Published in final edited form as:

Mach Learn Med Imaging. 2017 ; 10541: 159–167. doi:10.1007/978-3-319-67389-9_19.

Sparse Multi-view Task-Centralized Learning for ASD Diagnosis

Jun Wang^{1,2}, Qian Wang³, Shitong Wang², and Dinggang Shen^{1,4}

¹Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

²School of Digital Media, Jiangnan University, Wuxi 214122, Jiangsu, China

³School of Biomedical Engineering, Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China

⁴Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

Abstract

It is challenging to derive early diagnosis from neuroimaging data for autism spectrum disorder (ASD). In this work, we propose a novel sparse multi-view task-centralized (Sparse-MVTC) classification method for computer-assisted diagnosis of ASD. In particular, since ASD is known to be age- and sex-related, we partition all subjects into different groups of age/sex, each of which can be treated as a classification task to learn. Meanwhile, we extract multi-view features from functional magnetic resonance imaging to describe the brain connectivity of each subject. This formulates a multi-view multi-task sparse learning problem and it is solved by a novel Sparse-MVTC method. Specifically, we treat each task as a central task and other tasks as the auxiliary ones. We then consider the task-task and view-view relations between the central task and each auxiliary task. We can use this task-centralized strategy for a highly efficient solution. The comprehensive experiments on the ABIDE database demonstrate that our proposed Sparse-MVTC method can significantly outperform the existing classification methods in ASD diagnosis.

1 Introduction

Autism spectrum disorder (ASD) is a major mental disorder for childhood. The computer-assisted diagnosis of ASD using neuroimaging data is highly desired, especially before abnormal behaviors of the patients turn observable. In recent years, resting-state functional magnetic resonance imaging (rs-fMRI) has become a pivotal tool in understanding the mechanism of ASD, as researchers can identify many disease-related biomarkers. However, it is non-trivial to extract features from the noisy rs-fMRI data, select the effective biomarkers, and learn the classification model for the diagnosis of the disease [1].

Functional connectivity (FC), which measures the correlation of the temporal BOLD signals in different brain regions and encodes the functional connection of the structurally segregated brain regions, is often adopted for feature extraction in rs-fMRI analysis [2]. By computing FCs between many regions, one may derive a large matrix to represent the FC

network. Since the pathology of ASD may disrupt the FC network, we are expected to utilize these correlation-based FC measures to diagnose the disease.

Note that the correlation-based FC is not the only metric to present the brain functional network. For example, Zhang *et al.* have proposed high-order functional connectivity (HOFC) [3], 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, i.e., following the conventional FC analysis method, to record its correlations with all other regions in the brain. After that, a high-level correlation is calculated between the correlation feature vectors of two regions, thus resulting in their in-between HOFC. This HOFC measure has shown its advantages, including the sensitivity to group difference, individual variability, and prominent modular structures [3]. To this end, for diagnosing ASD, we may extract both FC and HOFC measures from the rs-fMRI data of a certain subject, and treat them as two different views of connectivity features.

Recent studies show that ASD is an age- and sex-related neurodevelopmental disorder. For example, ASD patients differ from normal controls in the intrinsic FCs of the default mode networks, while the differences vary with ages [4, 5]. Besides, Alaerts *et al.* have found that males and females have different neural expressions of ASD, i.e., characterized by the predominant hypo-connectivity patterns in males while the hyper-connectivity patterns in females [6]. In order to account for these findings, one may partition the subjects into different age/sex groups naturally. Then, each group can be treated as a task, such that the ASD diagnosis upon all groups can be solved by the multi-task learning.

Since FC and HOFC act as two different views of connectivity features, they can be utilized simultaneously in the multi-view learning [7, 8]. In this way, the computer-assisted ASD diagnosis can be formulated as a *multi-view multi-task* (MVMT) learning problem. Higher classification accuracy is presumably achievable, once the learning upon all tasks with coupled views of connectivity features is jointly conducted in a proper way.

The solution to the MVMT learning, however, is non-trivial. For example, Zhang *et al.* has proposed regMVMT for multi-task learning on the multi-view data, in which co-regularization and task-task relationships are considered simultaneously [9]. Although the solution of regMVMT is a close-form analytically, it involves an inverse operation of a large matrix that strongly limits the actual usage of this method, especially when the number of the tasks is large.

In this work, we propose a novel sparse multi-view task-centralized classification method (Sparse-MVTC) [9, 10], in which the task-centralized learning mechanism helps handle a large number of tasks. Specifically, we decompose the multi-view classification problem into multiple age/sex-related tasks, each of which can be solved by regressing the ASD diagnosis from imaging features via sparse learning. Then, we introduce the task-centralized strategy by picking up a certain task as a central task and treating all others as the auxiliary tasks. By further considering the relationship between the central task and each auxiliary task, the joint sparse learning will lead to a set of classifiers that correspond to individual central/auxiliary tasks. Moreover, since the central task can be selected in turn, the above

procedure of the joint sparse learning can be repeated for multiple times, thus resulting in multiple classifiers for each task. By integrating all the classifiers of each task, we are able to attain high-performance diagnosis accordingly. We do not find any similar work reported before.

We will detail the proposed Sparse-MVTC method and its solution in Sect. 2. In particular, the task-centralized strategy in our method considers the relationship between the central task and each auxiliary task only. The relationship across auxiliary tasks is neglected for easy computation. Regarding the task-task relationship, we devise novel terms in sparse learning to account for the joint feature selection and weighting between the central and the auxiliary tasks, as well as the consistency across the multi-view features. In Sect. 3, we apply our method to ASD diagnosis and present the experimental results on the multi-center ABIDE dataset. We conclude this work in Sect. 4.

2 Method

2.1 Sparse Multi-view Task-Centralized (Sparse-MVTC) Learning

For ASD diagnosis, we partition all subjects into T groups according to their age/sex information. Each group corresponds to a supervised learning task, as V views of features are assumed available per subject. In the t -th task ($t = 1, 2, \dots, T$), we have N_t subjects

$\{\mathbf{x}_i, y_i\}_{i=1}^{N_t}$ for training. Here, $\mathbf{x}_i = ((\mathbf{x}_i^1)', \dots, (\mathbf{x}_i^V)')' \in \mathbb{R}^D$ is a vector including all V -view

features for the i -th subject. Each view has D_v features ($D = \sum_{v=1}^V D_v$). $y_i \in \{-1, 1\}$ is the class label of $\mathbf{x}_i$ (e.g., “-1” for healthy controls, and “1” for ASD patients). Let $\mathbf{X}_t = [\mathbf{x}_1, \dots, \mathbf{x}_{N_t}]' \in \mathbb{R}^{N_t \times D}$ and $\mathbf{y}_t = (y_1, \dots, y_{N_t})$ be the data matrix and the training label vector for the t -th task, respectively. The vector $\mathbf{w}_t^v \in \mathbb{R}^{D_v}$ indicates the weights of all features in the v -th view

with respect to the t -th task. We further define $\mathbf{w}_t = ((\mathbf{w}_t^1)', \dots, (\mathbf{w}_t^V)')' \in \mathbb{R}^D$ and $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_T] \in \mathbb{R}^{D \times T}$.

We pick up one task as a central task and treat all others as the auxiliary tasks. Suppose that the t -th ($t = 1, 2, \dots, T$) task is the central task. The sparse learning upon all tasks (i.e., Sparse-MVTC) is formulated as the following optimization problem

$$\min_{\mathbf{W}} \left(\sum_{u=1}^T \frac{1}{N_u} \left\| \mathbf{y}_u - \sum_{v=1}^V \mathbf{X}_u^v \mathbf{w}_u^v \right\|^2 + R(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s \neq t}^T) \right), \quad (1)$$

where N_u is the number of the training subjects in the u -th task, and $R(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s \neq t}^T)$ is a regularization to encode the relationship between the central task and each auxiliary task in the multi-view setting. In particular, we devise three terms for the regularization

$R(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s \neq t}^T)$, i.e., $R = \gamma R_1 + \eta R_2 + \theta R_3$, where γ , η and θ are non-negative scalars to control the respective regularization terms, R_1 , R_2 , and R_3 .

Joint task-task feature selection—To conduct feature selection jointly between the central task and each auxiliary task, we adopt

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

where $\mathbf{W}_{s,t}^v = [\mathbf{w}_s^v, \mathbf{w}_t^v] \in \mathbb{R}^{D_v \times 2}$ and $c_{s,t}$ measures the similarity between the two tasks s and t . If two tasks are similar, they should select similar features by sparse learning. To this end,

we compute the coefficient as $c_{s,t} = \exp \left(-\left\| \bar{\mathbf{X}}_t - \bar{\mathbf{X}}_s \right\|^2 / \sigma^2 \right)$, where

$$\bar{\mathbf{X}}_t = \sum_{i=1}^{N_t} \mathbf{x}_i / N_t, \bar{\mathbf{X}}_s = \sum_{i=1}^{N_s} \mathbf{x}_i / N_s \text{ and } \sigma^2 = \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} \mathbf{x}_i - \mathbf{x}_j^2 / N_s N_t.$$

Joint task-task feature weighting—Besides joint feature selection, we assume that, if the central task is similar with a certain auxiliary task, their jointly selected features should share similar weights when regressing the disease labels. To this end, we propose the regularization term R_2 for the joint task-task feature weighting:

$$R_2 \left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s \neq t}^T \right) = \sum_{s=1, s \neq t}^T c_{s,t} \sum_{v=1}^V \left\| \mathbf{w}_s^v - \mathbf{w}_t^v \right\|^2. \quad (3)$$

Note that $c_{s,t}$ encodes the similarity between the central task and the auxiliary task, as defined above.

Consistency of view-view predictability—It is important to effectively integrate the two views of features for ASD diagnosis. In the multi-view learning, the discriminant functions for different views of a subject tend to yield the same classification result. Therefore, we devise the following regularization term to ensure the consistency:

$$R_3 \left(\mathbf{w}_t; \{\mathbf{w}_s\}_{s=1, s \neq t}^T \right) = \sum_{u=1}^T \frac{1}{N_u} \sum_{i,j=1}^v \left\| \mathbf{X}_u^i \mathbf{w}_u^i - \mathbf{X}_u^j \mathbf{w}_u^j \right\|^2. \quad (4)$$

2.2 Iterative Optimization in Sparse-MVTC

With the above three regularization terms, we now aim to minimize the following objective function for the proposed Sparse-MVTC method:

$$\begin{aligned}
J(\mathbf{w}_t, \{\mathbf{w}_s\}) = & \sum_{u=1}^T \frac{1}{N_u} \left\| \mathbf{y}_u - \sum_{v=1}^V \mathbf{X}_u^v \mathbf{w}_u^v \right\|^2 + \gamma \sum_{\substack{s=1 \\ s \neq t}}^T, c_{s,t} \sum_{v=1}^V \text{tr}(\mathbf{W}_{s,t}^{v'} \Lambda_{s,t}^v \mathbf{W}_{s,t}^v) + \\
& \eta \sum_{\substack{s=1 \\ s \neq t}}^T, c_{s,t} \sum_{v=1}^V \left\| \mathbf{w}_s^v - \mathbf{w}_t^v \right\|^2 + \theta \sum_{u=1}^T \frac{1}{N_k} \sum_{i,j=1}^V \left\| \mathbf{X}_u^i \mathbf{w}_u^i - \mathbf{X}_u^j \mathbf{w}_u^j \right\|^2
\end{aligned}
\tag{5}$$

Here, we replace $\left\| \mathbf{W}_{s,t}^v \right\|_{2,1}$ with $\text{tr}(\mathbf{W}_{s,t}^{v'} \Lambda_{s,t}^v \mathbf{W}_{s,t}^v)$, since the minimization upon the two terms are shown to be equivalent [11]. Specifically, $\Lambda_{s,t}^v (v=1, \dots, V; s=1, \dots, T; s \neq t)$ is a $D_v \times D_v$ diagonal matrix with its i -th diagonal element being computed as $1/2 \left\| (\mathbf{e}_i)' \mathbf{W}_{s,t}^v \right\|$. Note that $\mathbf{e}_i$ is a column vector with its i -th element as 1 and all others as zero.

In Eq. (5), the feature weights $\mathbf{w}_t$ for the central task and $\{\mathbf{w}_s\}$ for the auxiliary tasks, as well as $\{\Lambda_{s,t}^v\}$, can be optimized alternatively, i.e., fixing any two of them and then optimizing the third one. The details are provided in the below.

1. Fixing $\mathbf{w}_t$ and $\{\mathbf{w}_s\}$, we compute $\Lambda_{s,t}^v$ with its i -th diagonal element as $1/2 \left\| (\mathbf{e}_i)' \mathbf{W}_{s,t}^v \right\|$.
2. Fixing $\{\Lambda_{s,t}^v\}$ and $\mathbf{w}_t$, we compute $\mathbf{P}_s^v$, $\mathbf{Q}_s^v$ and $\mathbf{r}_s^v$ for each auxiliary task following:

$$\mathbf{P}_s^v = \left(\gamma c_{s,t} \Lambda_{s,t}^v + \eta c_{s,t} \mathbf{I} + 2\theta \frac{1}{N_s} V (\mathbf{X}_s^v)' \mathbf{X}_s^v \right), \tag{6}$$

$$\mathbf{Q}_s^v = \left(\frac{1 - 2\theta}{N_s} \right) (\mathbf{X}_s^v)', \mathbf{r}_s^v = \frac{1}{N_s} (\mathbf{X}_s^v)' \mathbf{y}_s + \eta c_{s,t} \mathbf{w}_t^v.$$

We are then able to update $\{\mathbf{w}_s\}$ by:

$$\begin{aligned}
\hat{\mathbf{P}}_{s=1}^{\text{Eij}} = & \begin{pmatrix} \mathbf{P}_s^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{P}_s^V \end{pmatrix}, \hat{\mathbf{Q}}_{s=1}^{\text{Eij}} = \begin{pmatrix} \mathbf{Q}_s^1 \mathbf{X}_s^1 & \cdots & \mathbf{Q}_s^1 \mathbf{X}_s^V \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_s^V \mathbf{X}_s^1 & \cdots & \mathbf{Q}_s^V \mathbf{X}_s^V \end{pmatrix}, \hat{\mathbf{r}}_{s=1}^{\text{Eij}} = (\mathbf{r}_s^1 \quad \cdots \quad \mathbf{r}_s^V)'
\end{aligned}
\tag{7}$$

$$\mathbf{w}_s = (\tilde{\mathbf{P}}_s + \tilde{\mathbf{Q}}_s)^{-1} \tilde{\mathbf{r}}_s. \quad (8)$$

3. Similarly, we can fix $\{\Lambda_{s,t}^v\}$ and $\{\mathbf{w}_s\}$, and optimize $\mathbf{w}_t$ by:

$$\mathbf{P}_t^v = \left(\gamma \sum_{\substack{s=1 \\ s \neq t}}^T c_{s,t} \Lambda_{s,t}^v + \eta \sum_{\substack{s=1 \\ s \neq t}}^T c_{s,t} \mathbf{I} + 2\theta \frac{1}{N_t} M(\mathbf{X}_t^v)' \mathbf{X}_t^v \right), \quad (9)$$

$$\mathbf{Q}_t^v = \frac{1-2\theta}{N_t} (\mathbf{X}_t^v)' \sum_{i=1}^V \mathbf{X}_t^i \mathbf{w}_t^i, \mathbf{r}_t^v = \frac{1}{N_t} (\mathbf{X}_t^v)' \mathbf{y}_t + \eta \sum_{\substack{s=1 \\ s \neq t}}^T c_{s,t} \mathbf{w}_s^v,$$

$$\begin{aligned} \tilde{\mathbf{P}}_t &= \begin{pmatrix} \mathbf{P}_t^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{P}_t^V \end{pmatrix}, \\ \tilde{\mathbf{Q}}_t &= \begin{pmatrix} \mathbf{Q}_t^1 \mathbf{X}_t^1 & \cdots & \mathbf{Q}_t^V \mathbf{X}_t^V \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_t^V \mathbf{X}_t^1 & \cdots & \mathbf{Q}_t^V \mathbf{X}_t^V \end{pmatrix}, \tilde{\mathbf{r}}_t = (\mathbf{r}_t^1 \quad \cdots \quad \mathbf{r}_t^V)' \end{aligned} \quad (10)$$

$$\mathbf{w}_t = (\tilde{\mathbf{P}}_t + \tilde{\mathbf{Q}}_t)^{-1} \tilde{\mathbf{r}}_t. \quad (11)$$

The optimization of the objective function in Eq. (5) is conducted by iteratively calling the three steps above. We abort from the optimization, once $\mathbf{W}$ tends to be stable.

Note that both $(\tilde{\mathbf{P}}_t + \tilde{\mathbf{Q}}_t)$ and $(\tilde{\mathbf{P}}_s + \tilde{\mathbf{Q}}_s)$ are $D \times D$ matrices. The bottleneck of the Sparse-MVTC method is thus the inverse of matrices in $\mathbf{w}_s = (\tilde{\mathbf{P}}_s + \tilde{\mathbf{Q}}_s)^{-1} \tilde{\mathbf{r}}_s$ (Eq. 8) and $\mathbf{w}_t = (\tilde{\mathbf{P}}_t + \tilde{\mathbf{Q}}_t)^{-1} \tilde{\mathbf{r}}_t$ (Eq. 11), each of which has the time complexity of $O(D^3)$. Compared to regMVMT [9] that also involves an inverse of a $TD \times TD$ matrix, Sparse-MVTC costs less computation time and less memory space, thus making Sparse-MVTC more suitable for MVMT applications with many tasks.

2.3 Ensemble Implementation of Sparse-MVTC

The training process in Sparse-MVTC determines the discriminant weights $\mathbf{W}$ for multi-view features. Once the weights are estimated, a testing subject can acquire the diagnosis accordingly. Moreover, in order to tackle the bias caused by the arbitrary selection of the

central task, we pick up the central task in turn and repeat Sparse-MVTC for multiple times. In this way, for each task, there are T classifiers. Given the testing subject with known age/sex information, we can easily identify the classifiers corresponding to the respective group/task. The final diagnosis is thus obtained by averaging all the outputs of the respective classifiers, to which the testing subject belongs.

3 Experimental Results

3.1 Experimental Settings

In this study, we consider rs-fMRI scans acquired from two different imaging centers (i.e., NYU and UM-1 of ABIDE, http://fcon_1000.projects.nitrc.org/indi/abide/index.html). Both male and female subjects are included in our study and their ages vary from 6 to 40. We partition these subjects into different groups according to their age and sex information. We further extract FCs from rs-fMRI using the standard pipeline provided by DPARSF (<http://rfmri.org/DPARSF>). To measure the FCs between ROIs, the pairwise Pearson correlation coefficients are computed, which results in a 116×116 correlation matrix for each subject. After that, HOFC is computed following [3].

In our study, ASD diagnosis is performed by combining the two views of FC and HOFC. For both views, the upper triangles of the matrices are utilized due to the symmetry of the matrices. These measures are reshaped into a vector with 6670 elements. Prior to training, simple feature selection is conducted. Specifically, we select 300 FC features and 300 HOFC features using t -tests. All these features are further normalized regarding the z-score and used for subsequent computation.

We consider ASD diagnosis as the binary classification problem, with label '+1' for ASD patients and label '-1' for the healthy controls. We adopt a 10-fold nested cross-validation strategy to evaluate the performance of our proposed method. The nested cross-validation consists of a 5-fold inner loop that is used to determine the optimal parameters including γ , η , and θ automatically. Finally, we repeat the 10-fold nested cross-validation for 10 times to avoid the arbitrary bias of fold partitioning. We report the average statistics of all 10 repetitions.

3.2 Comparisons with State-of-the-Art Methods

We compare Sparse-MVTC with several popular classifiers. We consider the following experimental settings for CSVC [12] and M2SVC [13] under comparison (see Table 1).

Besides, we also compare our method with regMVMT, which is a typical MVMT learning method proposed in [9].

We report the classification performances of all methods in Table 2. One may observe that the proposed Sparse-MVTC method could achieve the best classification accuracy. Concerning all 279 subjects from both clinical centers, the performances are mostly comparable across the centers for our method.

Note that, in CSVC, both FC and HOFC features are concatenated into a long feature vector, by simply ignoring the view-view relationship. On the contrary, M2SVC fuses the multi-view features for the classification. However, both CSVC and M2SVC fail to consider the multi-task learning setting. Different from CSVC and M2SVC, our Sparse-MVTC method *not only* treats FC/HOFC features jointly, *but also* fully considers the task-task relations through the task-centralized strategy. In this way, we are able to obtain much better results than CSVC and M2SVC.

Sparse-MVTC performs joint feature selection and weighting between the central task and each auxiliary task. The regMVMT method also considers the task-task and the view-view relations. However, regMVMT does not require the selected features to be sparsely distributed. The experimental results show that Sparse-MVTC outperforms regMVMT significantly in terms of classification accuracy. This implies that the sparse feature selection helps overcome the overfitting problem on small training datasets, and contributes to higher performances of the discriminant classifiers.

Regarding the memory requirement, it might be hard to deploy regMVMT on an ordinary PC, since it needs an inverse operation of a 10800×10800 matrix. However, Sparse-MVTC reduces the scale to 600×600 matrices. The run-time is about 2 s for Sparse-MVTC and about 10 s for regMVMT on a computer with X5670 processor and 48 GB memory. It is evident to see the advantages of the task-centralized strategy on multi-task learning, especially for the case with many tasks.

4 Conclusion

ASD is a complex neurodevelopmental disorder, as the disease-related alterations in brain's functional connectivity vary according to age and sex. In this work, we formulate the computer-assisted ASD diagnosis as an MVMT learning problem, and propose a novel Sparse-MVTC method to enforce the strong regularizations upon the task-task and view-view relations. An iterative optimization is further schemed for Sparse-MVTC. Using ABIDE database, we achieve a significant performance improvement for diagnosing ASD, compared with state-of-the-art classification methods.

References

1. Anagnostou E, Taylor MJ. Review of neuroimaging in autism spectrum disorders: what have we learned and where we go from here. *Mol. Autism*. 2011; 2:1–9. [PubMed: 21247446]
2. Greicius M. Resting-state functional connectivity in neuropsychiatric disorders. *Curr. Opin. Neurol*. 2008; 21:424–430. [PubMed: 18607202]
3. Zhang H, Chen X, Shi F, et al. Topographical information-based high-order functional connectivity and its application in abnormality detection for mild cognitive impairment. *J. Alzheimer's Dis*. 2016; 54:1–18. [PubMed: 27567881]
4. Wiggins JL, Peltier SJ, Ashinoff S, et al. Using a self-organizing map algorithm to detect age-related changes in functional connectivity during rest in autism spectrum disorders. *Brain Res*. 2011; 1380:187–197. [PubMed: 21047495]
5. Wiggins JL, Bedoyan JK, Peltier SJ, 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*. 2012; 59:2760–2770. [PubMed: 22032950]

6. Alaerts K, Swinnen SP, Wenderoth N. Sex differences in autism: a resting-state fMRI investigation of functional brain connectivity in males and females. *Soc. Cogn. Affect. Neurosci.* 2016; 11:1002–1016. [PubMed: 26989195]
7. Sun S. A survey of multi-view machine learning. *Neural Comput. Appl.* 2013; 23:2031–2038.
8. Jiang Y, Chung F-L, Wang S, et al. Collaborative fuzzy clustering from multiple weighted views. *IEEE Trans. Cybern.* 2015; 45:688–701. [PubMed: 25069132]
9. Zhang, J., Huan, J. Inductive multi-task learning with multiple view data; Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2012. p. 543-551.
10. He, J., Lawrence, R. A graph-based framework for multi-task multi-view learning; Proceedings of the 28th International Conference on Machine Learning (ICML 2011); 2011. p. 25-32.
11. Nie F, Huang H, et al. Efficient and robust feature selection via joint ℓ_2, ℓ_1 -norms minimization. *Advances in Neural Information Processing Systems.* 2010:1813–1821.
12. Chang, C-C., Lin, C-J. LIBSVM: a library for support vector machine, 2001. 2012. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
13. Zhang D, Wang Y, Zhou L, et al. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage.* 2011; 55:856–867. [PubMed: 21236349]

Table 1

Summary of the methods under comparison.

Method	Description on the method	Data preparation
CSVC-S	C-SVC in LibSVM. Linear kernel is adopted	FC and HOFc features of each subject are concatenated
CSVC-J		
M2SVC-S	Multi-modal classification proposed by [13]	FC and HOFc features are regarded as one modality in M2SVC
M2SVC-J		
regMVMT	The regularized MVMT method proposed by [9]	FC and HOFc features are regarded as one view, respectively, and each group is regarded as one task
Sparse-MVTC	Our proposed method	

(1) CSVC-S and M2SVC-S: All groups are trained and tested separately. That is, there are T generated classifiers, each of which corresponds to a task.

(2) CSVC-J and M2SVC-J: All groups are combined together for joint training and testing. There is only a single classifier that is applied to all T groups.

Table 2

Performances of the nested 10-fold classifications for different methods.

	ACC	SEN	SPE	AUC	<i>p</i> -value
CSVC-S	0.588	0.729	0.402	0.566	<1e−3
CSVC-J	0.676	0.775	0.544	0.66	0.045
M2SVC-S	0.599	0.716	0.443	0.58	<1e−3
M2SVC-J	0.656	0.719	0.571	0.645	0.030
regMVMTC	0.594	0.644	0.527	0.585	0.003
Sparse-MVTC	0.688	0.736	0.625	0.680	