbf{w}_S\right\}_{S=1, s \neq t}^{T}\right)$ enforces the regularization regarding the relationships between the central task and all auxiliary tasks in the multiview setting.

The task-centralized learning in (1) treats the central task and the auxiliary tasks with their separate loss functions. Several methods can be used for computing $L_t(\cdot)$ and $L_s(\cdot)$, including least squared error loss, hinge loss or logistic loss. In this paper, we use least squared error loss to compute $L_t(\cdot)$ and $L_s(\cdot)$ particularly

$$L_t(\mathbf{X}_t, \mathbf{w}_t) = \frac{1}{N_t}\left\|\mathbf{y}_t - \sum_{v=1}^{V}\mathbf{X}_t^v\mathbf{w}_t^v\right\|_2^2 \quad (2)$$

$$L_s(\mathbf{X}_s, \mathbf{w}_s) = \frac{1}{N_S}\left\|\mathbf{y}_s - \sum_{v=1}^{V}\mathbf{X}_s^v\mathbf{w}_s^v\right\|_2^2, s = 1, 2, \ldots, T, s \neq t. \quad (3)$$

The denominators N_t and N_s are the numbers of the training subjects in the respective central and auxiliary tasks.

The well-defined regularization can produce a better posed problem for optimization. In this multiview task-centralized learning model, we devise three parts for the regularization $R\left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s\neq t}^{T}\right)$ in (1), i.e., $R = \gamma R_1 + \eta R_2 + \theta R_3$. These three terms ($R_1$, R_2, and R_3) will be detailed below.

1) Central-Auxiliary Joint Feature Selection: In the classical sparse MTL, a small number of features are selected and then shared by all tasks [34], [35]. In these methods, the joint feature selection is often implemented by the $l_{2,1}$ –norm regularization upon the weights of the features across all the tasks. In this paper of MVTC learning, however, we only consider the joint feature selection between the central and the auxiliary tasks. We propose the central-auxiliary joint feature selection strategy by introducing the following regularization into our Sparse-MVTC model:

$$R_1\left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s\neq t}^{T}\right) = \sum_{\substack{s=1 \\ s\neq t}}^{T} c_{s,t}\sum_{v=1}^{V}\|\mathbf{W}_{s,t}^v\|_{2,1} \quad (4)$$

where $\mathbf{W}_{s,t}^v = \left[\mathbf{w}_s^v, \mathbf{w}_t^v\right] \in \mathbb{R}^{D_v \times 2}$, and $c_{s,t}$ measures the relationship between the two tasks.

If the tth central task and the sth auxiliary task are more similar, the joint feature selection using $\sum_{v=1}^{V}\|\mathbf{W}_{s,t}^v\|_{2,1}$ across the two tasks should have a larger weight $c_{s,t}$. To this end, we compute the coefficient $c_{s,t} = \exp\left(-\|\overline{\mathbf{X}}_t - \overline{\mathbf{X}}_s\|^2/\sigma^2\right)$, where $\overline{\mathbf{X}}_t = \left(1/N_t\right)\sum_{i=1}^{N_t}\mathbf{x}_t$, $\overline{\mathbf{X}}_s = \left(1/N_S\right)\sum_{j=1}^{N_s}\mathbf{x}_j$, and $\sigma^2 = \left[1/\left(N_s \times N_t\right)\right]\sum_{i=1}^{N_s}\sum_{j=1}^{N_t}\|\mathbf{x}_i - \mathbf{x}_j\|^2$. Minimizing (4) ensures joint selection of a small number of common features across the two tasks, by enforcing every auxiliary task to share the commonly selected features with the central task according to their varied similarity measures with respect to the central task. In this sense, the central-auxiliary joint feature selection strategy is quite different from the conventional $l_{2,1}$-norm method.

2) Consistency of Central-Auxiliary Feature Weighting: The weights of the selected features are essentially important to derive the linear discriminant function that is capable of diagnosing a new subject. Although the improved joint feature selection regularization in $R_1\left(\mathbf{w}_t; \{\mathbf{w}_S\}_{S=1, s\neq t}^{T}\right)$ requires each auxiliary task sharing similar features with the central task with different weights, we go one step further to ensure that, if the auxiliary task is highly similar with the central task, the weights learned for the selected features in the auxiliary task should also be similar with those in the central task. Otherwise, they will be less similar. In the multiview setting, we can formulate the above constraint as follows:

$$R_2\left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s\neq t}^{T}\right) = \sum_{\substack{s=1 \\ s\neq t}}^{T} c_{s,t} \sum_{v=1}^{V} \|\mathbf{w}_s^v - \mathbf{w}_t^v\|_2^2 \quad (5)$$

where $c_{s,t} = \exp\left(-\|\overline{\mathbf{X}}_t - \overline{\mathbf{X}}_s\|^2/\sigma^2\right)$ is the coefficient of the term $\|\mathbf{w}_s^v - \mathbf{w}_t^v\|_2^2$, both $\overline{\mathbf{X}}_t = \left(1/N_t\right)\sum_{i=1}^{N_t}\mathbf{x}_i$ and $\overline{\mathbf{X}}_s = \left(1/N_S\right)\sum_{j=1}^{N_s}\mathbf{x}_j$ are the average data vector in the tth and sth tasks, respectively, and $\sigma^2 = \left[1/\left(N_s \times N_t\right)\right]\|\sum_{i=1}^{N_s}\sum_{j=1}^{N_t}\mathbf{x}_i - \mathbf{x}_j\|^2$. If an auxiliary task (the sth task) is similar with a central task (the tth task), $\overline{\mathbf{X}}_t$ and $\overline{\mathbf{X}}_S$ will be similar and $c_{s,t}$ thus takes larger values. Intuitively, minimizing (5) makes $\mathbf{w}_S^v$ and $\mathbf{w}_t^v$ be close to each other in such a case. On the other hand, if the auxiliary task is unrelated with the central task, $c_{s,t}$ will take very small values and then $\mathbf{w}_S^v$ and $\mathbf{w}_t^v$ tend to be different. Conclusively, the introduction of $c_{s,t}$ allows less contribution of $\sum_{v=1}^{V}\|\mathbf{w}_s^v - \mathbf{w}_t^v\|_2^2$ to the objective function if the sth task is less similar with the central one. The regularization also allows us to skip the assumption in many existing MTL methods such as *IteM*^{2} [30] and *regMVMT* [31] that all tasks should be similar. Equation (5) is also different from [36, eq. (2)], where the relatedness of all tasks are controlled by a fixed user defined value λ_2 and should thus be in the same level. The relatedness of multiple tasks is various in our method and they are determined adaptively by the data distribution of different tasks.

3) Consistency Across Views: Although the conventional FC is widely used for biomarker identification in many neuroscience applications, HOFC shows its merit by providing additional pathology information effectively [20]. Therefore, we consider FC and HOFC as two different views of the same brain and take the relation between them into account. Since the discriminant functions for different views of a subject tend to yield the same classification results, the difference between the outputs of the two views should be minimized (i.e., a small value for $\|\mathbf{X}_u^i\mathbf{w}_u^i - \mathbf{X}_u^j\mathbf{w}_u^j\|^2, u = 1, 2, \ldots, T$). Therefore, with u for the task index, we derive the respective regularization term

$$R_3\left(\mathbf{w}_t; \{\mathbf{w}_S\}_{S=1, s\neq t}^{T}\right) = \sum_{u=1}^{T} \frac{1}{N_u} \sum_{i,j=1}^{V} \|\mathbf{X}_u^i \mathbf{w}_u^i - \mathbf{X}_u^j \mathbf{w}_u^j\|^2. \quad (6)$$

In summary, we have the following overall regularization term:

$$R\left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s\neq t}^{T}\right) = \gamma R_1\left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s\neq t}^{T}\right) + \eta R_2\left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s\neq t}^{T}\right) + \theta R_3\left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s\neq t}^{T}\right) \quad (7)$$

where γ, η and θ are the non-negative scalar weights. In this way, we derive the final model for Sparse-MVTC as follows:

$$J\left(\mathbf{w}_t, \{\mathbf{w}_s\}\right) = \sum_{u=1}^{T} \frac{1}{N_u} \|\mathbf{y}_u - \sum_{v=1}^{V} \mathbf{X}_u^v \mathbf{w}_u^v\|_2^2 + \gamma \sum_{\substack{s=1 \\ s\neq t}}^{T} c_{s,t} \sum_{v=1}^{V} \|\mathbf{W}_{s,t}^v\|_{2,1} + \eta \sum_{\substack{s=1 \\ s\neq t}}^{T} c_{s,t} \sum_{v=1}^{V} \|\mathbf{w}_s^v - \mathbf{w}_t^v\|^2 + \theta \sum_{u=1}^{T} \frac{1}{N_k} \sum_{i,j=1}^{V} \|\mathbf{X}_u^i \mathbf{w}_u^i - \mathbf{X}_u^j \mathbf{w}_u^j\|^2. \quad (8)$$

Notice that some existing works such as [30], [31], [34], and [35] use the similar regularizations. However, the view and task regularizations in (4) and (5) are indeed different from those in the conventional MTL settings. Besides, all these previous works used just one or two of these regularizations and did not put all of them together.

B. Solution to Sparse-MVTC—The solution to $\min_{\mathbf{W}_{s,t}^v} \|\mathbf{W}_{s,t}^v\|_{2,1}$ [the second term in (8)] is equivalent to $\min_{\mathbf{w}_{s,t}^v} tr\left(\mathbf{W}_{s,t}^{v'} \mathbf{\Lambda}_{s,t}^v \mathbf{W}_{s,t}^v\right)$, [37] where $\mathbf{\Lambda}_{s,t}^v$ $(v = 1, \ldots, V)$ is a $D_v \times D_v$ diagonal matrix with the ith diagonal element $\left(\lambda_{s,t}^v\right)_i^i$ being computed as

$$\left(\lambda_{s,t}^v\right)_i^i = \frac{1}{2\|(\mathbf{e}_i)' \mathbf{W}_{s,t}^v\|_2}. \quad (9)$$

Accordingly, optimizing $\min_{\mathbf{W}_{s,t}^v} \|\mathbf{W}_{s,t}^v\|_{2,1}$ can be solved by optimizing $\mathbf{\Lambda}_{s,t}^v$ and $\mathbf{W}_{s,t}^v$ in $\text{tr}\left(\mathbf{W}_{s,t}^{v'} \mathbf{\Lambda}_{s,t}^v \mathbf{W}_{s,t}^v\right)$.

Since the central task and the auxiliary tasks should be associated with different linear discriminant functions for classification in the proposed model, we treat them separately for optimization. Given $\mathbf{\Lambda}_{s,t}^v$, we first fix $\mathbf{w}_t$ and update $\mathbf{w}_s$s for all auxiliary tasks. Next, we

compute $\mathbf{w}_t$ when $\{\mathbf{w}_s, s = 1, 2,\ldots,T, s \neq t$ are fixed. After all weights are updated, we compute $\mathbf{\Lambda}^v_{s,t}$ with (9). These processes iterate until the objective function converges. We present the implementation details of our algorithm in the following and provide its convergence proof in the Appendix.

The implementation of Sparse-MVTC is straightforward. For each auxiliary task, we calculate the matrices $\mathbf{P}^v_s$ and $\mathbf{Q}^v_S$ and the column vector r^v_s for each view. Then we obtain w_s as in (16) by computing the inverse of the matrix $\left(\widetilde{\mathbf{P}}_s + \widetilde{\mathbf{Q}}_s\right)$. After that, we calculate the matrices $\mathbf{P}^v_t$ and $\mathbf{Q}^v_t$ and the column vector $\mathbf{r}^v_t$ for each view in the central task, and then obtain w_t as in (23) by computing the inverse of the matrix $\left(\widetilde{\mathbf{P}}_t + \widetilde{\mathbf{Q}}_t\right)$. Notice that both $\left(\widetilde{\mathbf{P}}_s + \widetilde{\mathbf{Q}}_s\right)$ and $\left(\widetilde{\mathbf{P}}_s + \widetilde{\mathbf{Q}}_t\right)$ are positive definite and hence reversible. The parameters γ, η, and θ can be determined using nested cross-validation.

C. Sparse-MVTC-E: Ensemble Solution to ASD Diagnosis—As different groups are treated as the central task in turn, we perform the Sparse-MVTC algorithm for Z ($Z \leq T$) times. In this way, we are able to construct Z classifiers for each task. When there is a new test subject with its age and sex information available, we can easily identify the corresponding classifiers that are learned following the proposed Sparse-MVTC algorithm. The outputs of individual classifiers will be integrated by an ensemble scheme.

The selective ensemble algorithm GASEN [38] can be utilized for integrating multiple Sparse-MVTC classifiers. Specifically, we bootstrap-sample a validation dataset from the training set. Then, given two classifiers indexed by t_1 and t_2, we compute the correlation of the errors of their classification outputs by

$$C_{t_1 t_2} = \sum_i p(\mathbf{x}_i)\left(f_{t_1}(\mathbf{x}_i) - y_i\right)\left(f_{t_2}(\mathbf{x}_i) - y_i\right) \quad (24)$$

where $f_{t_1}(\mathbf{x}_i)$ and $f_{t_2}(\mathbf{x}_i)$ are the actual outputs of the two classifiers on the validation subject $\mathbf{x}_i$, respectively, y_i is the ground-truth label of $\mathbf{x}_i$, and $p(\mathbf{x}_i)$ is the possibility of $\mathbf{x}_i$'s occurrence in the validation dataset. Let $\alpha = (\alpha_1, \ldots, \alpha_T)$ record the weights for integrating individual classifiers. The optimal weights can be learned by optimizing

$$\alpha_{\text{opt}} = \underset{\boldsymbol{\alpha}}{argmin} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \alpha_{t_1} \alpha_{t_2} C_{t_1 t_2}. \quad (25)$$

Finally, we select K classifiers and obtain the final decision by weighted voting. Note that the original GASEN proposed in [38] could select the number of classifiers automatically. For easy implementation, we only record the top $K = 3$ classifiers with the largest weights in our Sparse-MVTC-E implementation. In this way, the ensemble classifier Sparse-MVTC-E

Algorithm 1

Sparse-MVTC

Input: $\mathbf{X}_t^v$, $v = 1,\ldots, V$, $t = 1,\ldots, \boldsymbol{T}$; $\mathbf{y}_t$, $t = 1,\ldots, \boldsymbol{T}$; γ, η, θ, ε

Output: $\mathbf{W}$

Set $l = 0$, and initialize $\mathbf{w}_t^v (v = 1, \ldots, V, t = 1, \ldots, T)$ to generate the initial $\mathbf{W}$ with random values;

repeat

Compute $\mathbf{\Lambda}_{s,t}^v, \left(\lambda_{s,t}^v\right)_i^i = \frac{1}{2\|\mathbf{e}_i \mathbf{W}_{s,t}^v\|_2}, v = 1, 2, \ldots, V, s = 1, 2, \ldots, T, s \neq t$

for $s = 1, 2, \ldots, T, s \neq t$

Compute $\mathbf{P}_s^v$, $\mathbf{Q}_s^v$ and $\mathbf{r}_s^v (v = 1, 2, \ldots, V)$ using Eqs. (10)–(12) for each auxiliary task, respectively;

$$\mathrm{P}_s^v = (\gamma c_{s,t} \mathbf{\Lambda}_{s,t}^v \ + \ \eta c_{s,t} \mathrm{I} + 2\theta \frac{1}{N_s} V(\mathrm{X}_s^v)' \ \mathrm{X}_s^v) \quad (10)$$

$$\mathbf{Q}_s^v = \left(\frac{1 - 2\theta}{N_s}\right)\left(\mathbf{X}_s^v\right)' \quad (11)$$

$$\mathbf{r}_s^v = \frac{1}{N_s}\left(\mathbf{X}_s^v\right)' \mathbf{y}_s + \eta c_{s,t} \mathbf{w}_t^v \quad (12)$$

Compute $\mathbf{w}_s$ with Eqs. (13)-(16):

$$\widetilde{\mathbf{P}}_s = \begin{pmatrix} \mathbf{P}_s^1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{P}_s^v \end{pmatrix} \quad (13)$$

$$\widetilde{\mathbf{Q}}_s = \begin{pmatrix} \mathbf{Q}_s^1 \mathbf{X}_s^1 & \cdots & \mathbf{Q}_s^1 \mathbf{X}_s^V \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_s^V \mathbf{X}_s^1 & \cdots & \mathbf{Q}_s^V \mathbf{X}_s^V \end{pmatrix} \quad (14)$$

$$\tilde{\mathbf{r}}_s = \left(\mathbf{r}_s^1 \ \cdots \ \mathbf{r}_s^V\right)' \quad (15)$$

$$\mathbf{w}_s = \left(\widetilde{\mathbf{P}}_s + \widetilde{\mathbf{Q}}_s\right)^{-1}\tilde{\mathbf{r}}_s \quad (16)$$

end for

Compute $\mathbf{P}_t^v$, $\mathbf{Q}_t^v$ and $\mathbf{r}_t^v$ using Eqs. (17)-(19) for the central task, respectively

$$\mathbf{P}_t^v = \left(\gamma \sum_{\substack{s=1 \\ s \neq t}}^{T} c_{s,t}\, \mathbf{\Lambda}_{s,t}^v + \eta \sum_{\substack{s=1 \\ s \neq t}}^{T} c_{s,t}\mathbf{I} + 2\theta \frac{1}{N_t} M\left(\mathbf{X}_t^v\right)' \mathbf{X}_t^v\right) \quad (17)$$

$$\mathbf{Q}_t^v = \frac{1-2\theta}{N_t}\left(\mathbf{X}_t^v\right)' \sum_{i=1}^{V} \mathbf{X}_t^i \mathbf{w}_t^i \quad (18)$$

$$\mathbf{r}_t^v = \frac{1}{N_t}\left(\mathbf{X}_t^v\right)' \mathbf{y}_t + \eta \sum_{\substack{s=1 \\ s \neq t}}^{T} c_{s,t}\mathbf{w}_s^v \quad (19)$$

Compute $\mathbf{w}_t$ with Eqs. (20)-(23);

$$\widetilde{\mathbf{P}}_t = \begin{pmatrix} \mathbf{P}_t^1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{P}_t^V \end{pmatrix} \quad (20)$$

$$\widetilde{\mathbf{Q}}_t = \begin{pmatrix} \mathbf{Q}_t^1 \mathbf{X}_t^1 & \cdots & \mathbf{Q}_s^1 \mathbf{X}_t^V \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_t^V \mathbf{X}_t^1 & \cdots & \mathbf{Q}_t^V \mathbf{X}_t^V \end{pmatrix} \quad (21)$$

$$\tilde{\mathbf{r}}_t = \tilde{\mathbf{r}}_t = \left(\mathbf{r}_t^1 \ \ldots \ \mathbf{r}_t^V\right)' \quad (22)$$

$$\mathbf{w}_t = \left(\widetilde{\mathbf{P}}_t + \widetilde{\mathbf{Q}}_t\right)^{-1}\tilde{\mathbf{r}}_t \quad (23)$$

Compute $\mathbf{W}(l) = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_T\right]$;

if $\|\mathbf{W}(l) - \mathbf{W}(l-1)\|_2 > \varepsilon, l = l + 1$, continue; else, abort;

end

can achieve better classification results than the average performance of Sparse-MVTC with different central tasks.

D. Time Complexity of Sparse-MVTC-E—The complexity of Sparse-MVTC-E is highly related to that of Sparse-MVTC. In Sparse-MVTC, both $\left(\widetilde{\mathbf{P}}_t + \widetilde{\mathbf{Q}}_t\right)$ and $\left(\widetilde{\mathbf{P}}_s + \widetilde{\mathbf{Q}}_s\right)$ are $(D \times D)$ matrices, where $D = \sum_{v=1}^{V} D_v$ is the total number of features. Although constructing these matrices is nontrivial, the real bottleneck of the Sparse-MVTC algorithm is the inverse of both matrices in (16) and (23), each of which has the time complexity of $O(D^3)$. There will be T inverse computation tasks for one run of Sparse-MVTC. If we select Z central tasks and the maximal iteration number is L, the time complexity of Sparse-MVTC-E will be $O(LZTD^3)$. Moreover, Sparse-MVTC-E can be easily parallelized by executing multiple Sparse-MVTC tasks simultaneously. In this way, the computation efficiency of Sparse-MVTC-E can be further improved. On the other hand, the conventional MVMT methods such as *regMVMT* + [31] involve an inverse operation of a $(TD \times TD)$ matrix, which has the time complexity of $O(L(TD)^3)$, with L as the maximal iteration number. Therefore, Sparse-MVTC-E requires less computation time and memory space, which makes our method efficient for the MVMT applications with many tasks.

III. EXPERIMENTS

A. Image Acquisition and Data Preparation

Our experiments are based on the autism brain imaging data exchange (ABIDE, http://fcon_1000.projects.nitrc.org/indi/abide/index.html database, which includes rs-fMRI data from 1112 ASD patients and healthy controls from multiple imaging centers. In this paper, we consider rs-fMRI scans acquired from two different imaging centers (i.e., NYU and UM-1) of ABIDE. Both male and female subjects are included in this paper and their ages span from 6 to 40 years old. Since the rs-fMRI images in different imaging centers were acquired using different scanners and parameters, they should be treated as separate tasks in classification. Therefore, we partition the subjects into several groups according to their age, sex and imaging center information. Since subjects do not distribute uniformly according to

their ages, we group the subjects with different age gaps to make sure that each group has enough subjects for learning. Finally, we have 18 groups (tasks). The demographic information and the scanning parameters of all groups are summarized in Tables I and II, respectively.

We extract features from rs-fMRI with the data processing assistant for resting-state fMRI. Specifically, the first ten acquired rs-fMRI volumes of each subject are discarded. Then, slice timing and head motion correction are performed. All rs-fMRI images are normalized to the Montreal Neurological Institute space with the resolution of $3 \times 3 \times 3$ mm^3. Nuisance variable regression is further conducted [39]. The resulted rs-fMRI images are parcellated into 116 regions-of-interest (ROIs) according to the automated anatomical labeling (AAL) template [40]. The band-pass filtering (0.005–0.1 Hz) is applied to the rs-fMRI time series of each ROI. Scrubbing is further performed based on the filtered time series, and the volumes with 0.5-mm or larger frame-wise displacement are removed. Also, two volumes before and one volume after the volumes with excessive motion are removed. Finally, the subjects with fewer than 3-min remaining data after scrubbing are excluded. To measure the FCs between ROIs, pairwise Pearson correlation coefficients are computed to yield the values between −1 and 1 for every individual ROI pair. The above processing results in a 116×116 FC matrix for each subject. After that, HOFC is computed according to [20] and [41].

B. Experimental Settings

In this paper, ASD diagnosis is performed by combining FC and HOFC, which are both extracted from rs-fMRI. For both FC and HOFC, the upper triangles of the connectivity matrices are utilized due to the symmetry of the matrices. These measures are reshaped into a vector with 6670 elements (corresponding to 6670 distinct region pairs). Prior to training, simple feature selection is conducted. Specifically, we select 300 FC features and 300 HOFC features using t-tests. All the features are further normalized regarding the z-score and used for subsequent computation.

We consider ASD diagnosis as the binary classification problem, by labeling ASD patients as +1 and the healthy controls as –1. We adopt a tenfold nested cross-validation strategy to evaluate the performance of our proposed method. Specifically, we divided each group in Table I into ten parts, one for testing and the others for training. Based on the training data of each fold, we select one central task in turn and use the rest as the auxiliary tasks. The above procedure iterates to pick up the central task nonrepeatedly, which results in T classifiers for ensemble learning upon a specific task. Moreover, the nested cross-validation consists of a fivefold inner loop, aiming to determine the optimal values for γ, η, and θ automatically. After obtaining T optimal classifiers for each task, we further utilize the selective ensembling strategy in Section II-C to find $K = 3$ best classifiers, and combine the outputs of these best classifiers for the testing subjects. In this way, the final result is produced for each testing subject. We repeat the above procedure for ten times to avoid the arbitrary bias of fold partition. We report the average of all ten repetitions as final outcome.

We use four metrics, including accuracy (ACC), sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC), to quantitatively evaluate the

performances of all competing methods. To confirm whether our method performs statistically better in the comparisons, we perform paired t-tests on the ACCs of our method and other methods and report the p-values.

All the experiments were conducted on a computer cluster, which is a Linux-based computing system available to researchers across University of North Carolina at Chapel Hill.

C. Effect of the Regularization Terms in Sparse-MVTC—A well-defined regularization helps pose the optimization problem in Sparse-MVTC. To validate the regularization terms in our method, we compare the performances where the three terms in (8) are temporarily disabled (e.g., by setting γ, η, and θ to zero, respectively). Since there are 18 groups/tasks as in Table I, we denote Group 1 as the central task without loss of generality. The classification accuracies of Sparse-MVTC, as well as the contributions of individual regularization terms, are plotted in Fig. 2. We can observe that the three regularization terms in Sparse-MVTC can effectively improve the classification performance. Similar results can be observed when other groups are selected as the central tasks.

D. Comparing Sparse-MVTC-E With Popular Machine Learning Methods—In this section, we report the classification results of Sparse-MVTC-E *with respect to* a large number of tasks. The subjects coming from NYU and UM-1 are combined together and grouped into 18 tasks, with the details given in Table I. We pick up each group as a central task in turn to train Sparse-MVTC. In this way, we have $Z = 18$ classifiers. We also select the top $K = 3$ classifiers for the ensembling of each task.

We compare the performance of Sparse-MVTC-E with several popular classifiers, which are summarized in Table III. We particularly define the experimental settings for CSVC [42], M2SVC [43] and random forest (RF) as follows.

1) The image data from each group are trained and tested separately. In this case, there are T classifiers, each of which corresponds to an independent group. We denote these methods as CSVC-S, M2SVC-S, and RF-S, respectively.

2) The data from T groups are combined for joint training and testing. That is, there is a single and common classifier for all T groups. We denote these methods as CSVC-J, M2SVC-J, and RF-J, respectively.

3) The classifier is trained with a single training group and then transferred to all groups for testing. In this way, T classifiers can be acquired, while each testing subject can be tested for T times. We obtain the final classification result by majority voting. And we denote these methods as CSVC-E, M2SVC-E and RF-E, respectively.

Besides, we also compare our method with *IteM*^{2}, *regMVMT*, and *regMVMT+*, which are representative MVMT methods in [30] and [31].

The Sparse-MVTC does not consider the relatedness among all the tasks simultaneously but find the relatedness to a central task. This is similar with a transfer learning strategy by improving the performance of a target task with some source tasks. To compare the performance of the task-centralized learning with the transfer learning, we further perform Sparse-MVTC in a transfer learning way, which is named as Sparse-MVTC-Tr. Specifically, we consider the central task as the learning task in the target domain and the auxiliary tasks as those in the source domains. We run Sparse-MVTC with different central tasks in turn and the final prediction results are obtained by combining the prediction results of different central tasks. We also show the performance of Sparse-MVTC, in which the ensemble learning is disabled and the final results were obtained by simply averaging all the outputs of the respective classifiers. For fair comparisons, the same tenfold nested cross-validation is applied to all competing methods.

Tables IV and V report the comparison results on this multicenter dataset, respectively, after 10 repetitions of the tenfold nested cross-validation.

From both Tables IV and V, we found that the proposed Sparse-MVTC-E method achieved the best classification accuracy on both imaging centers, although these subjects were collected using different scanners and parameters.

CSVC combines the FC and HOFC features into a long vector, thus ignoring the correlation between them. On the contrary, M2SVC treats the FC and HOFC features as two views and fuses the multiview data for classification. Both of them are linear kernel methods and fail to consider the MTL settings. On the other hand, MTMFJL is a typical MTL method that jointly learns common features and parameters in related tasks. Different from these aforementioned methods, our Sparse-MVTC-E learning method introduces both the task-task and view-view regularizations in the task-centralized learning mechanism, and obtains much better results than CSVC, M2SVC and MTMFJL. This fully demonstrates the superiority of the joint MVMT learning in Sparse-MVTC-E. Besides, Sparse-MVTC-E performs joint feature selection between the central task and each auxiliary task, thus providing useful information to help ASD diagnosis.

Note that CSVC-E, M2SVC-E and RF-E are the ensemble versions of CSVC, M2SVC, and RF, respectively. The experimental results suggest that they are not always better compared with the cases without ensembling. A possible reason is that the classifiers prior to selection and ensembling have weak generalization capability with insufficient training data. In the task-centralized strategy of our Sparse-MVTC-E learning method, the auxiliary tasks provide additional knowledge to the central task, which greatly improves the generalization capability of classifiers on all tasks and solves the computational bottleneck. Comparing the Sparse-MVTC-E with Sparse-MVTC (see Tables IV and V), we conclude that the utilization of the ensemble learning improves the final results significantly.

The *IteM*^{2}*, regMVMT,* and *regMVMT +* also consider both task-task and the view-view relationships. However, these methods do not have $l_{2,1}$-norm regularization; therefore, it could not select the discriminative FC/HOFC features jointly for the ASD diagnosis. We found that Sparse-MVTC-E significantly outperforms *IteM*^{2}*, regMVMT,* and *regMVMT +*

in terms of diagnosis accuracy. This is because our method uses (5) to model the task relations whereas both *IteM*^{2} and *regMVMT* assume that all the tasks should be similar. Although *regMVMT* + learns the task relations adaptively, it depends on suitable initialization of the algorithm without providing prior knowledge on task relations. Moreover, the interaction between the tasks in *regMVMT* + are very complex and it is not easy for the algorithm to find the optimal solutions given too many tasks. Different from *regMVMT*+, our method integrates prior knowledge on the task relations using $c_{s,t} = \exp\left(-\|\overline{\mathbf{X}}_t - \overline{\mathbf{X}}_s\|^2/\sigma^2\right)$ and simplifies the interaction between the tasks using task-centralized mechanism. These two factors help Sparse-MVTC-E to find better solutions than *regMVMT* +. On the other hand, Sparse-MVTC-E only involves the inverse of a 600 × 600 matrix, which is more efficient than *regMVMT* and *regMVMT* +, with the latter two involving the inverse of a 10 800 × 10 800 matrix. Therefore, regarding the memory requirement, *regMVMT* and *regMVMT* + could not work efficiently on a computer with limited physical memory.

The iteration number L, the selected central task number Z and the total task number T have linear relations with the time complexity of Sparse-MVTC-E, which are often considered as ideal properties in algorithm analysis. Since the time complexity of Sparse-MVTC-E is cubic to the feature number D, we vary the total feature number of the two views from 300 to 1200 and further compare the training time for *Sparse-MVTC-E, regMVMT,* and *regMVMT* + classifiers in Fig. 3. Given all 18 tasks to learn as in Table I, we plot the total training time of *Sparse-MVTC-E* when every task serves as the central task in turn. One may observe from the figure that Sparse-MVTC-E has the lowest training time when the total feature number becomes large, which fully demonstrates the advantages of the proposed method on the MVMT learning problems, especially with a large number of learning tasks and features.

To further demonstrate the superiority of the Sparse-MVTC-E over the Sparse-MVTC, we run Sparse-MVTC for multiple times with different central tasks selected; the results are plotted and compared in Fig. 4. Note that the classification results of Sparse-MVTC are slightly different as different central tasks were selected. The underlying reason is that different central tasks have different sizes and relationships with the auxiliary tasks, which makes the performance of Sparse-MVTC unstable. Using the selective ensemble strategy, the optimal classifiers can be selected by Sparse-MVTC-E. In this way, the outcomes of the Sparse-MVTC-E are generally more precise and stable.

E. Discriminative FC and HOFC—We report the most discriminative FC/HOFC features identified to separate the ASD patients from healthy controls. By checking the features selected by Sparse-MVTC in each cross-validation fold, we find that the selected FC/HOFC features are mostly stable yet with slight variation. Quantitatively, we summarize top 15/20 frequently selected FC/HOFC features over all cross-validation folds, and show them in Tables VI and VII, respectively. The number in the parenthesis indicates the index of the structural region in the AAL template.

From both Tables VI and VII, we can conclude the following major characteristics of the discriminative FCs and HOFCs, in terms of their hemispheric distributions and attributions. First, neither FCs nor HOFCs contributed to accurate ASD diagnosis are limited to the same hemisphere, but across both hemispheres. Second, regarding the hemispheric distribution of the brain regions involved in the HOFCs, there are significantly more regions in the right hemisphere than in the left, which is consistent with the results reported in [45]. Third, regarding the brain regions related with these FCs and HOFCs, the selected regions include multiple cortical regions and subcortical structures, which have been found to be related with ASD in [7]. Fourth, we note several symmetric FCs and HOFCs between the left and right hemispheres. The symmetric connections of bilateral middle temporal gyrus with the amygdala in the right hemisphere can be found in both tables, which implies their important roles in ASD diagnosis.

As a summary, our proposed Sparse-MVTC-E method can find the most discriminative FC by utilizing FC and HOFC jointly. These evidences can be inferred by the modified $l_{2,1}$-norm sparse regularization term for the task-centralized learning. The selected FC are effective for ASD diagnosis with a joint multiview task-centralized classification.

IV. DISCUSSION

A. Comparing Sparse-MVTC With Transfer Learning

The proposed Sparse-MVTC works under the framework of MTL, and shares the spirit of inductive transfer learning [46]. However, the inductive transfer learning methods often focus on the target task only by transferring the knowledge from other (source) tasks. We note that Sparse-MVTC tries to learn the target and the source tasks simultaneously by utilizing their relationship, which is different from transfer learning. Moreover, the task-centralized learning strategy in sparse-MVTC diversifies the tentative results for the subsequent ensemble learning in Sparse-MVTC-E, where the overall learning performance of individual sparse-MVTC methods can be improved.

One alternative ensemble solution other than the Sparse-MVTC-E is to perform Sparse-MVTC in a way of inductive transfer learning. That is, one can select every task as the central task in turn, and run Sparse-MVTC for T times of all possible central tasks. The final result can then be obtained by simply combining the predictions associated with the different central tasks. However, this strategy requires Sparse-MVTC to be performed for T times, which introduces a heavy computation burden, especially with a large number of learning tasks. In the proposed Sparse-MVTC-E however, it is not necessary to call Sparse-MVTC for T times by selecting each task as the central task in turn. Instead, one can select Z ($Z < T$, especially when T is large) representative tasks as the central tasks and combine the results of Z individual Sparse-MVTC calls by ensemble learning. In this way, the efficiency of the whole learning process, as well as the final performance of the method, can be improved.

B. Comparison With Existing Multitask Neuroimaging-Based Diagnosis Methods

There are many reported works on neural disorder diagnoses using MTL [34], [35], [47]–[49]. Wang *et al.* [35] proposed a joint classification and regression learning model to

identify the disease-sensitive-and-quantitative trait-relevant biomarkers from heterogeneous imaging genetic data. Similarly, Zhu *et al.* proposed a novel feature selection method by embedding the inherent relationships of the observations into a sparse MTL framework. Our proposed Sparse-MVTC-E method, while also falls in the category of MTL-based disease diagnosis, is different from the existing methods.

First: In the previous methods, each task is treated equally and the classification performance is effectively improved by utilizing the task-task relationships. We use a graph to describe the designs of the methods (see Fig. 5, where the vertices indicate tasks and the edges are for the task-task relationships considered in learning. In our task-centralized learning method, the graph is reduced to several trees, the root of which is the central task and the leaves are the auxiliary tasks. In this way, the computation of our method can be greatly simplified in the optimization process. By selecting the central tasks in turn, the selective ensemble procedure also helps to obtain better results.

Second: In the previous methods, the learning of each response variable is treated as one task and thus requires the same numbers of subjects in all tasks. In our proposed task-centralized method, the classification of each age/sex group is regarded as a certain task. The number of the subjects can *vary* across different tasks, which makes our method more flexible than other methods.

Third: The task-task and view-view relationships are not considered simultaneously in the previous works. In our proposed method, both the task-task and view-view relationships are integrated seamlessly by regularization terms, so that both tasks and views become complementary to contribute toward the feature selection and the classification simultaneously. The comparison results of Sparse-MVTC, M2SVC, and MTMFJL in experimental section fully verify this conclusion.

V. CONCLUSION

ASD is a complex neurodevelopmental disorder, where the autism-related alterations in brain's FC are different with respect to different age and gender. However, existing computer-assisted ASD diagnosis methods often focus on the relatively narrow age ranges (e.g., from late childhood to adolescence), ignoring the age-related changes of brain functions that occur over the course of development. They also fail to consider the sex differences of both ASD patients and healthy controls. Besides, these methods only utilize the conventional FC, without involving the high-level knowledge in the complex brain functional associations. To this end, we formulate the problem of ASD diagnosis as an MVMT learning problem, and propose a novel Sparse-MVTC-E learning method to solve the problem. using ABIDE datasets, we achieve a significant performance improvement for ASD classification compared with the state-of-the-art classification methods for ASD diagnosis.

The contribution of this paper can be summarized as follows.

1) Different from the conventional ASD studies that consider age-related changes or sex differences separately, this paper considers both factors for ASD

classification in a unified framework. To the best of our knowledge, this is the *first work* to consider the two factors in ASD diagnosis.

2) HOFC defined by correlations of correlations is utilized for ASD diagnosis, and several discriminant HOFC are discovered accordingly using our Sparse-MVTC-E method.

3) Different from the conventional MVMT solutions that treat each task in the same way, this paper proposes a novel multiview task-centralized learning framework by treating the tasks as a central task and auxiliary tasks and handling them differently in the MTL model.

Compared with most recent MVMT methods, the overall complexity of our method is greatly reduced and the computation efficiency is thus improved, especially in the situation, where a large number of tasks and features are involved. Additionally, the utilization of ensemble learning further improves the performance of our method, which is based on the prediction results of multiple Sparse-MVTC classifiers.

Although our proposed method is successful on multitask ASD diagnosis using FC and HOFC, it considers only the inherent linear relationships inside the subjects. In the future work, we will develop the respective nonlinear versions for neuroimaging classification.

Acknowledgments

This work was supported in part by the National Institutes of Health under Grant 1U01MH110274, Grant EB006733, Grant MH100217, and Grant AG041721, in part by the National Natural Science Foundation of China under Grant 61300151, Grant 81471733, Grant 61473190, and Grant 61702225, in part by the National Key Research and Development Program under Grant 2017YFC0107602, in part by the Science and Technology Commission of Shanghai Municipality under Grant 16410722400, and Grant 16511101100, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20151299, Grant BK20151358, Grant BK20160187, and Grant BK20161268. This paper was recommended by Associate Editor D. Tao.

APPENDIX

In the Appendix, we analyze the convergence property of Sparse-MVTC. Since the solution to $\min_{\mathbf{W}^v_{s,t}} \|\mathbf{W}^v_{s,t}\|_{2,1}$ [for the second term in (8)] is equivalent to $\min_{\mathbf{W}^v_{s,t}} tr\left(\mathbf{W}^{v'}_{s,t}\,\mathbf{\Lambda}^v_{s,t}\,\mathbf{W}^v_{s,t}\right)$ [37], minimizing the objective function in (8) can be attained by optimizing $\mathbf{\Lambda}^v_{s,t}$, $\mathbf{w}_t$ and $\mathbf{w}_s$ in the following:

$$J\left(\mathbf{w}_t, \{\mathbf{w}_s\}, \left\{\mathbf{\Lambda}^v_{s,t}\right\}\right) = \sum_{u=1}^{T} \frac{1}{N_u} \|\mathbf{y}_u - \sum_{v=1}^{V} \mathbf{X}^v_u \mathbf{w}^v_u\|_2^2 + \gamma \sum_{\substack{s=1 \\ s\neq t}}^{T} c_{s,t} \sum_{v=1}^{V} tr\left(\mathbf{W}^{v'}_{s,t}\,\mathbf{\Lambda}^v_{s,t}\,\mathbf{W}^v_{s,t}\right) \quad (26)$$

$$+ \eta \sum_{\substack{s=1, \\ s\neq t}}^{T} c_{s,t} \sum_{v=1}^{V} \|\mathbf{w}^v_s - \mathbf{w}^v_t\|^2 + \theta \sum_{s=1}^{T} \frac{1}{N_s} \sum_{i,j=1}^{V} \|\mathbf{x}^i_s \mathbf{w}^i_s - \mathbf{X}^j_s \mathbf{w}^j_s\|^2.$$

In the next, we prove that Sparse-MVTC makes the value of the objective function in (26) decrease monotonically.

Theorem 1:

Given $\mathbf{\Lambda}_{s,t}^{v}, v = 1, \ldots, V,$ and $\mathbf{w}_s, s = 1, \ldots, Ts \neq t,$ for the case that the auxiliary tasks are fixed, (26) is minimized if and only if $\mathbf{w}_t$ for the central task is computed from (20) to (23).

Proof:

By taking the derivatives of (26) with respect $\mathbf{w}_t^v (v = 1, 2, \ldots, V)$ and set them to zeros, we have

$$\mathbf{P}_t^v \mathbf{w}_t^v + \mathbf{Q}_t^v \sum_{i=1}^{V} \mathbf{X}_t^i \mathbf{w}_t^i = \mathbf{r}_t^v \quad (27)$$

where $\mathbf{P}_t^v$, Q_t^v, and $\mathbf{r}_t^v$ are computed with (17)−(19), respectively. For each view in the tth task, we can have an equation in the form of (27), and learn all $\mathbf{w}_t^v$ jointly as

$$\left(\widetilde{\mathbf{P}}_t + \widetilde{\mathbf{Q}}_t\right)\mathbf{w}_t = \tilde{\mathbf{r}}_t \quad (28)$$

where $\widetilde{\mathbf{P}}_t$, $\widetilde{\mathbf{Q}}_t$, and $\tilde{\mathbf{r}}_t$ are the block matrices in (20)−(22). The analytic solution of $\mathbf{w}_t^v (t = 1, \ldots, T, v = 1, \ldots, V)$ can be easily obtained from (23). This proves that (23) is the necessary condition of minimizing (26). On the other hand, the Hessian matrix of the objective function in (26) is in the form of

$$\mathbf{H}_t = \begin{pmatrix} \mathbf{H}_t^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{H}_t^V \end{pmatrix}$$

where $\mathbf{H}_t^v = \left(1/N_t\right)\left(\mathbf{X}_t^v\right)'\mathbf{X}_t^v + \gamma \sum_{\substack{s=1 \\ s \neq t}}^{T} c_{s,t} \mathbf{\Lambda}_{s,t}^v + \eta \sum_{\substack{s=1 \\ s \neq t}}^{T} c_{s,t} + 2\theta(V-1)\left(1/N_t\right)\left(\mathbf{X}_t^v\right)'\mathbf{X}_t^v$ is positive definite for the central task when $\gamma > 0$, $\eta > 0$, $\theta > 0$ and $V \geq 1$. Thus, the Theorem 1 is proved.

Theorem 2 (Auxiliary Tasks):

Given $\mathbf{\Lambda}_{s,t}^{v}, v = 1, \ldots, V,$ and $\mathbf{w}_t, t = 1, \ldots, T,$ for the case that the central task is fixed, (26) is minimized if and only if $\mathbf{w}_s, s = 1, \ldots, T, s \neq t,$ for the auxiliary tasks are computed from (13) to (16).

Proof:

Similarly, we may easily find that (16) is the necessary condition of minimizing (26). On the other hand, the Hessian matrix of the objective function in (26) is in the form of

$$\mathbf{H}_s = \begin{pmatrix} \mathbf{H}_s^1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{H}_s^V \end{pmatrix}$$

where

$$\mathbf{H}_s^v = \frac{1}{N_s}\left(\mathbf{X}_s^v\right)'\mathbf{X}_s^v + \gamma c_{s,t}\mathbf{\Lambda}_{s,t}^v + \eta c_{s,t}\mathbf{I} + 2\theta\frac{1}{N_s}(V-1)\left(\mathbf{X}_s^v\right)'\mathbf{X}_s^v$$

is positive definite for each auxiliary task when $\gamma > 0$, $\eta > 0$, $\theta > 0$ and $V \geq 1$. Thus, the Theorem 2 is proved. ■

Lemma 1 [37]:

For any nonzero vectors $\mathbf{u}, \mathbf{u}_t \in \mathbb{R}^d$, the following inequality holds:

$$\|\mathbf{u}\|_2 - \frac{\|\mathbf{u}\|_2^2}{2\|\mathbf{u}_t\|_2} \leq \|\mathbf{u}_t\|_2 - \frac{\|\mathbf{u}_t\|_2^2}{2\|\mathbf{u}_t\|_2}. \quad (29)$$

Theorem 3:

In each iteration, Algorithm 1 monotonically decreases the objective function value in (26).

Proof:

In the 1th iteration, $l = 1, 2, \ldots$, (26) can be rewritten as follows:

$$J\left(\mathbf{w}_t(l), \{\mathbf{w}_S(l)\}, \{\mathbf{\Lambda}_{s,t}^v(l)\}\right) = \sum_{s=1}^{T}\frac{1}{N_S}\|\mathbf{y}_s - \sum_{v=1}^{V}\mathbf{X}_s^v\mathbf{w}_s^v(l)\|_2^2 + \gamma\sum_{\substack{s=1 \\ s\neq t}}^{T} c_{s,t}\sum_{v=1}^{V} tr\left(\mathbf{W}_{s,t}^v(l)'\,\mathbf{\Lambda}_{s,t}^v(l)\mathbf{W}_{s,t}^v(l)\right)$$
$$+\eta\sum_{\substack{s=1 \\ s\neq t}}^{T} c_{s,t}\sum_{v=1}^{V}\|\mathbf{w}_s^v(l) - \mathbf{w}_t^v(l)\|^2 + \theta\sum_{s=1}^{T}\frac{1}{N_s}\sum_{i,j=1}^{V}\|\mathbf{X}_s^i\mathbf{w}_s^i(l) - \mathbf{X}_s^j\mathbf{w}_s^j(l)\|^2 = J_1\left(\mathbf{w}_t(l), \{\mathbf{w}_s(l)\}, \{\mathbf{\Lambda}_{s,t}^v(l)\}\right)$$
$$+\gamma\sum_{\substack{s=1 \\ s\neq t}}^{T} c_{s,t}\sum_{v=1}^{V} tr\left(\mathbf{W}_{s,t}^v(l)'\,\mathbf{\Lambda}_{s,t}^v(l)\mathbf{W}_{s,t}^v(l)\right)$$

where

$$J_1\left(\mathbf{w}_t(l), \{\mathbf{w}_s(l)\}\right) = \sum_{s=1}^{T} \frac{1}{N_s} \|\mathbf{y}_s - \sum_{v=1}^{V} \mathbf{X}_s^v \mathbf{w}_s^v(l)\|_2^2 + \eta \sum_{\substack{s=1 \\ s \neq t}}^{T} c_{s,t} \sum_{v=1}^{V} \|\mathbf{w}_s^v(l) - \mathbf{w}_t^v(l)\|^2 \quad (30)$$
$$+ \theta \sum_{s=1}^{T} \frac{1}{N_s} \sum_{i,j=1}^{V} \|\mathbf{X}_s^i \mathbf{w}_s^i(l) - \mathbf{X}_s^j \mathbf{w}_s^j(l)\|^2 .$$

From Theorem 1, we can infer that $\mathbf{w}_t = \left(\widetilde{\mathbf{P}}_t + \widetilde{\mathbf{Q}}_t\right)^{-1} \tilde{\mathbf{r}}_t$ is the local minimum of $J(\mathbf{w}_t(l), \{\mathbf{w}_s(l)\}, \{\Lambda_{s,t}^v(l)\})$ when $\Lambda_{s,t}^v, v = 1, \cdots, V$ and $\mathbf{w}_s, s = 1, \ldots, T, s \neq t$ are fixed. Thus, we have

$$J\left(\mathbf{w}_t(l+1), \{\mathbf{w}_s(l)\}, \{\Lambda_{s,t}^v(l)\}\right) \leq J\left(\mathbf{w}_t(l), \{\mathbf{w}_s(l)\}, \{\Lambda_{s,t}^v(l)\}\right) \quad (31)$$

From Theorem 2, we can infer that $\mathbf{w}_s = \left(\widetilde{\mathbf{P}}_s + \widetilde{\mathbf{Q}}_s\right)^{-1} \tilde{\mathbf{r}}_s$ is the local minimum of $J(\mathbf{w}_t(l), \{\mathbf{w}_s(l)\}, J(\mathbf{w}_t(l), \{\mathbf{w}_s(l)\}, \{\Lambda_{s,t}^v(l)\})$ when $\Lambda_{s,t}^v, v = 1, \ldots, V$, and $\mathbf{w}_t$ are fixed. Thus, we have

$$J\left(\mathbf{w}_t(l+1), \{\mathbf{w}_s(l+1)\}, \{\Lambda_{s,t}^v(l)\}\right) \leq J\left(\mathbf{w}_t(l+1), \{\mathbf{w}_s(l)\}, \{\Lambda_{s,t}^v(l)\}\right) \quad (32)$$

We can easily have

$$J\left(\mathbf{w}_t(l+1), \{\mathbf{w}_s(l+1)\}, \{\Lambda_{s,t}^v(l)\}\right) \leq J\left(\mathbf{w}_t(l), \{\mathbf{w}_s(l)\}, \{\Lambda_{s,t}^v(l)\}\right) \quad (33)$$

$$J_1\left(\mathbf{w}_t(l+1),\left\{\mathbf{w}_s(l+1)\right\}\right)+\gamma\sum_{\substack{s=1\\ s\neq t}}^{T}c_{s,t}\sum_{v=1}^{V}\sum_{d=1}^{D_v}\left(\frac{\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l+1)\right\|_2^2}{2\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2}-\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l+1)\right\|_2^2+\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l+1)\right\|_2^2\right)\leq J_1\left(\mathbf{w}_t(l),\left\{\mathbf{w}_s(l)\right\}\right)$$

$$+\gamma\sum_{\substack{s=1\\ s\neq t}}^{T}c_{s,t}\sum_{v=1}^{V}\sum_{i=1}^{D_v}\left(\frac{\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2^2}{2\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2}-\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2^2+\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2^2\right)J_1$$

$$\left(\mathbf{w}_t(l+1),\left\{\mathbf{w}_s(l+1)\right\}\right)+\gamma\sum_{\substack{s=1\\ s\neq t}}^{T}c_{s,t}\sum_{v=1}^{V}tr\left(\mathbf{W}_{s,t}^{v}(l+1)'\,\mathbf{\Lambda}_{s,t}^{v}(l)\mathbf{W}_{s,t}^{v}(l+1)\right)$$

$$\leq J_1\left(\mathbf{w}_t(l),\left\{\mathbf{w}_s(l)\right\}\right)+\gamma\sum_{\substack{s=1\\ s\neq t}}^{T}c_{s,t}\sum_{v=1}^{V}tr\left(\mathbf{W}_{s,t}^{v}(l)'\,\mathbf{\Lambda}_{s,t}^{v}(l)\mathbf{W}_{s,t}^{v}(l)\right). \tag{34}$$

By substituting (9) into the above and also changing the trace form into the form of summation, we have

$$J_1\left(\mathbf{w}_t(l+1),\left\{\mathbf{w}_s(l+1)\right\}\right)+\gamma\sum_{\substack{s=1\\ s\neq t}}^{T}c_{s,t}\sum_{v=1}^{V}\sum_{d=1}^{D_v}\frac{\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l+1)\right\|_2^2}{2\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2}\leq J_1\left(\mathbf{w}_t(l),\left\{\mathbf{w}_s(l)\right\}\right)$$

$$+\gamma\sum_{\substack{s=1\\ s\neq t}}^{T}c_{s,t}\sum_{v=1}^{V}\sum_{i=1}^{D_v}\frac{\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2^2}{2\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2}.$$

With a simple modification, we have (34) as shown at the bottom of the previous page. Since $\sum_{d=1}^{D_v}\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2^2=\left\|\mathbf{W}_{s,t}^{v}(l)\right\|_{2,1}$, with (29) and (34), we can easily arrive at

$$J_1\left(\mathbf{w}_t(l+1),\left\{\mathbf{w}_s(l+1)\right\}\right)+\gamma\sum_{\substack{s=1\\ s\neq t}}^{T}c_{s,t}\sum_{v=1}^{V}\sum_{d=1}^{D_v}\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l+1)\right\|_2^2\leq J_1\left(\mathbf{w}_t(l),\left\{\mathbf{w}_s(l)\right\}\right) \tag{35}$$

$$+\gamma\sum_{\substack{s=1\\ s\neq t}}^{T}c_{s,t}\sum_{v=1}^{V}\sum_{i=1}^{D_v}\left\|\left(\mathbf{e}_i\right)'\mathbf{W}_{s,t}^{v}(l)\right\|_2^2x$$

i.e.,

$$J\left(\mathbf{w}_t(l+1), \{\mathbf{w}_s(l+1)\}, \{\Lambda^v_{s,t}(l+1)\}\right) \leq J\left(\mathbf{w}_t(l), \{\mathbf{w}_s(l)\}, \{\Lambda^v_{s,t}(l)\}\right).$$

That is, the algorithm monotonically decreases the objective function value in (8), and the proposed iterative algorithm converges to its local optimum.

Biography

Jun Wang (M'14) received the Ph.D. degree in pattern recognition and intelligence systems from the School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, China, in 2011.

He has been a Research Assistant with the Department of Computing, Hong Kong Polytechnic University, Hong Kong, and a Post-Doctoral Research Fellow with the Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. He is currently an Associate Professor with the School of Digital Media, Jiangnan University, Wuxi, China. He has published over 50 articles in international/national journals. His current research interests include machine learning, fuzzy systems, and medical image classification.

Qian Wang (M'14) received the Ph.D. degree in computer science from the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, in 2013.

He is currently a Research Scientist of the Institute for Medical Imaging Technology and the Med-X Research Institute, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include medical image analysis, computer vision, machine learning, artificial intelligence, and translational medical studies.

Han Zhang (SM'18) received the Ph.D. degree in cognitive neuroscience from Beijing Normal University, Beijing, China, in 2011.

He is an Instructor with the Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, where he is currently researching with the Image Display, Enhancement, and Analysis Laboratory and the Biomedical Research Imaging Center. He was a Faculty Investigator with the Hangzhou Normal University, Hangzhou, China, and Center for Cognition and Brain Disorders, Hangzhou. His current research

interests include modeling dynamics and high-order functional connectivity of the brain functional networks and their applications to abnormal aging and infant brain development.

Dr. Zhang is currently a member of International Society for Magnetic Resonance in Medicine and Organization for Human Brain Mapping.

Jiawei Chen (M'16) received the B.S. degree in information computation science from the Xi'an University of Posts and Telecommunication, Xi'an, China, in 2010, and the Ph.D. degree from Xidian Univeristy, Xi'an, in 2016.

He is currently a Post-Doctoral Research Fellow with the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. His current research interests include sparse feature learning and medical image analysis.

Shitong Wang received the M.S. degree in computer science from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1987.

He visited London University, London, U.K., and Bristol University, Bristol, U.K., Hiroshima International University, Higashihiroshima, Japan, the Hong Kong University of Science and Technology, Hong Kong, Hong Kong Polytechnic University, Hong Kong, and Hong Kong City University, Hong Kong, as a Research Scientist, for over five years. He is currently a Full Professor with the School of Digital Media, Jiangnan University, Wuxi, China. He has published over 150 papers in international/national journals and has authored/co-authored seven books. His current research interests include AI, neuron-fuzzy systems, pattern recognition, and image processing.

Dinggang Shen (M'00-SM'07-F'18) received the Ph.D. degree in electronic communication from Shanghai Jiao Tong University, China, in 1995.

He is a Jeffrey Houpt Distinguished Investigator, and a Professor of Radiology, Biomedical Research Imaging Center (BRIC), Computer Science, and Biomedical Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, where he is currently directing the Center for Image Analysis and Informatics, the Image Display, Enhancement, and Analysis, Laboratory with the Department of Radiology, and also the medical image analysis core with the BRIC. He was a Tenure-Track Assistant Professor with the University of Pennsylvanian, Philadelphia, PA, USA, and a Faculty Member with the Johns Hopkins University, Baltimore, MD, USA. He has published over 800 papers in the international journals and conference proceedings. His current research interests include medical image analysis, computer vision, and pattern recognition.

Dr. Shen serves as an editorial board member for eight international journals. He has also served in the Board of Directors, The Medical Image Computing and Computer Assisted Intervention Society (MICCAI), from 2012 to 2015, and will be the General Chair for MICCAI 2019. He is a fellow of the American Institute for Medical and Biological Engineering, and also a fellow of the International Association for Pattern Recognition.

REFERENCES

[1]. Anagnostou E and Taylor MJ, "Review of neuroimaging in s spectrum disorders: What have we learned and where we go from here," Mol. Autism, vol. 2, no. 1, pp. 1–9, 2011. [PubMed: 21247446]

[2]. Greicius MD, Krasnow B, Reiss AL, and Menon V, "Functional connectivity in the resting brain: A network analysis of the default mode hypothesis," Proc. Nat. Acad. Sci. USA, vol. 100, no. 1, pp. 253–258, 2003. [PubMed: 12506194]

[3]. Kennedy DP and Courchesne E, "The intrinsic functional organization of the brain is altered in autism," Neuroimage, vol. 39, no. 4, pp. 1877–1885, 2008. [PubMed: 18083565]

[4]. Monk CS et al., "Abnormalities of intrinsic functional connectivity in autism spectrum disorders," Neuroimage, vol. 47, no. 2, pp. 764–772, 2009. [PubMed: 19409498]

[5]. Weng S-J et al., "Alterations of resting state functional connectivity in the default network in adolescents with autism spectrum disorders," Brain Res, vol. 1313, pp. 202–214, 2 2010. [PubMed: 20004180]

[6]. Tyszka JM, Kennedy DP, Paul LK, and Adolphs R, "Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism," Cerebral Cortex, vol. 24, no. 7, pp. 1894–1905, 2014. [PubMed: 23425893]

[7]. Hull JV, Jacokes ZJ, Torgerson CM, Irimia A, and Van Horn JD, "Resting-state functional connectivity in autism spectrum disorders: A review," Front. Psych, vol. 7, p. 205, 1 2017.

[8]. Wee CY et al., "Resting-state multi-spectrum functional connectivity networks for identification of MCI patients," PLoS ONE, vol. 7, 5 2012, Art. no. e37828.

[9]. Lee MH, Smyser CD, and Shimony JS, "Resting-state fMRI: A review of methods and clinical applications," Amer. J. Neuroradiology, vol. 34, no. 10, pp. 1866–1872, 2013.

[10]. Greicius M, "Resting-state functional connectivity in neuropsychiatric disorders," Current Opin. Neurol, vol. 21, no. 4, pp. 424–430, 2008.

[11]. Just MA, Cherkassky VL, Keller TA, and Minshew NJ, "Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity," Brain, vol. 127, pp. 1811–1821, 8 2004. [PubMed: 15215213]

[12]. Minshew NJ and Keller TA, "The nature of brain dysfunction in autism: Functional brain imaging studies," Current Opin. Neurol, vol. 23, no. 2, p. 124, 2010.

[13]. Müller R-A et al., "Underconnected, but how? A survey of functional connectivity MRI studies in autism spectrum disorders," Cerebral Cortex, vol. 21, no. 10, pp. 2233–2243, 2011. [PubMed: 21378114]

[14]. Vissers ME, Cohen MX, and Geurts HM, "Brain connectivity and high functioning autism: A promising path of research that needs refined models, methodological convergence, and stronger behavioral links," Neurosci. Biobehav. Rev, vol. 36, no. 1, pp. 604–625, 2012. [PubMed: 21963441]

[15]. Alaerts K et al., "Age-related changes in intrinsic function of the superior temporal sulcus in autism spectrum disorders," Soc. Cogn. Affect. Neurosci, vol. 10, no. 10, pp. 1413–14123, 2015. [PubMed: 25809403]

[16]. Wiggins JL et al., "Using a self-organizing map algorithm to detect age-related changes in functional connectivity during rest in autism spectrum disorders," Brain Res, vol. 1380, pp. 187–197, 3 2011. [PubMed: 21047495]

[17]. Wiggins JL et al., "The impact of serotonin transporter (5-HTTLPR) genotype on the development of resting-state functional connectivity in children and adolescents: A preliminary report," Neuroimage, vol. 59, no. 3, pp. 2760–2770, 2012. [PubMed: 22032950]

[18]. Fombonne E, "Epidemiology of autistic disorder and other pervasive developmental disorders," J. Clin. Psychi, vol. 66, no. 10, pp. 3–8, 2005.

[19]. Alaerts K, Swinnen SP, and Wenderoth N, "Sex differences in autism: A resting-state fMRI investigation of functional brain connectivity in males and females," Soc. Cogn. Affect. Neurosci, vol. 11, no. 6, pp. 1002–1016, 2016. [PubMed: 26989195]

[20]. Zhang H et al., "Topographical information-based high-order functional connectivity and its application in abnormality detection for mild cognitive impairment," J. Alzheimer's Disease, vol. 54, no. 3, pp. 1095–1112, 2016. [PubMed: 27567817]

[21]. Argyriou A, Pontil M, Ying Y, and Micchelli CA, "A spectral regularization framework for multi-task structure learning," in Proc. Adv. Neural Inf. Process. Syst, 2007, pp. 25–32.

[22]. Argyriou A, Evgeniou T, and Pontil M, "Multi-task feature learning," in Proc. Adv. Neural Inf. Process. Syst, 2007, pp. 41–48.

[23]. Kim S and Xing EP, "Tree-guided group lasso for multi-task regression with structured sparsity," Haifa, Israel, 2010, pp. 543–550.

[24]. Ji Y and Sun S, "Multitask multiclass support vector machines: Model and experiments," Pattern Recognit, vol. 46, no. 3, pp. 914–924, 2013.

[25]. Fang M, Yin J, Hall LO, and Tao D, "Active multitask learning with trace norm regularization based on excess risk," IEEE Trans. Cybern, vol. 47, no. 11, pp. 3906–3915, 11 2017. [PubMed: 27479984]

[26]. Pan S, Wu J, Zhu X, Long G, and Zhang C, "Task sensitive feature exploration and learning for multitask graph classification," IEEE Trans. Cybern, vol. 47, no. 3, pp. 744–758, 3 2017. [PubMed: 26978839]

[27]. Sun S, "A survey of multi-view machine learning," Neural Comput. Appl, vol. 23, nos. 7–8, pp. 2031–2038, 2013.

[28]. Xu C, Tao D, and Xu C, "A survey on multi-view learning," arXiv preprint arXiv:13045634, 2013.

[29]. Jiang Y et al., "Collaborative fuzzy clustering from multiple weighted views," IEEE Trans. Cybern, vol. 45, no. 4, pp. 688–701, 4 2015. [PubMed: 25069132]

[30]. He J and Lawrence R, "A graph-based framework for multi-task multi-view learning," in Proc. 28th Int. Conf. Mach. Learn. (ICML), 2011, pp. 25–32.

[31]. Zhang J and Huan J, "Inductive multi-task learning with multiple view data," in Proc. 18th ACM SIGKDD Int. Conf Knowl. Disc. Data Min, Beijing, China, 2012, pp. 543–551.

[32]. Jin X, Zhuang F, Wang S, He Q, and Shi Z, "Shared structure learning for multiple tasks with multiple views," in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Disc. Databases, Prague, Czech Republic, 2013, pp. 353–368.

[33]. Liu T, Tao D, Song M, and Maybank SJ, "Algorithm-dependent generalization bounds for multi-task learning," IEEE Trans. Pattern Anal. Mach. Intell, vol. 39, no. 2, pp. 227–241, 2 2017. [PubMed: 27019472]

[34]. Zhu X et al., "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," Med. Image Anal, vol. 38, pp. 205–214, 5 2017. [PubMed: 26674971]

[35]. Wang H et al., "Identifying disease sensitive and quantitative traitrelevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning," Bioinformatics, vol. 28, no. 12, pp. i127–i136, 2012. [PubMed: 22689752]

[36]. Evgeniou T and Pontil M, "Regularized multi-task learning," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min, Seattle, WA, USA, 2004, pp. 109–117.

[37]. Nie F, Huang H, Cai X, and Ding CH, "Efficient and robust feature selection via joint l2,1-norms minimization," in Proc. Adv. Neural Inf. Process. Syst, Vancouver, BC, Canada, 2010, pp. 1813–1821.

[38]. Zhou Z-H, Wu J, and Tang W, "Ensembling neural networks: Many could be better than all," Artif. Intell, vol. 137, nos. 1–2, pp. 239–263, 2002.

[39]. Friston KJ, Williams S, Howard R, Frackowiak RS, and Turner R, "Movement-related effects in fMRI time-series," Magn. Reson. Med, vol. 35, no. 3, pp. 346–355, 1996. [PubMed: 8699946]

[40]. Tzourio-Mazoyer N et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," Neuroimage, vol. 15, no. 1, pp. 273–289, 2002. [PubMed: 11771995]

[41]. Zhang H, Chen X, Zhang Y, and Shen D, "Test-retest reliability of 'high-order' functional connectivity in young healthy adults," Front. Neurosci, vol. 11, p. 439, 8 2017. [PubMed: 28824362]

[42]. Chang C-C and Lin C-J. (2012). LIBSVM: A Library for Support Vector Machine, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[43]. Zhang D et al., "Multimodal classification of Alzheimer's disease and mild cognitive impairment," Neuroimage, vol. 55, no. 3, pp. 856–867, 2011. [PubMed: 21236349]

[44]. Li Y, Tian X, Liu T, and Tao D, "Multi-task model and feature joint learning," in Proc. 24th Int. Joint Conf. Artif. Intell, 2015, pp. 3643–3649.

[45]. Yahata N et al., "A small number of abnormal brain connections predicts adult autism spectrum disorder," Nat. Commun., vol. 7, 4 2016, Art. no. 11254.

[46]. Pan SJ and Yang Q, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng, vol. 22, no. 10, pp. 1345–1359, 10 2010.

[47]. Zhang D and Shen D, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," Neuroimage, vol. 59, no. 2, pp. 895–907, 2012. [PubMed: 21992749]

[48]. Zhang Y et al., "Sparse Bayesian multiway canonical correlation analysis for EEG pattern recognition," Neurocomputing, vol. 225, pp. 103–110, 2 2017.

[49]. Jie B, Zhang D, Cheng B, and Shen D, "Manifold regularized multitask feature learning for multimodality disease classification," Human Brain Mapping, vol. 36, no. 2, pp. 489–507, 2015. [PubMed: 25277605]

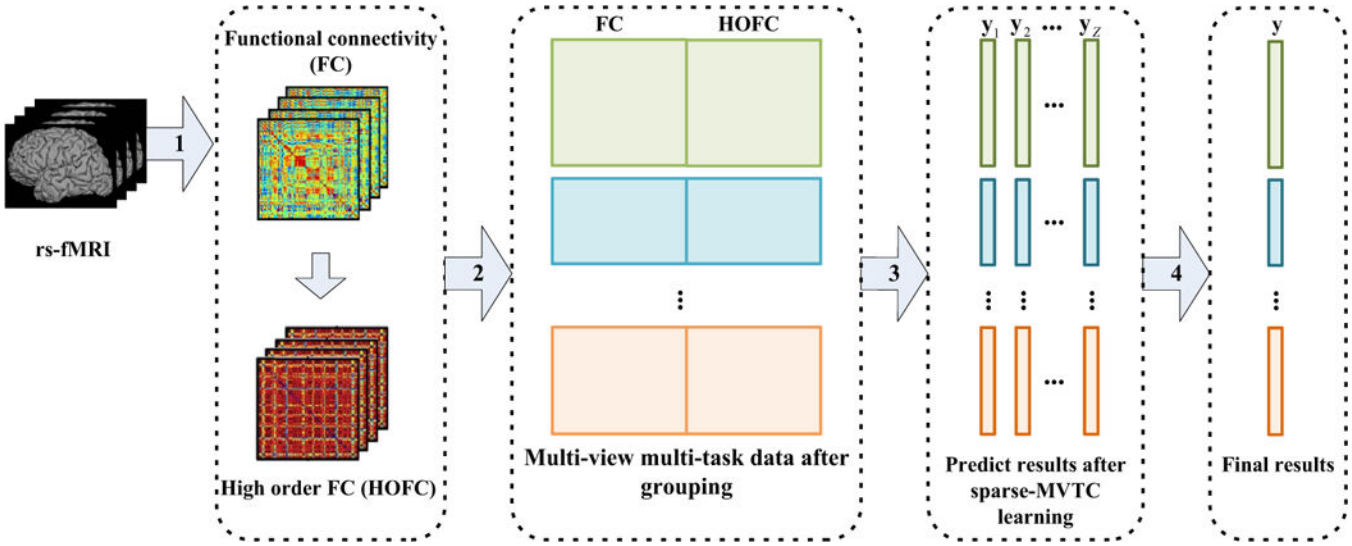

Fig. 1.

Framework of our Sparse-MVTC-E classification, which includes four main steps: 1) extraction of the FC and the HOFC features; 2) subject grouping or task partitioning according to both age and sex information; 3) Sparse-MVTC learning upon Z tasks; and 4) an ensemble strategy to make the final decision.

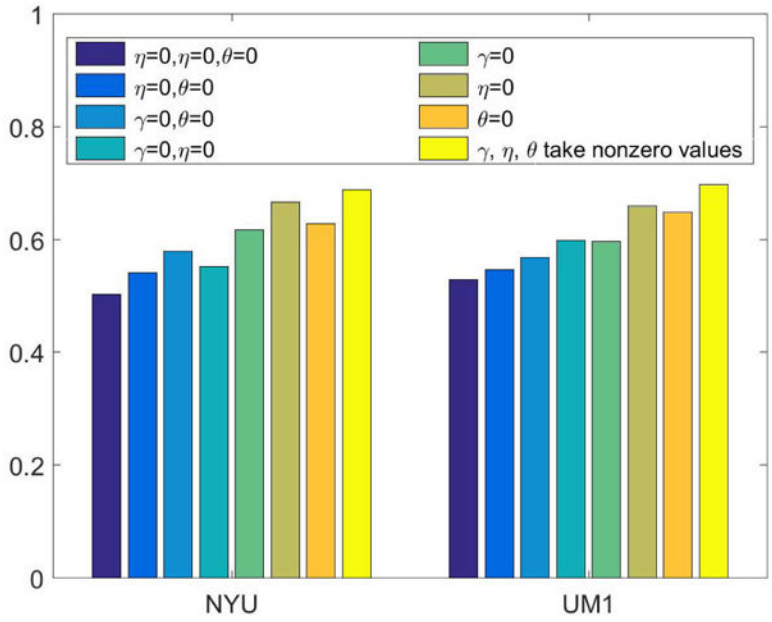

Fig. 2.

Comparisons of ACCs for Sparse-MVTC and its variants when turning off individual regularization term(s).

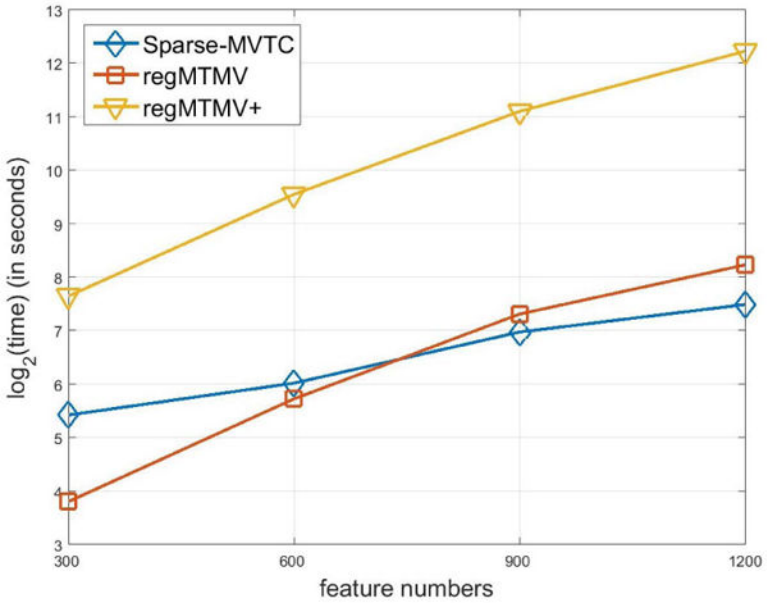

Fig. 3.
Training time for three different MVMT learning methods.

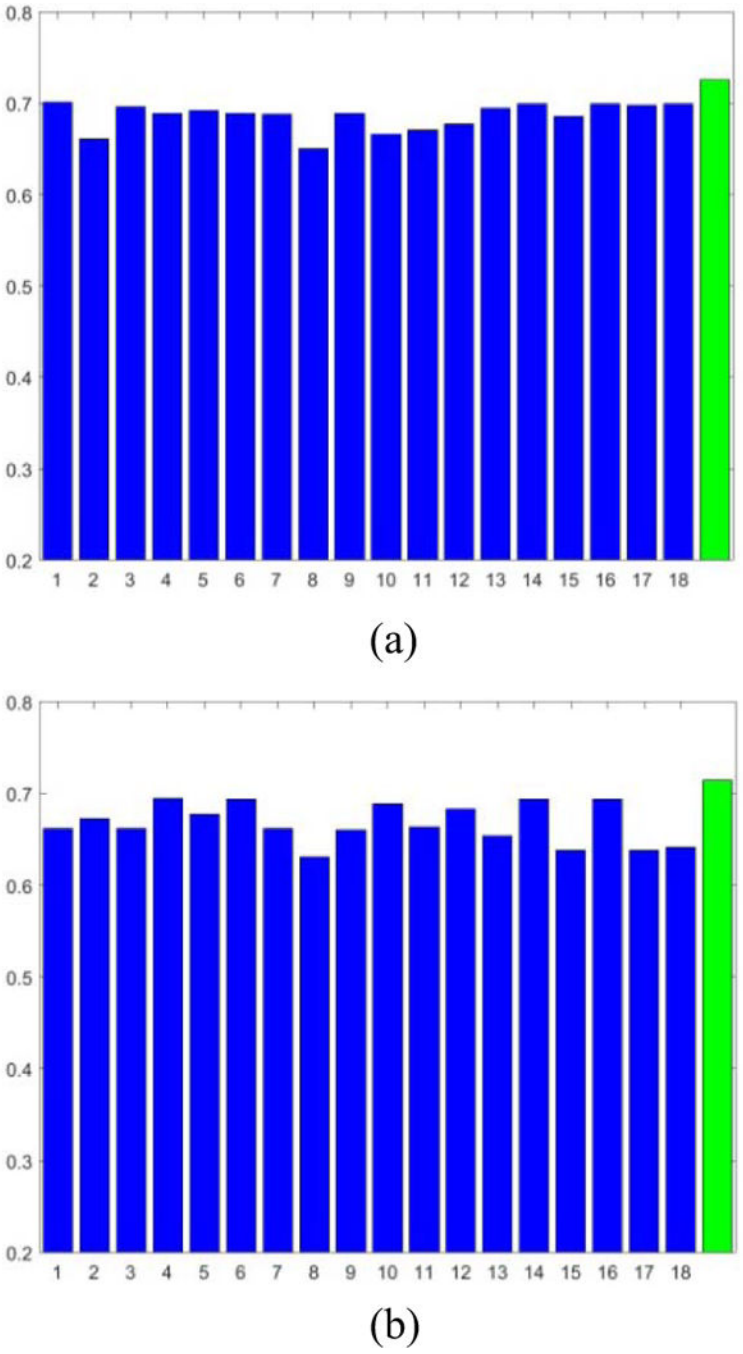

Fig. 4.

Comparison of accuracy for Sparse-MVTC-E and Sparse-MVTC, with different central tasks on imaging centers, i.e., NYU and UM-1. The green bar indicates the accuracy for Sparse-MVTC-E and the blue bars indicate the accuracies of Sparse-MVTC with different central tasks, respectively. The number on the x-axis under each blue bar denotes the respective central task ID of Sparse-MVTC.

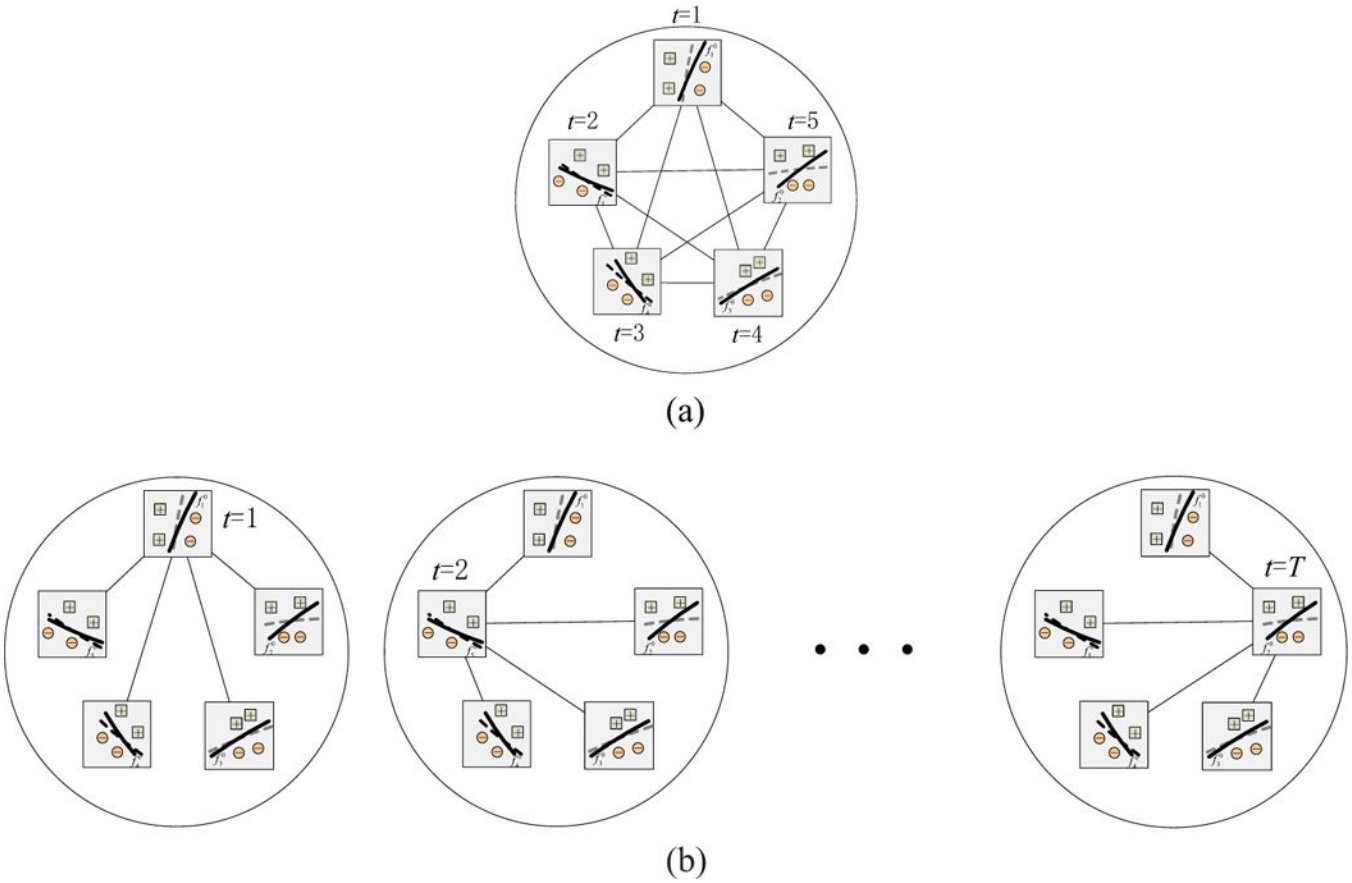

Fig. 5.

(a) Full connection graph can be utilized to describe the relationships between tasks in the existing MTL frameworks, in which the tasks are correlated with each other. Each learning task is described by a rectangle in the figure, in which the dotted line and the solid line represent the true and computed separate planes of the classifier, respectively. (b) The tree structure can be utilized to describe the relationships between tasks in the proposed task-centralized learning framework, in which the root of the tree is the central task and the leaves are the auxiliary tasks. The ensemble process enhances the performance of the classifiers and helps to obtain better results.

TABLE I

DEMOGRAPHIC INFORMATION OF GROUPS IN NYU and UM-1

Imaging center	Group ID	Age range	Sex	Number of Subjects	Patients/ Controls
NYU	1	(6,12]	F	12	3/9
	2	(12,18]	F	14	4/10
	3	(18,40]	F	11	4/7
	4	(6,10]	M	32	21/11
	5	(10,12]	M	27	14/13
	6	(12,14]	M	20	9/11
	7	(14,16]	M	17	6/11
	8	(16,18]	M	12	3/9
	9	(18,22]	M	13	4/9
	10	(22,25]	M	12	5/7
	11	(25,40]	M	13	5/8
UM-1	12	(8,14]	F	12	5/7
	13	(14,20]	F	12	2/10
	14	(8,10]	M	9	5/4
	15	(10,11]	M	11	3/8
	16	(11,12]	M	10	6/4
	17	(12,15]	M	19	14/5
	18	(15,20]	M	23	7/16

TABLE II

ACQUISITION PARAMETERS OF RS-FMRI IN NYU and UM-1

Center	NYU	UM-1
Make (model)	Siemens Magnetom (Allegra)	GE (Signa)
Voxel size (mm^3)	3.0×3.0×4.0	3.438×3.438×3.0
Flip angle (deg)	90	90
TR(ms)	2000	2000
TE (ms)	15	30
Bandwidth (Hz/Px)	3906	NA

TABLE III

SUMMARY OF THE METHODS UNDER COMPARISON

Method	Description	Data preparation
CSVC-S	C-SVC in LibSVM. Linear kernel was adopted in C-SVC.	FC and HOFC features of each subject were concatenated.
CSVC-J		
CSVC-E		
M2SVC-S	Multi-modal classification proposed by Zhang *et al.* [43].	FC and HOFC features were regarded as different views inM2SVC.
M2SVC-J		
M2SVC-E		
RF-S	Random forest algorithm in.	The features from different views of each subject were concatenated.
RF-J		
RF-E		
MTMFJL	Multi-Task Model and Feature Joint Learning method proposed in [44]	Each group is regarded as one task. FC and HOFC features of each subject were concatenated into one view.
IteM^2	MVMT method proposed by He *et al.* [30]	FC and HOFC features were regarded as different views, respectively. Each group is regarded as one task.
regMVMT	Regularized multi-view multi-task learning method proposed by Zhang *et al.* [31]	FC and HOFC features were regarded as different views, respectively. Each group is regarded as one task.
regMVMT+	Enhanced version of regMVMT proposed by Zhang *et al.* [31]	FC and HOFC features were regarded as one view, respectively. Each group is regarded as one task.
Sparse-MVTC-Tr	Sparse-MVTC in a transfer learning way	FC and HOFC features were regarded as different views, respectively. The central task in Sparse MVTC is considered as the learning task in the target domain and the auxiliary tasks are considered as the learning tasks in the source domains.
Sparse-MVTC-E	Our proposed ensemble classifier.	FC and HOFC features were regarded as different views, respectively. Each group is regarded as one task.

Table IV

CLASSIFICATION RESULTS ON NYU

Method	ACC	SEN	SPE	AUC	p-value
CSVC-S	0.607	0.725	0.449	0.587	<le-3
CSVC-J	0.683	0.804	0.521	0.662	0.001
CSVC-E	0.600	0.958	0.118	0.538	<le-3
M2SVC-S	0.596	0.730	0.414	0.572	<le-3
M2SVC-J	0.675	0.756	0.565	0.661	0.001
M2SVC-E	0.603	**0.979**	0.097	0.538	<le-3
RF-S	0.584	0.679	0.456	0.568	<le-3
RF-J	0.650	0.766	0.494	0.630	<le-3
RF-E	0.577	0.910	0.128	0.519	<le-3
MTMFJL	0.656	0.762	0.513	0.637	<le-3
IteM^2	0.557	0.552	0.564	0.558	<le-3
regMVMT	0.593	0.608	0.574	0.591	<le-3
regMVMT+	0.612	0.629	0.590	0.609	<le-3
Sparse-MVTC	0.691	0.763	0.593	0.678	<le-3
Sparse-MVTC-Tr	0.672	0.752	0.564	0.658	<le-3
Sparse-MVTC-E	**0.726**	0.790	**0.640**	**0.715**	

Table V

CLASSIFICATION RESULTS ON UM-1

Method	ACC	SEN	SPE	AUC	p-value
CSVC-S	0.552	0.737	0.314	0.526	<le-3
CSVC-J	0.661	0.719	0.588	0.653	0.006
CSVC-E	0.564	0.900	0.131	0.515	<le-3
M2SVC-S	0.605	0.689	0.498	0.593	<le-3
M2SVC-J	0.619	0.648	0.581	0.615	<le-3
M2SVC-E	0.562	**0.915**	0.110	0.512	<le-3
RF-S	0.597	0.724	0.433	0.579	<le-3
RF-J	0.633	0.679	0.574	0.627	0.003
RF-E	0.582	0.831	0.262	0.547	<le-3
MTMFJL	0.604	0.667	0.524	0.595	<le-3
IteM^2	0.572	0.685	0.429	0.557	<le-3
regMVMT	0.594	0.715	0.438	0.576	<le-3
regMVMT+	0.604	0.722	0.452	0.587	<le-3
Sparse-MVTC	0.672	0.706	0.629	0.668	<le-3
Sparse-MVTC-Tr	0.646	0.667	0.619	0.643	<le-3
Sparse-MVTC-E	**0.714**	0.743	**0.677**	0.710	

Table VI

COMMON REGION-TO-REGION FC SELECTED BY SPARSE-MVTC-E

	ROI1	ROI2
1	Frontal_Inf_Tri_L (13)	Parietal_Sup_L (59)
2	Supp_Motor_Area_L(19)	Cerebelum_Crus2_L (93)
3	Frontal_Med_Orb_L(25)	Cingulum_Post_R (36)
4	Frontal_Med_Orb_R (26)	ParaHippocampal_L (39)
5	Rectus_L (27)	Precuneus_R(68)
6	Cingulum_Post_L (35)	Parietal_Inf_R (62)
7	ParaHippocampal_L (39)	Pallidum_L (75)
8	ParaHippocampal_R (40)	Putamen_L(73)
9	ParaHippocampal_R (40)	Pallidum_L (75)
10	Amygdala_R (42)	Temporal_Mid_L (85)
11	Amygdala_R (42)	Temporal_Mid_R (86)
12	Amygdala_R (42)	Temporal_Pole_Mid_R(88)
13	Putamen_R (74)	Vermis_8(114)
14	Pallidum_R (76)	Vermis_8(114)
15	Cerebelum_6_L_(99)	Vermis_10(116)

TABLE VII

COMMON REGION-TO-REGION *High-Order* FC SELECTED BY SPARSE-MVTC-E

	ROI1	ROI2
1	Precentral_R (2)	Cerebelum_9_L (105)
2	Frontal_Inf_Tri_L (13)	Supp_Motor_Area_L (19)
3	Frontal_Inf_Tri_R (14)	Rectus_R (28)
4	Hippocampus_R (38)	Calcarine_L (43)
5	Hippocampus_R (38)	Temporal_Mid_R (86)
6	ParaHippocampal_R (40)	Pallidum_L (75)
7	Amygdala_L (41)	Vermis_6 (112)
8	Amygdala_R (42)	Temporal_Mid_L (85)
9	Amygdala_R (42)	Temporal_Mid_R (86)
10	Parietal_Inf_L (61)	Caudate_L (71)
11	SupraMarginal_R (64)	Precuneus_L (67)
12	SupraMarginal_R (64)	Precuneus_R(68)
13	Putamen_R (74)	Vermis_8 (114)
14	Pallidum_L (75)	Temporal_Pole_Mid_L (87)
15	Pallidum_L (75)	Temporal_Pole_Mid_R (88)
16	Pallidum_L (75)	Temporal_Inf_L (89)
17	Pallidum_L (75)	Temporal_Inf_R (90)
18	Pallidum_R (76)	Temporal_Inf_R (90)
19	Thalamus_R(78)	Temporal_Inf_R (90)
20	Heschl_L (79)	Temporal_Inf_R(90)