Fused Sparse Network Learning for Longitudinal Analysis of Mild Cognitive Impairment

Peng Yang, Feng Zhou, Dong Ni, Yanwu Xu, *Senior Member, IEEE*, Siping Chen, Tianfu Wang, and Baiying Lei, *Senior Member, IEEE*

***Abstract*—Alzheimer's disease (AD) is a neurodegenerative disease with an irreversible and progressive process. To understand the brain functions and identify the biomarkers of AD and early stages of the disease [also known as, mild cognitive impairment (MCI)], it is crucial to build the brain functional connectivity network (BFCN) using resting-state functional magnetic resonance imaging (rs-fMRI). Existing methods have been mainly developed using only a single time-point rs-fMRI data for classification. In fact, multiple time-point data is more effective than a single time-point data in diagnosing brain diseases by monitoring the disease progression patterns using longitudinal analysis. In this article, we utilize multiple rs-fMRI time-point to identify early MCI (EMCI) and late MCI (LMCI), by integrating the fused sparse network (FSN) model with parameter-free centralized (PFC) learning. Specifically, we first construct the FSN framework by building multiple time-point BFCNs. The multi-task learning via PFC is then leveraged for longitudinal analysis of EMCI and LMCI. Accordingly, we can jointly learn the multiple time-point features constructed from the BFCN model. The proposed PFC method can automatically balance the contributions of different time-point information via learned specific and common features. Finally, the selected multiple time-point features are fused by a similarity network fusion (SNF) method. Our proposed method is evaluated on the public AD neuroimaging initiative phase-2 (ADNI-2) database. The experimental results demonstrate that our method can achieve quite promising performance and outperform the state-of-the-art methods.**

***Index Terms*—Fused sparse network (FSN), longitudinal analysis, mild cognitive impairment (MCI), parameter-free centralized (PFC) learning, similarity network fusion (SNF).**

I. Introduction

ALZHEIMER'S disease (AD) is one of the most frequent forms of dementia, which leads to approximately 60%–80% of dementia cases in the world [1]. AD patients' experience gradually decline in memory and cognitive functions as their disease progresses [2], and eventually die in 3–10 years [3]. To date, there is no evidence that the pathophysiologic progression due to AD can be reversed or cured [4], [5]. According to the recent statistics, there are a total number of 47 million people suffering from this disease and the number will increase to more than 131 million by 2050 [6]. The average probability of converting mild cognitive impairment (MCI) [e.g., early MCI (EMCI) and late MCI (LMCI)] to AD is 15%. Hence, it is necessary to treat and monitor AD progression before onset [7]. However, the majority of the existing computer-aided diagnosis methods do not meet the clinical diagnosis requirement, especially for MCI diagnosis.

Resting-state functional MRI (rs-fMRI) is capable of building the brain functional connectivity network (BFCN) among brain regions, by examining the functional impairment of brain networks caused by MCI [8], [9]. In fact, BFCN represents the temporal correlation of blood-oxygenation-level-dependent (BOLD) time series among the brain areas [10], and the reduced BOLD connectivity is related to hypoperfusion in AD [11]. The BFCN [12] approaches have been successfully used to diagnose MCI and AD [13]–[16] to understand the functional interactions among the brain areas, and have received an increasing attention in the neurocognitive area [17].

In the last decade, numerous methods for constructing BFCNs have been proposed, including the Pearson's correlation (PC)-based methods [18], graphical models [19], and sparse representation methods [20]. The PC method is one of the typical BFCN modeling algorithms based on the pairwise correlation among different brain regions, which can be enhanced by a high-order functional connectivity [21]. However, the BFCN constructed by the PC method is too dense to clearly express the connection between the brain

Manuscript received March 16, 2019; revised July 20, 2019 and August 5, 2019; accepted August 12, 2019. Date of publication September 30, 2019; date of current version December 22, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61871274, Grant 61801305, and Grant 81571758, in part by the National Natural Science Foundation of Guangdong Province under Grant 2017A030313377, in part by the Shenzhen Peacock Plan under Grant KQTD2016053112051497 and Grant KQTD2015033016104926, and in part by the Shenzhen Key Basic Research Project under Grant JCYJ20170413152804728, Grant JCYJ20180507184647636, Grant JCYJ20170818142347251, and Grant JCYJ20170818094109846. This article was recommended by Associate Editor H. A. Abbass. *(Corresponding author: Baiying Lei.)*

P. Yang, D. Ni, T. Wang, and B. Lei are with the 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 518060, China (e-mail: leiby@szu.edu.cn).

F. Zhou is with the Department of Industrial and Manufacturing Systems Engineering, University of Michigan–Dearborn, Dearborn, MI 48128 USA.

Y. Xu is with the Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo 530031, China.

S. Chen is with the School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China.

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

Digital Object Identifier 10.1109/TCYB.2019.2940526

regions that are most related to the disease [22]. The graphical models are another popular ways to enhance the PC method [19]. For example, Rosa *et al.* [23] proposed a sparse network-based discriminative modeling framework based on the Gaussian graphical models. However, these methods require a great amount of prior knowledge, which may be infeasible [20], [24], [25]. In contrast, it is known that sparse learning is sparsely constructed, which can show biological significant findings. Therefore, it is more appealing to construct a BFCN via sparse learning. For example, Wee *et al.* [22] proposed a group-constrained sparse (GCS) model to construct BFCN for MCI identification. As the disease progresses, brain function changes are reflected by the regions of interest (ROIs) relationships of the BFCN at different time points. However, these attempts fail to consider the multiple time-point information and fused group constraints to characterize the brain function. Recently, the techniques of dynamical graph theory networks have been proposed and achieved a good performance in studying dementia diseases [26]–[29]. The dynamical graph theory networks can capture the dynamical information in brain to replace the static information. However, those methods increase computational complexity. The previous works mainly utilize the single time-point rs-fMRI data [30], [31]. Longitudinal analysis with multiple time-point data can be more appealing for MCI disease diagnosis [32]. To overcome the deficiencies of the previous studies and enhance the diagnostic performance, we propose a fused sparse learning method to construct the BFCN using multiple time points (i.e., FSN). The FSN can utilize the group and time constraints to characterize complex brain functions as well.

For the BFCN-based methods, the high feature dimension and low subject size suffer from overfitting [33]. To address this drawback, feature selection from the BFCN is an effective way. For example, Zhu *et al.* [25] used the Lasso method to reduce feature dimension. Yu *et al.* [34] adopted the *t*-test to select similar subnetworks, and the weighted sparse groups method was developed for MCI classification. Guo *et al.* [20] used the Kolmogorov–Smirnov test to select features from a hypernetwork. However, these are single-task methods, which are often disadvantageous compared to the multitask learning methods. Also, these methods contain too many parameters, which make them difficult to tune for the clinical applications. To overcome the aforementioned problems, we propose a parameter-free centralized (PFC) multitask learning strategy for feature selection. It can describe the interaction of the BFCN constructed at different time points and perform effective longitudinal analysis.

The PFC multitask learning method can identify specific and common BFCN features at multiple time points. This method leverages the underlying development of the MCI based on the rs-fMRI data. The PFC explicitly identifies the BFCN features with multiple time points, which have different contributions for the MCI classification in a multitask learning framework. These BFCN features are regularized to be sparse and centralized, which can be considered as common features to be shared across different time points. The PFC model exploits a square root objective function to discover the differences among multiple time points. The importance of each task is automatically learned without tuning parameters (i.e., parameter-free). As a result, such a learning strategy is efficient and robust to outliers. In addition, our method makes use of these features across multiple time points collaboratively. It further improves the classification performance of each task by exploiting more informative features unlike the ones obtained only from a single time point. These specific and common features in different time points from the PFC model are representative to facilitate the disease progression analysis and diagnosis. To further improve the effectiveness of classification, we fuse the features in multiple time points by the similarity network fusion (SNF) method [35], [36]. We evaluate our method in the rs-fMRI database of the AD Neuroimaging Initiative Phase-2 (ADNI-2), which includes 29 normal control (NC), 29 EMCI, and 18 LMCI in the two time points, baseline, and year 1. The experimental results demonstrate that our proposed framework achieves a quite promising diagnostic performance. The best identification performance of the proposed method for LMCI versus NC, EMCI versus NC, and EMCI versus LMCI is 87.23%, 82.76%, and 80.85%, respectively.

II. Methodology

A. Proposed Framework

The overview of our proposed method is shown in Fig. 1. The input is the rs-fMRI data with multiple time points. First, we use the preprocessed rs-fMRI data in multiple time points to build BFCNs by the FSN model. Second, we use the proposed PFC feature selection method to jointly learn the features constructed from the BFCN model. Finally, the selected features are fused by the SNF method, and the fused features are sent to the support vector machine (SVM) for classification [37].

B. Subjects and Data Acquisition

The data in this article is obtained from the ADNI database (http://adni.loni.usc.edu/). Specific and sensitive markers for the detection of early AD progression are designed to help scholars and clinical experts to develop innovative therapies and monitor their efficacy, which can effectively reduce the cost and time of clinical diagnosis. A large number of private companies and academic institutions are working together to build the ADNI database and subjects were recruited from more than 50 sites in the USA and Canada [38]–[40].

The data used in this article is acquired from the ADNI-2 database. The MRI scan data is obtained from the ADNI site, and the spatial distortion caused by B1-field inhomogeneity and gradient nonlinearity is corrected.

C. Image Preprocessing

All subjects are scanned by the 3.0T Philips Achieva at different centers and the parameters are defined as follows: TR/TE = 3000/30 mm, rollover angle = 80°, 140 volumes, imaging matrix = 64 × 64, 48 slices, and body thickness = 3.3 mm. We obtain the rs-fMRI data in 4-D spatiotemporal neuroimaging informatics technology initiative (NIFTI)

format. The data is preprocessed with the standard preprocessing procedure based on the statistical parametric mapping (SPM12) and data processing assistant for resting-state fMRI (DPARSF) software.

The preprocessing includes the following stages.

1) Discarding the first 10 rs-fMRI volumes of each subject before any further processing to keep the magnetization equal.
2) The remaining 130 volumes are corrected by the staggered sequence of slice collection, which takes advantage of the echo planar scan to ensure that the data on each slice corresponds to the unanimous point in time.
3) The interpolation time point is set as TR/2 so that the relative error of every TR is minimized. A rigid body spatial transformation and a least square approach are used to realign the slice timing to correct the rs-fMRI time series of each subject after the acquisition of time delay.
4) Removing head motion, the second last volume of each subject is used as the reference to which all subsequent volumes are realigned [41].
5) The rs-fMRI is divided into 90 brain regions using the automatic anatomical labeling (AAL) template.
6) A high-pass filter is used to refine the average rs-fMRI time series of each brain region. Furthermore, we regress out the head movement parameters, the cerebrospinal fluid, and mean BOLD time series of the white matter. We extract the mean of the BOLD signal in 90 ROIs as the original rs-fMRI signal (i.e., 90 nodes) [42].

D. Fused Sparse Network

In this article, the matrices are represented in bold capital letters, the vectors are in bold lowercase letters, and the scalars are in normal italic letters. Assuming that there are N subjects and $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_r, \ldots, \mathbf{x}_R] \in \mathbb{R}^{R\times N}$ denotes our input data, the AAL template is utilized to divide the brain into R ROIs and the rth ROI with a BOLD regional mean time series (M length) is represented by a response vector $\mathbf{x}_r^n = [x_{1r}^n, x_{2r}^n, \ldots, x_{Mr}^n]$. $\mathbf{A}_r^{\mathbf{n}}$ denotes all ROIs signal matrix except $\mathbf{x}_r^n$, $\mathbf{A}_r^{\mathbf{n}} = [\mathbf{x}_1^n, \ldots \mathbf{x}_{r-1}^n, \ldots \mathbf{x}_{r+1}^n, \ldots, \mathbf{x}_R^n]$, $\boldsymbol{w}_r^n \in \mathbb{R}^{R-1}$ is a weighting regression coefficient vector, and $\mathbf{W}_r = [\boldsymbol{w}_r^1, \ldots \boldsymbol{w}_r^n, \ldots, \boldsymbol{w}_r^N]$. The sparse networks are used to represent the brain functional connectivity that can be constructed using GCS and is defined as

$$J(\mathbf{W}_r) = \min_{W_r} \frac{1}{2}\sum_{n=1}^{N} \left\|\mathbf{x}_r^n - \mathbf{A}_r^n \boldsymbol{w}_r^n\right\|_2^2 + R_g(\mathbf{W}_r) \tag{1}$$

where $R_g(\mathbf{W}_r)$ is a group regularization given by

$$R_g(\mathbf{W}_r) = \lambda_1 \|\mathbf{W}_r\|_{2,1} = \lambda_1 \sum_{d=1}^{R-1} \left\|\boldsymbol{w}_r^d\right\|_2 \tag{2}$$

where λ_1 is the group regularization parameter, and $\|\mathbf{W}_r\|_{2,1}$ is the summation of the l_2-norm of $\boldsymbol{w}_r^n$. Specifically, we use the l_2-norm on the row vectors of the rth feature in the entire subjects by imposing the weights. We further adopt the l_1-norm to jointly select information by $R-1$ ROIs' weights. $\boldsymbol{w}_r^d$ is the dth row vector of $\mathbf{W}_r$. As a sparse regression network method, GCS ensures all models in the unequal group with identical connections. The l_2-norm is imposed on identical elements across the unequal matrix $\mathbf{W}_r$, which forces the weight corresponding to connections across the different subjects to be grouped together. The constraint imposes a common connection topology among the subjects and leverages the variation of connection weights among them. Therefore, the model is able to rebuild the target ROI using the remaining ROIs. Moreover, the reconstruction of each ROI is independent from others. However, the existing GCS model with penalty ignores the smoothing properties of different subjects within the model. To overcome this drawback, a novel model is devised to jointly learn the shared functional brain networks of each subject by the group sparse regularization and the smoothness regularization. The objective function is

$$J(\mathbf{W}_r) = \min_{W_r} \frac{1}{2}\sum_{n=1}^{N} \left\|\mathbf{x}_r^n - \mathbf{A}_r^n \boldsymbol{w}_r^n\right\|_2^2 + R_g(\mathbf{W}_r) + R_s(\mathbf{W}_r) \tag{3}$$

where $R_g(\mathbf{W}_r)$ is the group regularization, and $R_s(\mathbf{W}_r)$ denotes the smoothness regularization, which is denoted as

$$R_s(\mathbf{W}_r) = \lambda_2 \sum_{n=1}^{N-1} \left\|\boldsymbol{w}_r^n - \boldsymbol{w}_r^{n+1}\right\|_1 \tag{4}$$

where λ_2 is the parameter of smoothness regularization. The second term $\|\boldsymbol{w}_r^n - \boldsymbol{w}_r^{n+1}\|_1$ constrains the diversity between two consecutive weighting vectors from the same groups to be as small as possible. When the smoothness regularization parameter λ_2 is zero, the proposed method reproduces the GCS original method. Because of the l_1-norm utilized in the smoothness term, the weight vectors difference is encouraged since lots of zero components appear in the imparity vectors of weighting. In other words, due to the smoothing constraints, a large number of components from the adjacent weight vectors will be the same. The informative features will be selected due to nonzero weights in our task. We introduce the smoothness terms to smooth the connectivity coefficients of the subjects. In addition, we fuse the regularization terms in this learning model to impose a high level of constraints. We call this sparse learning model as FSN.

E. Optimization Algorithm

Both group and smoothness regularization are included in our objective function simultaneously. We use the iterative projected gradient descent algorithm to solve this equation. Specifically, the objective function in (3) is divided into the smoothness term

$$s(\mathbf{W}_r) = \min_{\mathbf{W}_r} \frac{1}{2}\sum_{n=1}^{N} \left\|\mathbf{x}_r^n - \mathbf{A}_r^n \boldsymbol{w}_r^n\right\|_2^2 \tag{5}$$

and the nonsmoothness term

$$n(\mathbf{W}_r) = \lambda_1 \|\mathbf{W}_r\|_{2,1} + \lambda_2 \sum_{n=1}^{N-1} \left\|\boldsymbol{w}_r^n - \boldsymbol{w}_r^{n+1}\right\|_1. \tag{6}$$

In iteration k, two steps are contained in the projected gradient descent. Let the gradient of $s(\mathbf{W}_r)$ at $\mathbf{W}_r^k$ denote $s'(\mathbf{W}_r^k)$,

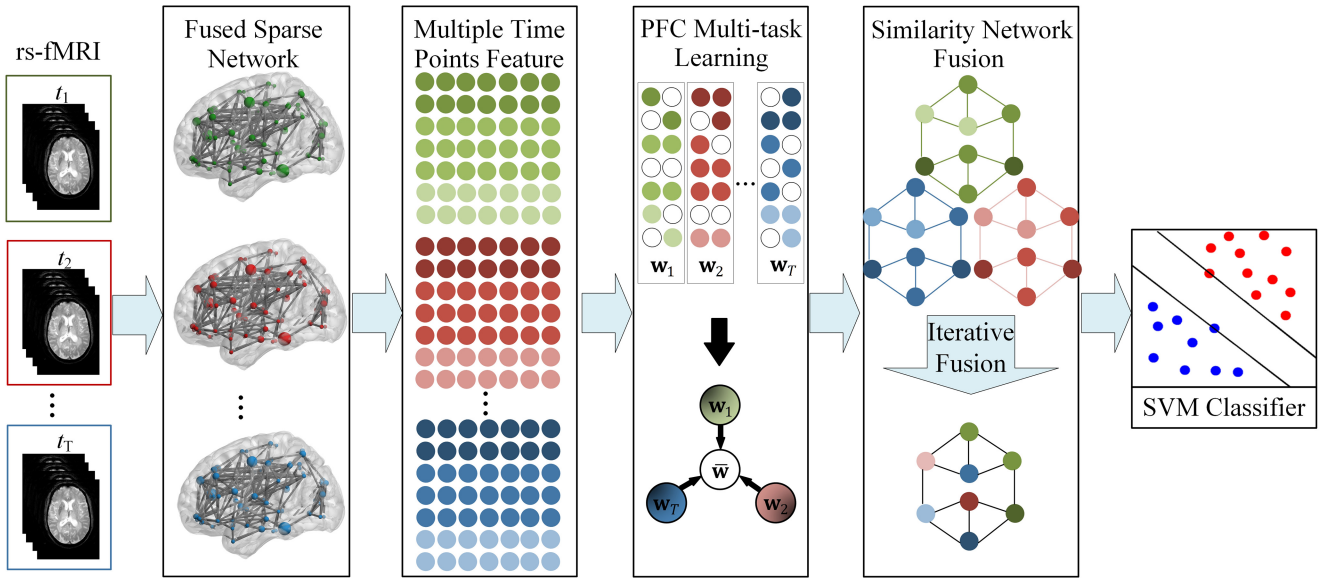

Fig. 1. Flowchart of the proposed method.

and the step size is denoted as γ_k and is determined via the line search. The first step is denoted as

$$\mathbf{V}_r^{(k)} = \mathbf{W}_r^k - \gamma_k s'\left(\mathbf{W}_r^k\right) \tag{7}$$

and the second step is represented as

$$\mathbf{W}_r^{k+1} = \arg\min \frac{1}{2}\left\|\mathbf{W}_r - \mathbf{V}_r^k\right\|_2^2 + n(\mathbf{W}_r). \tag{8}$$

For the nonsmoothness term $n(\mathbf{W}_r)$ in (8), we can sequentially calculate the proximal operator related to the group Lasso and fused Lasso constraints [43], [44].

An accelerated gradient descent method is used to further accelerate the above gradient [45]. We compute the search point $\mathbf{S}_r^k$ instead of performing gradient descent based on $\mathbf{W}_r^k$

$$\mathbf{S}_r^k = \mathbf{W}_r^k + \alpha_k\left(\mathbf{W}_r^k - \mathbf{W}_r^{k-1}\right) \tag{9}$$

where α_k is a predefined variable and $\mathbf{V}_r^k$ is

$$\mathbf{V}_r^k = \mathbf{S}_r^k - \gamma_i s'\left(\mathbf{S}_r^k\right). \tag{10}$$

Finally, the new approximate solution is obtained. The detailed solution is listed in Algorithm 1.

Algorithm 1: Solving (3)

Input: $\mathbf{x}_r^n \in \mathbb{R}^{M\times 1}$, $\mathbf{A}_r^{\mathbf{n}} \in \mathbb{R}^{M\times(R-1)}$, λ_1, λ_2;
Output: $\mathbf{W}_r \in \mathbb{R}^{(R-1)\times N}$;
Initialize $k = 0$, $\mathbf{S}_r^k$, $\mathbf{W}_r^k$ by an identity matrix;
Repeat
  Update matrix $\mathbf{V}_r^k$ by solving Eq. (7);
  Update matrix $\mathbf{S}_r^k$ by solving Eq. (9);
  Update matrix $\mathbf{W}_r^k$ by solving Eq. (8);
  $k = k + 1$;
Until it converges or a stop criterion is satisfied.;

F. Feature Extraction

The asymmetrical BFCN does not contribute to the final classification accuracy. Hence, $\mathbf{W}^* = (\mathbf{W}_n + \mathbf{W}_n^{\mathrm{T}})/2$ is defined to obtain the symmetry [46]. Local clustering coefficients of weighted graphs are utilized to extract features from each established BFCN. Specifically, the clustering coefficient of each node is computed in the network to quantify the probability of the nodes being connected to the neighboring nodes. Assuming that there are R nodes in a network, $\boldsymbol{w}_{ij}$ is the weight of the edge, which connects vertex i to vertex j. $\mathbf{v}_i$ denotes the set of vertices directly connected to vertex i, and the number of elements in $\mathbf{v}_i$ is defined as $|\mathbf{v}_i|$, we define the cluster coefficient of vertex i as

$$\rho_i = \frac{2\sum_{j:j\in\mathbf{v}_i}\left(\boldsymbol{w}_{ij}\right)^{\frac{1}{3}}}{|\mathbf{v}_i|(|\mathbf{v}_i|-1)}. \tag{11}$$

In this article, the AAL atlas with 90 ROIs is utilized for brain parcellation. Therefore, each network generates a feature vector containing 90 clustering coefficients [21], [22].

G. Parameter-Free Centralized Multitask Learning

The multitime point information for the longitudinal analysis is more desirable for improving the diagnostic accuracy of the disease as compared with the single time-point data. For this reason, we jointly learn features with multiple time points to simultaneously identify specific and common features, which helps to analyze the development of the disease. The specific features of different time points characterize the differences in the disease progression, while the common features describe the interaction effect of the disease progression. These groups of different time points collaboratively help each other to improve identification performance. It is possible to discover more biomarkers of biological significance unidentified in a model built with only one time-point data.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_t, \ldots, \mathbf{X}_T \in \mathbb{R}^{N\times R}$ be the FSN feature matrices at T time points. We divide the features of each time point

into a group. Each group of data is regarded as a task, and there are T tasks. Let $\boldsymbol{y}_t \in \mathbb{R}^N$ be a label vector at the t-th task, N is the number of subjects, and R is the dimension of feature, the weight matrix is $\mathbf{W} = [\boldsymbol{w}_1, \ldots \boldsymbol{w}_t, \ldots, \boldsymbol{w}_T] \in \mathbb{R}^{R\times T}$. The multitask learning via the least square loss function to select the informative features is given as

$$\min_{\boldsymbol{w}_t} \sum_{t=1}^{T} \left\|\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{w}_t\right\|_2^2 + \gamma\|\mathbf{W}\|_1 \tag{12}$$

where γ denotes a non-negative parameter controlling the sparsity ratio. In this article, the centralized regularization is used to penalize the variance of the coefficient vector, which can make different tasks to help each other collaboratively. The objective function is computed as

$$\min_{\boldsymbol{w}_t, \overline{\boldsymbol{w}}} \sum_{t=1}^{T} \sqrt{\left\|\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{w}_t\right\|_2^2 + \|\boldsymbol{w}_t - \overline{\boldsymbol{w}}\|_{2,1}} + \gamma\|\mathbf{W}\|_1. \tag{13}$$

To facilitate the optimization and find the variables $\boldsymbol{w}_t$ and $\overline{\boldsymbol{w}}$, we reformulate (13) as

$$\min_{\boldsymbol{w}_t, \overline{\boldsymbol{w}}} \sum_{t=1}^{T} \alpha_t\left(\left\|\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{w}_t\right\|_2^2 + \|\boldsymbol{w}_t - \overline{\boldsymbol{w}}\|_{2,1}\right) + \gamma\|\mathbf{W}\|_1 \tag{14a}$$

$$\alpha_t = \frac{1}{2\sqrt{\left\|\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{w}_t\right\|_2^2 + \|\boldsymbol{w}_t - \overline{\boldsymbol{w}}\|_{2,1}}}. \tag{14b}$$

α_t is optimized via sequentially taking the derivative with respect to the variables $\boldsymbol{w}_t$ and $\overline{\boldsymbol{w}}$. Therefore, we solve (13) by alternatively updating (14a) and (14b). To optimize them easily, we reformulate (14a) as

$$\min_{\boldsymbol{w}_t, \overline{\boldsymbol{w}}} \sum_{t=1}^{T} \alpha_t\left(\left\|\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{w}_t\right\|_2^2 + \beta_t\|\boldsymbol{w}_t - \overline{\boldsymbol{w}}\|_2^2\right) + \gamma\|\mathbf{W}\|_1 \tag{15a}$$

$$\beta_t = \frac{1}{2\sqrt{\|\boldsymbol{w}_t - \overline{\boldsymbol{w}}\|_2}} \tag{15b}$$

where $\overline{\boldsymbol{w}}$ is initialized as $(\mathbf{X}_t^T\mathbf{X}_t + 0.001\times\mathbf{I})^{-1}\mathbf{X}_t^T\boldsymbol{y}_t$, $\mathbf{I} \in \mathbb{R}^{(R\times R)}$ is the identity matrix, $\overline{\boldsymbol{w}}$ is the average of all $\boldsymbol{w}_t$, and $\alpha_t = (1/T)$, $\beta_t = (1/T)$, $t = 1, 2, \ldots, T$. α_t and β_t can be regarded as the weights of the centralized regularization. Note that both α_t and β_t are automatically acquired without tuning parameters. $\alpha_t\beta_t$ is the weight to reduce the variance of $\boldsymbol{w}_t$ in the term $\alpha_t\beta_t\|\boldsymbol{w}_t - \overline{\boldsymbol{w}}\|_2^2$, which makes all tasks similar. $\alpha_t\beta_t$ is used to measure the diversity and the flexibility of $\mathbf{X}_t$. If $\mathbf{X}_t$ is more similar to other tasks, the value of $\alpha_t\beta_t$ should be large enough to push $\boldsymbol{w}_t$ closer to $\overline{\boldsymbol{w}}$. In this case, $\mathbf{X}_t$ has less flexibility and more stability. Let $\widehat{\boldsymbol{y}}_t = [\sqrt{\alpha_t}\boldsymbol{y}_t^{\mathrm{T}}, \sqrt{\alpha_t\beta_t}\overline{\boldsymbol{w}}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{(n+R)\times 1}$, $\widehat{\mathbf{X}}_t = [\sqrt{\alpha_t}\mathbf{X}_t^{\mathrm{T}}, \sqrt{\alpha_t\beta_t}\mathbf{I}]^{\mathrm{T}} \in \mathbb{R}^{(n+R)\times R}$, (15) can be reformulated to

$$\min \sum_{t=1}^{T} \left\|\widehat{\boldsymbol{y}}_t - \widehat{\mathbf{X}}_t\boldsymbol{w}_t\right\|_2^2 + \gamma\|\mathbf{W}\|_1. \tag{16}$$

Equation (16) is a standard objective function of sparse multitask learning, which can be solved by the MALSAR toolbox [47], [48].

H. Similarity Network Fusion

Let the selected features of the ith and jth subjects be $\mathbf{f}_i$ and $\mathbf{f}_j$, respectively. The similarity network is represented as a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ and $\mathbf{E}$ denote the vertices and edges weighted by the similarity among the subjects, respectively. The pair of sample similarity (i.e., matrix $\mathbf{S}$) is calculated as

$$\mathbf{S}(i,j) = \exp\left(-\left(\frac{\rho^2(\mathbf{f}_i, \mathbf{f}_j)}{\mu\varepsilon_{i,j}}\right)\right) \tag{17}$$

$$\rho(\mathbf{f}_i, \mathbf{f}_j) = \sqrt{(\mathbf{f}_{i,1} - \mathbf{f}_{j,1})^2 + \cdots + (\mathbf{f}_{i,d} - \mathbf{f}_{j,d})^2} \tag{18}$$

where $\mathbf{S}(i,j)$ and $\rho(\mathbf{f}_i, \mathbf{f}_j)$ indicate the similarity and Euclidean distance between subjects i and j, respectively. μ is an empirically set hyperparameter and $\varepsilon_{i,j}$ is introduced to solve the scaling problem using

$$\varepsilon_{i,j} = \frac{\text{mean}(\rho(\mathbf{f}_i, \mathbf{N}_i)) + \text{mean}\big(\rho(\mathbf{f}_j, \mathbf{N}_j)\big) + \rho(\mathbf{f}_i, \mathbf{f}_j)}{3} \tag{19}$$

where $\mathbf{N}_i$ denotes a set of $\mathbf{f}_i$ and its neighbors in the graph, and $\text{mean}(\rho(\mathbf{f}_i, \mathbf{N}_i))$ is the average Euclidean distance of $\mathbf{f}_i$ from its neighbors. We construct a full kernel and a sparse kernel on the vertex set $\mathbf{V}$. A normalized weight matrix $\mathbf{P}$ with $\sum_j \mathbf{P}(i,j) = 1$ is denoted as

$$\mathbf{P}(i,j) = \begin{cases} \frac{\mathbf{S}(i,j)}{2\sum_{k\neq i}\mathbf{S}(i,k)} & i \neq j \\ \frac{1}{2} & i = j. \end{cases} \tag{20}$$

The local affinity using means of K-nearest neighbors is computed by

$$\mathbf{K}(i,j) = \begin{cases} \frac{\mathbf{S}(i,j)}{\sum_{in\mathbf{N}_i}\mathbf{S}(i,k)} & j \in \mathbf{N}_i \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

The similarity values of the non-neighboring points are set based on the assumption that the remote similarities are less reliable than the local ones. Here, $\mathbf{P}$ has the full information while $\mathbf{K}$ only contains the most similar ones for each subject. Supposing that there are T time points data with full and sparse kernels $\mathbf{P}^{(t)}$ and $\mathbf{K}^{(t)}$, $t = 1, 2, \ldots, T$, and $\mathbf{P}^{(t)}_{m=0} = \mathbf{P}^{(t)}$ is the initial state matrices at $m = 0$, the full kernel matrices are then updated iteratively as follows:

$$\mathbf{P}^{(t)}_{m+1} = \mathbf{K}^{(t)} \times \left(\frac{\sum_{l\neq t}\mathbf{P}^{(l)}_m}{T-1}\right) \times \left(\mathbf{K}^{(t)}\right)^{\mathrm{T}}, \ t, \ l = 1, 2, \ldots, T \tag{22}$$

where T is the transpose operation. $\mathbf{P}^{(t)}_{m+1}$ denotes the state matrix of the t-th time points data after M iterations. Then, the state matrix of all time points is calculated by

$$\mathbf{P} = \frac{\sum_{t=1}^{T}\mathbf{P}^{(t)}_M}{T}. \tag{23}$$

SNF starts from $\mathbf{P}$ as the initial state and use $\mathbf{K}$ as the kernel matrix in the fusion process to capture the local structure of graphs and computational efficiency.

TABLE I
DEMOGRAPHIC INFORMATION OF THE USED SUBJECTS

Group	NC	EMCI	LMCI
Baseline_Sex(male/female)	29(12/17)	29(14/15)	18(13/5)
Baseline_Age(mean $\pm$ SD)	74.6 $\pm$ 5.26	72.5 $\pm$ 7.17	70.6 $\pm$ 8.48
Year1_Sex(male/female)	29(12/17)	29(14/15)	18(13/5)
Year1_Age(mean $\pm$ SD)	75.6 $\pm$ 5.22	73.5 $\pm$ 7.17	71.6 $\pm$ 8.46

III. EXPERIMENTS AND RESULTS

A. Experimental Setup

We conduct numerous experiments on the public ADNI dataset (http://adni.loni.usc.edu/) to verify the performance of the proposed method. There are 76 subjects (including 47 MCI patients and 29 NC) acquired from the ADNI-2 database. Out of the 47 MCI patients, 18 are LMCI patients (an impairment on cognitive testing is usually defined as the performance of 1.5 standard deviations (SD) below the normative mean) and 29 are EMCI patients (an impairment on cognitive testing is usually defined as the performance between 1.0 and 1.5 SD below the normative mean on a standard test). All subjects have the rs-fMRI data with two time points (baseline and year1). The features are the BOLD signal in every ROI. In this article, we jointly identify the different classes (e.g., LMCI, EMCI, and NC) [49]. Table I summarizes all the demographic information of all used subjects.

Since we have a small amount of data, the leave-one-out cross-validation (LOOCV) strategy is used to assess our proposed method. For N subjects, one of them is left out for testing, and the remaining $N-1$ subjects are utilized for training. The hyperparameters in each method are empirically set by the greedy search to identify the optimal parameters. For example, we obtain the values of λ_1 and λ_2 through the exhaustive search strategy from 10^{-5} to 10^2. Five quantitative metrics are utilized to evaluate the diagnosis performance: 1) accuracy (ACC); 2) area under receiver operating characteristic curve (AUC); 3) Precision (PRE); 4) Recall (REC); and 5) balanced accuracy (BAC). The SLEP and LibSVM toolboxes are used to construct sparse representation and classification, respectively [50]. To confirm whether our method performs statistically better than the selected methods, we perform paired t-tests on the ACCs of our method and other methods. The p-values (P) are reported.

There are $N-1$ corresponding testing subjects and $N-1$ different training subsets. The combination of regularization parameters that give the best performance is selected as the optimal parameters. $N-1$ classifiers are used to classify the completely unknown testing subject. A majority voting strategy is used for the final classification decision. Each subject in the dataset will be picked out for testing, so that the above process is repeated by N times. Then, the overall cross-validation classification results are computed [34].

To assess the effectiveness of our proposed framework, we conduct three groups (LMCI versus NC, EMCI versus NC, and EMCI versus LMCI) of experiments on the multiple time points (baseline and year1) data of the ADNI-2 dataset. In addition, we compare our results with the recent related methods in literature, namely, PCR, PCL, PCI, PCP, PCAE, PCPF, GCSR, GCSL, GCSI, GCSP, GCSAE, GCSPF, FSNR, FSNL, FSNI, FSNP, FSNAE, and FSNPF.

PCR: PCR is a framework, which uses PC to construct the BFCN and the regression method to select the effective features for classification.

PCL: PCL denotes a framework that uses PC to construct the BFCN and Lasso for feature selection.

PCI: PCI is a method that utilizes PC to generate the BFCN and independent component analysis (ICA) to reduce the data dimension for classification.

PCP: PCP denotes a method, which uses PC to construct the BFCN and principal components analysis (PCA) to reduce the features for training classifier.

PCAE: PCAE is a framework that uses PC to construct the BFCN and the autoencoder (AE) method to reduce the feature dimension for classification.

PCPF: PCPF is a BFCN constructed from PC and feature selection is selected by the PFC method. The PFC jointly learns the features with multiple time points.

GCSR: GCSR denotes a framework that uses GCS to construct the BFCN and adopts the regression method to select the effective features to train the classifier.

GCSL: GCSL is a framework that the BFCN is constructed by GCS and the feature is selected by Lasso for training classifier.

GCSI: GCSI is a method that utilizes GCS to generate the BFCN and ICA is used to reduce the data dimension for classification.

GCSP: GCSP denotes a framework that the BFCN is constructed by GCS and PCA is used to reduce the features for training classifier.

GCSAE: GCSAE denotes a framework that GCS is used to construct the BFCN and the AE method is used to reduce the feature dimensions to train the classifier.

GCSPF: GCSPF is a framework that the BFCN is constructed by GCS and feature is selected by the PFC multiple task model.

FSNR: FSNR denotes a framework that FSN is used to construct the BFCN and the regression method is used to select the effective features.

FSNL: FSNL is a framework that the BFCN is constructed by the proposed FSN and the features are selected by the Lasso method.

FSNI: FSNI is a method that BFCN is constructed by FSN and ICA is used to reduce the data dimension for classification.

FSNP: FSNP denotes a method that FSN is used for the BFCN construction and PCA is used to reduce the features for training classifier.

FSNAE: FSNAE is a framework that the BFCN is constructed by the proposed FSN and the AE method is adopted to reduce the feature dimensions.

FSNPF: FSNPF is a framework that the BFCN is built by FSN. The PFC feature selection method is employed to jointly learn the features with multiple time points, which is our proposed method.

B. Effect of the FSN Functional Connectivity Network

A good BFCN model helps to construct relationships among the ROIs. To reveal the effect of our proposed FSN

TABLE II
CLASSIFICATION RESULTS OF THE PROPOSED METHOD AND THE COMPETING METHODS ON BASELINE, YEAR1, AND FUSION (%)

Data	BFCN	Feature Selection	LMCI vs. NC					EMCI vs. NC					EMCI vs. LMCI				
			ACC	AUC	PRE	REC	BAC	ACC	AUC	PRE	REC	BAC	ACC	AUC	PRE	REC	BAC
Baseline	PC	Regression	55.32	60.34	42.30	56.25	51.15	53.45	58.03	57.51	53.95	53.45	51.06	47.13	34.14	48.26	46.65
		Lasso	63.83	48.66	41.13	49.19	53.83	55.17	59.93	60.56	54.88	55.17	59.57	57.47	45.18	54.51	52.49
		ICA	61.07	54.21	18.90	26.97	50.00	53.45	57.19	52.59	53.54	53.45	61.70	54.21	44.94	48.50	50.00
		PCA	53.19	38.12	34.90	42.82	49.43	53.45	53.86	51.73	51.90	53.45	55.32	38.12	38.58	42.82	49.04
		AE	55.32	53.83	40.22	52.31	51.15	50.00	51.72	52.52	50.85	50.00	57.45	52.68	39.51	51.62	51.82
		PFC	68.09	79.12	61.61	67.59	59.39	60.34	60.40	61.62	55.11	60.34	61.70	34.67	31.64	40.74	50.00
	GCS	Regression	63.83	53.45	45.91	52.08	56.99	60.34	61.83	57.54	55.82	60.34	61.70	51.34	40.94	50.81	56.32
		Lasso	70.21	67.24	48.64	60.42	64.27	63.79	65.64	55.58	57.69	63.79	65.96	54.41	44.11	52.66	58.72
		ICA	61.07	52.11	40.69	51.27	50.00	46.55	48.28	51.15	49.15	46.55	61.07	40.23	31.97	44.10	50.00
		PCA	61.07	50.00	10.81	19.79	50.00	48.28	46.61	50.06	48.33	48.28	65.96	50.38	45.90	50.23	57.66
		AE	65.96	52.68	46.34	51.62	61.88	53.45	52.91	53.90	51.43	53.45	63.83	63.79	49.18	58.33	58.05
		PFC	70.21	65.52	53.89	59.37	61.11	55.17	56.12	55.58	53.01	55.17	65.96	43.49	41.63	46.06	56.61
	FSN	Regression	70.21	68.77	55.74	61.34	65.33	62.07	66.23	64.74	57.98	62.07	59.57	42.53	34.45	45.49	49.33
		Lasso	78.72	78.93	37.94	59.38	**82.76**	72.41	**76.81**	**70.85**	**63.18**	72.41	70.21	60.15	46.50	56.13	75.86
		ICA	74.47	67.05	57.16	60.30	68.77	58.62	60.76	56.09	55.29	58.62	61.70	50.96	25.30	26.50	50.00
		PCA	74.47	72.41	57.04	63.54	72.99	70.69	76.81	69.96	63.18	70.69	65.96	54.79	45.90	52.89	61.88
		AE	72.34	76.25	55.50	65.86	68.10	62.07	66.71	63.01	58.21	62.07	68.09	76.44	**60.41**	65.97	63.60
		PFC	**82.98**	**94.06**	**71.43**	**76.62**	79.89	**77.59**	68.49	70.33	59.09	**77.59**	**76.60**	**76.82**	59.91	**66.20**	**81.03**
Year1	PC	Regression	61.70	55.56	44.71	53.36	53.16	43.10	31.03	37.04	40.68	43.10	55.32	50.57	38.09	50.35	47.99
		Lasso	68.09	63.79	51.77	58.33	59.39	62.07	63.50	58.76	56.63	62.07	57.45	56.51	42.73	53.94	51.82
		ICA	61.07	45.79	31.87	47.45	50.00	31.03	38.17	43.56	44.18	31.03	61.07	47.32	42.41	48.38	50.00
		PCA	63.83	34.67	28.97	40.74	52.78	62.07	72.06	65.11	60.84	62.07	63.83	48.08	44.63	48.84	54.89
		AE	63.83	59.58	48.04	55.79	58.05	51.72	41.94	40.83	46.05	51.72	61.70	58.81	55.39	56.71	59.48
		PFC	72.34	53.26	47.52	51.97	63.89	55.17	55.77	55.17	52.83	55.17	63.83	64.94	51.38	59.03	54.89
	GCS	Regression	65.96	61.88	44.05	57.18	61.88	60.34	57.19	52.56	53.54	60.34	61.70	66.28	49.89	59.84	50.00
		Lasso	76.60	74.33	56.33	64.70	71.55	67.24	58.86	52.30	54.35	67.24	63.83	70.50	53.07	62.38	55.94
		ICA	61.07	39.27	32.47	43.52	50.00	58.62	45.18	46.17	47.63	58.62	61.07	45.02	35.90	46.99	50.00
		PCA	65.96	67.62	57.49	60.65	55.56	58.62	62.19	57.57	55.99	58.62	63.83	50.57	45.59	50.35	56.99
		AE	70.21	69.92	58.60	62.04	63.38	67.24	65.76	61.72	57.74	67.24	59.57	49.23	41.92	49.54	52.49
		PFC	76.60	65.90	55.30	59.61	69.44	70.69	74.08	69.12	61.84	70.69	68.09	58.62	49.57	55.21	60.44
	FSN	Regression	72.34	74.52	56.09	64.81	67.05	65.52	76.93	73.72	63.24	65.52	61.70	62.26	51.37	57.41	56.32
		Lasso	80.85	90.23	68.58	74.31	**84.48**	74.14	89.54	**79.71**	69.43	74.14	70.21	70.50	51.78	62.38	66.38
		ICA	78.72	85.06	65.62	71.18	75.38	70.69	80.14	74.53	64.82	70.69	68.09	70.50	55.38	62.38	62.55
		PCA	76.60	92.53	68.86	75.69	71.55	72.41	79.90	70.99	64.70	72.41	65.96	62.64	50.33	57.64	58.72
		AE	74.47	83.52	62.67	70.25	75.10	68.97	70.87	66.87	60.26	68.97	68.09	61.11	55.39	56.71	60.44
		PFC	**85.11**	**93.87**	**71.88**	**76.50**	80.56	**81.03**	**90.13**	79.58	**69.73**	**81.03**	**78.72**	**94.06**	**71.48**	**76.62**	**80.65**
Fusion	PC	Regression	61.70	38.12	33.12	42.82	52.11	56.90	59.93	59.67	54.88	56.90	57.45	50.96	37.54	50.58	47.61
		Lasso	53.19	34.67	30.65	40.74	44.16	48.28	52.56	52.36	51.26	48.28	61.70	45.98	24.19	38.08	50.00
		ICA	61.70	45.59	32.80	47.34	50.00	60.34	59.69	51.54	54.76	60.34	61.07	49.43	40.02	49.65	50.00
		PCA	61.70	42.15	31.43	45.25	50.00	51.72	49.82	46.09	49.91	51.72	61.70	50.57	36.36	50.35	50.00
		AE	61.70	29.89	29.29	37.85	50.00	53.45	49.23	51.43	49.62	53.45	61.07	47.32	41.53	48.38	50.00
		PFC	68.09	57.85	47.66	54.75	59.39	65.52	64.92	60.69	57.33	65.52	61.70	48.23	33.85	44.10	51.05
	GCS	Regression	61.70	42.72	30.44	45.60	50.00	51.72	45.66	47.96	47.87	51.72	61.70	25.10	26.81	34.95	50.00
		Lasso	59.57	42.72	31.69	45.60	48.28	51.72	45.18	45.28	47.63	51.72	68.09	51.34	41.33	50.81	64.66
		ICA	63.38	48.66	39.26	49.19	52.78	53.45	52.68	59.18	51.32	53.45	61.70	41.57	28.39	44.91	50.00
		PCA	61.70	5.36	12.67	23.03	50.00	53.45	48.51	46.23	49.27	53.45	61.70	60.34	38.65	56.25	50.00
		AE	61.70	26.82	21.56	36.00	50.00	56.90	51.72	60.42	58.80	59.60	61.70	47.70	35.87	48.61	50.00
		PFC	61.70	26.82	22.48	34.72	50.00	56.90	67.90	53.08	53.01	56.90	61.70	48.23	25.63	36.46	51.05
	FSN	Regression	74.47	80.27	62.36	68.29	67.72	68.97	72.77	68.45	61.19	68.97	63.83	51.53	39.31	50.93	59.10
		Lasso	82.98	89.08	67.92	73.61	80.94	75.86	83.83	77.44	66.63	75.86	74.47	75.86	64.38	65.62	66.67
		ICA	63.83	49.62	43.90	49.77	52.78	68.97	78.00	70.64	63.76	68.97	72.34	67.43	50.24	60.53	67.05
		PCA	76.60	83.33	68.59	70.14	77.87	77.59	86.44	77.15	67.91	77.59	70.21	57.28	49.05	54.40	64.27
		AE	76.60	89.66	65.79	73.96	71.55	72.41	85.02	76.39	67.21	72.41	70.21	67.05	57.55	60.30	61.11
		PFC	**87.23**	**92.34**	**69.13**	**75.58**	**87.55**	**82.76**	**88.23**	**79.38**	**68.79**	**82.76**	**80.85**	**84.87**	**68.25**	**71.06**	**75.00**

network, we compare the performance to the typical FC networks, such as PC and GCS. Table II summarizes the comparison results in terms of various metrics (boldfaces represent the best performances). Obviously, the FSN model achieves the best performance. In the classification task of LMCI versus NC, the highest accuracy of our FSN model is 82.98% and 85.11% on the baseline and year1 datasets, respectively, which are higher than the PC and GCS models, respectively.

Similarly, in the classification task of EMCI versus NC, our FSN model achieves an accuracy of 77.59% and 81.03% on the baseline and year1 datasets, respectively, which are higher than the PC and GCS models. For the classification task of EMCI versus LMCI, our FSN model obtains an accuracy of 76.00% and 78.72% on the baseline and year1 datasets, respectively, which are higher than the PC and GCS models. The above results demonstrate that our FSN model is effective and outperforms the other competing methods.

C. Effect of the PFC Feature Selection Method

Dimensionality reduction is a very important step in biomedical applications. To evaluate the effectiveness of the PFC feature selection method, we compare our method with the typical feature selection methods, such as Lasso, regression, ICA, PCA, and AE (Table II). In the classification task of LMCI versus NC, the accuracies of our PFC feature selection method are 82.98% and 85.11% on the baseline and year 1 datasets, respectively. In the classification task of EMCI versus NC, our FSNPF model achieves 77.59% and 81.03% in the baseline and year1 datasets, respectively. In the classification task of EMCI versus LMCI, the highest accuracies of our FSNPF model are 76.00% and 78.72% in the baseline and year1 dataset, respectively. The results demonstrate that the PFC method is quite effective.

The above results demonstrate that the effectiveness of our PFC method in selecting the most discriminative features. Among the top-10 brain regions selected from LMCI versus NC classification, the discovered common regions on the baseline and year1 data are parahippocampal gyrus, amygdala, middle frontal gyrus, median cingulate, superior frontal gyrus, and the special regions are percental gyrus, superior parietal gyrus, olfactory cortex, inferior temporal gyrus, and cuneus.

Among the top-10 brain regions selected from EMCI versus NC classification, the discovered common regions on the baseline and year1 datasets are anterior cingulate, thalamus, paracentral lobule, lenticular nucleus, and supplementary motor area and the special regions are superior parietal gyrus, inferior occipital gyrus, inferior frontal gyrus, postcentral gyrus, and inferior temporal gyrus. It is worth noting that these brain regions are consistent with the previous MCI classification studies (e.g., inferior frontal gyrus, olfactory cortex, and anterior cingulate in [34], and middle frontal gyrus, amygdala, and inferior temporal gyrus in [22]). The common regions have a strong correlation with MCI, which may be the potential factors causing the disease.

D. Fusion via SNF

It is obvious that an effective fusion method can improve the classification accuracy. To reveal the fusion results via SNF, we compare the performance based on two different time points, baseline, and year1 datasets. From Table II, we can see that our proposed FSNPF framework yields a better classification performance than the single time-point data. Specifically, the highest classification accuracy for LMCI versus NC, EMCI versus NC, and EMCI versus LMCI is 87.23%, 82.76%, and 80.85%, respectively. The fusion results demonstrate that the SNF method is quite effective. Fig. 2 shows the accuracy results of various time points for different methods using SVM classifiers. Fig. 3 shows the ROC curves of various methods and Fig. 4 indicates the Precision–Recall curves of different methods using SVM. To confirm whether our method performs statistically better than the selected methods, we perform the *t*-tests on both our method and other methods in terms of

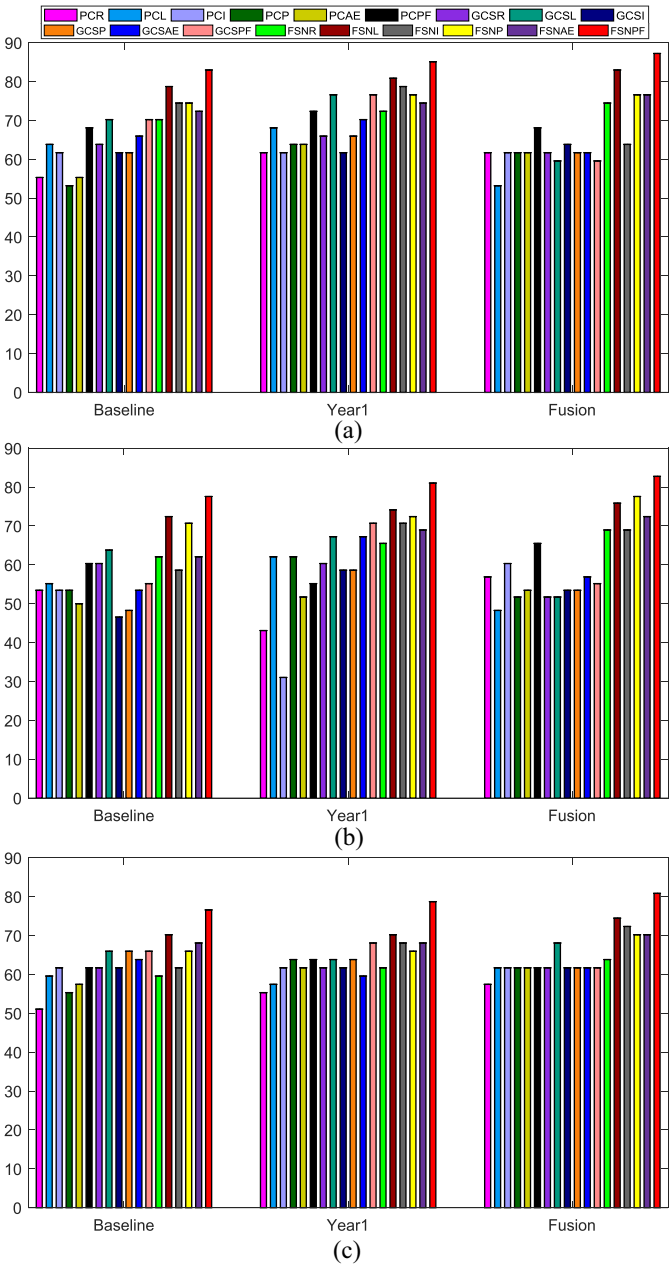

Fig. 2. Classification accuracy results of different methods. (a) LMCI versus NC. (b) EMCI versus NC. (c) EMCI versus LMCI.

ACC results under different scenarios. In our statistical analysis results, all the *P* values in all scenarios are lower than 0.01, which indicates that our method is statistically better than the listed method.

We can see that our proposed FSN model outperforms PC and GCS in terms of classification results based on single time point and multiple time points. From the comparison results of different feature selection methods, it is observed that the PFC feature selection method is the most effective one. The main reason is that we use the PFC method for feature selection, which not only can achieve a high accuracy but also can analyze lesion areas in multiple time-point data.

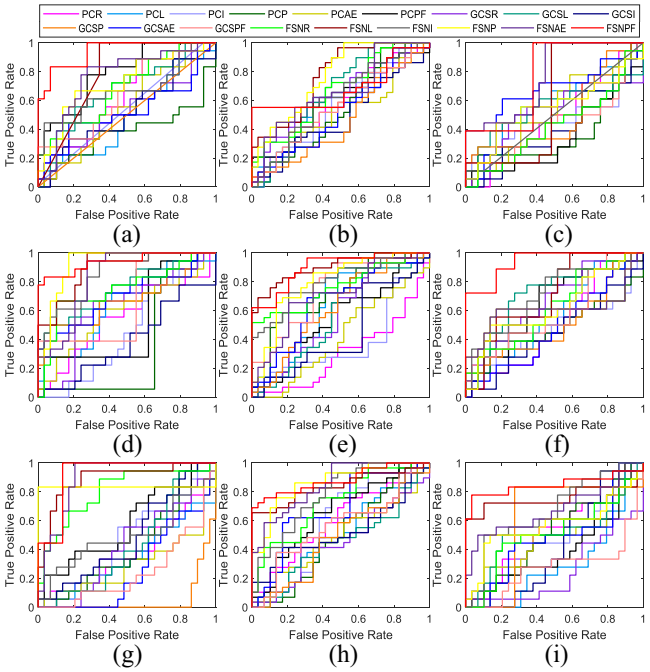

Fig. 3. ROC curves of various methods. (a) LMCI versus NC–Baseline. (b) EMCI versus NC–Baseline. (c) EMCI versus LMCI–Baseline. (d) LMCI versus NC–Year1. (e) EMCI versus NC–Year1. (f) EMCI versus LMCI–Year1. (g) LMCI versus NC–Fusion. (h) EMCI versus NC–Fusion. (i) EMCI versus LMCI–Fusion.

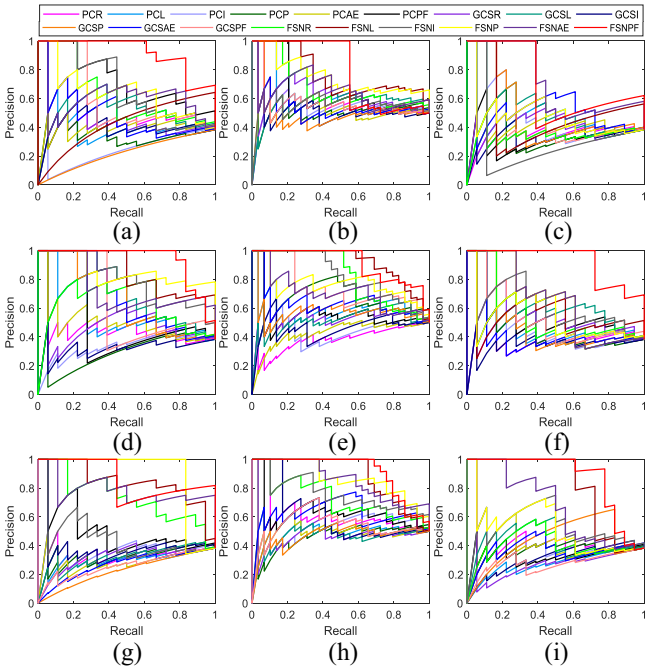

Fig. 4. Precision–Recall curves of various methods. (a) LMCI versus NC–Baseline. (b) EMCI versus NC–Baseline. (c) EMCI versus LMCI–Baseline. (d) LMCI versus NC–Year1. (e) EMCI versus NC–Year1. (f) EMCI versus LMCI–Year1. (g) LMCI versus NC–Fusion. (h) EMCI versus NC–Fusion. (i) EMCI versus LMCI–Fusion.

The experimental results reveal that the features fused with multitime points are more effective than those with single time points in brain disease diagnosis. The primary explanation for the decrease in classification accuracy is that the PC and GCS

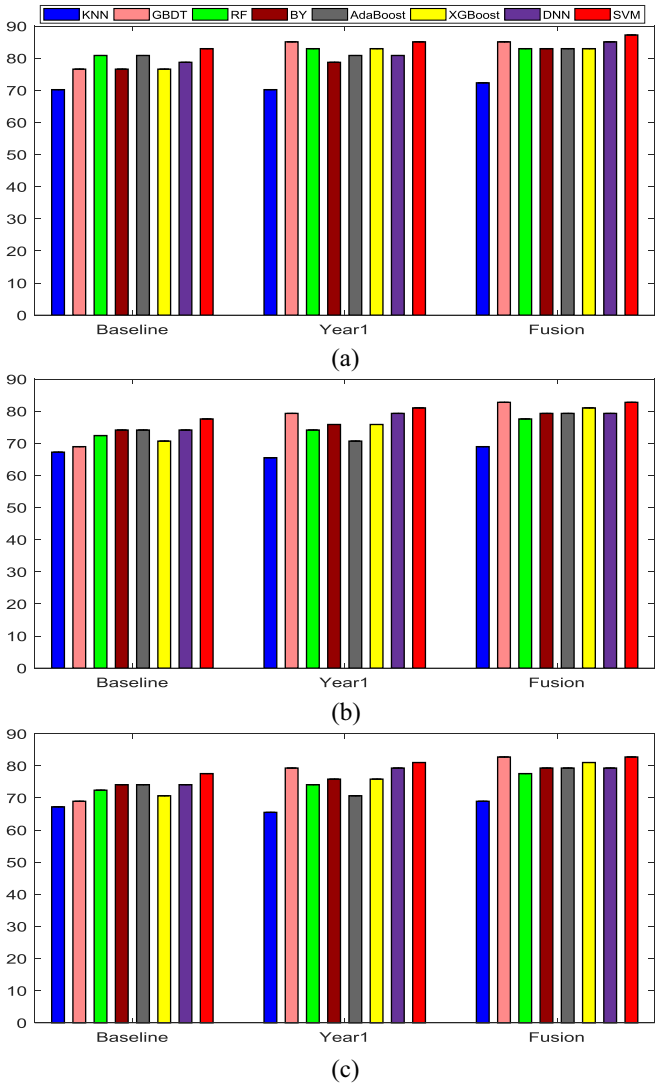

Fig. 5. Classification accuracy results of different classifiers. (a) LMCI versus NC. (b) EMCI versus NC. (c) EMCI versus LMCI.

models cannot accurately and effectively express the relationship between brain ROIs. This also proves that our proposed FSN model is promising. However, not all fused multitime features by SNF are useful for the disease diagnosis. Some single time-point features extracted from PC and GCS functional connectivity networks can gain higher classification accuracy than the fused multitime point features.

E. Classifiers Comparison

Eight classifiers are used in our experiments, including K-nearest neighbor (KNN), gradient boosting decision tree (GBDT), random forest (RF), Bayes (BY), adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), deep neural networks (DNNs), and SVM [51]–[53]. We use the FSNPF method to learn the features that are then sent to different classifiers.

The results are shown in Table III. Fig. 5 shows the FSNPF method on different classifiers' accuracy results in terms of various time points. Fig. 2 shows the different methods on

TABLE III
CLASSIFICATION RESULTS OF THE PROPOSED METHOD WITH DIFFERENT CLASSIFIERS ON BASELINE, YEAR1, AND FUSION (%)

Data	Classifier	LMCI vs. NC					EMCI vs. NC					EMCI vs. LMCI				
		ACC	AUC	PRE	REC	BAC	ACC	AUC	PRE	REC	BAC	ACC	AUC	PRE	REC	BAC
Baseline	KNN	70.21	66.38	35.11	44.79	66.38	67.24	67.24	42.00	47.63	67.24	68.09	65.71	33.05	45.25	65.71
	GBDT	76.60	82.76	56.62	61.46	71.55	68.97	80.08	68.94	62.89	68.97	74.47	73.56	51.81	60.07	72.99
	RF	80.85	85.34	67.28	70.72	78.16	72.41	**88.64**	**77.02**	**67.86**	72.41	72.32	68.10	55.17	59.49	67.05
	BY	76.60	80.46	54.84	68.40	78.93	74.14	81.81	74.22	65.63	74.14	76.60	**82.95**	**64.58**	**69.91**	75.77
	AdaBoost	80.85	87.36	67.89	72.57	**80.27**	74.14	75.15	62.09	62.36	74.14	72.34	70.31	54.10	62.27	70.21
	XGBoost	76.60	80.08	63.11	68.17	71.55	70.69	68.25	62.75	58.97	70.69	72.34	74.52	55.84	64.81	68.10
	DNN	78.72	74.54	54.72	59.72	75.38	74.14	85.02	73.85	60.60	74.14	74.47	81.32	56.92	67.82	71.93
	SVM	**82.98**	**94.06**	**71.43**	**76.62**	79.89	**77.59**	68.49	70.33	59.09	**77.59**	**76.60**	76.82	59.91	66.20	**81.03**
Year1	KNN	70.21	66.38	35.11	44.79	66.38	65.52	65.52	41.08	46.46	65.52	68.09	67.82	32.56	47.57	67.82
	GBDT	85.11	83.33	68.59	70.14	**84.77**	79.31	83.17	69.95	63.65	79.31	76.60	76.82	44.39	58.56	75.77
	RF	82.98	86.30	66.07	70.72	79.89	74.14	80.20	73.32	63.59	74.14	78.72	88.70	67.31	72.34	75.38
	BY	78.78	92.15	70.86	75.46	79.60	75.86	86.56	72.90	67.97	75.86	68.09	73.56	60.47	64.24	68.87
	AdaBoost	80.85	**96.17**	**72.53**	**77.89**	79.21	70.69	74.55	69.41	62.07	70.69	76.60	79.89	62.08	68.06	72.36
	XGBoost	82.98	90.04	67.43	74.19	81.99	75.86	81.09	75.45	65.28	75.86	74.47	77.59	57.48	66.67	68.77
	DNN	80.85	86.78	58.17	71.30	80.27	79.31	83.17	54.30	61.25	79.31	74.47	76.82	56.83	62.85	70.88
	SVM	**85.11**	93.87	71.88	76.50	80.56	**81.03**	**90.13**	**79.38**	**68.79**	**81.03**	**78.72**	**94.06**	71.48	**76.62**	**80.65**
Fusion	KNN	72.34	68.10	37.40	45.83	68.10	68.97	68.97	42.90	48.80	68.97	76.60	75.77	40.24	54.40	75.77
	GBDT	85.11	83.33	68.59	70.14	84.77	82.76	83.23	60.41	61.72	82.76	78.72	83.62	51.90	64.12	78.54
	RF	82.98	87.64	66.84	71.88	80.94	77.59	86.50	75.49	66.80	77.59	80.85	86.40	66.44	71.18	77.11
	BY	82.98	83.05	46.17	61.46	83.05	79.31	79.31	54.35	56.28	79.31	78.72	78.54	41.86	57.06	78.54
	AdaBoost	82.98	90.23	67.80	74.31	83.05	79.31	89.89	**80.06**	69.61	79.31	76.60	76.25	61.10	65.86	74.71
	XGBoost	82.98	90.04	67.43	74.19	81.99	81.03	**91.20**	79.61	70.25	81.03	78.72	81.23	62.85	68.87	75.38
	DNN	85.11	**94.06**	**71.21**	**76.62**	84.77	79.31	83.71	74.33	66.57	79.31	78.72	**86.97**	63.44	**72.34**	**78.54**
	SVM	**87.23**	92.34	69.13	75.58	**87.55**	**82.76**	88.23	79.58	**69.73**	**82.76**	**80.85**	84.87	**68.25**	71.06	75.00

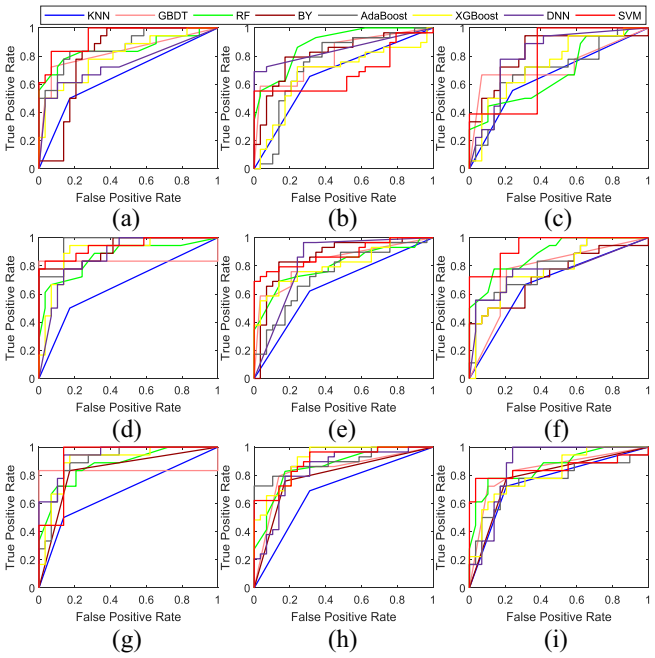

Fig. 6. ROC curves of various classifiers. (a) LMCI versus NC–Baseline. (b) EMCI versus NC–Baseline. (c) EMCI versus LMCI–Baseline. (d) LMCI versus NC–Year1. (e) EMCI versus NC–Year1. (f) EMCI versus LMCI–Year1. (g) LMCI versus NC–Fusion. (h) EMCI versus NC–Fusion. (i) EMCI versus LMCI–Fusion.

SVM classifiers' accuracy results in terms of multiple time-point datasets. Fig. 6 shows the ROC curves of the FSNPF method on different classifiers. Fig. 7 indicates the Precision–Recall curves of the FSNPF method on different classifiers. We can clearly see that the classification performance of SVM classifier is more stable.

TABLE IV
ALGORITHM COMPARISONS (%)

Algorithm	Subject	Classifier	Modality	ACC
Wee *et al.* [41]	29EMCI+30NC	SVM	rs-fMRI	79.66
Guo *et al.* [20]	33EMCI+28NC	SVM	rs-fMRI	72.10
Guo *et al.* [20]	32LMCI+28NC	SVM	rs-fMRI	78.63
Prasad *et al.* [54]	74EMCI+38LMCI	SVM	MRI	64.30
Proposed	18LMCI+29NC	SVM	rs-fMRI	**87.23**
Proposed	29EMCI+29NC	SVM	rs-fMRI	**82.76**
Proposed	29EMCI+18LMCI	SVM	rs-fMRI	**80.85**

SVM obtains an accuracy of 87.23%, 82.76%, and 80.85% for LMCI versus NC, EMCI versus NC, and EMCI versus LMCI classification, respectively. In most of the cases, the SVM classifier can get the highest classification accuracy. SVM classifiers are more suitable for the classification of small datasets. Although some quantitative measurements of other classifiers are also very good, we choose SVM as the final classifier.

F. Algorithm Comparison

The state-of-the-art competing methods are compared with our proposed FSNPF framework based on the ADNI-2 database. The LMCI versus NC, EMCI versus NC, and EMCI versus LMCI classification results are listed in Table IV. The FSNPF obtains an accuracy of 87.23%, 82.76%, and 80.85% for LMCI versus NC classification, EMCI versus NC, and EMCI versus LMCI classification, respectively. Note that our method achieves the best classification performance compared with the listed methods. Overall, the FSNPF achieves the best accuracy compared with the related algorithms. It is worth noting that our method takes an advantage of the PFC feature selection method, which can simultaneously learn the features

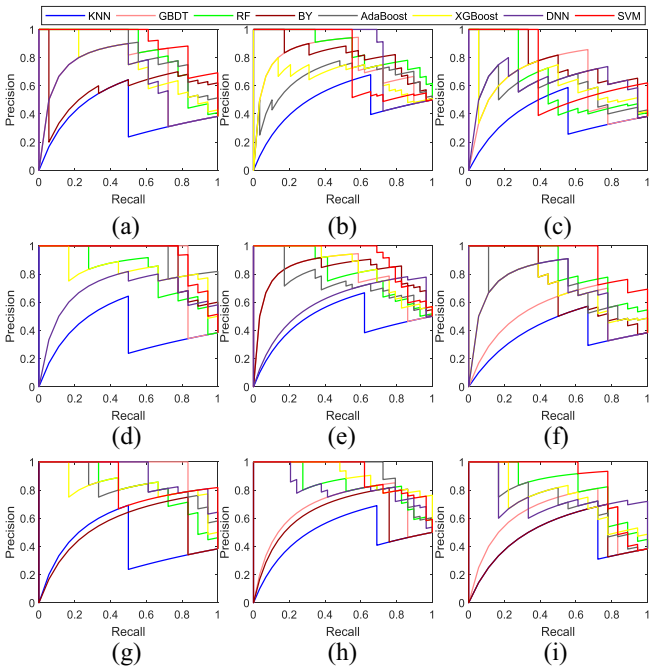

Fig. 7. Precision–Recall curves of various classifiers. (a) LMCI versus NC–Baseline. (b) EMCI versus NC–Baseline. (c) EMCI versus LMCI–Baseline. (d) LMCI versus NC–Year1. (e) EMCI versus NC–Year1. (f) EMCI versus LMCI–Year1. (g) LMCI versus NC–Fusion. (h) EMCI versus NC–Fusion. (i) EMCI versus LMCI–Fusion.

of different time points and select the most representative features extracted by the parameter-free adjustment.

IV. Discussion

Since the brain area interacts mainly with only a few other regions, each region can be expressed accurately and effectively using a small number of regions [22]. Fig. 8 shows the brain functional connection networks, where the line thickness represents the weights of brain area. Fig. 9 illustrates the BFCN with the most discriminative 30 brain regions, where the green and blue lines denote the baseline and year1 datasets, respectively.

We set the same threshold at two time points to pick out the brain function connection. We randomly select LMCI and EMCI patients from the database to compare the BFCN at different time points. Obviously, the baseline connection is thicker than the year1 connection. As time progresses, the condition continues to deteriorate, and the brain lesion regions are also changed, which verifies that our proposed FSNPF model is quite effective.

The importance of different brain regions is illustrated using the histogram and exploiting BrainNet Viewer [55] to visualize the regional importance on the brain, as shown in Fig. 10. The colors in different brain regions represent the importance to identify the LMCI and EMCI. The histograms show the different ROI's weight for diagnosing the diseases. The experimental results show that several common ROIs are selected by our PFC method as an important features for LMCI and EMCI classification on baseline and year1 datasets. The

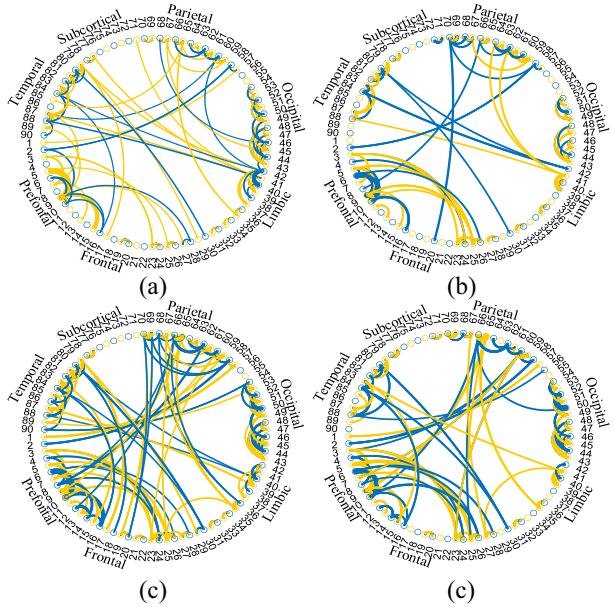

Fig. 8. Brain functional connection networks. (a) LMCI–Baseline. (b) LMCI–Year1. (c) EMCI–Baseline. (d) EMCI–Year1.

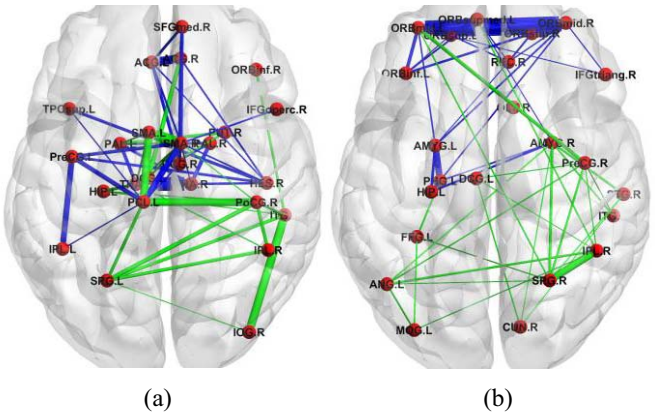

Fig. 9. Top 30 brain regions connectivity. (a) LMCI versus NC. (b) EMCI versus NC.

identified brain biomarkers are consistent with the findings in the previous studies [22], [34].

Although the proposed framework gains a promising performance, there are still two main limitations in this article. One limitation of this article is evaluated on small number of subjects, which may make it hard to generalize our results. In fact, the analysis of multiple time points requires that each subject has the corresponding data. Hence, the subjects are limited. In addition, the MCI patients with more than two time-point data are too few. For example, there are more than 2000 MCI subjects in the ADNI dataset, while there are only less than 100 MCI subjects with rs-fMRI data at two time points. The second limitation of the proposed method is that, we only consider one modality. In fact, the multimodal data can boost the classification performance. However, subjects with more than one modality are limited. Yet, we plan to explore the efficacy of our method on multimodal data in our future work.

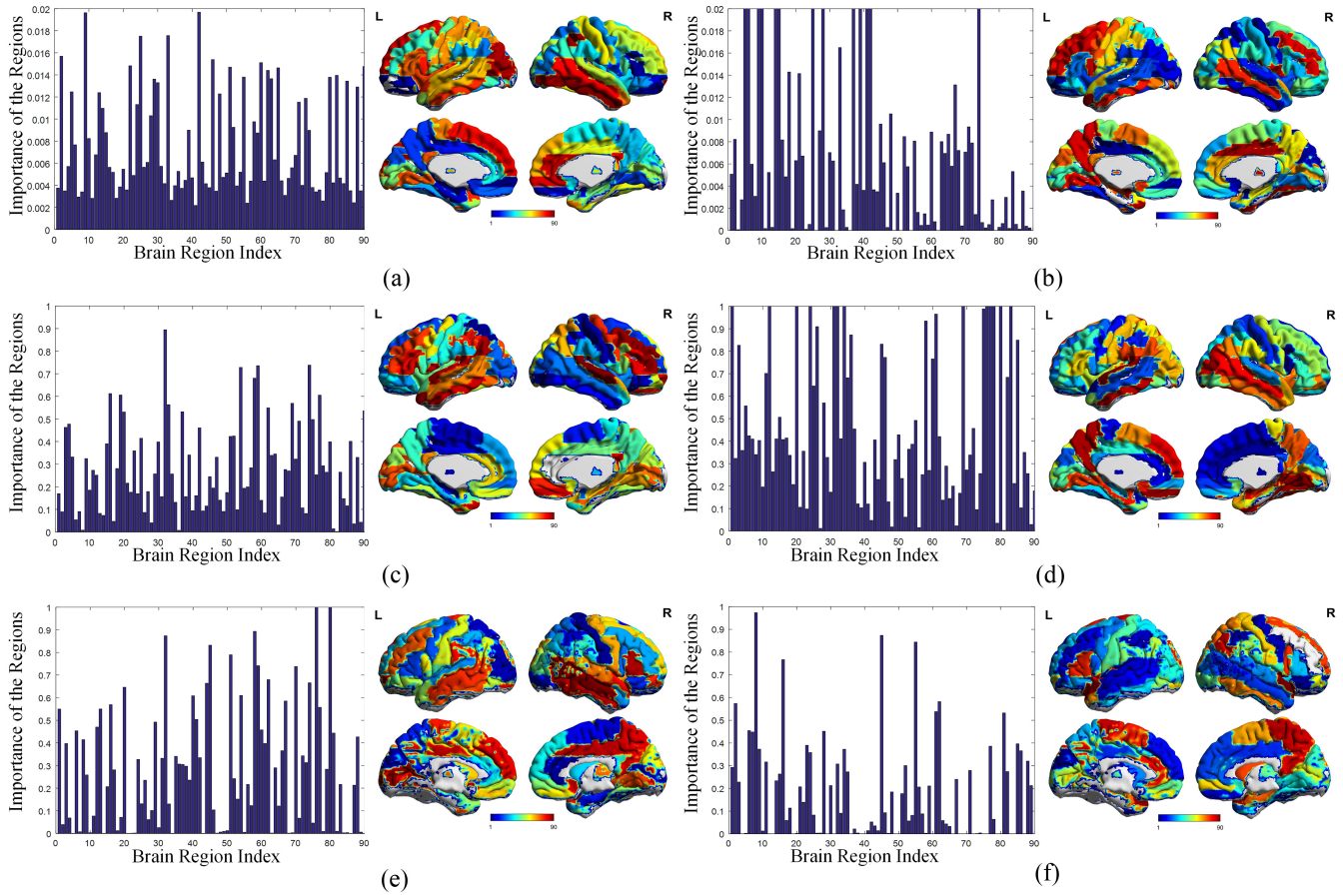

Fig. 10. Weights of different brain regions for classification. (a) LMCI versus NC–Baseline. (b) LMCI versus NC–Year1. (c) EMCI versus NC–Baseline. (d) EMCI versus NC–Year1. (e) EMCI versus LMCI–Baseline. (f) EMCI versus LMCI–Year1.

V. Conclusion

In this article, we propose an FSNPF framework to fuse the rs-fMRI data across multiple time points, which is effective in diagnosing both EMCI and LMCI. First, we devise an FSN model to build BFCNs with multiple time points, which can effectively construct brain functional network. Then, the PFC multitask learning is used for the longitudinal analysis of MCI. We jointly learn the features of two time points extracted from the BFCNs model (baseline and year1). The novel FSN model achieves more promising performance compared with PC and GCS. We jointly learn the multiple time points features extracted from the BFCNs by our PFC method. This method can automatically balance the importance of different time points features without parameter tuning. The representative common and specific features are exploited for disease progression analysis and diagnosis. In addition, we fuse the selected features by the SNF method to enhance the representation ability and classification accuracy. In identifying EMCI and LMCI, our method is clearly shown from the experimental results that it is superior to the related algorithms. In addition, our method outperforms the state-of-the-art methods as well. In our future work, more modalities and smoothing constraints will be considered to enhance the MCI diagnosis accuracy. Dynamic brain function network and high-order network can also be incorporated into our framework to enhance the performance of the entire framework as well.

References

[1] T. Tong, K. Gray, Q. Gao, L. Chen, and D. Rueckert, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern Recognit.*, vol. 63, pp. 171–181, Mar. 2017.

[2] Q. Lin, M. D. Rosenberg, K. Yoo, T. W. Hsu, T. P. O'Connell, and M. M. Chun, "Resting-state functional connectivity predicts cognitive impairment related to Alzheimer's disease," *Front. Aging Neurosci.*, vol. 10, pp. 1–10, Apr. 2018.

[3] O. Zanetti, S. B. Solerte, and F. Cantoni, "Life expectancy in Alzheimer's disease (AD)," *Archives Gerontol. Geriatrics*, vol. 49, no. S1, pp. 237–243, 2009.

[4] Alzheimer's Association, "2015 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 11, no. 3, pp. 332–384, 2015.

[5] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's Dementia*, vol. 3, no. 3, pp. 186–191, 2007.

[6] M. Prince, A. Comas-Herrera, M. Knapp, M. Guerchet, and M. Karagiannidou, "World Alzheimer report 2016: Improving healthcare for people living with dementia: Coverage, quality and costs now and in the future," Alzheimer's Disease Int., London, U.K., Rep., pp. 1–140, Sep. 2016. [Online]. Available: https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf

[7] M. S. Albert *et al.*, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 270–279, 2011.

[8] L. Qian, L. Zheng, Y. Shang, Y. Zhang, and Y. Zhang, "Intrinsic frequency specific brain networks for identification of MCI individuals using resting-state fMRI," *Neurosci. Lett.*, vol. 664, pp. 7–14, Jan. 2018.

[9] Y. Jin *et al.*, "3D tract specific local and global analysis of white matter integrity in Alzheimer's disease," *Human Brain Map.*, vol. 38, no. 3, pp. 1191–1207, 2016.

[10] R. M. Hutchison *et al.*, "Dynamic functional connectivity: Promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, Oct. 2013.

[11] J. Göttler *et al.*, "Reduced blood oxygenation level dependent connectivity is related to hypoperfusion in Alzheimer's disease," *J. Cereb. Blood Flow Metab.*, vol. 39, no. 7, pp. 1314–1325, 2019. doi: 10.1177/0271678X18759182.

[12] Y. Zhou, L. Qiao, W. Li, L. Zhang, and D. Shen, "Simultaneous estimation of low-and high-order functional connectivity for identifying mild cognitive impairment," *Front. Neuroinformat.*, vol. 12, pp. 1–8, Feb. 2018.

[13] B. Jie, M. Liu, D. Zhang, and D. Shen, "Sub-network kernels for measuring similarity of brain connectivity networks in disease diagnosis," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2340–2353, May 2018.

[14] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius, "Network analysis of intrinsic functional brain connectivity in Alzheimer's disease," *PLoS Comput. Biol.*, vol. 4, no. 6, pp. 1–11, 2008.

[15] Y. He, Z. Chen, and A. Evans, "Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease," *J. Neurosci.*, vol. 28, no. 18, pp. 4756–4766, 2008.

[16] P. Liang, Z. Li, G. Deshpande, Z. Wang, X. Hu, and K. Li, "Altered causal connectivity of resting state brain networks in amnesic MCI," *PLoS ONE*, vol. 9, no. 3, pp. 1–11, 2014.

[17] M. Richardson, "Current themes in neuroimaging of epilepsy: Brain networks, dynamic phenomena, and clinical relevance," *Clin. Neurophysiol.*, vol. 121, no. 8, pp. 1153–1175, 2010.

[18] L. Q. Uddin, A. M. Kelly, B. B. Biswal, F. X. Castellanos, and M. P. Milham, "Functional connectivity of default mode network components: Correlation, anticorrelation, and causality," *Human Brain Map.*, vol. 30, no. 2, pp. 625–637, 2009.

[19] J. del Etoile and H. Adeli, "Graph theory and brain connectivity in Alzheimer's disease," *Neuroscientist*, vol. 23, no. 6, pp. 616–626, 2017.

[20] H. Guo, F. Zhang, J. Chen, Y. Xu, and J. Xiang, "Machine learning classification combining multiple features of a hypernetwork of fMRI data in Alzheimer's disease," *Front. Neurosci.*, vol. 11, pp. 1–22, Nov. 2017.

[21] X. Chen, H. Zhang, Y. Gao, C. Y. Wee, G. Li, and D. Shen, "High-order resting-state functional connectivity network for MCI classification," *Human Brain Map.*, vol. 37, no. 9, pp. 3282–3296, 2016.

[22] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, "Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification," *Brain Struct. Funct.*, vol. 219, no. 2, pp. 641–656, 2014.

[23] M. J. Rosa *et al.*, "Sparse network-based models for patient classification using fMRI," *Neuroimage*, vol. 105, pp. 493–506, Jan. 2015.

[24] X. Zhu, X. Li, S. Zhang, Z. Xu, L. Yu, and C. Wang, "Graph PCA hashing for similarity search," *IEEE Trans. Multimedia*, vol. 19, no. 9, pp. 2033–2044, Sep. 2017.

[25] X. Zhu, S. Zhang, Y. Li, J. Zhang, L. Yang, and Y. Fang, "Low-rank sparse subspace for spectral clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 8, pp. 1532–1543, Aug. 2019.

[26] A. Tahmassebi *et al.*, "Dynamical graph theory networks techniques for the analysis of sparse connectivity networks in dementia," in *Proc. SPIE*, Anaheim, CA, USA, 2017, pp. 1–9.

[27] A. Tahmassebi, A. M. Amani, K. Pinker-Domenig, and A. Meyer-Baese, "Determining disease evolution driver nodes in dementia networks," in *Proc. SPIE*, Houston, TX, USA, 2018, pp. 1–8.

[28] A. Meyer-Bäse *et al.*, "Dynamical graph theory networks methods for the analysis of sparse functional connectivity networks and for determining pinning observability in brain networks," *Front. Comput. Neurosci.*, vol. 11, pp. 1–10, Oct. 2017.

[29] B. Jie, M. Liu, and D. Shen, "Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease," *Med. Image Anal.*, vol. 47, pp. 81–94, Jul. 2018.

[30] B. Hart, I. Cribben, and M. Fiecas, "A longitudinal model for functional connectivity networks using resting-state fMRI," *NeuroImage*, vol. 178, pp. 687–701, Sep. 2018.

[31] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, and D. Shen, "Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest," *Neurobiol. Aging*, vol. 46, pp. 180–191, Oct. 2016.

[32] B. Lei, F. Jiang, S. Chen, D. Ni, and T. Wang, "Longitudinal analysis for disease progression via simultaneous multi-relational temporal-fused learning," *Front. Aging Neurosci.*, vol. 9, pp. 1–17, Mar. 2017.

[33] J. Shi, Q. Jiang, Q. Zhang, Q. Huang, and X. Li, "Sparse kernel entropy component analysis for dimensionality reduction of biomedical data," *Neurocomputing*, vol. 168, pp. 930–940, Nov. 2015.

[34] R. Yu, H. Zhang, L. An, X. Chen, Z. Wei, and D. Shen, "Connectivity strength-weighted sparse group representation-based brain network construction for MCI classification," *Human Brain Map.*, vol. 38, no. 5, pp. 2370–2383, 2017.

[35] B. Wang *et al.*, "Similarity network fusion for aggregating data types on a genomic scale," *Nat. Methods*, vol. 11, no. 3, pp. 333–337, 2014.

[36] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT," *Inf. Fusion*, vol. 42, pp. 102–110, Jul. 2018.

[37] X.-A. Bi, Q. Shu, Q. Sun, and Q. Xu, "Random support vector machine cluster analysis of resting-state fMRI in Alzheimer's disease," *PLoS ONE*, vol. 13, no. 3, pp. 1–17, 2018.

[38] B. Jie, M. Liu, J. Liu, D. Zhang, and D. Shen, "Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 238–249, Jan. 2017.

[39] B. Lei, P. Yang, T. Wang, S. Chen, and D. Ni, "Relational regularized discriminative sparse learning for Alzheimer's disease diagnosis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1102–1113, Apr. 2017.

[40] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 173–183, Jan. 2018.

[41] C.-Y. Wee, S. Yang, P.-T. Yap, and D. Shen, "Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification," *Brain Imag. Behav.*, vol. 10, no. 2, pp. 342–356, 2016.

[42] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[43] J. Liu, L. Yuan, and J. Ye, "An efficient algorithm for a class of fused lasso problems," in *Proc. KDD*, Washington, DC, USA, 2010, pp. 323–332.

[44] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 68, no. 1, pp. 49–67, 2006.

[45] A. Beck and M. Teboulle, "A fast iterative shrinkage thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[46] L. Qiao, H. Zhang, M. Kim, S. Teng, L. Zhang, and D. Shen, "Estimating functional brain networks by incorporating a modularity prior," *Neuroimage*, vol. 141, pp. 399–407, Nov. 2016.

[47] J. Zhou, J. Chen, and J. Ye, *MALSAR: Multi-Task Learning Via Structural Regularization*, vol. 21, Arizona State Univ., Tempe, AZ, USA, pp. 1–50, 2011.

[48] X. Zhu, H. Li, and Y. Fan, "Parameter-free centralized multi-task learning for characterizing developmental sex differences in resting state functional connectivity," in *Proc. AAAI*, Palo Alto, CA, USA, 2018, pp. 2660–2667.

[49] F. Jessen *et al.*, "AD dementia risk in late MCI, in early MCI, and in subjective memory impairment," *Alzheimer's Dementia*, vol. 10, no. 1, pp. 76–83, 2014.

[50] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[51] A. Tahmassebi, A. H. Gandomi, I. McCann, M. H. J. Schulte, A. E. Goudriaan, and A. Meyer-Baese, "Deep learning in medical imaging: fMRI big data analysis via convolutional neural networks," in *Proc. PEARC*, Pittsburgh, PA, USA, 2018, pp. 1–4.

[52] A. Tahmassebi *et al.*, "Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients," *Invest. Radiol.*, vol. 54, no. 2, pp. 110–117, 2018.

[53] A. Tahmassebi, A. H. Gandomi, M. H. J. Schulte, A. E. Goudriaan, S. Y. Foo, and A. Meyer-Baese, "Optimized naive-Bayes and decision tree approaches for fMRI smoking cessation classification," *Complexity*, vol. 2018, pp. 1–24, May 2018. [Online]. Available: https://doi.org/10.1155/2018/27408170

[54] G. Prasad, S. H. Joshi, T. M. Nir, A. W. Toga, and P. M. Thompson, "Brain connectivity and novel network measures for Alzheimer's disease classification," *Neurobiol. Aging*, vol. 36, pp. S121–S131, Jan. 2015.

[55] M. Xia, J. Wang, and H. Yong, "BrainNet viewer: A network visualization tool for human brain connectomics," *PLoS ONE*, vol. 8, no. 7, pp. 1–15, 2013.

Peng Yang received the B.Sc. degree in biomedical engineering from Northeastern University, Shenyang, China, in 2016. He is currently pursuing the Ph.D. degree with the School of Biomedical Engineering, Shenzhen University, Shenzhen.

His current research interests include medical image analysis, brain connectivity analysis, machine learning, pattern recognition, and computer vision.

Feng Zhou received the first Ph.D. degree in human factors engineering from Nanyang Technological University, Singapore, in 2011, and the second Ph.D. degree in engineering design from the George W. Woodruff School of Mechanical Engineering, Georgia Tech, Atlanta, GA, USA, in 2014.

He is currently an Assistant Professor with the Department of Industrial and Manufacturing Systems Engineering, University of Michigan–Dearborn, Dearborn, MI, USA. His current research interests include on physiological computing, engineering design, sentiment analysis, and human factors issues in automated driving. He has published various peer-reviewed papers in the areas of engineering design, human–computer interaction, and human factors engineering.

Dong Ni received the Ph.D. degree in computer science and engineering from the Chinese University of Hong Kong, Hong Kong, in 2009.

He is currently an Associate Professor with the School of Biomedical Engineering, Shenzhen University, Shenzhen, China. His current research interests include ultrasound image analysis, image-guided surgery, and pattern recognition.

Yanwu Xu (SM'15) received the B.Eng. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2004 and 2009, respectively.

He is the Chief Architect/Scientist with the AI Innovation Business Department, Baidu Inc., Sunnyvale, CA, USA. He is also an Adjunct Professor with the Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China. His current research interests include ocular imaging, medical image analysis, computer vision, and machine learning.

Dr. Xu has been an Executive Member of IEEE Engineering in Medicine and Biology Society and Intelligent Transportation Systems Society, Singapore Chapter, since 2014. He has been an Associate Editor of *BMC Medical Imaging* since 2015. He also served as an Area Chair for the International Conference on Medical Image Computing and Computer-Assisted Intervention from 2017 to 2019.

Siping Chen received the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 1987.

He is currently a Professor with the School of Biomedical Engineering, and the Director of the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Shenzhen University, Shenzhen, China. His current research interests include ultrasound imaging, digital signal processing, and pattern recognition.

Tianfu Wang received the Ph.D. degree in biomedical engineering from Sichuan University, Chengdu, China, in 1997.

He is currently a Professor with the School of Biomedical Engineering, and the Associate Chair of the Health Science Center, Shenzhen University, Shenzhen, China. His current research interests include ultrasound image analysis, medical image processing, pattern recognition, and medical imaging.

Baiying Lei (SM'18) received the M.Eng. degree in electronics science and technology from Zhejiang University, Hangzhou, China, in 2007, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2013.

She is currently with the School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China. She has coauthored over 120 scientific articles, such as the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON MEDICAL IMAGING, the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, *Pattern Recognition*, and *Information Sciences*. Her current research interests include medical image analysis, machine learning, and pattern recognition.

Dr. Lei serves as an Editorial Board Member of Scientific Reports, *Frontiers in Neuroinformatics*, *Frontiers in Aging Neuroscience*, and an Academic Editor of *PLoS ONE*.