Article:
Lei, B, Cheng, N, Frangi, AF orcid.org/0000-0002-2675-528X et al. (7 more authors) (2020) Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease. Medical Image Analysis, 61. 101652. ISSN 1361-8415

https://doi.org/10.1016/j.media.2020.101652

Self-calibrated Brain Network Estimation and Joint Non-Convex Multi-Task Learning for Identification of Early Alzheimer's Disease

Baiying Lei[a], Nina Cheng[a], Alejandro F Frangi[b], Ee-Leng Tan[c], Jiuwen Cao[d], Peng Yang[a], Ahmed Elazab[a,e], Jie Du[a], Yanwu Xu[e*], Tianfu Wang[a*]

[a] *National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China,518055 Tel: 86-755-86713997 (Correspondence Tianfu Wang:email:tfwang@szu.edu.cn).*

[b]*CISTIB Centre for Computational Imaging & Simulation Technologies in Biomedicine, School of Computing and the School of Medicine, University of Leeds, Leeds, LS2 9JT Leeds, United Kingdom.*

[c]*School of Electric and Electronic Engineering, Nanyang Technological University, Singapore, 639798*

[d]*Artificial Intelligence Institute, Hangzhou Dianzi University, Zhejiang, China, 310010.*

[e]*Computer Science Department, Misr Higher Institute of Commerce and Computers, Mansoura, Egypt, 35516*

[e] *Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo, China, 530031 (Correspondence Yanwu Xu:email: ywxu@nimte.ac.cn).*

Abstract Detection of early stages of Alzheimer's disease (AD) (i.e., mild cognitive impairment (MCI)) is important to maximize the chances to delay or prevent progression to AD. Brain connectivity networks inferred from medical imaging data have been commonly used to distinguish MCI patients from normal controls (NC). However, existing methods still suffer from limited performance, and classification remains mainly based on single modality data. This paper proposes a new model to automatically diagnosing MCI (early MCI (EMCI) and late MCI (LMCI)) and its earlier stages (i.e., significant memory concern (SMC)) by combining low-rank self-calibrated functional brain networks and structural brain networks for joint multi-task learning. Specifically, we first develop a new functional brain network estimation method. We introduce data quality indicators for self-calibration, which can improve data quality while completing brain network estimation, and perform correlation analysis combined with low-rank structure. Second, functional and structural connected neuroimaging patterns are integrated into our multi-task learning model to select discriminative and informative features for fine MCI analysis. Different modalities are best suited to undertake distinct classification tasks, and similarities and differences among multiple tasks are best determined through joint learning to determine most discriminative features. The learning process is completed by non-convex regularizer, which effectively reduces the penalty bias of trace norm and approximates the original rank minimization problem. Finally, the most relevant disease features classified using a support vector machine (SVM) for MCI identification. Experimental results show that our method achieves promising performance with high classification accuracy and can effectively discriminate between different sub-stages of MCI.

Keywords: Early stage of Alzheimer's disease (AD); Brain network estimation; Self-calibration; Multi-modal classification; Joint non-convex multi-task learning

---

1 Introduction

AD is a common chronic neurodegenerative disease and the only disease in the world's ten most deadly diseases that cannot be prevented and cured (2018). AD accounts for about 60-70% of the causes for dementia. AD not only brings suffering to patients, severely degrading their life quality, but also puts heavy mental and economic stress on the family and society. Therefore, detection of the early stage of AD and then intervention or even prevention for AD show important significance for the patients, family and society. According to some research, SMC and MCI may be the stages between normal aging and AD and have attracted much interest due to their close connection with the development of AD (Jessen et al., 2014; Lei et al., 2017; Lian et al., 2018; Peng et al., 2019; Shi et al., 2017; Zhou et al., 2018a; Zhou et al., 2019a; Zhou et al., 2019b; Zhu et al., 2014). Thus, developing methods to detect early stages of AD such as SMC and MCI is very necessary. There is a growing corpus of work on automatic diagnosis of MCI: the earlier the diagnosis of MCI, the more meaningful is to understand the affected brain connections to target prevention or slowing down disease progression. Hence, this paper studies the early AD stages, including SMC, EMCI and LMCI.

Although SMC and MCI present mild symptoms, making it more difficult for the detection, many neuroimaging-based techniques such as magnetic resonance imaging (MRI) (Davatzikos et al., 2011), diffusion tensor imaging (DTI) (Cherubini et al., 2010), and resting state fMRI (rs-fMRI) (Teipel et al., 2017), have proven to be powerful tools for the classification of AD and its early stages (Gao et al., 2015; Lei et al., 2017; Wee et al., 2012; Yang et al., 2016; Zhu et al., 2018). Different neuroimaging techniques can obtain different brain information, such as structural and functional information. Among these imaging techniques, DTI provides local microscopic features of water diffusion, while rs-fMRI characterizes fluctuations in spontaneous blood oxygen level signals associated with disease-specific neural circuits, which can be used to assess structural separation and functionally specific brain-linked networks (Chen et al., 2011a; Lian et al., 2018). Often, it is advantageous to combine the information

obtained from multiple technologies to give more accurate and objective results. A large number of previous studies have shown that integrating multi-modal information can effectively improve classification performance (Shi et al., 2017; Zhang et al., 2011). Therefore, the method based on multi-modal data is proposed for detection of the early stages of AD, including SMC, MCI, which contains EMCI and LMCI.

There are a large number of studies on automatic diagnosis of MCI. For instance, Wee *et al.* (Wee et al., 2014a) proposed a group-constrained sparse network connectivity model and then used the minimum redundancy and maximum relevance methods (mRMR) for feature selection for the MCI identification. Such a group constrained sparse network can simultaneously make a common region of interest (ROI) selection across subjects and optimally estimate the ROI time series under consideration. Li *et al.* (Li et al., 2018) used a new sparse constrained connectivity inference method and an elastic multilayer perceptron classifier (MPC) to identify MCI. The topology of the network connection is effectively identified by considering the weak derivative information of the data. Zhou *et al.* (Zhou et al., 2018b) proposed a high-order functional brain network (FBN) connection with improved sparsity and modularity for MCI classification. Li *et al.* (Li et al., 2019) proposed a multi-modal super-connectivity functional network using functionally weighted LASSO for MCI identification. Both bold fMRI and ASL fMRI's hyper-connected network are integrated to simultaneously evaluate brain networks with high temporal resolution and low temporal resolution fMRI. These studies are devoted to the automated diagnosis of MCI, but in the estimation of brain networks, the effects of noise and artifacts are not considered, which may degrade classification performance. Most functional brain network estimation methods can be interpreted under the regularization framework (Qiao et al., 2016; Wee et al., 2014b), but such a framework requires an accurate estimation model to effectively encode brain prior information while fitting data. The actual observation data is often affected by the noise of complex interference sources. Although the data is preprocessed, it is difficult to eliminate all noise, and new noise may be introduced in the preprocessing stage. Therefore, it is very important to improve the data quality before brain network estimation, which directly affects the accuracy of brain network estimation.

In this paper, we propose a framework for detection of MCI and its early stage. We first develop a new functional brain network estimation method, then introduce data quality indicators for self-calibration, which can improve data quality while completing brain network estimation, and perform correlation analysis combined with low-rank module structure. Second, functional and structural connected neuroimaging patterns are integrated into our multi-task learning model to select discriminative and informative features for MCI analysis. Different modalities are best suited to undertake different tasks, and similarities and differences among multiple tasks are determined through joint learning to determine the most discriminative features. The learning process is completed by non-convex regularization, which effectively reduces the penalty bias of trace norm and approximates the original rank minimization problem. Finally, the most relevant features of the disease are sent to the SVM classifier for MCI identification.

In our experiments, different modalities perform different tasks but also contribute any shared information with other modalities. By exploring the relationships among different tasks, we select the most discriminative features. In details, the feature selection process obtains weights through a multi-task learning method and then selects the most relevant discriminative features according to the weights. To reduce the penalty deviation of the tracking norm in the learning process, we use the joint non-convex regularization method to learn the subspace. Finally, the selected features are imported into the SVM model for early diagnosis. The contributions and main advantages of the proposed framework are as follows.

1) A network construction method is proposed, which encodes the brain modular structure based on low rank constraints, and automatically calibrates the data quality to remove low quality data points and noise.

2) A feature selection method based on multi-modal data is proposed. By using the multi-task learning framework, the most discriminative features are selected in the structural and functional information.

3) A joint non-convex regularizer is proposed for subspace learning, which learns the potential related information between multi-tasks and effectively approximates the original rank minimization

problem.

To evaluate the performance of our method, we validate our method using the Alzheimer's disease Neuroimaging Initiative (ADNI) (https://ida.loni.usc.edu/) public database (Jack Jr et al., 2008). The proposed framework is validated via a leave one out (LOO) strategy (Cawley and Talbot, 2003) using rs-fMRI and DTI data from 170 subjects, including NC, SMC, EMCI, and LMCI. The experimental results show that the model has higher classification accuracy for the diagnosis of MCI and SMC. Compared with the existing methods, this model has better performance, which detects brain abnormalities more accurately and better pictures the pathological abnormalities of patients.

The rest of this paper is organized as follows. In the second section, we introduce the process of data acquisition and preprocessing. A detailed description of the proposed method is provided in the third section. Details of experimental results and comparisons are given in the fourth section. Finally, in the discussion and conclusion sections, we discuss the strengths and limitations of our work, and future research directions.

2 Methodology

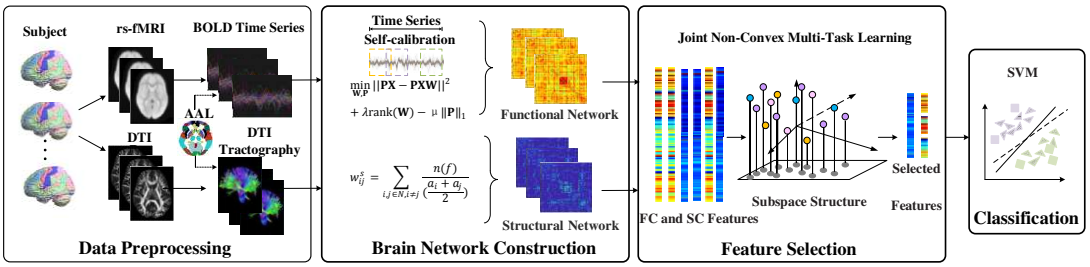

Figure 1: Illustration of the proposed method.

The flowchart of the proposed method is shown in Figure 1. First, the multi-modal image data (rs-fMRI and DTI) are preprocessed via automatic anatomical landmark (AAL) template. Second, we construct the brain function network using our proposed self-calibration method via low-rank learning. Third, the functional and structural networks are extracted as inputs, and the joint non-convex multi-task learning model is used for feature selection. Finally, the selected features are sent into the SVM for classification. In the following, we will describe the proposed method.

2.1 Notations

In this study, a bold uppercase letter is used to represent a matrix, bold lowercase letters are used to represent vectors, and ordinary italic letters to indicate scalars. For a matrix $\mathbf{X}$, where $\boldsymbol{x}^i$ and $\boldsymbol{x}_j$ denote its i-th row and j-th column, respectively. The norm of vector is defined as $\|\mathbf{x}\|$. $\mathbf{X}^{\mathrm{T}}$, rank($\mathbf{X}$) and $\mathbf{X}^{-1}$ denotes the transpose operator, the order and inverse of the matrix $\mathbf{X}$, respectively. We summarize all the notations used in this paper in Table 1.

Table 1: The summarization of notations.

Notation	Description
d	Feature dimension of data
K	The number of tasks, in this case $K = 2$ (two modalities)
n_i	Number of subjects in task i
$\mathbf{X} \in \mathbf{R}^{n \times d}$	Size of $n \times d$ matrix
$\|\mathbf{x}\|$	The norm of vector $\mathbf{x}$
x^i, x_j	i-th row and j-th column of $\mathbf{X}$
$\sigma_i(\mathbf{X}) = \sigma_i$	The i-th singular value of $\mathbf{X}$
$\lambda_i(\mathbf{X}) = \lambda_i$	The i-th eigenvalue value of $\mathbf{X}$
$\mathbf{u}_i(\mathbf{X}) = \mathbf{u}_i$	The i-th left singular vector of $\mathbf{X}$
$\mathbf{v}_i(\mathbf{X}) = \mathbf{v}_i$	The i-th right singular vector of $\mathbf{X}$
$\big\|\|\mathbf{x}\|\big\|_1 = \sum_{i=1}^{n} \|\mathrm{x}_i\|$	ℓ_1-norm of vector $\mathbf{x}$
$\|\|\mathbf{x}\|\|_2 = \sqrt{\sum_{i=1}^{n} \mathrm{x}_i^2}$	ℓ_2-norm of vector $\mathbf{x}$
$\|\|\mathbf{x}\|\|_2^2 = \sum_{i=1}^{n} \mathrm{x}_i^2$	ℓ_2-norm of vector $\mathbf{x}$ squared
$\|\|\mathbf{X}\|\|_* = \sum_{i=1} \sigma_i(\mathbf{X})$	Nuclear norm of $\mathbf{X}$
$\|\|\mathbf{X}\|\|_{\mathrm{F}} = \sqrt{\sum_i \|\|\mathbf{x}_i\|\|_2^2}$	Frobenius norm of $\mathbf{X}$
$\|\|\mathbf{X}\|\|_{\mathrm{F}}^2 = \mathbf{Tr}(\mathbf{X}^{\mathrm{T}}\mathbf{X})$	Frobenius norm of $\mathbf{X}$ squared

2.2 Brain network estimation

There are many methods (i.e., Pearson's correlation (PC) and sparse representation (SR)) for FBN estimation, most of which can be explained in the regularization framework including data fitting terms and regularization terms. Commonly used regularization terms include sparsity (Lee et al., 2011), group sparsity (Wee et al., 2014a) and low rank (Liu et al., 2013), which can encode brain organization

and improve FBN estimation results. The data fitting term is also very important for the estimation of FBN, because the noise is difficult to eliminate in the preprocessing and the noise caused by the preprocessing process itself will also affect the data quality, resulting in poor data fitting results. Here, we review several commonly used brain network construction methods, which will be compared with our proposed methods in our experiments.

2.2.1 Pearson's correlation (PC)

After data preprocessing, each subject's brain is divided into *N* ROI based on a template (commonly used AAL template). Each ROI corresponds to a time series $\mathbf{x}_i \in \mathbf{R}^m, i = 1, 2, \dots, N$, where m represents the number of time points in each time series. Then, the edge weights matrix is defined as follows:

$$\mathbf{W}_{ij}^{(PC)} = \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^{\mathrm{T}}(\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^{\mathrm{T}}(\mathbf{x}_i - \bar{\mathbf{x}}_i)}\sqrt{(\mathbf{x}_j - \bar{\mathbf{x}}_j)^{\mathrm{T}}(\mathbf{x}_j - \bar{\mathbf{x}}_j)}}. \quad (1)$$

After centralizing and normalizing the original $\mathbf{x}_i$ with $\mathbf{x}_i - \bar{\mathbf{x}}_i$ and $\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^{\mathrm{T}}(\mathbf{x}_i - \bar{\mathbf{x}}_i)}$, Eq.(1) can be simply expressed as $\mathbf{W}_{ij}^{(PC)} = \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j$. It is can be solved via the following optimization problem:

$$\min_{\mathbf{W}} ||\mathbf{W} - \mathbf{X}^{\mathrm{T}}\mathbf{X}||_{\mathrm{F}}^2, \quad (2)$$

where $\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_N]^{\mathbf{T}} \in \mathbf{R}^{m \times N}$ represents the data matrix, and $\mathbf{W} \in \mathbf{R}^{N \times N}$ represents the edge weight matrix.

2.2.2 Partial correlation and sparse representation (SR)

PC is the simplest completely correlated modeling method, which cannot exclude the hybrid problems caused by other brain regions. Therefore, some relevant methods are proposed to overcome the influence of confounding effect. Estimation based on the inverse covariance matrix has been widely used to calculate the partial correlation (Meinshausen and Bühlmann, 2006). However, the estimation is not appropriate when the number of network nodes is larger than the number of time point *m*. Therefore, the ℓ_1 regularization is introduced into the model. Here, we describe the SR because it is one of the methods compared with our method, which is denoted as:

$$\min_{\mathbf{W}} \sum_{i,j=1}^{N} \left(\left\| \mathbf{x}_i - \sum_{j \neq i} \mathbf{W}_{ij}\mathbf{x}_j \right\|^2 + \lambda \sum_{j \neq i} |\mathbf{W}_{ij}| \right). \quad (3)$$

Eq. (3) can be simplified into the following matrix form:

$$\min_{\mathbf{W}} ||\mathbf{X} - \mathbf{XW}||_{\mathrm{F}}^2 + \lambda||\mathbf{W}||_1. \quad (4)$$

Other regularizers can also be used in estimation models, for space limitation, we won't discuss them here. We can simply define the FBN of SR as: $\mathbf{W}^{(\mathrm{SR})} = \frac{\mathbf{W}^* + \mathbf{W}^{*\mathrm{T}}}{2}$.

2.2.3 Matrix-regularized FBN Estimation Framework

Unlike estimates of PC that lack reasonable biological significance, the matrix regularization framework realizes the effective fitting of data and the effective coding of prior knowledge. Many FBN estimation models can be expressed by the regularization framework, which is defined as follows:

$$\min_{\mathbf{W}} f(\mathbf{X}, \mathbf{W}) + \lambda \mathrm{R}(\mathbf{W}), \quad (5)$$

where $f(\mathbf{X}, \mathbf{W})$ are data fitting terms for capturing statistics of data and $\mathrm{R}(\mathbf{W})$ are matrix regularization terms for encoding a prior information.

2.2.4 The proposed method

The fitting of the data and the encoding of the prior information are very important. The quality of the data has a great impact on the estimation of the FBN. Therefore, the learning task needs to measure the data quality. In this paper, we introduce a parameter p_t to control data quality under the regularization framework, which improves data quality while estimating FBN, and completes self-calibrated FBN estimation. The conventional method of encoding prior information through a regularization term cannot control the quality of the rs-fMRI sequence time point. To this end, we propose a new method for controlling the quality of time points during FBN estimation, which eliminates the omission noise of the traditional preprocessing pipeline and the noise generated by the preprocessing itself, without excessive deletion. Based on the matrix regularization network estimation framework, we propose a new FBN estimation model that can estimate FBN more accurately through combining low rank and self-calibration.

Let $\mathbf{X} \in \mathbf{R}^{T \times N}$ and $\mathbf{P} \in \mathbf{R}^{T \times T}$ denote the data matrix and a diagonal matrix of N ROI and T time points in each series, respectively. Our model is:

$$\min_{\mathbf{W},p_t\in[0,1]} \sum_{i=1}^{N}\sum_{t=1}^{T} p_t f\left(\mathbf{X}^{(t)},\mathbf{W}\right) + \lambda \mathrm{rank}(\mathbf{W}) - \mu \sum_{t=1}^{T} |p_t| \quad (6)$$

where $\mathbf{X} = [\mathbf{x}_1;\ \mathbf{x}_2;\ \ldots;\ \mathbf{x}_N]^{\mathrm{T}}$, $\mathbf{P} = \mathrm{diag}(p_1, p_2, \ldots, p_T)$, $p_1, p_2, \ldots, p_T$ are the indicators, $p_t \in [0,1]$. When $p_t = 0$, the *t*-th time point will be discarded and is not used for the estimation of FBN. If p_t is equal to 1 at all times, Eq. (6) is reduced to Eq. (5). **W** is the edge weight matrix while λ and μ are two regularization parameters in the objective function. Most of the time points will be discarded when μ has a small value close to 0. Conversely, when the value of μ is large, most of the time points will be retained. In other words, by controlling parameter μ, the model can learn indicator p_t, measure data quality, remove poor quality data, and complete self- calibrated FBN estimation.

Here, we use partial correlation method for data fitting, which is defined as

$$\min_{\mathbf{W},p_t\in[0,1]} \sum_{i=1}^{N}\sum_{t=1}^{T} p_t(\mathbf{x}_i^{(t)} - \sum_{i\neq j}^{n} \mathrm{W}_{ij}\mathbf{x}_j^{(t)})^2 + \lambda \mathrm{rank}(\mathbf{W}) - \mu \sum_{t=1}^{T}|p_t|, \quad (7)$$

where $\mathbf{x}_i^{(t)}$is the time point *t* corresponding to the *i*-th ROI. The abbreviated Eq. (7) becomes:

$$min_{\mathbf{W},\mathbf{P}}\ ||\mathbf{PX} - \mathbf{PXW}||^2 + \lambda \mathrm{rank}(\mathbf{W}) - \mu\|\mathbf{P}\|_1. \quad (8)$$

The low-rank constraint incorporates prior information, which makes the brain network have a modular structure and conform to the inherent structural FBN features. By controlling the parameter μ in the last term of the objective function, the indicator is learned from the data and then the quality of the data is controlled for self-calibration. The algorithm for solving Eq. (8) is summarized in Algorithm 1. The network construction procedure is graphically illustrated in Figure 2.

Algorithm 1 FBN estimation via self-calibration low rank regularization

Input: **X**, λ, μ
Output: **W**, **P**

1. Initialize **P**
2. while *not converged*
   Repeat
   $\mathbf{W} \leftarrow \mathbf{W} - \alpha(-2\mathbf{X}^{\mathrm{T}}\mathbf{P}^{\mathbf{T}}\mathbf{PX} + 2\mathbf{X}^{\mathrm{T}}\mathbf{P}^{\mathbf{T}}\mathbf{PXW})$
   $\mathbf{W} \leftarrow \mathrm{prox}_{\lambda||\cdot||_1}(\mathbf{W}) = \mathrm{Udiag}(\max\{\sigma_1 - \lambda_2, 0\}, \ldots, \max\{\sigma_n - \lambda_2, 0\})\mathbf{V}^{\mathrm{T}}$
   end
3. $p_t = \begin{cases} 1, & ||\mathbf{X}^{(t)} - \mathbf{X}^{(t)}\mathbf{W}||^2 < \mu \\ 0, & otherwise \end{cases}$
4. end
5. return **W**, **P**

We initialize **P** to an identity matrix. For the two variables **P** and **W** in Eq. (8), we use the alternating convex search (ACS) (Sherali and Shetty, 2006) to complete. First, fix **P**, and we use the proximal

approach for optimization. According to the gradient descent criterion, the updating formula of data fitting term is: $\mathbf{W}_{k+1} = \mathbf{W}_k - \alpha_k \nabla_{\mathbf{W}} f(\mathbf{X}, \mathbf{W}_k)$, $\nabla_{\mathbf{W}} f(\mathbf{X}, \mathbf{W}) = 2\mathbf{X}^{\mathrm{T}}\mathbf{P}^{\mathbf{T}}\mathbf{P}\mathbf{X} - 2\mathbf{X}^{\mathrm{T}}\mathbf{P}^{\mathbf{T}}\mathbf{P}\mathbf{X}\mathbf{W}$. The proximal operator is defined:

$$\mathrm{prox}_{\lambda||\cdot||_1}(\mathbf{W}) = \mathbf{U}\mathrm{diag}(\max\{\sigma_1 - \lambda_2, 0\}, \dots, \max\{\sigma_n - \lambda_2, 0\})\mathbf{V}^{\mathrm{T}} \tag{9}$$

where $\mathbf{U}\mathrm{diag}(\sigma_1, \dots, \sigma_n)\mathbf{V}^{\mathrm{T}}$ is the singular value decomposition (SVD) of matrix $\mathbf{W}$.

After that, we fix $\mathbf{W}$ and update $\mathbf{P}$. This is a linear programming problem that is easy to get the optimal solution. If $||\mathbf{X}^{(t)} - \mathbf{X}^{(t)}\mathbf{W}||^2 < \mu$, the t-th time point will be labeled as 1 and will be retained, otherwise, the time-point will be labeled as 0.

For the construction of the DTI structural brain network, the AAL template aligned to each subject's DTI fiber bundle imaging was converted to a common space as a node. The PANDA is used to calculate the ROI and the structural connection between them as the total number of standardized connected fibers between the ROI pairs to construct a DTI network. The connection of white matter fiber bundles between each pair of brain regions is considered the edge of the brain structure network. The average FA of the connections between network nodes is defined as the connection weights in the DTI network. Finally, a 90 by 90 weighting matrix is obtained.

$$w_{ij}^{s} = \sum_{i,j \in N, i \neq j} \frac{n(f)}{\left(\frac{a_i + a_j}{2}\right)}, \tag{10}$$

where $i, j \in N = \{1, 2, \dots, 90\}$ and $i \neq j$, $n(f)$ is the total number of fibers linking ROIs i and j, a_i is the surface area of ROI i between gray matter and white matter. $\frac{a_i + a_j}{2}$ is used to correct deviations in structural connection strength estimates caused by different ROI sizes.

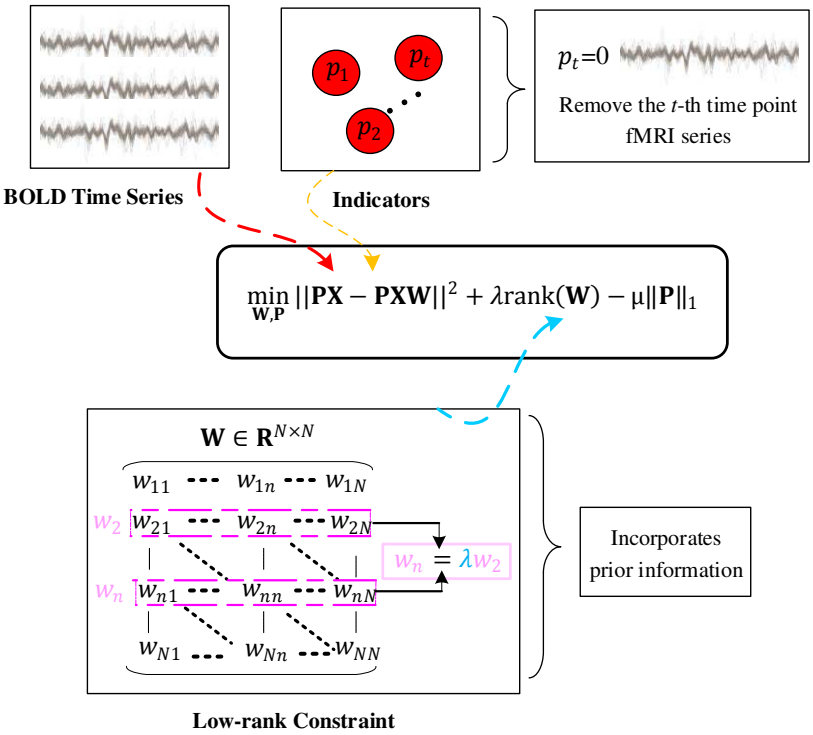

Figure 2: Illustration of the FBN construction procedure.

2.3 Joint non-convex multi-task learning

After completing the functional and structural brain network estimation, we treat different modal information as different tasks and multi-task learning method is used for feature selection. Specifically, we design a joint non-convex multi-task learning model, which can automatically learn the relationship between different modalities, reduce the influence of different noise levels of different modal information, and obtain the most stable and discriminative features. Multi-modal data have similarities and differences, the fusion of multi-modal information can enhance the expression of complementary information and useful features, improve generalization performance and classification performance. Therefore, we learn multiple modal data simultaneously in the framework and expect these tasks to work together to obtain useful features, which cannot be done in a single task model.

We use a low-rank regularizer to perform subspace learning of samples to learn about potentially relevant information between different tasks. To facilitate optimization, the current commonly used method is to treat the rank norm as a trace norm. But the difference between the two norms and the

penalty bias caused by the trace norm will cause the algorithm to degenerate, and the low-rank structure of coefficient matrix cannot be learned well. We propose to use a joint non-convex regularizer for multi-task learning, the original rank minimization problem is approximated by non-convex regularizer.

$\mathbf{X}_i \in \mathbb{R}^{n_i \times d}$ is the training set in K tasks, $\mathbf{y}_i \in \mathbb{R}^{n_i \times 1}$ is its associated label vector. In multi-task learning, the coefficient vector $\mathbf{w}_i \in \mathbb{R}^{d \times 1}$ in each task can be jointly learned, and all coefficient vectors form a coefficient matrix $\mathbf{W} \in \mathbb{R}^{d \times K}$. In this case, n and d represent the number of subjects and feature dimensions, respectively. $\mathbf{Y} \in \mathbb{R}^{n_i \times K}$ is a matrix of all task labels, the least squares loss function is used to measure the relationship between X and Y. $\sigma_i(\mathbf{W})$ is the i-th singular value of $\mathbf{W}$, γ is a parameter balancing two terms. Our objective function is:

$$\min_{\mathbf{W}} \sum_{i=1}^{K} ||\mathbf{X}_i \mathbf{w}_i - \mathbf{y}_i||_2^2 + \gamma \sum_{i=1}^{d} \log(\sigma_i(\mathbf{W}) + 1). \tag{11}$$

Here, $\sum_{i=1}^{d} \log(\sigma_i(\mathbf{W}) + 1)$ is a new regularization, it is between the ℓ_0-norm and the ℓ_1-norm of $\sigma(\mathbf{W})$, which can reduce the relaxation problem and the minimization problem of the original rank.

Regarding the optimization of the objective function, we define a new function: $\mathrm{J}(\mathbf{W}) = \mathbf{W}\mathbf{W}^{\mathrm{T}}$.Then the function can be reformulated as:

$$\min_{\mathbf{W}} \sum_{i=1}^{K} ||\mathbf{X}_i \mathbf{w}_i - \mathbf{y}_i||_2^2 + \gamma \sum_{i=1}^{d} r\left(\lambda_i\left(\mathrm{J}(\mathbf{W})\right)\right), \tag{12}$$

where $r(\mathrm{x}) = \log(x^{\frac{1}{2}} + 1)$, λ_i is the eigenvalue of $\mathrm{J}(\mathbf{W})$. Further, it can be written as:

$$\min_{\mathbf{W}} \mathrm{P}(\mathbf{W}) = \mathrm{L}(\mathbf{W}) + \gamma \mathrm{R}(\mathbf{W}), \tag{13}$$

where $\mathrm{L}(\mathbf{W}) = \sum_{i=1}^{K} ||\mathbf{X}\mathbf{W} - \mathbf{Y}||_2^2$, $\mathrm{R}(\mathbf{W}) = \sum_{i=1}^{d} r(\lambda_i(\mathrm{J}(\mathbf{W})))$. Supposing $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$ is the eigenvalue decomposition of $\mathrm{J}(\mathbf{W})$, the derivative of $\mathrm{R}(\mathbf{W})$ is: $\frac{\partial R(\mathbf{W})}{\partial \mathbf{W}} = 2\mathbf{D}^{\mathrm{T}}\mathbf{W}$, where $\mathbf{D} = \sum_{i=1} r'(\lambda_i)\mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$ is the weighted matrix. For $\mathrm{L}(\mathbf{W})$, $\frac{\partial \mathrm{L}(\mathbf{W})}{\partial \mathbf{W}} = 2\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{W} - \mathbf{Y})$, combining these two terms, we have:

$$\sum_{i=1}^{K} (2\mathbf{X}_i^{\mathrm{T}}(\mathbf{X}_i \mathbf{w}_i - \mathbf{y}_i) + 2\gamma \mathbf{D}^{\mathrm{T}} \boldsymbol{w}_i) = 0. \tag{14}$$

We can obtain the optimal solution of $\boldsymbol{w}_i$ via solving the following problem:

$$\mathbf{X}_i^{\mathrm{T}}(\mathbf{X}_i \mathbf{w}_i - \mathbf{y}_i) + \gamma \mathbf{D}^{\mathrm{T}} \boldsymbol{w}_i = 0. \tag{15}$$

The calculation of **D** depends on the variable **W**, and **D** can be calculated when **W** is fixed. Each column of **W** is computed by solving Eq. (15). Hence, we utilize an iterative method to find the solution of Eq. (12). In $(t+1)$-th iteration, updating **W** equals solve the following problem:

$$\min_{\mathbf{W}} \mathrm{F}(\mathbf{W}) = \sum_{i=1}^{K} ||\mathbf{X}_i\mathbf{w}_i - \mathbf{y}_i||_2^2 + \gamma \mathrm{Tr}((\mathbf{D}^t)^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}). \tag{16}$$

The entire iterative process is outlined in Algorithm 2. Figure 3 is a one-dimensional illustration of our proposed new regularization term, it is closer to the ℓ_0-norm than the ℓ_1-norm. Therefore, better sample subspace learning can be carried out in the low-rank feature selection process.

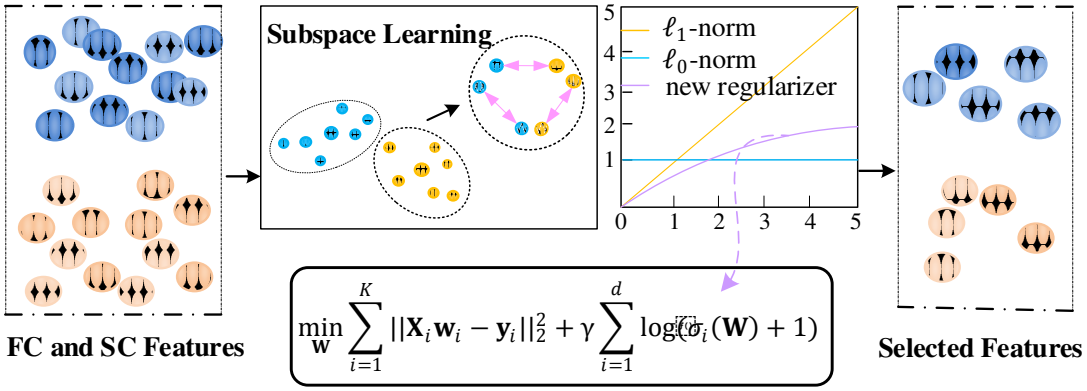

Figure 3: Illustration of the feature selection procedure.

Algorithm 2

Input: **X**, **Y**

Output: **W**

Initialize $\{\mathbf{w}_i\}_1^K = \{\mathbf{X}_i^T\mathbf{y}_i\}_1^K$

while *not converged*

1. Eigenvalue decomposition: $\mathrm{h}(\mathbf{W}) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathbf{T}}$
2. Update $\mathbf{D} = \sum_{i=1} r'(\lambda_i)\mathbf{u}_i\mathbf{u}_i^T$
3. Update **W** by solving the Eq. (16)
4. Until convergence, where the convergence condition is:

$$\frac{|F(\mathbf{W}_t) - F(\mathbf{W}_{t+1})|}{|F(\mathbf{W}_t)|} < 1e-6$$

end

2.4 Classification

We do not perform classifier research in this paper, which is beyond the scope of our discussion. Therefore, we use the simplest classifier with linear SVM (c=1) with default parameters for classifica-

tion. Complex classification pipelines may confuse the validation of the method, so a simple SVM is used to complete the classification in all our experiments.

3 Experiments

3.1 Dataset and pre-processing

The data source in this article is the ADNI public database. Both rs-fMRI and DTI data are collected from the 3.0T Siemens MRI scanner. For rs-fMRI images, the data acquisition parameters are set as follows: Pulse Sequence = EP; imaging matrix =64×64 with 48 slices; 180 volumes; voxel thickness c= 3.3 mm; flip angle=9 degree; TE=30 ms; TR=3000 ms. For DTI images, the data acquisition parameters are set as follows: b=0 and 1000 s/mm^2; Pulse Sequence=EP; imaging matrix =64×64 with 48 slices; 180 volumes; gradient directions =54; voxel thickness=3.3 mm; flip angle=90 degree; TE=56 ms; TR=7200 ms. During the scan process, all subjects were instructed to open their eyes and stare at a fixed cross in the middle of the screen to prevent them from falling asleep and avoiding the saccade-related activation caused by closing their eyes.

126 MCI patients are obtained from the ADNI dataset, including 44 EMCI (22M/22F) and 38 LMCI (19M/19F), 44 SMC patients (17M/27F), and 44 demographically matched NCs (22M/22F). The total number of subjects was 170. Table 2 summarizes the demographic data of each group.

Table 2: Demographic details of the used database.

Group	**NC(44)**	**SMC(44)**	**EMCI(44)**	**LMCI(38)**
Male/Female	22M/22F	17M/27F	22M/22F	19M/19F
Age(mean±SD)	76.4±4.5	76.3±5.4	76.5±6.1	76.0±7.7

For rs-fMRI data preprocessing, we use the widely used rs-fMRI data analysis software: statistical parameter mapping (SPM12) software package (Ashburner et al., 2014), rs-fMRI data processing assistant (DPARSF) toolbox (Yan and Zang, 2010) and analysis toolkit (REST). The specific processing process is as follows: In the first step, to maintain magnetization balance, we discard the data of the first 10 time points of each subject. The second step is hierarchical correction, which corrects the scanning layer section at all remaining time points to the intermediate layer, to solve the hierarchical disorder caused by the staggered sequence of odd layers in the data collection process. In the third step, head

movement correction was carried out to align the data at all time points of each subject with the data at the first time point, to remove the head movement artifact and solve the possible influence caused by each subject shaking their head during the scanning process. The fourth step is to register and register all the data of the subjects to the standard template. Here, the standard template of the Montreal Neurological Institute (MNI) is used, and the downsampled voxel size is $3\times3\times3mm^3$. The fifth step is spatial smoothing. A 4 mm full width half-maximum Gaussian kernel is adopted for smoothing, to eliminate noise interference. The sixth step performs a regression covariate on the smoothed data, i.e., removes the interfering signal from the time series of each voxel to reduce the effects of non-neuronal fluctuations, including white matter signals, cerebrospinal fluid signals, and head motion signals. The seventh step of time filtering uses a band pass filter (0.01 Hz $\leq$ f $\leq$0.08 Hz) to filter the time series, to minimize the effects of low-frequency drift and high-frequency noise. Then, by aligning the AAL map (Craddock et al., 2012) with the rs-fMRI image, the brain space is divided into 90 ROIs (excluding 26 ROIs from the cerebellum). Finally, we obtain the time series of each ROI of each subject through the time series of all voxels in that specific ROI.

Preprocessing of DTI images uses the PANDA toolbox (Cui et al., 2013) based on the FMRIB software library (FSL, https://fsl.fmrib.ox.ac.uk/fsl/), which generates global brain determinism fiber bundle imaging by using the FACT algorithm (Kamal and Burk, 1996) with default parameters. First use the bet command to perform brain tissue extraction and eddy distortion correction, then use the dtifit command to calculate the diffusion tensor, and use the fact command in the FSL to perform deterministic beam raying, thus in the assumed white matter tissue (fractional anisotropy (FA) > 0.2) generate all possible fibers with an angle threshold = 45° and two seeds per voxel. Each subject's T1 image is first co-registered to its respective T2-weighted image and then spatially registered to the standard MNI space as the same as used for rs-fMRI registration. The resulting deformation field is applied to map the brain parcellation atlas from the MNI space to the native space of each individual. Finally, the brain space is divided into 90 ROIs by aligning AAL to each image, where each ROI represents a network node. Deterministic fiber bundle imaging is then performed to generate all possible fiber bundles with-

in the white matter tissue. The average FA of links between network nodes is defined as the connection weight in the DTI network.

3.2 Experimental setup

The data imaging matrix used in our experiment is 64×64 with 48 slices, details are given in the data set and preprocessing section. After data preprocessing, rs-fMRI is 170×90 time series and DTI is 90×90 structural brain network. After using the proposed method to construct functional brain network, rs-fMRI is 90×90 functional brain network. They are then used together as input for feature selection, and because of the symmetry of the matrix, the upper trigonometry of the join matrix is used. The 90×90 network of each subject was taken as a column in the upper triangle and reshaped into a vector with 4005 elements, and the feature dimension was 4005× the number of subjects (Wang et al., 2018). After feature selection, we extract the top 50 features with the largest weight, obtain the feature dimension of 50× the number of subjects, and input them into the classifier for classification.

In our study, we conduct six binary classification experiments: NC vs. SMC, NC vs. EMCI, NC vs. LMCI, SMC vs. EMCI, SMC vs. LMCI and EMCI vs. LMCI. Our experiments are conducted using LOO cross-validation algorithm to fairly assess the classification performance.used as . LOO cross-validation algorithm is computationally expensive, the sample utilization rate is the highest and also suitable for small samples as in our tasks.

Since the regularization parameters involved in the network estimation and feature selection process may affect the classification results, we select the optimal parameters by performing a grid search over a wide range. For each regularization parameter λ (including SR, LR and SLR), we use eleven candidate values in $[2^{-5}, 2^{-4}, \dots, 2^{-0}, \dots, 2^{4}, 2^{5}]$. For the regularization parameters λ and μ in our method, the candidate values in range $[2^{-5}, 2^{-4}, \dots, 2^{-0}, \dots, 2^{4}, 2^{5}]$ and $[0.1, 0.2, \dots, 0.9, 1]$, respectively. For the fairness of comparison, we use eleven sparse levels as threshold parameters in the PC, levels from $[1\%, 10\%, \dots, 90\%, 100\%]$, where 100% means that all edges are preserved and 90% means that 10% of the weak edges in the FBN are filtered out. In the feature selection process, we also use eleven candidate values in $[10^{-5}, 10^{-4}, \dots, 10^{-0}, \dots, 10^{4}, 10^{5}]$. In the experiments, where the parameters are giv-

en, the training samples are used to select features and train the classifier, and then the remaining 1 sample is used to verify the classification performance. The average accuracy obtained from the internal LOO is used as the a good metric to verify the best performance. Since the optimal network parameters may vary with the training set, we re-select features and train the classifier based on the current training set of optimal network parameters. Finally, we classify the test samples using the selected features and trained classifiers.

3.3 Comparison methods

To demonstrate the effectiveness of our proposed method, we perform experiments based on the ADNI3 database and compare it with other methods. Different network estimation methods and multi-task feature selection methods are compared in the experiments. Specifically, network estimation methods include PC, low rank (LR), SR and sparse low rank (SLR), which have been described in the Methods section. Multi-task feature selection methods include least absolute shrinkage and selection operator method (LASSO) (Vorlíčková, 2017) clustered multi-task learning (CMTL) (Jacob et al., 2009), and robust multi-task feature learning (rMTFL) (Chen et al., 2011b). LASSOLASSO is one of the most commonly used feature selection regularization techniques, but it does not perform multi-task feature selection process and selects all features for comparison. The other two comparison methods are commonly used multi-task learning methods, which can effectively learn multiple tasks simultaneously and complete the selection of high-dimensional features.

3.4 Performance evaluation

For the classification performance evaluation, we use classification accuracy (the disease state of the subject is correctly classified as the actual disease state of each category of subjects) (ACC), sensitivity (SEN), specificity (SPE) and area under the receiver operating characteristic (ROC) curve (AUC) defined:

$$\mathrm{ACC} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP}) \tag{17}$$

$$\mathrm{SEN} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{18}$$

$$\mathrm{SPE} = \mathrm{TN}/(\mathrm{TN} + \mathrm{FP}) \tag{19}$$

where TP, TN, FP, and FN denote the true positive, true negative, false positive, and false negative, respectively.

3.5 Classification Performance

Tables 3 and 4 respectively show the performance of different network estimation methods and different multi-task feature selection methods on the same dataset, including the classification results of NC vs. SMC, NC vs. EMCI, NC vs. LMCI, SMC vs. EMCI, SMC vs. LMCI, EMCI vs. LMCI. We include NC and the three sub-categories of MCI in our classification tasks. ROC curves for the above discussed binary classification problems are also compared and shown in Figure 4. The radar diagram in Figure 5 illustrates the classification performance compared with other feature selection methods. It can be seen that our method performs best in terms of all the four evaluation metrics. Through the experimental results, we have these findings.

Table 3: Classification performance of various methods in NC vs. SMC, NC vs. EMCI and NC vs. LMCI. (Boldface denotes best performance.)

BFCN	Feature Selection	NC vs. SMC				NC vs. EMCI				NC vs. LMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
PC	LASSO	60.23	68.18	52.27	64.45	59.09	45.45	72.73	72.31	65.85	68.42	63.64	71.83
	rMTFL	60.23	65.91	54.55	64.72	68.18	65.91	70.45	80.22	57.32	50.00	**63.64**	57.19
	CMTL	63.64	61.36	65.91	69.34	67.05	**68.18**	65.91	71.41	58.54	55.26	61.36	52.00
	Ours	**73.86**	**72.73**	**75.00**	**80.81**	**69.32**	65.91	**72.73**	**81.16**	**67.07**	**76.32**	59.09	**79.26**
SR	LASSO	76.14	77.27	75.00	84.92	65.50	63.64	61.36	70.40	73.17	81.58	65.91	70.75
	rMTFL	76.14	**81.82**	70.45	85.28	70.45	65.91	75.00	78.25	74.39	63.16	84.09	83.43
	CMTL	75.00	72.73	77.27	83.32	79.55	81.82	77.27	84.40	67.07	63.16	70.45	71.05
	Ours	**78.41**	75.00	**81.82**	**86.11**	**82.95**	**84.09**	**81.82**	**86.05**	**84.15**	**84.21**	**84.09**	**89.89**
LR	LASSO	73.86	65.91	81.82	74.48	65.91	68.18	63.64	73.66	68.29	71.05	65.91	72.19
	rMTFL	63.64	54.55	72.73	82.95	76.14	75.00	77.27	86.47	65.85	57.89	72.73	88.82
	CMTL	67.05	63.64	70.45	75.88	80.68	**86.36**	75.00	90.55	62.20	60.53	63.64	71.41
	Ours	**77.27**	**72.73**	**81.82**	**84.14**	**81.82**	79.55	**84.09**	**93.60**	**74.39**	**73.68**	**75.00**	**90.73**
SLR	LASSO	65.91	63.64	68.18	77.07	60.23	56.82	63.64	65.70	64.63	60.53	68.18	69.68
	rMTFL	72.73	63.64	**81.82**	84.19	77.27	75.00	79.55	87.04	70.73	73.68	68.18	77.57
	CMTL	70.45	63.64	77.27	78.62	80.68	77.27	**84.09**	91.63	70.73	68.42	72.73	86.30
	Ours	**78.41**	**79.55**	77.27	**88.02**	**82.95**	**86.36**	79.55	**92.20**	**82.93**	**81.58**	**84.09**	**96.29**
Ours	LASSO	77.27	79.55	75.00	84.71	79.55	81.82	77.27	87.04	79.27	81.58	77.27	86.78
	rMTFL	71.59	70.45	72.73	77.84	81.82	86.36	77.27	92.10	82.93	81.58	84.09	90.67
	CMTL	75.00	72.73	77.27	84.30	81.82	84.09	79.55	91.27	80.49	78.95	81.82	91.93
	Ours	**82.95**	**88.64**	**77.27**	**89.82**	**85.23**	**86.36**	**84.09**	**93.54**	**87.80**	**84.21**	**90.91**	**98.86**

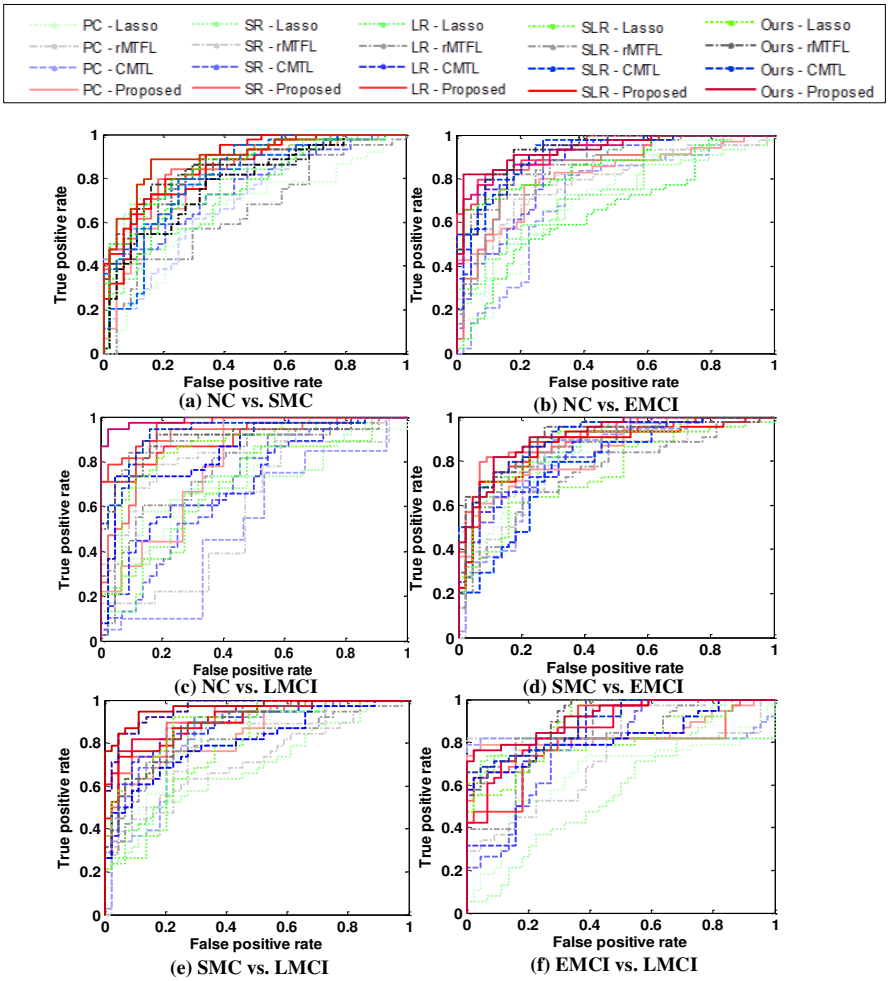

Figure 4: ROC plots comparison of competing methods on different classification tasks.

Table 4: Classification performance of various methods in SMC vs. EMCI, SMC vs. LMCI and EMCI vs. LMCI. (Boldface denotes best performance.)

BFCN	Feature Selection	SMC vs. EMCI				SMC vs. LMCI				EMCI vs. LMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
PC	LASSO	62.50	63.64	61.36	74.67	67.07	65.79	68.18	81.18	58.54	**63.16**	54.55	64.71
	rMTFL	64.77	65.91	63.64	75.26	54.88	60.53	50.00	79.42	54.88	57.89	52.27	82.78
	CMTL	62.50	65.91	59.09	77.27	67.07	57.89	**75.00**	81.10	52.44	55.26	50.00	82.06
	Ours	**69.32**	**70.45**	**68.18**	**79.62**	**68.29**	**71.05**	65.91	**84.15**	**65.85**	57.89	**72.73**	**84.39**
SR	LASSO	72.73	75.00	70.45	83.32	62.20	**71.05**	54.55	68.36	53.66	42.11	63.64	58.01
	rMTFL	78.41	77.27	79.55	89.05	63.41	60.53	65.91	71.23	64.63	**65.79**	63.64	74.88
	CMTL	76.14	81.82	70.45	84.35	65.85	65.79	65.91	88.40	56.10	57.89	54.55	77.87
	Ours	**82.95**	**84.09**	**81.82**	**90.96**	**67.07**	63.16	**70.45**	**90.31**	**67.07**	60.53	**72.73**	**80.50**
LR	LASSO	76.14	75.00	77.27	87.71	74.39	76.32	72.73	81.88	75.61	81.58	70.45	79.01
	rMTFL	77.27	**81.82**	72.73	84.87	78.05	84.21	72.73	85.59	76.83	71.05	**81.82**	84.93
	CMTL	73.86	63.64	84.09	84.45	79.27	76.32	81.82	87.98	75.61	78.95	72.73	83.73
	Ours	**80.68**	77.27	**84.09**	**88.43**	**82.93**	**84.21**	**81.82**	**90.25**	**78.05**	**81.58**	75.00	**85.41**
SLR	LASSO	65.91	61.36	70.45	74.23	60.98	57.89	63.64	75.90	60.98	57.89	63.64	83.07
	rMTFL	67.05	61.36	72.73	74.38	75.61	78.95	72.73	80.26	74.39	**84.21**	65.91	85.65

	CMTL	73.86	72.73	75.00	77.43	79.27	**81.58**	77.27	80.80	68.29	71.05	65.91	82.42
	Ours	**81.82**	**84.09**	**79.55**	**86.26**	**85.37**	79.95	**90.91**	**91.93**	**78.05**	78.95	**77.27**	**87.56**
Ours	LASSO	78.41	77.27	79.55	87.76	79.27	78.95	79.55	89.77	74.39	68.42	79.55	88.76
	rMTFL	80.68	**90.91**	70.45	89.82	78.05	78.95	77.27	87.74	76.83	**81.58**	72.73	**91.21**
	CMTL	79.55	72.73	86.36	90.70	85.37	86.84	84.09	96.11	76.83	73.68	79.55	89.47
	Ours	**84.09**	81.82	**86.36**	**91.27**	**90.24**	**89.47**	**90.91**	**96.65**	**81.71**	78.95	**84.09**	92.11

Table 5: Classification performance with related works in NC vs. MCI. (Boldface denotes best performance.)

Method	Modality	Subjects	FBN construction	Feature selection	ACC	SEN	SPE	AUC
Wee et al. (Wee et al., 2014a)	fMRI	50	Group-constrained sparse	mRMR	84.00	84.00	84.00	86.56
Li et al.(Li et al., 2018)	fMRI	73	Ultra-group LASSO	MPC	80.82	80.56	81.08	80.82
Zhou et al. (Zhou et al., 2018b)	fMRI	137	High-order FCN	t-test	86.13	83.82	88.41	-
Li et al. (Li et al., 2019)	ASL+ BOLD fMRI	61	Ultra-Weighted-LASSO	M2TFS (Jie et al., 2015)	86.90	82.10	90.90	90.00
Proposed	fMRI+ DTI	170	Low Rank Self-calibrated	Joint Non-Convex	**87.80**	**84.21**	**90.91**	**98.86**

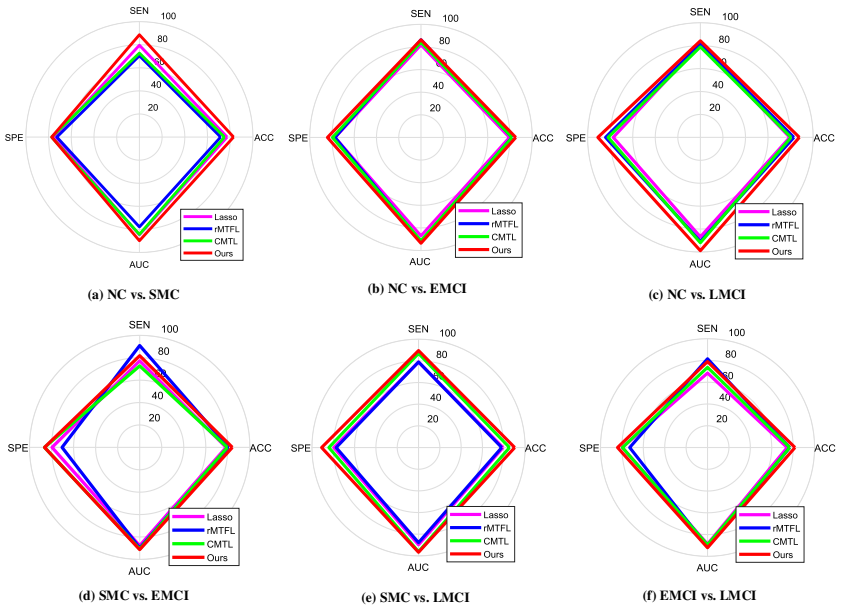

Figure 5: In six classification problems, the performance comparison of different feature selection methods after using our method to construct the network.

Among the six classification problems, the proposed method is superior to the most commonly used comparison methods. Our model has classification accuracies of–82.95%, 85.23%, 87.80%, 84.09%, 90.24% and 81.71% on the classification of NC vs. SMC, NC vs. EMCI, NC vs. LMCI, SMC vs.

EMCI, SMC vs. LMCI and EMCI vs. LMCI, respectively. This suggests that our proposed method can accurately classify MCI and SMC patients from NCs or in several subcategories. In addition, Tables 3 and 4 show not only the classification accuracy obtained by the classification model, but also the sensitivity, specificity, and AUC. Among the six classification problems, the sensitivity reaches 88.64%, 86.36%, 84.21%, 81.82%, 89.47% and 78.95%, respectively. The specificity reaches 77.27%, 84.09%, 90.91%, 86.36%, 90.91 and 84.09%, respectively. Higher specificity indicates a higher true negative rate, while higher sensitivity indicates a higher true positive rate.

The results indicate that effective results can be achieved in the diagnosis of NCs and in the selection of all disease samples. The AUC reaches 89.82%, 93.54%, 98.86%, 91.27%, 96.65% and 92.11%, respectively, in the six classification problems. The larger the AUC, the better the classification effect of the classifier.

Comparing experiments with different network estimation methods using the same feature selection method reveals that PC is usually the worst performing method. Because the PC can only detect complete correlation when modeling, it cannot rule out the confounding effects of other brain regions. Hence, some redundant and unimportant associations are included in modeling, which brings difficulties to the brain network analysis. There are no significant differences when comparing to the performance of the other three partial-correlation network estimation methods under this regularization framework (including SR, LR, and SLR), which all improve the confounding problems involved in PC, and incorporate prior information, which can be effectively constructed brain networks. Our method has superior performance when compared to other network estimation methods implemented herThe proposed method does not only effectively encode the prior information, but also identifies the data quality, reduces the amount of bad data points, and limits the impact of noise and artifacts during the network estimation process. The results prove that the improvement of the network estimation process enhances the classification accuracy rate and is more conducive to brain network analysis.

A comparison of different feature selection methods using the same network estimation method reveals that LASSO is the worst performing method in most cases. The main reason is that other comparison methods are multi-task feature selection methods, but LASSO is to select all features simultane-

ously. LASSO only selects features with sparse regularization, which is not enough to correctly classify neuroimaging features. Observing the results of the comparative experiments, we find that our method is superior to several other multi-task feature selection methods. Among all classification problems, there is an increase on classification accuracy of 1.14%~10.22%. This indicates that the proposed multi-task joint non-convex feature selection method can indeed improve the performance of MCI clinical recognition.

Table 5 shows the comparison between our method and the classification results of other MCI automatic diagnosis studies, and our results are optimal in four classification measures. This is only a comparison of NC and MCI. Other studies of AD in earlier stages have been conducted in our experiments.

In addition, we perform an ablation experiment to study the performance impact of the improved method. The results are shown in Table 6. In all classification problems, we study the effects of multi-modal data learning, data quality indicators, low rank constraints and regularization terms of feature selection processes. But they do affect performance, and improvements to traditional solutions can improve the performance of the classification.

Table 6: Comparison of ablation results in six classification problems.(Boldface denotes best performance).

Modality	NC vs. SMC				NC vs. EMCI				NC vs. LMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
DTI	60.23	65.91	54.55	70.22	71.59	63.64	79.55	84.92	67.07	65.79	68.18	73.33
fMRI	80.00	81.82	78.95	85.17	81.82	79.55	84.09	89.82	85.37	78.95	90.91	96.47
$p_t = 0$	76.14	72.73	**79.55**	82.85	75.00	79.55	70.45	83.73	78.05	81.58	75.00	83.76
$\lambda = 0$	77.27	79.55	75.00	83.99	76.14	84.09	68.18	86.57	79.27	81.58	77.27	88.70
$\gamma = 0$	78.41	79.55	77.27	85.80	77.27	75.00	79.55	85.43	82.93	84.21	81.82	87.38
Ours	**82.95**	**88.64**	77.27	**89.82**	**85.23**	**86.36**	**84.09**	**93.54**	**87.80**	**84.21**	**90.91**	**98.86**
Modality	SMC vs. EMCI				SMC vs. LMCI				EMCI vs. LMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
DTI	54.14	54.55	47.73	62.23	70.73	71.05	70.45	81.28	53.66	57.89	50.00	52.60
fMRI	80.68	77.27	84.09	90.55	85.37	81.58	88.64	93.24	78.05	73.68	81.82	85.71
$p_t = 0$	67.05	72.73	61.36	80.00	69.51	73.68	65.91	89.54	63.41	71.05	56.82	82.48
$\lambda = 0$	75.61	**84.21**	68.18	85.74	78.41	75.00	81.82	89.65	78.05	76.32	79.55	84.27
$\gamma = 0$	61.36	54.55	68.18	68.40	77.27	72.73	81.82	87.55	74.39	71.05	77.27	85.59
Ours	**84.09**	81.82	**86.36**	**91.27**	**90.24**	**89.47**	**90.91**	**96.65**	**81.71**	**78.95**	**84.09**	**92.11**

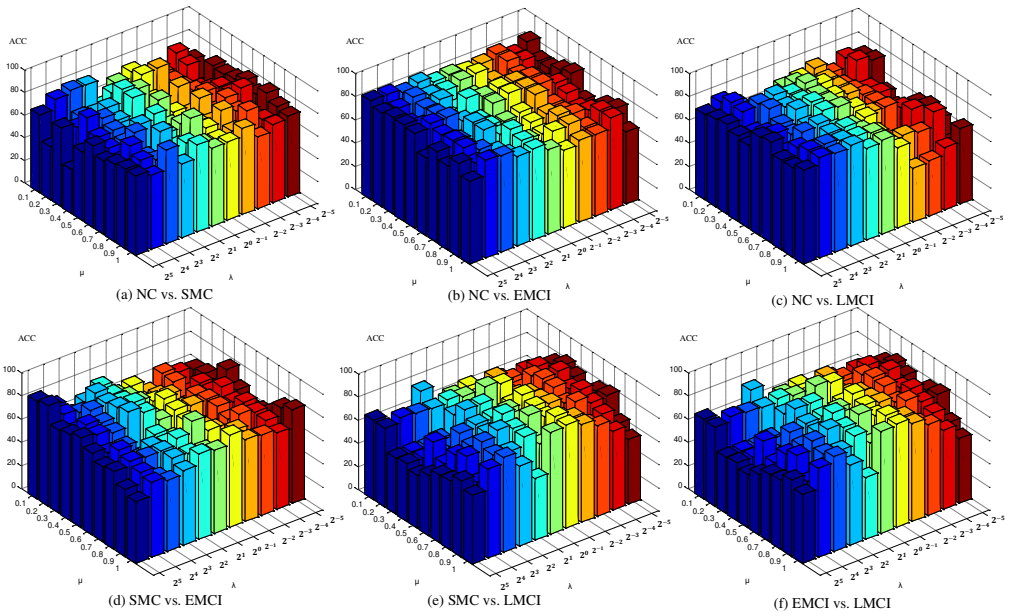

Figure 6: The accuracy in all classification tasks based on the networks estimated by the proposed method with different regularized parametric values in the interval.

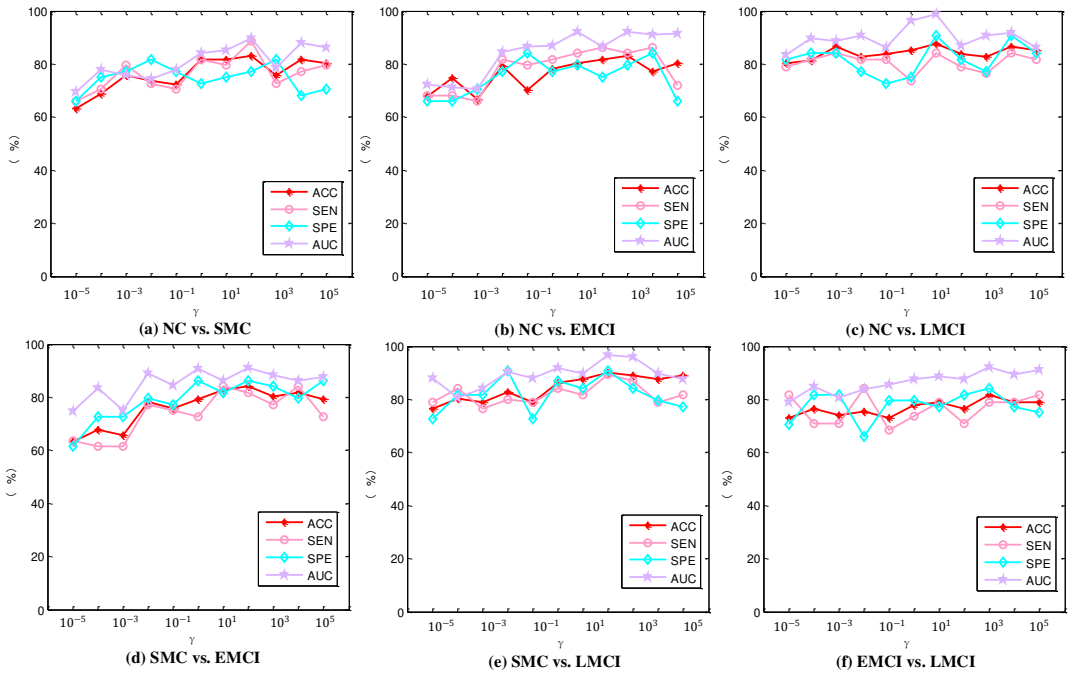

Figure 7: The accuracy in all classification tasks based on feature selection by the proposed method with different regularized parametric values in the interval.

3.6 Results Summary

Figure 6 is the sensitivity of the network model parameters in all classification problems, showing the classification accuracy of the proposed network construction method for different parameter selection combinations. It can be concluded that the choice of network model parameters has a great impact on the final classification accuracy. LOO test is performed on all subjects. Similarly, we perform sensitivity experiments on all parameters in the feature selection process, and the classification results are shown in Figure 7. The selected connections and corresponding brain regions in each experiment are recorded. Although different parameter settings affect the selected connections and corresponding brain regions, some brain regions appear frequently in the same classification task. For example, in NC vs. LMCI, SOG.L and MOG.R appeared in all parameter selection, which provide new findings for disease detection. In Figure 8, we use five different methods to construct the edge weight matrix of the brain network for visualization. For SR, LR, and SLR, we uniformly set the regularized parameter to $\lambda = 1$, and for our method, $\lambda = 1$ and $\mu = 0.5$. From the results in Figure 8, our method removes weak connections and noise while maintaining important connections. Figure 9 shows the functional connectivity networks of NC, SMC, EMCI and LMCI, respectively. Several functionally connected networks exhibit relatively different modes in terms of network topology and strength. In several subcategories of MCI, the three types are progressively distributed, and the degree of cognitive impairment is gradually increased. It can be found that some connections are destroyed and the number of connections is gradually reduced compared with the NC subjects. This follows previous research (Zhou et al., 2013). Figure 9 is created using the Matlab function *circularGraph* shared by Paul Kassebaum. The color of each arc in the graph is random and meaningless, and is allocated only for better visualization.

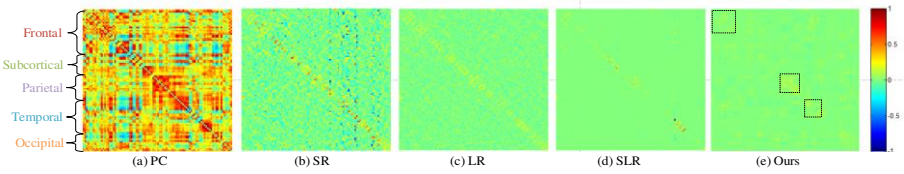

Figure 8: The FBN edge weight matrix of a subject estimated by 5 methods.

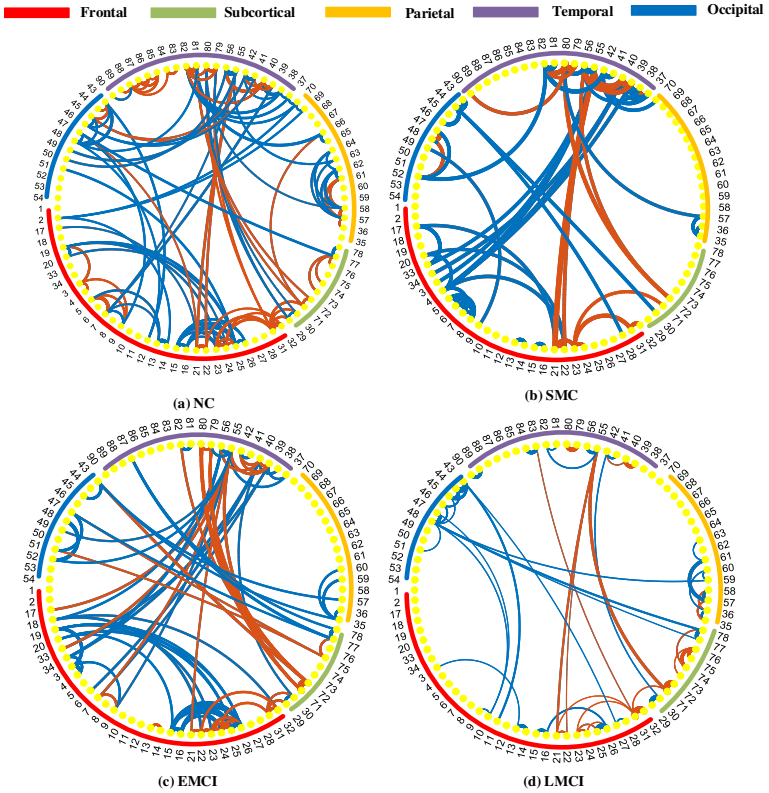

Figure 9: The structure of four networks for (a)NC ;(b)SMC; (c) EMCI; and (d)LMCI.

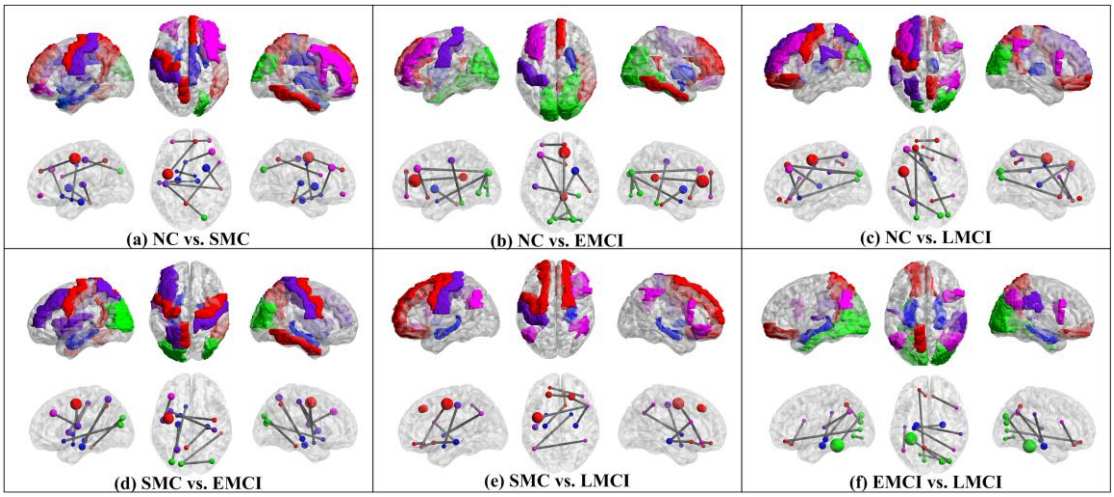

Figure 10: Of the six classification questions, the top 10 most discriminative connections and associated ROI brain regions.

Table 7: Indices and names of recognized brain ROIs related to the top 10 connectivity from our classification tasks in NC vs. SMC, NC vs. EMCI and NC vs. LMCI using ALL template with 90 ROIs.

NC vs. SMC		NC vs. EMCI		NC vs. LMCI	
ROIs index	ROI names	ROIs index	ROI names	ROIs index	ROI names
9, 10	ORBmid.L,ORBmid.R	9, 24	ORBmid.L,SFGmed.R	3, 49	SFGdor.L, SOG.L
21, 39	OLF.L, PHG.L	7, 22	MFG.L, OLF.R	5, 6	ORBsup.L,ORsup.R
8, 21	MFG.R, OLF.L	32, 36	ACG.R, PCG.R	21, 52	OLF.L, MOG.R
50, 57	SOG.R, PoCG.L	43, 50	CAL.L, SOG.R	62, 68	IPL.R, PCUN.R
78, 79	THA.R, HES.L	56, 57	FFG.R, PoCG.L	7, 22	MFG.L, OLF.R
34, 41	DCG.R, AMYG.L	74, 90	PUT.R, ITG.R	15, 34	ORBinf.L, DCG.R
12, 67	IFGoperc.R, PCUN.L	45, 48	CUN.L, LING.R	12, 27	IFGoperc.R, REC.L
74, 90	PUT.R, ITG.R	46, 54	CUN.R, IOG.R	3, 78	SFGdor.L, THA.R
81, 90	STG.L, ITG.R	7, 50	MFG.L, SOG.R	50, 63	SOG.R, SMG.L
1, 24	PreCG.L, SFGmed.R	42, 49	AMYG.R, SOG.L	1, 59	PreCG.L, SPG.L

Table 8: Indices and names of recognized brain ROIs related to the top 10 connectivity from our classification tasks in SMC vs. EMCI, SMC vs. LMCI and EMCI vs. LMCI using ALL template with 90 ROIs.

SMC vs. EMCI		SMC vs. LMCI		EMCI vs. LMCI	
ROIs index	ROI names	ROIs index	ROI names	ROIs index	ROI names
1, 17	PreCG.L, ROL.L	22, 37	OLF.R, HIP.L	50, 55	SOG.R, FFG.L
49, 52	SOG.L, MOG.R	3, 4	SFGdor.L, SFGdor.R	39, 64	PHG.L, SMG.R
38, 67	HIP.R, PCUN.L	12, 28	IFGoperc.R, REC.R	37, 40	HIP.L, PHG.R
39, 59	PHG.L, SPG.L	5, 16	ORBsup.L, ORBinf.R	37, 66	HIP.L, ANG.R
2, 41	PreCG.R, AMYG.L	41, 57	AMYG.L, PoCG.L	5, 43	ORBsup.L, CAL.L
29, 58	INS.L, PoCG.R	65, 76	ANG.L, PAL.R	12, 28	IFGoperc.R, REC.R
49, 90	SOG.L, ITG.R	1, 12	PreCG.L, IFGoperc.R	47, 54	LING.L, IOG.R
7, 17	MFG.L, ROL.L	62, 65	IPL.R, ANG.L	65, 79	ANG.L, HES.L
21, 51	OLF.L, MOG.L	22, 28	OLF.R, REC.R	5, 27	ORBsup.L, REC.L
58, 75	PoCG.R, PAL.L	12, 38	IFGoperc.R, HIP.R	52, 67	MOG.R, PCUN.L

We use LOO cross-validation algorithm to evaluate the performance of the proposed framework and select different subsets of features for MCI classification. Therefore, the connection feature with the highest number of selected frequencies in LOO cross-validation is considered the most discriminative connection for MCI classification. Tables 7 and 8 list the top 10 most discriminative connections and related ROI brain regions in the six classification problems of the experiment and visualized using BrainNet Viewer (Xia et al., 2013), as shown in Figure 10.

The brain regions are associated with the top ten connections selected by the NC vs. SMC classifications using our proposed method are middle frontal gyrus, orbital part (ORBmid.L and ORBmid.R), olfactory cortex (OLF.L), parahippocampal gyrus (PHG.L), middle frontal gyrus (MFG.R), superior occipital gyrus (SOG.R), postcentral gyrus (PoCG.L), thalamus (THA.R), heschl gyrus (HES.L), median cingulate and paracingulate gyri (DCG.R), amygdala (AMYG.L), inferior frontal gyrus, opercular

part (IFGoperc.R), precuneus (PCUN.L), lenticular nucleus, putamen (PUT.R), inferior temporal gyrus (ITG.R), superior temporal gyrus (STG.L), precentral gyrus (PreCG.L), and superior frontal gyrus, medial (SFGmed.R).

The top ten connections related ROIs selected from NC vs. EMCI classification using our proposed method are ORBmid.L, SFGmed.R, middle frontal gyrus (MFG.L), olfactory cortex(OLF.R), anterior cingulate and paracingulate gyri (ACG.R), posterior cingulate gyrus (PCG.R), calcarine fissure and surrounding cortex (CAL.L), superior occipital gyrus (SOG.R, SOG.L), fusiform gyrus (FFG.R), PoCG.L, lenticular nucleus, putamen (PUT.R), ITG.R, cuneus (CUN.L, CUN.R), lingual gyrus (LING.R), inferior occipital gyrus (IOG.R), and amygdala (AMYG.R) . For NC vs. LMCI classification, the most informative connections related ROIs are superior frontal gyrus, dorsolateral (SFGdor.L), SOG.L, superior frontal gyrus, orbital part (ORBsup.L, ORBsup.R), OLF.L, middle occipital gyrus (MOG.R), inferior parietal, but supramarginal and angular gyri (IPL.R), precuneus (PCUN.R), MFG.L, OLF.R, inferior frontal gyrus, orbital part (ORBinf.L), DCG.R, IFGoperc.R, gyrus rectus (REC.L), THA.R, SOG.R, supramarginal gyrus (SMG.L), PreCG.L, and superior parietal gyrus (SPG.L). For SMC vs. EMCI classification, the most informative connections related ROIs are precentral gyrus (PreCG.L, PreCG.R), rolandic operculum (ROL.L), SOG.L, MOG.R, hippocampus (HIP.R), PCUN.L, PHG.L, superior SPG.L, AMYG.L, insula (INS.L), postcentral gyrus (PoCG.R), ITG.R, MFG.L, ROL.L, OLF.L, middle occipital gyrus (MOG.L), and lenticular nucleus, pallidum (PAL.L). For the SMC vs. LMCI and EMCI vs. LMCI classification, the relevant brain regions are OLF.R, hippocampus (HIP.L), superior frontal gyrus, dorsolateral (SFGdor.L, SFGdor.R), IFGoperc.R, gyrus rectus (REC.R), ORBsup.L, inferior frontal gyrus, orbital part (ORBinf.R), AMYG.L, PoCG.L, angular gyrus (ANG.L), lenticular nucleus, pallidum (PAL.R), PreCG.L, IPL.R, ANG.L, HIP.R, and SOG.R, fusiform gyrus (FFG.L), PHG.L, supramarginal gyrus (SMG.R), HIP.L, PHG.R, angular gyrus (ANG.R), ORBsup.L, CAL.L, IFGoperc.R, REC.R, lingual gyrus (LING.L), IOG.R, ANG.L, HES.L, REC.L, MOG.R, PCUN.L, respectively.

In all classification tasks, there are many overlapping related brain regions, such as olfactory cortex, parahippocampal gyrus, superior occipital gyrus, postcentral gyrus, amygdala, precuneus, angular

gyrus, etc., which appear multiple times in different classifications. These areas are often reported to be highly correlated with AD / MCI pathology (Salvatore et al., 2015; Sun et al., 2012; Xu et al., 2016). For example, the olfactory cortex is the most frequently occurring brain region in the experiment, which follows previous studies (Vasavada et al., 2015). AD / MCI preferentially attacks the central olfactory structure, and the symptoms of olfactory defects usually precede clinical cognitive defects and memory defects. Parahippocampal gyrus plays an important role in memory function, where lesions often cause memory impairment, which is a highly relevant brain region of AD / MCI. It has been found that motor dysfunction is associated with MCI, and postcentral gyrus as a somatic sensory center is also an important brain area for disease. Areas such as superior occipital gyrus and posterior cingulate gyrus have been proven to be related to MCI. On the other hand, in our experiments, there are some newly discovered brain regions such as the lenticular nucleus, pallidum, fusiform gyrus, and supramarginal gyrus, which play an important role in the recognition of diseases, possibly in memory with MCI and SMC patients. It is related to cognitive impairment and compensation mechanisms. The connections in the brain provide a broader perspective for analyzing disease progression and provide new insights into disease research, which can help improve the effectiveness of computer-aided diagnosis.

4 Discussion

The typical brain network analysis framework mainly includes three steps: brain network construction, feature learning and classification prediction. Based on this, we propose a low-rank self-calibration brain network estimation and joint non-convex multi-task learning in this paper for early MCI diagnosis. In the stage of functional brain network construction, a new model is proposed, which introduces self-calibrated indicators and completes brain network estimation while controlling data quality. It can be seen from the results of the comparison experiments that the method of controlling the data quality is superior to other methods because the removal of noise can improve the performance of the model. In the feature selection stage, non-convex regularizer is used to complete the subspace learning of samples. Different from the traditional optimization method, our method can reduce the

penalty deviation and learn the low-rank structure of the coefficient matrix well, which integrates the functional connection and structural connection for joint multi-task learning.

The automatic diagnosis framework is based on multi-task learning to improve their performance, which leverages the useful information between multiple related learning tasks. The relationship between different modalities of the subjects is used to conduct multi-task learning, to obtain the most discriminating disease-related features to realize the automatic diagnosis of early MCI. In the classification experiment, MCI and SMC are classified with better results, which is helpful for the auxiliary physician to improve the diagnostic accuracy. According to the connection relationship between brain regions, some biomarkers sensitive to diseases can be explored and biologically meaningful reliability measurements are added, which reveals potential application for biomarker recognition based on medical images.

Although the framework proposed in this study has achieved classification satisfactory results, there are still deficiencies. In this paper, the structural and functional brain network are constructed separately, and the process is relatively independent. The information integration of different modalities is realized in the feature learning process. Building a functional and structural connection network and integrating the network-learning framework may obtain higher classification accuracy of MCI and SMC, which is the focus of our future research. In future research, we will increase the amount of data for the problem of the small number of subjects in this paper. In addition, the brain network functional connection analysis method used is based on a single subject, which fails to consider the relationship between subjects. Our future work will consider the relationship among different subjects.

5 Conclusions

In this study, we propose a framework for automated diagnosis of the early stages of AD to distinguish between MCI patients, earlier SMC patients and healthy subjects. The structural and functional brain networks are first constructed, and then the multi-modal features are selected based on the joint non-convex multi-task learning method. An SVM is used for classification to diagnose patients. The method uses multi-modal neuroimaging data to classify six dichotomous problems. The experiments on

the ADNI dataset demonstrate the effectiveness of the method compared with other methods. The results based on cross-validation show that our method obtains better results in classification performance. The brain regions associated with disease also demonstrate the effectiveness of the method. It can be proved that the method can effectively integrate multi-modal data information for multi-task learning, which helps computer-aided diagnosis to determine the early stages of AD and predicts the progress of disease development.

Acknowledgements

This work was supported partly by National Natural Science Foundation of China (Nos.61871274, U1909209, 61801305 and 81571758), Key Laboratory of Medical Image Processing of Guangdong Province (No. K217300003), Guangdong Pearl River Talents Plan (2016ZT06S220), Shenzhen Peacock Plan (Nos. KQTD2016 053112051497 and KQTD2015033016104926), and Shenzhen Key Basic Research Project (Nos. JCYJ20180507184647636 and JCYJ20170818094109846). AFF was partially supported by the OCEAN project (EP/M006328/1) and MedIAN Network (EP/N026993/1) both funded by the Engineering and Physical Sciences Research Council (EPSRC), the European Commission FP7 Project VPH-DARE@IT (FP7-ICT-2011-9-601055), and a Royal Academy of Engineering Chair in Emerging Technology to AFF.

References

(2018). 2018 Alzheimer's disease facts and figures. *Alzheimers & Dementia* 14, 367-429.

Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., Moran, R., (2014). SPM12 manual. *Wellcome Trust Centre for Neuroimaging, London, UK*.

Cawley, G.C., Talbot, N.L., (2003). Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. *Pattern Recognition* 36, 2585-2592.

Chen, G., Ward, B.D., Xie, C., Li, W., Wu, Z., Jones, J.L., Franczak, M., Antuono, P., Li, S.-J., (2011a). Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. *Radiology* 259, 213-221.

Chen, J., Zhou, J., Ye, J., (2011b).Integrating low-rank and group-sparse structures for robust multi-task learning, in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 42-50.

Cherubini, A., Péran, P., Spoletini, I., Di Paola, M., Di Iulio, F., Hagberg, G.E., Sancesario, G., Gianni, W., Bossu, P., Caltagirone, C., (2010). Combined volumetry and DTI in subcortical structures of mild cognitive impairment and Alzheimer's disease patients. *Journal of Alzheimer's Disease* 19, 1273-1282.

Craddock, R.C., James, G.A., Holtzheimer III, P.E., Hu, X.P., Mayberg, H.S., (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping* 33, 1914-1928.

Cui, Z., Zhong, S., Xu, P., Gong, G., He, Y., (2013). PANDA: a pipeline toolbox for analyzing brain diffusion images. *Frontiers in human neuroscience* 7, 42.

Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q., (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging* 32, 2322.e2319-2322.e2327.

Gao, Y., Wee, C.-Y., Kim, M., Giannakopoulos, P., Montandon, M.-L., Haller, S., Shen, D., (2015).MCI identification by joint learning on multiple MRI data, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 78-85.

Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27, 685-691.

Jacob, L., Vert, J.-p., Bach, F.R., (2009).Clustered multi-task learning: A convex formulation, in *Advances in neural information processing systems*, pp. 745-752.

Jessen, F., Wolfsgruber, S., Wiese, B., Bickel, H., Mösch, E., Kaduszkiewicz, H., Pentzek, M., Riedel-Heller, S.G., Luck, T., Fuchs, A., (2014). AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. *Alzheimers Dement* 10, 76-83.

Jie, B., Zhang, D., Cheng, B., Shen, D., Initiative, A.s.D.N., (2015). Manifold regularized multitask feature learning for multimodality disease classification. *Human brain mapping* 36, 489-507.

Kamal, S., Burk, L., (1996). FACT: A new neural network-based clustering algorithm for group technology. *International journal of production research* 34, 919-946.

Lee, H., Lee, D.S., Kang, H., Kim, B.-N., Chung, M.K., (2011). Sparse brain network recovery under compressed sensing. *IEEE Transactions on Medical Imaging* 30, 1154-1165.

Lei, B., Yang, P., Wang, T., Chen, S., Ni, D., (2017). Relational regularized discriminative sparse learning for Alzheimer's disease diagnosis. *IEEE Trans Cybern* 47, 1102-1113.

Li, Y., Liu, J., Gao, X., Jie, B., Kim, M., Yap, P.-T., Wee, C.-Y., Shen, D., (2019). Multimodal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification. *Medical image analysis* 52, 80-96.

Li, Y., Yang, H., Lei, B., Liu, J., Wee, C.-Y., (2018). Novel Effective Connectivity Inference Using Ultra-Group Constrained Orthogonal Forward Regression and Elastic Multilayer Perceptron Classifier for MCI Identification. *IEEE transactions on medical imaging*.

Lian, C., Liu, M., Zhang, J., Shen, D., (2018). Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis using Structural MRI. *IEEE transactions on pattern analysis and machine intelligence*.

Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y., (2013). Robust recovery of subspace structures by low-rank representation. *IEEE transactions on pattern analysis and machine intelligence* 35, 171-184.

Meinshausen, N., Bühlmann, P., (2006). High-dimensional graphs and variable selection with the lasso. *The annals of statistics* 34, 1436-1462.

Peng, J., Zhu, X., Wang, Y., An, L., Shen, D., (2019). Structured sparsity regularized multiple kernel learning for Alzheimer's disease diagnosis. *Pattern recognition* 88, 370-382.

Qiao, L., Zhang, H., Kim, M., Teng, S., Zhang, L., Shen, D., (2016). Estimating functional brain networks by incorporating a modularity prior. *Neuroimage* 141, 399-407.

Salvatore, C., Cerasa, A., Battista, P., Gilardi, M.C., Quattrone, A., Castiglioni, I., (2015). Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Frontiers in neuroscience* 9, 307.

Sherali, H.D., Shetty, C., (2006). *Nonlinear programming: Theory and algorithms*. Wiley-Interscience.

Shi, J., Zheng, X., Li, Y., Zhang, Q., Ying, S., (2017). Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE journal of biomedical and health informatics* 22, 173-183.

Sun, G.H., Raji, C.A., MacEachern, M.P., Burke, J.F., (2012). Olfactory identification testing as a predictor of the development of Alzheimer's dementia: a systematic review. *The Laryngoscope* 122, 1455-1462.

Teipel, S.J., Wohlert, A., Metzger, C., Grimmer, T., Sorg, C., Ewers, M., Meisenzahl, E., Klöppel, S., Borchardt, V., Grothe, M.J., (2017). Multicenter stability of resting state fMRI in the detection of Alzheimer's disease and amnestic MCI. *NeuroImage: Clinical* 14, 183-194.

Vasavada, M.M., Wang, J., Eslinger, P.J., Gill, D.J., Sun, X., Karunanayaka, P., Yang, Q.X., (2015). Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment. *Journal of Alzheimer's disease* 45, 947-958.

Vorlíčková, J., (2017). Least Absolute Shrinkage and Selection Operator Method.

Wang, J., Wang, Q., Zhang, H., Chen, J., Wang, S., Shen, D., (2018). Sparse Multiview Task-Centralized Ensemble Learning for ASD Diagnosis Based on Age- and Sex-Related Functional Connectivity Patterns. *IEEE Transactions on Cybernetics*, 1-14.

Wee, C.-Y., Yap, P.-T., Zhang, D., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D., (2012). Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* 59, 2045-2056.

Wee, C.-Y., Yap, P.-T., Zhang, D., Wang, L., Shen, D., (2014a). Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Structure and Function* 219, 641-656.

Wee, C.Y., Yap, P.T., Zhang, D., Wang, L., Shen, D., (2014b). Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Structure & Function* 219, 641-656.

Xia, M., Wang, J., He, Y., (2013). BrainNet Viewer: a network visualization tool for human brain connectomics. *PloS one* 8, e68910.

Xu, L., Wu, X., Li, R., Chen, K., Long, Z., Zhang, J., Guo, X., Yao, L., (2016). Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers. *Journal of Alzheimer's Disease* 51, 1045-1056.

Yan, C., Zang, Y., (2010). DPARSF: a MATLAB toolbox for" pipeline" data analysis of resting-state fMRI. *Frontiers in systems neuroscience* 4, 13.

Yang, X., Jin, Y., Chen, X., Zhang, H., Li, G., Shen, D., (2016).Functional connectivity network fusion with dynamic thresholding for MCI diagnosis, in *MLMI*, Greece: Athens, pp. 246-253.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., Initiative, A.s.D.N., (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856-867.

Zhou, L., Wang, L., Liu, L., Ogunbona, P., Shen, D., (2013).Discriminative brain effective connectivity analysis for Alzheimer's disease: a kernel learning approach upon sparse Gaussian Bayesian network, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2243-2250.

Zhou, T., Thung, K.-H., Liu, M., Shen, D., (2018a). Brain-Wide Genome-Wide Association Study for Alzheimer's Disease via Joint Projection Learning and Sparse Regression Model. *IEEE Transactions on Biomedical Engineering* 66, 165-175.

Zhou, T., Thung, K.-H., Liu, M., Shi, F., Zhang, C., Shen, D., (2019a). Multi-modal Latent Space Inducing Ensemble SVM Classifier for Early Dementia Diagnosis with Neuroimaging Data. *Medical Image Analysis*, 101630.

Zhou, T., Thung, K.H., Zhu, X., Shen, D., (2019b). Effective feature learning and fusion of multimodality data using stage‐wise deep neural network for dementia diagnosis. *Human brain mapping* 40, 1001-1016.

Zhou, Y., Zhang, L., Teng, S., Qiao, L., Shen, D., (2018b). Improving Sparsity and Modularity of High-Order Functional Connectivity Networks for MCI and ASD Identification. *Frontiers in neuroscience* 12.

Zhu, X., Suk, H.-I., Shen, D., (2014). A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *NeuroImage* 100, 91-105.

Zhu, Y., Zhu, X., Kim, M., Yan, J., Kaufer, D., Wu, G., (2018). Dynamic Hyper-Graph Inference Framework for Computer-Assisted Diagnosis of Neurodegenerative Diseases. *IEEE transactions on medical imaging* 38, 608-616.