



Brain functional connectivity analysis based on multi-graph fusion

Jiangzhang Gan^{a,c,1}, Ziwen Peng^{b,1}, Xiaofeng Zhu^{a,c,*}, Rongyao Hu^c, Junbo Ma^d,
Guorong Wu^{d,e}

^a Center for Future Media and School of Computer Science and Technology, University of Electronic Science and Technology of China, Chengdu 611731, China

^b Center for the Study of Applied Psychology, Guangdong Key Laboratory of Mental Health and Cognitive Science and School of Psychology, South China Normal University, Guangzhou 510631, China

^c School of natural and Computational Science, Massey University Auckland Campus, Auckland 0745, New Zealand

^d Department of Psychiatry, University of North Carolina, Chapel Hill, NC 27599, USA

^e Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:

Received 25 August 2020

Revised 25 March 2021

Accepted 27 March 2021

Available online 9 April 2021

Keywords:

Brain functional connectivity network analysis

Data fusion

fMRI data

Feature selection

Classification

ABSTRACT

In this paper, we propose a framework for functional connectivity network (FCN) analysis, which conducts the brain disease diagnosis on the resting state functional magnetic resonance imaging (rs-fMRI) data, aiming at reducing the influence of the noise, the inter-subject variability, and the heterogeneity across subjects. To this end, our proposed framework investigates a multi-graph fusion method to explore both the common and the complementary information between two FCNs, i.e., a fully-connected FCN and a 1 nearest neighbor (1NN) FCN, whereas previous methods only focus on conducting FCN analysis from a single FCN. Specifically, our framework first conducts the graph fusion to produce the representation of the rs-fMRI data with high discriminative ability, and then employs the L1SVM to jointly conduct brain region selection and disease diagnosis. We further evaluate the effectiveness of the proposed framework on various data sets of the neuro-diseases, i.e., Fronto-Temporal Dementia (FTD), Obsessive-Compulsive Disorder (OCD), and Alzheimers Disease (AD). The experimental results demonstrate that the proposed framework achieves the best diagnosis performance via selecting reasonable brain regions for the classification tasks, compared to state-of-the-art FCN analysis methods.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Resting state functional magnetic resonance imaging (rs-fMRI) data is a powerful tool to measure the spontaneous nerve activity of the human brain in the resting state (Betzel and Bassett, 2017). Using the rs-fMRI data to construct the brain functional connectivity network (FCN) can reveal the pathological basis of brain disease and develop biomarkers (Greicius et al., 2003; Zhu et al., 2021). Recently, brain network analysis using the rs-fMRI data has been widely used in computer-aided diagnosis of various brain diseases such as Fronto-Temporal Dementia (FTD), Obsessive-Compulsive Disorder (OCD), and Alzheimers Disease (AD) (Zou and Yang, 2020; Shu et al., 2019a).

Machine learning techniques has been widely used to analyze rs-fMRI data. For example, Shen et al. (2017) focuses on first extracting the semantic information (i.e., features) and then building a linear model to identify functional brain connections. Recently,

deep learning methods (e.g., Yao et al., 2020; Liu et al., 2018) were shown to outperform traditional machine learning methods by conducting the end-to-end learning. However, they ignore to extract semantic features, which is very important for pathological analysis. Hence, this work focuses on designing a new traditional machine learning method to extract semantic features for interpretability and conduct disease diagnosis. Traditional diagnosis methods with the rs-fMRI data for the brain functional connectivity network analysis mainly include three steps, i.e., the FCN construction, the feature learning, and the disease diagnosis. The FCN construction step models the functional association patterns between brain regions as networks, in which nodes correspond to brain Regions Of Interest (ROIs) and edges represent functional connections between two brain ROIs. The popular methods of constructing FCNs have correlation-based approaches (Chen et al., 2016), Granger causality analysis (Seth et al., 2015), regularized inverse covariance estimation (Brier et al., 2015), etc. In the literature, the correlation-based methods has been pointed out to provide relatively higher sensitivity of the network connection (Scheinost et al., 2019; Shu et al., 2019b). The feature learning step first extracts semantic features (such as the clustering coefficient

* Corresponding author at: School of natural and Computational Science, Massey University Auckland Campus, Auckland 0745, New Zealand.

¹ Jiangzhang Gan and Ziwen Peng contributed equally to this work.

and the connection strength) from the FCNs, and then conducts feature selection to select the most discriminative feature subset for classification, such as the t -test method (Acharya et al., 2019; Zhu et al., 2020b) and the Least Absolute Shrinkage and Selection Operator (Lasso) (Tibshirani, 1996). The disease diagnosis step usually employs traditional machine learning methods (i.e., SVM) to distinguish healthy subjects from patients (Rubbert et al., 2019). In the literature, most studies for the brain FCN analysis focus on the FCN construction step and the feature learning step. However, due to the reasons such as noise, the individual variability for each subject, and the inter-group heterogeneity across subjects, many issues should be further addressed for the brain FCN analysis.

On the one hand, FCNs constructed from the rs-fMRI data plays a vital role in the diagnosis of brain neuro-diseases. However, due to the influence of various aspects such as noise and the curse of dimensionality in the high-dimensional representation (Zhu et al., 2021; 2020a), it is still a challenging task to construct a FCN that can accurately reflect the functional connectivity of the brain. First, the fully-connected FCNs constructed by the correlation-based methods easily suffer from the influence of the noise or the false connection, which cannot explain the correlation between two brain ROIs well and will affect the diagnosis performance. To reduce the noisy connection as well as the network complexity, previous methods focused on constructing a sparse brain network based on the kNN method (Zhang et al., 2018). Specifically, every node connects a subset of nodes (i.e., its k nearest neighbors, kNN graph for short) in the sparse brain networks. Moreover, the nearest neighbor is obtained based on the similarity measurement. For example, Yang et al. (2016) proposed to first calculate the mean FCN matrix of all training subjects within the same time-series block to construct a kNN graph. Yao et al., 2020 proposed to first calculate Pearson correlation among the nodes within the individual brain and then to connect each brain region with 8 neighbours for all subjects. However, in real applications, considering the heterogeneity across different brain regions, the connection number (i.e., the value of k) of each brain region may be different (Zhang et al., 2017b; 2018). Hence, it is unfeasible to connect every node with the same number of neighbors. Second, the brain is the most complex system in the human body. A single FCN may not be able to capture the subtle disruption of the brain functional tissues caused by neurological diseases, because each network can only capture a part of the differences between groups (Huang et al., 2019; Kong et al., 2020). Previous methods (e.g., Wee et al., 2012a; Wu et al., 2016) explored the construction of multiple networks based on the rs-fMRI data for the disease classification. However, these methods did not take into account either the common information or the complementary information among multiple FCNs, which has been demonstrated to strengthen the disease diagnosis in medical image analysis (Shen et al., 2021). Third, the FCN construction of one subject is not related to the construction of others. In this way, the FCNs of all subjects may heterogeneous, so the independent process for the FCN construction ignores to consider the group effect, which may result in outputted representation has limited discriminative ability (Zhang et al., 2017c).

On the other hand, many fMRI-based methods extract functional features from FCNs to represent each subject, and then input these features into a predefined classification model for the disease diagnosis. The features are usually high-dimensional data, so that feature selection has been used to select the most discriminative features for the disease diagnosis. For example, Wee et al. (2012b) proposed to first extract the features from both the structural and functional connectivity networks, and then employed the t -test feature selection method for MCI identification. Liu et al. (2015) designed to first extract connectivity strength features from FCNs, and then to perform feature selection using the F-score method. These studies show that feature selection

can improve the diagnostic performance as well as help to discover biomarkers (Zhu et al., 2021; Shen et al., 2021). However, these methods regarded the feature selection task and the classification task as independent tasks, which easily results in that the optimally selected features are unsuitable for the classification task (Zhu et al., 2020b; Hu et al., 2020). Although some methods (e.g., Wang et al., 2019; Ma et al., 2017) were proposed to jointly perform feature learning and the classification task, the features extracted by these methods usually lack the interpretability.

To solve the above issues, this paper extend the conference version (Gan et al., 2020)² to propose a new framework of FCN analysis on the rf-fMRI data aiming to accurately identify patients of brain neuro-diseases. The key characteristics of our framework is to investigate a new fusion strategy to explore the complex structure of FCNs. Specifically, our framework first applies the kNN based method (e.g., Zhang et al., 2018; Zhang et al., 2017b) to construct multiple FCNs for each subject, i.e., the fully-connected FCN and the 1 nearest neighbor (1NN) FCN, and then designs a multi-fusion method to effectively integrate the information from multiple FCNs, which enhances the common intrinsic structure among all subjects and limits the error caused by the heterogeneity across the subjects. In particular, in the multi-fusion process, the number of edges of each node is updated iteratively, so that the number of the neighbors for each node is learned automatically. After yielding the fused FCN of each subject, our framework extracts the upper triangle of the FCNs as the representation of each subject, and then employs L1SVM (i.e., ℓ_1 -SVM in the Liblinear toolbox Fan et al., 2008) to jointly conduct feature selection (i.e., brain region selection) and the classification (i.e., disease diagnosis). The flowchart of the proposed framework is illustrated in Fig. 1.

The contributions of the proposed framework are two-fold. First, this work proposes a novel multi-graph fusion method to fuse FCNs and automatically learn the connections of brain regions. In addition, it also designs two regularization terms to achieve the group effect and solve the issue of the heterogeneity across the subjects by pushing the subjects in the same class close to each other and the subjects from different classes far away from each other. Second, the framework employs L1SVM to integrate the feature selection and the classification task in a unified framework. It is noteworthy that previous methods (Eavani et al., 2015; Zhang et al., 2017a) focused on separately conducting feature selection and classification. As a result, our proposed framework outperformed all comparison methods in terms of the classification performance, indicating its effectiveness in different neuro-disease diagnosis.

2. Method

Given the BOLD signal of the m -th subject among M subjects $\mathbf{B}^m \in \mathbb{R}^{n \times t}$ ($m = 1, \dots, M$) where n and t , respectively, represent the number of brain regions and the length of signals, in this paper, we first obtain multiple graphs (i.e., FCNs) $\mathbf{A}^{m,v} \in \mathbb{R}^{n \times n}$ ($v = 1, \dots, V$)³ by the Pearson correlation analysis where V is the graph number, and then propose to learn a sparse-connected FCN (sparse FCN for short) $\mathbf{S}^m \in \mathbb{R}^{n \times n}$ for each subject so that it could automatically learn the connection number of every node as well as is homogeneous and discriminative to other sparse FCNs $\mathbf{S}^{m'}$ ($m \neq m'$).

² Different from the conference version in Gan et al. (2020), this paper made the following changes: (1) Section 1 was rewritten; (2) Related work is new; (3) Appendix added the theoretical proof of the convergence of Algorithm 1; (4) Two more comparison methods (i.e., DIAL-GNN, CNHD) and one more data set (i.e., ADNI) were added in Section 3; and (5) The templates obtained by our proposed method and Discussion were added.

³ If the value of $\mathbf{A}^{m,v}$ is negative, we take its absolute value, i.e., $a_{i,j}^{m,v} = |a_{i,j}^{m,v}|$.

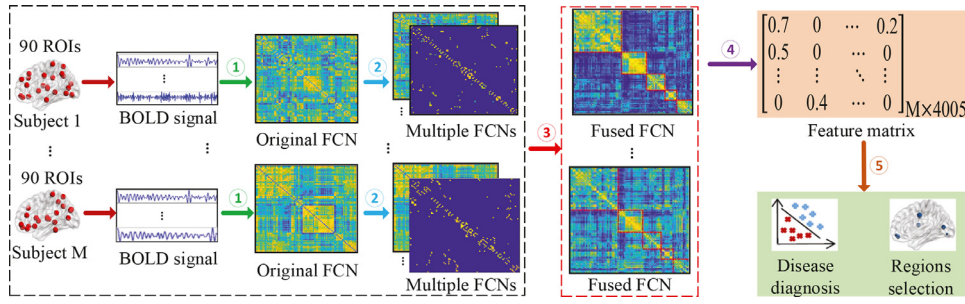


Fig. 1. The flowchart of the proposed framework for functional connectivity network analysis using the rs-fMRI data. (1) The original FCNs constructed by the Pearson correlation analysis; (2) The multi-FCN (i.e., a fully-connected FCN and a 1NN FCN) for every subject; (3) The proposed multi-graph fusion method, i.e., the key characteristics of the proposed framework; (4) The new feature matrix of all subjects, i.e., the upper triangle of each sparse FCN obtained by (3). Moreover, each row is the representation of one subject; (5) LISVM for joint disease diagnosis and brain regions selection. It is noteworthy that steps (1) and (2) are off-line.

2.1. Multi-graph fusion

Previous studies demonstrated that the sparse FCN is preferred in the representation learning of brain functional connectivity analysis (Zhang et al., 2019a; Karmonik et al., 2019), compared to the fully-connected FCN, due to that (1) the fully-connected FCN lacks the interpretability; (2) the connectivity between two nodes may contain noisy connectivity (i.e., either irrelevant or spurious connectivity) to affect brain functional connectivity analysis (Kong et al., 2016; Reyes et al., 2018); and (3) neurologically, a brain region predominantly interacts only with a part of brain regions. Existing methods of FCN analysis usually obtain sparse FCNs from the fully-connected FCNs. Specifically, they design different techniques to learn sparse FCNs based on the fully-connected FCNs, such as sparse learning (Zhang et al., 2019a; Eavani et al., 2015) and clustering (Wee et al., 2012b; Zhang et al., 2019b). However, these methods have limitations in the brain FCN analysis. First, existing methods usually assume that each node connects a fixed number of nodes. That is, the connection number is unchanged for all nodes. To achieve this, the sparse k -nearest neighbor (k NN) graph is constructed so that each node connects with k nodes. Such an assumption obviously ignores the fact that a brain region predominantly interacts only with a part of brain regions. Second, previous methods generate the sparse FCN of a subject independently from other subjects. On the one hand, by considering the heterogeneity across subjects, the FCNs obtained from these heterogeneous subjects possibly have different distributions. As a result, the robustness of the classifier constructed by these sparse FCNs will be affected. On the other hand, the independent process of representation learning makes it difficult to consider the group effect, e.g., the discriminative ability across classes or subjects.

Given the fully-connected FCN connecting each node with all nodes, we obtain an extreme sparse FCN, i.e., 1NN graph (excluding itself). By this way, we could obtain multiple graphs for each subject to solve the first issue of existing functional connectivity analysis. In this paper, we only use 2 graphs for every subject, i.e., a fully-connected FCN and an extremely sparse FCN, by taking the following observations into account. The fully-connected FCN contains all connectivity information (i.e., the most complex connectivity) and the extremely sparse FCN contain the least information (i.e., the simplest connectivity). We expect to obtain a flexible connection number for every node based on the data distribution in the range $[1, n]$ where n is the node number. To do this, we design the following objective function to automatically learn specific connection number for the m th subject \mathbf{S}^m by fusing the information from multiple graphs.

$$\min_{\mathbf{S}^m} \sum_{v=1}^V \|\mathbf{S}^m - \mathbf{A}^{m,v}\|_F^2 \quad (1)$$

$$s.t., \forall i, \mathbf{s}_{i,\cdot}^m \mathbf{1} = 1, \mathbf{s}_{i,i}^m = 0, \mathbf{s}_{i,j}^m \geq 0 \text{ if } j \in \mathcal{N}(i), \text{ otherwise } 0.$$

where $\|\cdot\|_F$ indicates the Frobenius norm. $\mathbf{s}_{i,\cdot}^m$ and $s_{i,j}^m$, respectively, represent the i th row of \mathbf{S}^m and the element in the i th row and the j th column of \mathbf{S}^m . $\mathbf{1}$ and $\mathcal{N}(i)$, respectively, indicate the all-one-element vector and the set of nearest neighbors of the i th node. After optimizing $\mathbf{s}_{i,\cdot}^m$ by our proposed optimization method in Appendix, we obtain different non-zero numbers for every row, i.e., $\mathbf{s}_{i,\cdot}^m$ in \mathbf{S}^m . This indicates that different nodes have different connection numbers for every subject.

Eq. (1) employs multiple graphs to conduct the feature or representation learning, aiming at selecting an optimal connection number between 1 and n . However, the optimization of \mathbf{S}^m is independent on the optimization of $\mathbf{S}^{m'}$ ($m \neq m'$), which explores the inter-subject variability, but does not touch the issue of the heterogeneity across subjects. To address this issue, we propose the following objective function.

$$\min_{\mathbf{S}^1, \dots, \mathbf{S}^M, \mathbf{H}, \mathbf{G}} \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^m - \mathbf{A}^{m,v}\|_F^2 + \alpha \mathcal{R}_1(\mathbf{H}, \mathbf{G}, \mathbf{S}^1, \dots, \mathbf{S}^M) + \beta \mathcal{R}_2(\mathbf{S}^1, \dots, \mathbf{S}^M) \quad (2)$$

$$s.t., \forall i, \mathbf{h}_{i,\cdot} \mathbf{1} = 1, \mathbf{h}_{i,i} = 0, \mathbf{h}_{i,j} \geq 0 \text{ if } j \in \mathcal{N}(i), \text{ otherwise } 0,$$

$$\mathbf{g}_{i,\cdot} \mathbf{1} = 1, \mathbf{g}_{i,i} = 0, \mathbf{g}_{i,j} \geq 0 \text{ if } j \in \mathcal{N}(i), \text{ otherwise } 0,$$

$$\mathbf{s}_{i,\cdot}^m \mathbf{1} = 1, \mathbf{s}_{i,i}^m = 0, \mathbf{s}_{i,j}^m \geq 0 \text{ if } j \in \mathcal{N}(i), \text{ otherwise } 0.$$

where $\mathbf{H} \in \mathbb{R}^{n \times n}$ and $\mathbf{G} \in \mathbb{R}^{n \times n}$ are two variables, $\mathcal{R}_1(\mathbf{H}, \mathbf{G}, \mathbf{S}^1, \dots, \mathbf{S}^M)$ and $\mathcal{R}_2(\mathbf{S}^1, \dots, \mathbf{S}^M)$ are regularization terms. We use the summation operator in the first term of Eq. (2) to learn the representations of all subjects in a unified framework, and design two regularization terms to achieve the group effect, e.g., discriminative ability across subjects. We list the details of two regularization terms as follows.

First, we expect that positive subjects are similar or close to the positive template \mathbf{G} while negative subjects are similar to the negative template \mathbf{H} . Hence, the subjects within the same class are close. It is noteworthy that \mathbf{G} and \mathbf{H} , respectively, can be regarded as the common information of the positive class and the negative class. Moreover, the outputted templates could be widely applied in medical imaging analysis, such as guiding parcellations for new subjects and measuring the group difference (Reyes et al., 2018). To achieve this, we design $\mathcal{R}_1(\mathbf{H}, \mathbf{G}, \mathbf{S}^1, \dots, \mathbf{S}^M)$ as follows

$$\mathcal{R}_1(\mathbf{H}, \mathbf{G}, \mathbf{S}^1, \dots, \mathbf{S}^M) = \begin{cases} \sum_{m=1}^{|\mathcal{D}|} \|\mathbf{S}^m - \mathbf{H}\|_F^2, & m \in \mathcal{D} \\ \sum_{m=1}^{|\mathcal{E}|} \|\mathbf{S}^m - \mathbf{G}\|_F^2, & m \in \mathcal{E} \\ 0, & m \in \mathcal{U} \end{cases} \quad (3)$$

where \mathcal{D} , \mathcal{E} , and \mathcal{U} , respectively, represent the set of negative subjects, positive subjects, and unlabeled subjects. Moreover, $|\mathcal{D}|$ and $|\mathcal{E}|$, respectively, indicate the cardinality of \mathcal{D} and \mathcal{E} .

Eq. (3) has at least two advantages: 1) preserving the global structure since all the subjects are close to their template and 2)

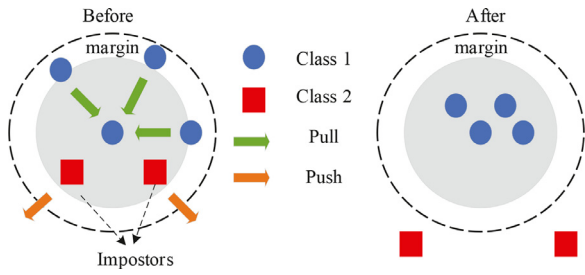


Fig. 2. The visualization of Eq. (4). The left figure is the original neighborhood structure among one subject (i.e., the centered point) and its neighbors. The right figure is the final status of the neighborhood structure about this subject after conducting the proposed multi-graph fusion method, where the subjects with the same label are close to each other and the subjects with different labels are far away from each other.

outputting practical templates. However, Eq. (3) does not take the local structure of the data, which has been regarded as the complementary of the global structure of the data (Weinberger and Saul, 2009; Shen et al., 2020). In this paper, we design $\mathcal{R}_2(\mathbf{S}^1, \dots, \mathbf{S}^M)$ as follows

$$\mathcal{R}_2(\mathbf{S}^1, \dots, \mathbf{S}^M) = \frac{\sum_{m=1}^M \sum_{p \in \mathcal{G}(m)} \|\mathbf{S}^m - \mathbf{S}^p\|_F^2}{\sum_{m=1}^M \sum_{q \in \mathcal{F}(m)} \|\mathbf{S}^m - \mathbf{S}^q\|_F^2} \quad (4)$$

where $\mathcal{G}(i)$ and $\mathcal{F}(i)$, respectively, are the set of the near neighbors and the set of the distant neighbors, of the i th subject. In the proposed framework, i.e., semi-supervised learning, the training subjects include labeled subjects and unlabeled subjects, we denote the set $\mathcal{G}(i)$ of the i th unlabeled subject as its k nearest neighbors including labeled subjects and unlabeled subjects, and the set $\mathcal{G}(i)$ of the i th labeled subject as its k nearest neighbors with the same label to the i th subject. We further define the set $\mathcal{F}(i)$ of the i th unlabeled subject as its k furthest subjects including labeled subjects and unlabeled subjects, and the $\mathcal{F}(i)$ of the i th labeled subject as its k nearest neighbors with different labels to the i th subject. It is noteworthy that the value of k is insensitive in our experiments, so we fixed $k = 10$ for all subjects. Eq. (4) minimizes the ratio of two terms, similar to linear discriminative analysis (Ye et al., 2005; Zhu et al., 2019). Specifically, the subjects have the same label with their nearest neighbors, while the subjects with far similarity have different labels. In this way, the local structure of the subjects is preserved. Fig. 2 visualizes the process of Eq. (4). The optimization of Eq. (4) is very challenging, so we follow Theorem 1 in Wang et al. (2014) to convert the minimization of Eq. (4) to minimize the following objective function:

$$\sum_{m=1}^M \left(\sum_{p \in \mathcal{G}(m)} \|\mathbf{S}^m - \mathbf{S}^p\|_F^2 - \lambda^m \sum_{q \in \mathcal{F}(m)} \|\mathbf{S}^m - \mathbf{S}^q\|_F^2 \right), \quad (5)$$

where λ^m can be updated as $\lambda^m = \frac{\sum_{p \in \mathcal{G}(m)} \|\mathbf{S}^m - \mathbf{S}^p\|_F^2}{\sum_{q \in \mathcal{F}(m)} \|\mathbf{S}^m - \mathbf{S}^q\|_F^2}$ in the implementation based on (Wang et al., 2014).

Compared to previous literature, Eq. (2) outputs the representation of every subject depending on other subjects as well as taking into account the following constraints, such as multi-graph information, and the preservations of the global and local structure among the subjects.

Theorem 1. The global solution of the following general optimization problem

$$\min_{\mathbf{v} \in \mathcal{C}} \frac{q(\mathbf{v})}{p(\mathbf{v})}, \text{ where } q(\mathbf{v}) > 0 \ (\forall \mathbf{v} \in \mathcal{C}) \quad (6)$$

can be calculated by the root of the following function:

$$h(\lambda) = \min_{\mathbf{v} \in \mathcal{C}} q(\mathbf{v}) - \lambda p(\mathbf{v}), \quad (7)$$

given that $q(\mathbf{v}) - \lambda p(\mathbf{v})$ is lower bounded.

Proof. Suppose \mathbf{v}^* is the global solution of the problem in Eq. (18), and λ^* is the corresponding global minimal objective value, the formulation $\frac{q(\mathbf{v}^*)}{p(\mathbf{v}^*)} = \lambda^*$ holds, so $\forall \mathbf{v} \in \mathcal{C}$, we have $\frac{q(\mathbf{v})}{p(\mathbf{v})} \geq \lambda^*$. By considering the characteristics of $p(\mathbf{v}) > 0$, we can yield $q(\mathbf{v}) - \lambda^* p(\mathbf{v}) \geq 0$, which means:

$$\min_{\mathbf{v} \in \mathcal{C}} q(\mathbf{v}) - \lambda^* p(\mathbf{v}) = 0 \iff h(\lambda^*) = 0 \quad (8)$$

That is, the global minimal objective value λ^* in Eq. (18) is the root of the function $h(\lambda)$. Hence, the proof of Theorem 1 has been completed. \square

2.2. Joint regions selection and disease diagnosis

Our proposed framework generates a sparse FCN \mathbf{S}^m ($m = 1, \dots, M$) from two graphs, i.e., a fully-connected FCN and a 1NN graph, for each subject. Moreover, we follow previous methods to transfer the matrix representation to its vector representation, i.e., extracting the upper triangle part of the symmetric matrix \mathbf{S}^m ($m = 1, \dots, M$) to form a row vector $\mathbf{x}_{m,\cdot} \in \mathbb{R}^{1 \times [n(n-1)/2]}$. As a result, we have the data matrix $\mathbf{X} \in \mathbb{R}^{M \times [n(n-1)/2]}$ and the corresponding label vector $\mathbf{y} \in \{-1, 1\}^{M \times 1}$.

Many existing studies separately conduct feature selection and disease diagnosis (i.e., classification) (Kong et al., 2020). The goal of feature selection is to remove the redundant features from high-dimensional data because the vector representation is a 4005-dimension vector for 90 nodes in our data sets. However, the optimal results of feature selection cannot guarantee to achieve the optimal classification in such two separate processes. In this paper, we employ L1SVM to jointly conduct feature selection and classification, where the result of feature selection will be iteratively updated by the optimized classifier so that finally outputting significant classification performance.

We list the pseudo of our proposed functional connectivity analysis framework in Algorithm 1 and report its optimization and convergence in Appendix. It is noteworthy that the multi-graph fusion model uses both labeled subjects and unlabeled subjects while the L1SVM uses the label subjects in training process.

3. Experiments

We experimentally evaluated our proposed method, compared to six state-of-the-art methods, on three real neuro-disease data sets with the rs-fMRI data, in terms of binary classification.

3.1. Experimental setting

3.1.1. Data sets

The data set fronto-temporal dementia (FTD) contains 95 FTD subjects and 86 age-matched healthy control (HC) subjects. FTD was derived from the NIFD database managed by the frontotemporal lobar degeneration neuroimaging initiative. The data set obsessive-compulsive disorder (OCD) has 20 HC subjects and 62 OCD subjects. The data set Alzheimer's Disease Neuroimaging Initiative (ADNI) includes 59 Alzheimer's disease (AD) subjects and 48 HC subjects.

Imaging data acquisition. We used A 3.0-Tesla MR system (Philips Medical Systems, USA) equipped with an eight-channel phased-array head coil to collect all rs-fMRI data. The parameters of gradient-echo Echo-Planar Imaging (EPI) sequence were listed as follows: Field Of View (FOV) = 208×180 mm, matrix = 104×90 , 72 slices, TR = 720 ms, TE = 33.1 ms, Flip Angle (FA) = 52° , and multi-band factor = 8, 1200 frames, 15 min/ru. The subjects' heads were fixed with a sponge pad for preventing head movement from affecting the experimental results. During

-
- 1 **Input:** \mathbf{B}^m ($m = 1, \dots, M$) and \mathbf{y} ;
 - 2 **Parameters:** C , α , and β ;
 - 3 **Output:** \mathbf{S}^m , \mathbf{H} , \mathbf{G} , and \mathcal{C} ;
 - 1: Obtain $\mathbf{A}^{m,v}$ ($v = 1, \dots, V$) by \mathbf{B}^m ;
 - 2: Initialize \mathbf{S}^m as the average of $\mathbf{A}^{m,v}$ ($v = 1, \dots, V$);
 - 3: **while** Eq.(2) *not converges* **do**
 - 4: Update $\lambda^m = \frac{\sum_{p \in \mathcal{G}(m)} \|\mathbf{S}^m - \mathbf{S}^p\|_F^2}{\sum_{q \in \mathcal{F}(m)} \|\mathbf{S}^m - \mathbf{S}^q\|_F^2}$;
 - 5: Update \mathbf{H} and \mathbf{G} via Eq. (22);
 - 6: Update \mathbf{S}^m ($m = 1, \dots, M$) via Eq. (19);
 - 7: **end while**
 - 8: Obtain \mathbf{X} by extracting the upper triangle of \mathbf{S}^m ;
 - 9: Run L1SVM on \mathbf{X} and \mathbf{y} to output the classifier \mathcal{C} ;
-

Algorithm 1. The pseudo of optimizing Eq. (2).

the scanning, the subjects need to close eyes, keep relax, and stay awake.

Functional imaging data preprocessing. We used the DPARSF toolbox⁴ to preprocess the rs-fMRI data. We first deleted the first 10 time points of each subject, and then conducted the steps including slice timing correction, motion correction, normalization and spatial smoothing, on the obtained rs-fMRI data, for adapting the subjects to the scanning environment.

For all imaging data, we followed the automated anatomical labeling (AAL) template (Tzourio-Mazoyer et al., 2002) to construct the functional connectivity network for each subject with 90 nodes. The region-to-region correlation was measured by the Pearson correlation coefficient.

3.1.2. Comparison methods

The comparison methods include the baseline method L1SVM, three popular methods for neuro-disease diagnosis, i.e., High-Order Functional Connectivity (HOFC) (Zhang et al., 2017a), Sparse Connectivity Pattern (SCP) (Eavani et al., 2015), and Connectivity Network Analysis method with Discriminative Hub Detection (CNHD) (Wang et al., 2019), and two deep learning methods, i.e., Simplify Graph Convolutional networks (SGC) (Wu et al., 2019), Deep Iterative and Adaptive Learning Graph Neural Networks (DIAL-GNN) (Chen et al., 2019). We list the details of the comparison methods as follows.

- L1SVM extracts the upper triangle part of the FCN of each subject as its representation, and then employs the ℓ_1 -norm regularization term to jointly conduct feature selection and classification.
- HOFC learns the sparse FCN based on the fully-connected FCN, whose element is the Pearson correlation coefficient, by taking into account the high-order information of the subjects.
- SCP searches the sparse FCNs from the fully-connected FCNs to effectively explore the heterogeneity across the subjects by taking into account the inter-subject variability among the subjects.
- CNHD first constructs functional connectivity networks based on the rs-fMRI data, and then conducts feature extraction and the classification task in a unified framework.
- SGC first regards the upper triangle part of the fully-connected FCN of each subject as its representation as well as uses the

fully-connected FCNs of all subjects to obtain the local structure of all subjects, and then designs a graph neural network by preserving the original local structure to update the representations of all subjects.

- DIAL-GNN first extracts the upper triangle part of the FCN of each subject as its representation to obtain a original graph, and then learns the new representation for each subject.

L1SVM, DIAL-GNN and SGC extract the upper triangle of the fully-connected FCN as the representation of each subject. The methods (e.g., HOFC, SCP, CNHD, and our proposed method) designed different methods to transfer fully-connected FCNs to sparse FCNs, followed by extracting the upper triangle part of the sparse FCN as the representation of each subject. It is noteworthy that all methods can be directly applied for supervised learning but only three methods (e.g., DIAL-GNN, SGC and our method) can be used for personalized classification.

3.1.3. Setting-up

In our experiments, we repeated the 10-fold cross-validation scheme 10 times for all methods to report the average results as the final results. In the model selection, we set $\alpha, \beta \in \{10^{-3}, 10^{-2}, \dots, 10^3\}$ in Eq. (2), and fixed $k = 10$ since the value of k is insensitive to the result of Eq. (2). We further set $C \in \{2^{-10}, 2^{-9}, \dots, 2^{10}\}$ for ℓ_1 -SVM. We followed the literature to set the parameters of the comparison methods so that they outputted the best results.

We designed the following experiments to evaluate all methods, i.e., the classification performance of both supervised learning and personalized classification, the effectiveness of the proposed multi-graph fusion method, the effectiveness of feature selection and visualization of the selected brain regions, and the visualization of the templates produced by our method. The evaluation metrics include ACCuracy (ACC), SENSitivity (SEN), SPECificity (SPE), and Area Under the ROC Curve (AUC). Besides, we conducted the paired-sample t-tests between our method and every comparison method, in terms of ACC, SEN, SPE, and AUC. Moreover, the symbols “*” and “***”, respectively, indicate that our method has statistically significant difference with $p < 0.05$ and $p < 0.001$ on the paired-sample t-tests at 95% significance level, compared to the comparison method. We report the results in Tables 1–3.

⁴ <http://rfmri.org/DPARSF>.

Table 1
Classification results (%) of all methods on FTD.

Methods	Accuracy	Sensitivity	Specificity	AUC
L1SVM	64.77 ± 3.89*	62.54 ± 1.56*	67.48 ± 2.87*	65.87 ± 1.95*
HOFC	79.37 ± 4.20*	78.59 ± 5.37*	82.16 ± 4.29**	79.52 ± 3.88*
SCP	84.75 ± 4.20**	82.59 ± 3.77*	85.56 ± 3.87**	83.93 ± 5.43*
CNHD	83.59 ± 3.67**	81.29 ± 2.58*	85.67 ± 3.33	82.67 ± 3.87*
DIAL-GNN	85.19 ± 1.59**	86.39 ± 1.14**	85.92 ± 2.43	84.33 ± 1.29**
SGC	84.55 ± 3.95**	84.86 ± 5.33*	84.59 ± 4.85	84.63 ± 4.33**
Proposed	86.98 ± 3.06	87.53 ± 3.89	84.14 ± 2.59	87.93 ± 3.59

Table 2
Classification results (%) of all methods on OCD.

Methods	Accuracy	Sensitivity	Specificity	AUC
L1SVM	76.67 ± 3.26*	73.29 ± 5.24*	77.77 ± 2.69*	78.59 ± 1.88*
HOFC	83.92 ± 2.36*	83.24 ± 5.08**	84.11 ± 2.12*	83.17 ± 1.26**
SCP	85.83 ± 3.42**	85.52 ± 4.11*	86.80 ± 3.87*	86.93 ± 3.10*
CNHD	86.83 ± 4.05**	86.98 ± 3.87**	86.64 ± 4.53**	84.56 ± 4.39*
DIAL-GNN	85.59 ± 1.55**	85.33 ± 2.58*	86.39 ± 2.77**	85.93 ± 2.47*
SGC	87.06 ± 2.43	85.52 ± 5.26*	87.56 ± 4.55*	86.15 ± 5.01**
Proposed	88.05 ± 4.21	87.52 ± 4.15	89.42 ± 3.56	88.48 ± 4.33

Table 3
Classification results (%) of all methods on ADNI.

Methods	Accuracy	Sensitivity	Specificity	AUC
L1SVM	76.88 ± 4.25*	77.83 ± 3.58*	76.10 ± 2.85*	74.95 ± 2.36*
HOFC	80.25 ± 1.72*	78.89 ± 2.09*	81.35 ± 2.12**	81.26 ± 3.78**
SCP	84.89 ± 3.98**	85.14 ± 3.21*	84.80 ± 2.83*	84.89 ± 3.25**
CNHD	85.97 ± 4.36**	87.21 ± 5.21	84.70 ± 3.95**	84.65 ± 4.47*
DIAL-GNN	86.97 ± 1.76**	86.88 ± 2.77**	87.02 ± 1.33	87.68 ± 2.73*
SGC	86.96 ± 2.81**	88.24 ± 2.85	86.15 ± 3.66**	88.78 ± 4.69**
Proposed	88.84 ± 3.22	89.55 ± 1.85	88.25 ± 2.49	90.22 ± 3.4

3.2. Result analysis

3.2.1. Supervised learning

In the experiments of supervised learning, we used all labeled subjects as the training set. We report the results of all methods in Tables 1-3 and list our observations as follows.

First, our proposed method achieved the best classification performance on all three data sets, in terms of four evaluation metrics, followed by SGC, DIAL-GNN, CNHD, SCP, HOFC, and L1SVM. Moreover, our proposed method has statistically significant difference at 95% significance level, compared to most of comparison methods, in terms of evaluation metrics including ACC, SEN, SPE, and AUC. Specifically, our method on average improved by 2.17%, 1.71%, and 1.68%, respectively, compared to the best comparison method SGC, on FTD, OCD, and AD, for all evaluation metrics. The possible reasons are that (i) our multi-graph fusion method takes the inter-subject variability, the heterogeneity across subjects, and the discriminative ability into account to output homogenous and discriminative representation, and (ii) our proposed method jointly selects features (i.e., brain regions) and conducts classification to avoid the influence of redundant features on high-dimensional data.

Second, L1SVM uses fully-connected FCNs, and other methods (i.e., our proposed method, DIAL-GNN, SGC, SCP, and HOFC) learns sparse FCNs. As a result, L1SVM obtained the worst classification performance. For example, the worst method for learning sparse FCNs, i.e., HOFC, on average improved by 14.74%, 7.03%, and 3.99%, respectively, compared to L1SVM, on FTD, OCD, and AD, in terms of all four evaluation metrics. In particular, our proposed method fuses a fully-connected FCNs with a 1NN FCN from each subject to output a sparse FCN for every subject, followed by employing the L1SVM to conduct the classification task. On the contrary, L1SVM directly regards a fully-connected FCN as the representa-

tion for each subject to conduct the classification task. Moreover, our method on average improved by 16.98%, compared to L1SVM, in terms of Sensitivity, on all three data sets. This indicates the reasonability of sparse FCNs. That is, the sparse FCN is more suitable than the fully-connected FCN for conducting the FCN analysis on the rs-fMRI data.

Third, the methods (e.g., HOFC, SCP, and our method) design different models to generate sparse FCNs. Specifically, they first generate the sparse FCNs in different ways and then convert the upper triangle parts of the derived FCNs as the new representation of the subjects. As a consequence, our method considers the heterogeneity across subjects to outperform other methods (e.g., HOFC and SCP). This demonstrates that it is reasonable for taking into account the heterogeneity across subjects. In addition, CNHD, SGC, and DIAL-GNN also considers the heterogeneity across subjects and outperforms either HOFC or SCP. This verifies the importance of the consideration of the heterogeneity across subjects again. Furthermore, our proposed method is the only one to fuse the information from multiple FCNs so that achieving the best classification performance. This shows that our multi-graph fusion method is feasible because it can use the common and complementary information among multiple FCNs to output the discriminative representation of all subjects.

3.2.2. Personalized classification

To verify the effectiveness of our proposed semi-supervised method, we randomly selected different percentages of labeled subjects (i.e., 20%, 40%, 60%, and 80%) from the whole data set as the training set. In this case, the methods (i.e., L1SVM, HOFC, SCP, and CNHD) only used labeled subjects to train the classifiers, while the methods (i.e., our method, SGC, and DIAL-GNN) used all subjects (i.e., labeled subjects and unlabeled subjects) to train

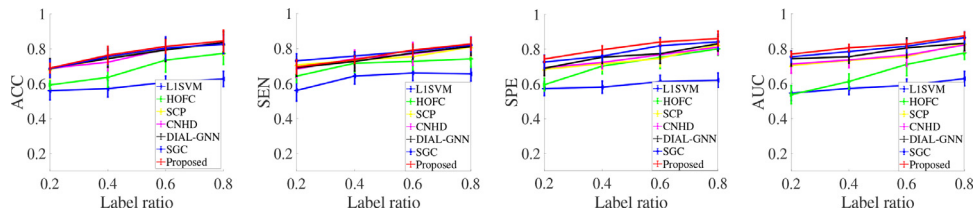


Fig. 3. Classification results (mean \pm standard deviation) of personalized classification on FTD.

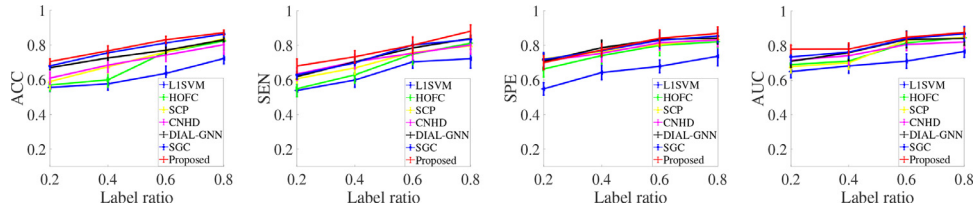


Fig. 4. Classification results (mean \pm standard deviation) of personalized classification on OCD.

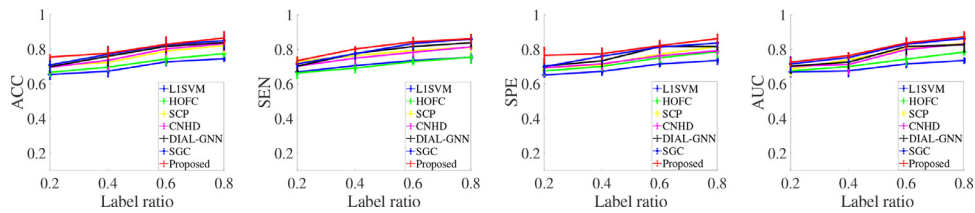


Fig. 5. Classification results (mean \pm standard deviation) of personalized classification on AD.

the classifiers. We report the classification results of all methods in Figs. 3–5.

Similar to the scenarios of supervised learning, our proposed method still achieved the best performance, followed by SGC, DIAL-GNN, CNHD, HOFC, SCP and L1SVM, in terms of semi-supervised learning. Moreover, the paired-sample t-tests between our method and every comparison method showed that our proposed method has statistically significant difference at 95% significance level, compared to every comparison method, in terms of evaluation metrics including ACC, SEN, SPE, and AUC, at different label ratios on each data set. For example, our method on average improved by 1.97% and 14.91%, respectively, compared to the best comparison method SGC and the worst comparison method L1SVM, on three data sets, in terms of all evaluation metrics. Besides, we have other observations as follows.

By comparing the semi-supervised learning methods (i.e., SGC, DIAL-GNN, and our method) with the supervised learning methods (i.e., CNHD, HOFC, SCP, and L1SVM), the former methods outperformed the latter methods. Specifically, the former methods on average improved by 8.74%, 6.52%, 6.05%, and 8.34%, respectively, compared to the latter methods, on all three data sets with all different percentages of labelled subjects, in terms of ACC, SEN, SPE, AUC. This reason is that the semi-supervised learning methods use more data (i.e., the unlabelled data) than the supervised learning methods during the training process. As a result, the semi-supervised learning methods may output more robust classifiers than the supervised learning methods. In particular, the improvement of the semi-supervised learning methods over the supervised learning methods achieves the maximum while the percentage of labeled subjects in the training set is small, i.e., 20%. For example, the classification accuracy of our proposed method improved by on average 3.08%, 3.71%, 4.15%, and 3.29%, respectively, compared to the performance of the best method of supervised learning, i.e., CNHD, in terms of 20%, 40%, 60%, and 80% of the percentage of labeled subjects in the training set, on all three data sets.

The percentage of labeled subjects in the training set is small, all methods achieved worse performance. Moreover, the more the percentage of labelled subjects is, the lower the improvement of our proposed method over the comparison methods is. For example, all methods achieved worse experiment results with only 20% label subjects for the training process. This contributes to the fact that it is difficult to build robust classifiers without sufficient labeled subjects. On the contrary, the classification performance increases with the increased percentage of the labeled subjects for the training process. For example, the classification accuracy of our method increased 7.41% from 20% to 40%, in terms of the percentage of labeled subjects, while improving by 4.47%, from 60% to 80%, on all three data sets. The main reason is that the limited labeled subjects is difficult to guarantee the discriminative ability of the classifiers.

3.2.3. Multi-graph fusion effectiveness

The novelty of our multi-graph fusion method is to automatically learn both the common information and the complementary information among multiple FCNs. To verify the effectiveness of our proposed fusion method, we used our method to first generate a sparse FCN for each subject followed by extracting its vector representation (i.e., the upper triangle part of the sparse FCN) as the new representation of the subjects, and then feed the new representation to the methods (e.g., L1SVM and SGC) to output the classification performance. We report the experimental results in Fig. 6. It is noteworthy that we only analyzed the best and the worst comparison methods due to the space limitations.

From Fig. 6, the classification performance of the methods (i.e., L1SVM and SGC) is better than the results of the corresponding methods in Tables 1–3. For example, L1SVM and SGC, respectively, on average improved by 16.2% and 4.3%, compared to the corresponding results in Tables 1–3, on all three data sets, in terms of all four evaluation metrics. This implies that the sparse FCNs outputted by our proposed multi-graph fusion method has strongly discriminative ability.

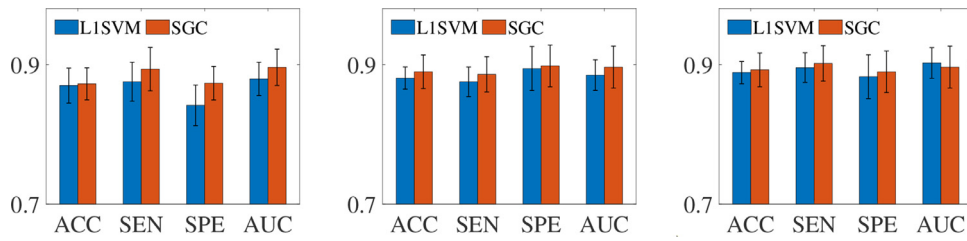


Fig. 6. Classification results of L1SVM and SGC using the sparse FCNs produced by our method on FTD (left), OCD (middle), and ADNI (right).

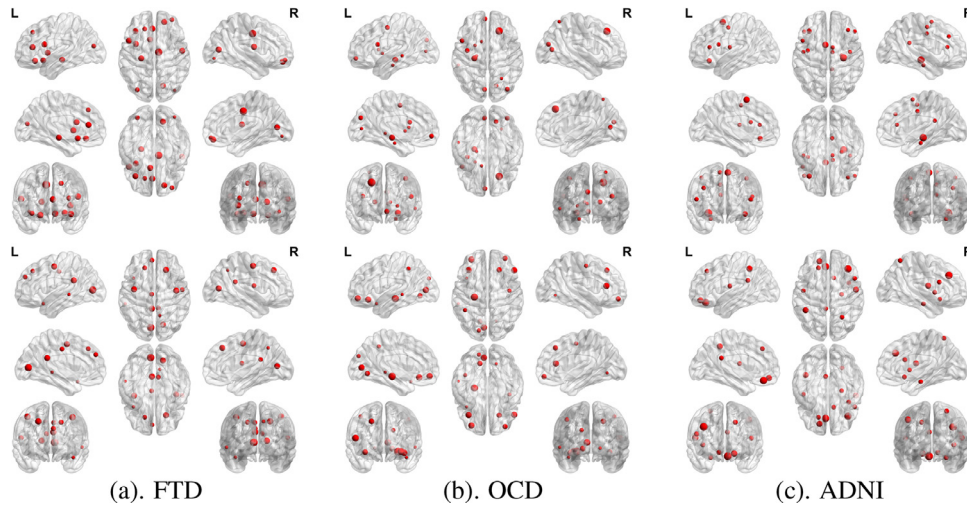


Fig. 7. Visualization of top selected brain regions selected and the connected regions by L1SVM (upper) and our method (bottom) on FTD, OCD, and ADNI.

3.2.4. Feature selection effectiveness

In this section, we designed two kinds of experiments to investigate the effectiveness of the selected features by our method. Specifically, in our experiments, we repeated the 10-fold cross validation scheme 10 times, and thus outputting 100 subsets of the selected features. We further calculated the selected frequency of all features over all 4005 features, and then reported the features with the frequency over 90 out of 100 times as the top features. As a result, our method selected 1270, 898, 923 nodes out of 4005 nodes, respectively, on FTD, OCD, and AD, while L1SVM selected 1477, 1058, and 1213 nodes, respectively. It is noteworthy that each node or feature is related to two brain regions.

We reported the top selected brain regions based on the selected features and plot the top selected brain regions of our method and L1SVM in Fig. 7. Based on the visualization of the top selected brain regions, many selected regions from both our method and L1SVM have been verified related to the neuro-diseases. Specifically, in Fig. 7(a), most of the nodes selected by our method occur in frontal and temporal lobes, which is consistent with the current neurobiological findings on FTD (de Haan et al., 2009). However, a large portion of nodes identified by L1SVM have low correlation with FTD. In Fig. 7(b), our method finds the brain regions, such as orbital-frontal cortex, caudate, thalamus, which are included in the cortical-striato-thalamic circuits, and is considered as the theoretical neuroanatomical network of OCD (Gillan et al., 2015; 2011). In particular, our method selected the brain regions throughout the whole brain because AD has been demonstrated to be associated with whole brain atrophy (Schott et al., 2008). On the contrary, L1SVM only selected the frontal regions on the data set ADNI.

In our experiments, we first obtained three data sets based on the original data, i.e., Feature 1, Feature 2, and Feature 3. Feature 1

represents the original data sets with high-dimensional data. Feature 2 is the data sets with the features selected by our method. Feature 3 is the data sets with the features unselected by our method. We fed these new data sets to L1SVM and reported the classification results in Fig. 8. From Fig. 8, Feature 2 achieved the best performance, followed by Feature 1 and Feature 3. Specifically, the classification accuracy of Feature 2 improved on average by 3.58%, compared to Feature 1, on all three data sets. The reason is that the original data contains redundant feature and noise, which affects classification performance. This illustrates (1) the effectiveness of the features selected by our method, and (2) feature selection can improve model performance.

3.2.5. Template visualization

In Eq. (3), we denoted \mathbf{G} and \mathbf{H} , respectively, as the positive and negative template, to make the outputted representation containing discriminative ability. In this section, we visualized the templates in Fig. 9 for all three data sets.

Obviously, the difference between the disease template and the healthy control template is significant. Moreover, the selected brain regions in the templates can be found as the top selection regions in Fig. 7. This indicates the outputted templates can make our proposed method have discriminative ability as well as are possibly used for guiding either the parcellations for new subjects or measuring the group difference in the study of medical image analysis.

4. Discussion

In this section, we discuss time complexity of all methods, the variations of our proposed method with different k values, and the variations of our proposed with different initialization, the variations of our proposed with different numbers of graphs.

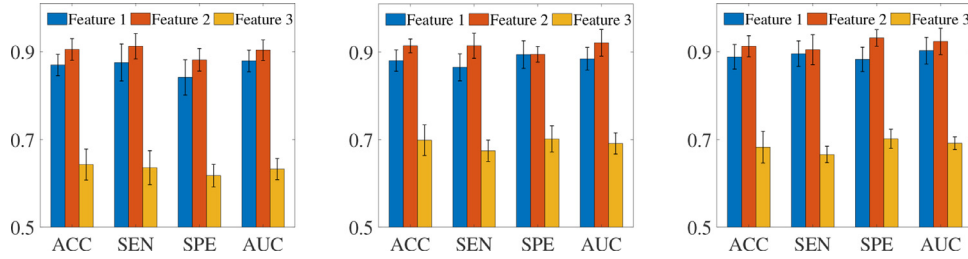


Fig. 8. Classification results of L1SVM using three kinds of data (i.e., Feature 1, Feature 2, and Feature 3) on FTD (left), OCD (middle), and ADNI (right).

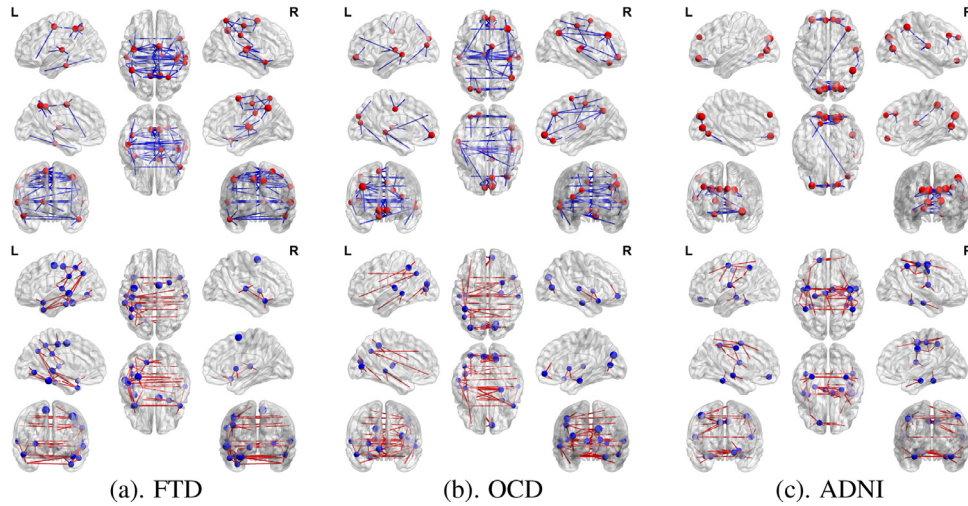


Fig. 9. Visualization of templates outputted by our method on FTD, OCD and AD (upper) and healthy control (bottom).

4.1. Time complexity analysis

The multi-graph generation is off-line. Hence, we ignore the calculation of the time complexity and the space complexity. The multi-graph fusion method takes a closed-form solution for the optimization of \mathbf{S}^m ($m = 1, \dots, M$), \mathbf{H} and \mathbf{G} . The time complexity of \mathbf{S}^m is $O(Mn^2)$ and the time complexity of either \mathbf{H} or \mathbf{G} is $O(n^2)$, where M and n , respectively, represent the number of the subjects and the number of brain regions. Hence, the time complexity of our multi-graph fusion method is $O(lMn^2)$, i.e., linear to the subject size, where l is the iteration number and is less than 50 in our experiments. Moreover, the proposed multi-graph fusion method needs to store \mathbf{S}^m ($m = 1, \dots, M$), \mathbf{H} , and \mathbf{G} in the memory with the space complexity $O(Mn^2)$. The time complexity of L1SVM is linear to the subject size, while its space time complexity is $O(Mn^2)$ (Fan et al., 2008). Moreover, based on (Fan et al., 2008), L1SVM fast achieves convergence.

The time complexity of DIAL-GNN and SGC, respectively, are $O(TdM^2)$ and $O(M^2d)$, where M , T , $d = \frac{n(n-1)}{2}$, and n , are the number of samples, the iterations, the features, the brain regions, respectively. The time complexity of HOFC, SCP, CNHD, and L1SVM, respectively, are $O(M(\omega - 1)n^2 + Md)$, $O(Mn^2 + Md)$, $O(Mn^3)$ and $O(Md)$, where ω is the window length.

More specifically, two deep learning methods (i.e., DIAL-GNN and SGC) is quadratic to the sample size, while four traditional methods (i.e., HOFC, SCP, CNHD, L1SVM, and our method) is linear to the sample size. However, in our data sets, the sample size is smaller than the number of the brain regions. Hence, two deep learning methods are faster than the traditional methods.

In Table 4, we report the training time of all methods on three data sets. HOFC needs the maximal training costs. Our proposed

Table 4

Training time (Seconds) of all methods on three data sets.

Dataset	FTD	OCD	AD
L1SVM	1997 s	951 s	1248 s
HOFC	3610 s	1857 s	2387s
SCP	2487 s	1293	1639 s
CNHD	2937	1433	2108 s
DIAL-GNN	872 s	407 s	625 s
SGC	668 s	272 s	438 s
Proposed	2616 s	1122 s	1753 s

method needs the second most time as it includes the process of L1SVM. As a result, our proposed multi-graph fusion model is efficient, i.e., linear to the sample size. For example, the time cost for our proposed multi-graph fusion model is 619 seconds, 171 seconds, and 505 seconds, respectively, for data sets FTD, OCD, and AD.

4.2. Sensitivity analysis of k values

We varied the values of k as $k = \{5, 10, 15, 20, 25\}$ and reported the classification accuracy of our proposed method with different k values on three data sets in Table 5. It is noteworthy that the data set OCD only takes the values of k as $k = \{5, 10, 15\}$ as it only has 20 healthy control subjects. As a result, our proposed method is insensitive to the k values as the gap of two different scenarios is not significant in terms of classification accuracy. For example, the difference between the case ' $k = 10$ ' and other cases (i.e., $k \neq 10$) varied on average by 1.12%, 0.26%, 0.79%, on data sets FTD, OCD, and AD, respectively, in terms of classification accuracy.

Table 5
Classification results (ACC%) of our proposed method with different k on three data sets.

k	5	10	15	20	25
FTD	84.25 ± 3.89	86.98 ± 1.56	86.98 ± 2.87	86.10 ± 1.95	86.10 ± 1.95
OCD	87.53 ± 4.20	88.05 ± 5.37	88.05 ± 4.29	–	–
AD	87.71 ± 4.20	88.84 ± 3.77	87.53 ± 3.87	87.53 ± 5.43	88.06 ± 1.95

Table 6
Classification results (ACC%) of our proposed method on five different initializations

	Initialization 1	Initialization 2	Initialization 3	Initialization 4	our method
FTD	86.25 ± 1.58	85.87 ± 4.25	86.18 ± 3.11	85.23 ± 1.45	86.18 ± 1.95
OCD	88.05 ± 3.78	88.05 ± 2.56	88.05 ± 2.66	87.49 ± 2.77	88.58 ± 3.16
AD	88.18 ± 3.15	87.46 ± 2.97	89.38 ± 4.58	87.53 ± 5.43	88.18 ± 2.54

4.3. Sensitivity analysis of initialization

In [Algorithm 1](#), we initialize \mathbf{S}^m ($m = 1, \dots, M$) as the average of $\mathbf{A}^{m,v}$ ($v = 1, \dots, V$), which makes the optimization of [Eq. \(2\)](#) converge within tens of iterations. Moreover, the result of [Eq. \(2\)](#) is insensitive to the initialization of \mathbf{S}^m ($m = 1, \dots, M$).

In the experiment, we used the Matlab function `rand()` to generate two uniformly distributed random matrices (i.e., Initialization 1, Initialization 2), and then used the Matlab function `randn()` to generate two random matrices that obey the Gaussian distribution (i.e., Initialization 3, Initialization 4). In particular, our method set the average of $\mathbf{A}^{m,v}$ ($v = 1, \dots, V$) as the initialization of \mathbf{S}^m . From [Table 6](#), our proposed method is insensitive to the initialization of \mathbf{S}^m . The main reason is that our optimization method is an iterative process that updates the value of \mathbf{S}^m at each iteration, so that even a poor initialization can finally achieve reasonable results after many iterations. In addition, our initialization method converges faster than other initialization methods. The possible reason is our initialization \mathbf{S}^m is the average of $\mathbf{A}^{m,v}$ ($v = 1, \dots, V$), which is closer to the final \mathbf{S}^m .

4.4. Effectiveness with different numbers of graphs

Our proposed method only explored the relationship between the fully-connected FCN and the extremely sparse FCN, i.e., the 1NN FCN, aiming at learning the suitable connection number of the brain regions for each subject. To do this, our proposed method took into account the following issues such as noise, individual variability for each subject, and the inter-group heterogeneity across subjects. Actually, we can combine much sparse FCNs with the current two FCNs to learn the connection number. Thus, we added five more FCNs (e.g., 3NN, 5NN, 8NN, 10NN, and 15NN) into the proposed objective function. As a result, the classification performance had no significant difference from the reported one in this work, i.e., only on average improving by about 0.25%. The possible reason is that two FCNs (such as the fully-connected FCN and the 1NN FCN) are enough for our proposed method to automatically learn the connection number between 1 and n (where n is the node number).

5. Conclusion

In this paper, we proposed a new framework for functional connectivity network analysis using the rs-fMRI data which can explore both the common and the complementary information among multiple FCNs for each subject to improve the discriminative ability of the learned representation from the rs-fMRI data. The experimental results on three real data sets verified the effectiveness of the proposed method, compared to four comparison methods, in terms of the classification performance.

Our method separately conducts the feature learning and the classification task to achieve the interpretability, i.e., conducting feature selection to select the top brain regions related to the neuro-diseases. However, in the two-stage process, the optimal results of the first stage (i.e., the feature learning) do not guarantee to output the optimal results in the second stage (i.e., the classification task). On the contrary, SGC combines two stages together to achieve the second best performance in our experiments because other comparison methods (e.g., HOF and SCP) are also the two-stage methods. This motivates us to design deep learning models to further improve our proposed framework. However, the interpretability is still very challenging in the study of deep learning. Hence, in the future work, we will extend the proposed framework to simultaneously take into account conducting both the feature learning and the classification tasks in a unified deep framework to achieve its interpretability.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

CRediT authorship contribution statement

Jiangzhang Gan: Conceptualization, Methodology, Software, Writing - original draft. **Ziwen Peng:** Methodology, Investigation, Writing - original draft, Writing - review & editing. **Xiaofeng Zhu:** Writing - review & editing, Supervision, Project administration. **Rongyao Hu:** Software, Validation, Investigation. **Junbo Ma:** Resources, Data curation, Visualization. **Guorong Wu:** Methodology, Writing - review & editing.

Acknowledgements

This work was partially supported by the [National Key Research and Development Program of China](#) (Grant no. 2018AAA0102200), the [National Natural Science Foundation of China](#) (Grants nos. 61876046 and 31871113), the NIH AG049089, AG068399, AG059065, the [Sichuan Science and Technology Program](#) (Grants nos. 2018GZDZX0032 and 2019YFG0535), and the [China Scholarship Council](#) (CSC).

Appendix A. Optimization

In this paper, we employ the alternating optimization strategy ([Daubechies et al., 2010](#)) to optimize \mathbf{S}^m ($m = 1, \dots, M$), \mathbf{H} , and \mathbf{G} .

(i) Update $\mathbf{S}^1, \dots, \mathbf{S}^M$ by fixing \mathbf{H} and \mathbf{G}

$\mathbf{S}^1, \dots, \mathbf{S}^M$ include the representations of positive subjects, negative subjects, and unlabeled subjects, so we explain the optimization process one by one.

When m th subject is a negative subject, we obtain the objective function with respect to \mathbf{S}^m as follows:

$$\begin{aligned} \min_{\mathbf{S}^m} & \sum_{v=1}^V \|\mathbf{S}^m - \mathbf{A}^{m,v}\|_F^2 + \alpha \|\mathbf{S}^m - \mathbf{H}\|_F^2 + \\ & \beta \left(\sum_{p \in \mathcal{G}(m)} \|\mathbf{S}^m - \mathbf{S}^p\|_F^2 - \lambda^m \sum_{q \in \mathcal{F}(m)} \|\mathbf{S}^m - \mathbf{S}^q\|_F^2 \right) \quad (9) \\ \text{s.t., } & \forall i, \mathbf{s}_{i,i}^m \mathbf{1} = 1, \mathbf{s}_{i,i}^m = 0, \mathbf{s}_{i,j}^m \geq 0 \text{ if } j \in \mathcal{N}(i), \text{ otherwise } 0. \end{aligned}$$

\mathbf{S}^m represents the neighbor relationship between two brain regions (we represent \mathbf{x}_i as the i th brain region), $\mathbf{s}_{i,i}^m$ represents the relationship between \mathbf{x}_i and other brain regions \mathbf{x}_j ($i \neq j$). $\mathbf{s}_{i,i}^m$ is only related to \mathbf{x}_i and \mathbf{x}_j ($i \neq j$), and is unrelated to \mathbf{S}^m . Therefore, in our optimization method, $\mathbf{s}_{i,i}^m$ is independent on \mathbf{S}^m ($i \neq j$), the objective function with respect to $\mathbf{s}_{i,i}^m$ is:

$$\begin{aligned} \min_{\mathbf{s}_{i,i}^m \mathbf{1}=1, \mathbf{s}_{i,i}^m=0, \mathbf{s}_{i,j}^m \geq 0} & \sum_{v=1}^V \|\mathbf{s}_{i,i}^m - \mathbf{a}_{i,i}^{m,v}\|_2^2 + \alpha \|\mathbf{s}_{i,i}^m - \mathbf{h}_{i,i}\|_2^2 \\ & + \beta \left(\sum_{p \in \mathcal{G}(m)} \|\mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^p\|_2^2 - \lambda^m \sum_{q \in \mathcal{F}(m)} \|\mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^q\|_2^2 \right) \quad (10) \end{aligned}$$

Expanding Eq. (10), we have

$$\begin{aligned} \min_{\mathbf{s}_{i,i}^m \mathbf{1}=1, \mathbf{s}_{i,i}^m=0, \mathbf{s}_{i,j}^m \geq 0} & \sum_{v=1}^V \text{Tr}(\mathbf{s}_{i,i}^m - \mathbf{a}_{i,i}^{m,v})^T (\mathbf{s}_{i,i}^m - \mathbf{a}_{i,i}^{m,v}) \\ & + \alpha \text{Tr}(\mathbf{s}_{i,i}^m - \mathbf{h}_{i,i})^T (\mathbf{s}_{i,i}^m - \mathbf{h}_{i,i}) \\ & + \beta \left(\sum_{p \in \mathcal{G}(m)} \text{Tr}(\mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^p)^T (\mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^p) \right. \\ & \left. - \lambda^m \sum_{q \in \mathcal{F}(m)} \text{Tr}(\mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^q)^T (\mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^q) \right) \quad (11) \end{aligned}$$

Expanding Eq. (11), we have

$$\begin{aligned} \min_{\mathbf{s}_{i,i}^m \mathbf{1}=1, \mathbf{s}_{i,i}^m=0, \mathbf{s}_{i,j}^m \geq 0} & \sum_{v=1}^V \text{Tr}(\mathbf{s}_{i,i}^{mT} \mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^{mT} \mathbf{a}_{i,i}^{m,v} - \mathbf{a}_{i,i}^{m,vT} \mathbf{s}_{i,i}^m + \mathbf{a}_{i,i}^{m,vT} \mathbf{a}_{i,i}^{m,v}) \\ & + \alpha \text{Tr}(\mathbf{s}_{i,i}^{mT} \mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^{mT} \mathbf{h}_{i,i} - \mathbf{h}_{i,i}^T \mathbf{s}_{i,i}^m + \mathbf{h}_{i,i}^T \mathbf{h}_{i,i}) \\ & + \beta \left(\sum_{p \in \mathcal{G}(m)} \text{Tr}(\mathbf{s}_{i,i}^{mT} \mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^{mT} \mathbf{s}_{i,i}^p - \mathbf{s}_{i,i}^{pT} \mathbf{s}_{i,i}^m + \mathbf{s}_{i,i}^{pT} \mathbf{s}_{i,i}^p) \right. \\ & \left. - \lambda^m \sum_{q \in \mathcal{F}(m)} \text{Tr}(\mathbf{s}_{i,i}^{mT} \mathbf{s}_{i,i}^m - \mathbf{s}_{i,i}^{mT} \mathbf{s}_{i,i}^q - \mathbf{s}_{i,i}^{qT} \mathbf{s}_{i,i}^m + \mathbf{s}_{i,i}^{qT} \mathbf{s}_{i,i}^q) \right) \quad (12) \end{aligned}$$

After that, we obtain:

$$\begin{aligned} \min_{\mathbf{s}_{i,i}^m \mathbf{1}=1, \mathbf{s}_{i,i}^m=0, \mathbf{s}_{i,j}^m \geq 0} & \text{Tr}((V + \alpha + \beta(k - \lambda^m k)) \mathbf{s}_{i,i}^{mT} \mathbf{s}_{i,i}^m \\ & - 2 \left(\sum_{v=1}^V \mathbf{a}_{i,i}^{m,vT} + \alpha \mathbf{h}_{i,i}^T + \beta \left(\sum_{p=1}^k \mathbf{s}_{i,i}^{pT} - \lambda^m \sum_{q=1}^k \mathbf{s}_{i,i}^{qT} \right) \mathbf{s}_{i,i}^m \right. \\ & \left. + \sum_{v=1}^V \mathbf{a}_{i,i}^{m,v} + \alpha \mathbf{h}_{i,i} + \beta \left(\sum_{p=1}^k \mathbf{s}_{i,i}^{pT} \mathbf{s}_{i,i}^p - \lambda^m \sum_{q=1}^k \mathbf{s}_{i,i}^{qT} \mathbf{s}_{i,i}^q \right) \right) \quad (13) \end{aligned}$$

After conducting mathematical transformation, we have

$$\min_{\mathbf{s}_{i,i}^m \mathbf{1}=1, \mathbf{s}_{i,i}^m=0, \mathbf{s}_{i,j}^m \geq 0} \|\mathbf{s}_{i,i}^m - \mathbf{f}_{i,i}^{m-}\|_2^2 \quad (14)$$

where

$$\mathbf{f}_{i,i}^{m-} = \frac{\sum_{v=1}^V \mathbf{a}_{i,i}^{m,vT} + \alpha \mathbf{h}_{i,i}^T + \beta \left(\sum_{p=1}^k \mathbf{s}_{i,i}^{pT} - \lambda^m \sum_{q=1}^k \mathbf{s}_{i,i}^{qT} \right)}{V + \alpha + \beta(k - \lambda^m k)} \in \mathbb{R}^{n \times 1} \quad (15)$$

We consider the Lagrangian Function of problem Eq. (14) as

$$\mathcal{L}(\mathbf{s}_{i,i}^m, \varphi_1, \omega) = \|\mathbf{s}_{i,i}^m - \mathbf{f}_{i,i}^{m-}\|_2^2 - \varphi_1 (\mathbf{s}_{i,i}^m \mathbf{1} - 1) - \omega \mathbf{s}_{i,i}^m \quad (16)$$

where φ_1 is a Lagrange multiplier and ω is a vector of nonnegative Lagrange multipliers. Conducting the differential with respect to $\mathbf{s}_{i,i}^m$ and setting the results as zero, we obtain:

$$\frac{d\mathcal{L}}{d\mathbf{s}_{i,i}^m} = \mathbf{s}_{i,i}^m - \mathbf{f}_{i,i}^{m-} - \frac{1}{2} \varphi_1 - \frac{1}{2} \omega_j = 0 \quad (17)$$

where $f_{i,i}^{m-}$ is the j th element of $\mathbf{f}_{i,i}^{m-}$. To facilitate the calculation, we set $\sigma_1 = \frac{1}{2} \varphi_1$, $\tau_j = \frac{1}{2} \omega_j$.

The complementary slackness of the Karush–Kuhn–Tucker (KKT) conditions implies that the condition $w\tau_j = 0$ holds while $\mathbf{s}_{i,i}^m > 0$. Thus, we have the closed-form solution for $\mathbf{s}_{i,i}^m$ as:

$$\mathbf{s}_{i,i}^m = (f_{i,i}^{m-} + \sigma_1)_+, j = 1, \dots, n \quad (18)$$

where $f_{i,i}^{m-}$ is the j th element of $\mathbf{f}_{i,i}^{m-}$.

By following the same process from Eqs. (9) to (18), we have

$$\mathbf{s}_{i,j}^m = \begin{cases} (f_{i,j}^{m-} + \sigma_1)_+, & m \in \mathcal{D} \\ (f_{i,j}^{m-} + \sigma_2)_+, & m \in \mathcal{E} \\ (f_{i,j}^{m-} + \sigma_3)_+, & m \in \mathcal{U} \end{cases} \quad (19)$$

where

$$\begin{cases} f_{i,j}^{m+} = \frac{\sum_{v=1}^V \mathbf{a}_{i,j}^{m,vT} + \alpha \mathbf{g}_{i,j}^T + \beta \left(\sum_{p=1}^k \mathbf{s}_{i,j}^{pT} - \lambda^m \sum_{q=1}^k \mathbf{s}_{i,j}^{qT} \right)}{V + \alpha + \beta(k - \lambda^m k)} \\ f_{i,j}^m = \frac{\sum_{v=1}^V \mathbf{a}_{i,j}^{m,v} + \beta \left(\sum_{p=1}^k \mathbf{s}_{i,j}^{pT} - \lambda^m \sum_{q=1}^k \mathbf{s}_{i,j}^{qT} \right)}{V + \beta(k - \lambda^m k)}. \end{cases} \quad (20)$$

σ_1 , σ_2 and σ_3 are the Lagrange multipliers.

(ii) Update \mathbf{H} and \mathbf{G} by fixing $\mathbf{S}^1, \dots, \mathbf{S}^M$

When $\mathbf{S}^1, \dots, \mathbf{S}^M$ are fixed, the objective function with respect to \mathbf{H} and \mathbf{G} are:

$$\begin{cases} \min_{\mathbf{h}_{i,i} \mathbf{1}=1, \mathbf{h}_{i,i}=0, \mathbf{h}_{i,j} \geq 0} \sum_{m=1}^{|\mathcal{D}|} \|\mathbf{S}^m - \mathbf{H}\|_F^2 \\ \min_{\mathbf{g}_{i,i} \mathbf{1}=1, \mathbf{g}_{i,i}=0, \mathbf{g}_{i,j} \geq 0} \sum_{m=1}^{|\mathcal{E}|} \|\mathbf{S}^m - \mathbf{G}\|_F^2 \end{cases} \quad (21)$$

According to the KKT conditions, we have:

$$\begin{cases} \mathbf{h}_{i,j} = (\hat{\mathbf{s}}_{i,j}^{m-} + \sigma_4)_+ \\ \mathbf{g}_{i,j} = (\hat{\mathbf{s}}_{i,j}^{m+} + \sigma_5)_+ \end{cases} \quad (22)$$

where $\hat{\mathbf{s}}_{i,j}^{m-} = (\sum_{m \in \mathcal{D}} \mathbf{s}_{i,j}^{mT}) / |\mathcal{D}|$, $\hat{\mathbf{s}}_{i,j}^{m+} = (\sum_{m \in \mathcal{E}} \mathbf{s}_{i,j}^{mT}) / |\mathcal{E}|$, σ_4 and σ_5 are Lagrange multipliers.

The values of the Lagrange multipliers σ_1 , σ_2 , σ_3 , σ_4 , and σ_5 , can be obtained by Lemma 1 (Duchi et al., 2008). For simplicity, we list the details of σ_3 as follows and the values of σ_1 , σ_2 , σ_4 , and σ_5 can be obtained by similar principles.

Lemma 1. By denoting $\mathbf{s}_{i,r}^{m*}$ the optimal solution in Eq. (18), letting r and u be two indices, and $f_{i,r}^m > f_{i,u}^m$, if $\mathbf{s}_{i,r}^{m*} = 0$, then $\mathbf{s}_{i,u}^{m*}$ must be equal to zero.

Based on Lemma 1, some integers $\mathbf{I} = [\rho]$, $1 \leq \rho \leq n$, meet the non-zero components of the sorted optimal solutions, i.e.,

$$\sigma_3 = \frac{1}{\rho} \left(1 - \sum_j^{\rho} f_{i,j}^m \right). \quad (23)$$

As a result, the optimal $\mathbf{s}_{i,i}^{m*}$ can be described as $\mathbf{s}_{i,i}^{m*} = \max\{f_{i,i}^m + \sigma_3, 0\}$, where the value of the optimal ρ is automatically obtained by Lemma 2 (Duchi et al., 2008).

Lemma 2. Let $\boldsymbol{\eta}$ represents the vector after sorting $\mathbf{f}_{i,i}^m$ in a descending order, the number of strictly non-negative elements in $\mathbf{s}_{i,i}^{m*}$ is $\rho = \max\{j \in [n] : \eta_j - \frac{1}{j} (\sum_{i=1}^j \eta_i - 1) > 0\}$.

Based on Lemma 2, the non-zero number in the i th row $\mathbf{s}_{i,\cdot}^m$, i.e., the number of brain regions connected to the i th brain region, is different from the non-zero number in the j th row $\mathbf{s}_{j,\cdot}^m$ ($i \neq j$). It is noteworthy that previous sparse methods set the same number of brain regions connected to each brain region. Obviously, our method is more flexible, compared to previous methods.

Appendix B. Convergence

Theorem 2 guarantees the convergence of Algorithm 1.

Theorem 2. Algorithm 1 decreases the objective value of Eq. (2) monotonically in each iteration until convergence.

Proof. We first denote $\mathbf{S}^{m(t+1)}$, $\mathbf{G}^{(t+1)}$, and $\mathbf{H}^{(t+1)}$, as the updated of $\mathbf{S}^{m(t)}$, $\mathbf{G}^{(t)}$, and $\mathbf{H}^{(t)}$, respectively, where t is the t th iteration. If $\mathbf{S}^1, \dots, \mathbf{S}^M$, and \mathbf{G} are fixed, \mathbf{H} has a closed form solution, so we have:

$$\begin{aligned} & \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^{m(t+1)} - \mathbf{A}^{m,v}\|_F^2 + \beta \mathcal{R}_2(\mathbf{S}^{1(t+1)}, \dots, \mathbf{S}^{M(t+1)}) \\ & + \alpha \mathcal{R}_1(\mathbf{H}^{(t+1)}, \mathbf{G}^{(t+1)}, \mathbf{S}^{1(t+1)}, \dots, \mathbf{S}^{M(t+1)}) \\ & \leq \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^{m(t)} - \mathbf{A}^{m,v}\|_F^2 + \beta \mathcal{R}_2(\mathbf{S}^{1(t)}, \dots, \mathbf{S}^{M(t)}) \\ & + \alpha \mathcal{R}_1(\mathbf{H}^{(t)}, \mathbf{G}^{(t)}, \mathbf{S}^{1(t)}, \dots, \mathbf{S}^{M(t)}) \end{aligned} \quad (24)$$

If $\mathbf{S}^1, \dots, \mathbf{S}^M$, and \mathbf{H} are fixed, \mathbf{G} has a closed form solution, so we have:

$$\begin{aligned} & \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^{m(t+1)} - \mathbf{A}^{m,v}\|_F^2 + \beta \mathcal{R}_2(\mathbf{S}^{1(t+1)}, \dots, \mathbf{S}^{M(t+1)}) \\ & + \alpha \mathcal{R}_1(\mathbf{H}^{(t)}, \mathbf{G}^{(t+1)}, \mathbf{S}^{1(t+1)}, \dots, \mathbf{S}^{M(t+1)}) \\ & \leq \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^{m(t)} - \mathbf{A}^{m,v}\|_F^2 + \beta \mathcal{R}_2(\mathbf{S}^{1(t)}, \dots, \mathbf{S}^{M(t)}) \\ & + \alpha \mathcal{R}_1(\mathbf{H}^{(t)}, \mathbf{G}^{(t)}, \mathbf{S}^{1(t)}, \dots, \mathbf{S}^{M(t)}) \end{aligned} \quad (25)$$

The optimization of \mathbf{S}^m is independent on the optimization of $\mathbf{S}^{m'} (m' \neq m)$. If \mathbf{H} and \mathbf{G} are fixed, \mathbf{S}^m has a closed form solution. That is,

$$\begin{aligned} & \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^{m(t+1)} - \mathbf{A}^{m,v}\|_F^2 + \beta \mathcal{R}_2(\mathbf{S}^{1(t+1)}, \dots, \mathbf{S}^{M(t+1)}) \\ & + \alpha \mathcal{R}_1(\mathbf{H}^{(t)}, \mathbf{G}^{(t)}, \mathbf{S}^{1(t+1)}, \dots, \mathbf{S}^{M(t+1)}) \\ & \leq \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^{m(t)} - \mathbf{A}^{m,v}\|_F^2 + \beta \mathcal{R}_2(\mathbf{S}^{1(t)}, \dots, \mathbf{S}^{M(t)}) \\ & + \alpha \mathcal{R}_1(\mathbf{H}^{(t)}, \mathbf{G}^{(t)}, \mathbf{S}^{1(t)}, \dots, \mathbf{S}^{M(t)}) \end{aligned} \quad (26)$$

Combining Eqs. (24) and (25) with Eq. (26), we have,

$$\begin{aligned} & \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^{m(t+1)} - \mathbf{A}^{m,v}\|_F^2 + \beta \mathcal{R}_2(\mathbf{S}^{1(t+1)}, \dots, \mathbf{S}^{M(t+1)}) \\ & + \alpha \mathcal{R}_1(\mathbf{H}^{(t+1)}, \mathbf{G}^{(t+1)}, \mathbf{S}^{1(t+1)}, \dots, \mathbf{S}^{M(t+1)}) \\ & \leq \sum_{m=1}^M \sum_{v=1}^V \|\mathbf{S}^{m(t)} - \mathbf{A}^{m,v}\|_F^2 + \beta \mathcal{R}_2(\mathbf{S}^{1(t)}, \dots, \mathbf{S}^{M(t)}) \\ & + \alpha \mathcal{R}_1(\mathbf{H}^{(t)}, \mathbf{G}^{(t)}, \mathbf{S}^{1(t)}, \dots, \mathbf{S}^{M(t)}) \end{aligned} \quad (27)$$

Eq. (27) indicates that Algorithm 1 decreases the objective value of Eq. (2) in each iteration. Hence, the proof of Theorem 2 is completed. \square

References

- Acharya, U.R., Fernandes, S.L., WeiKoh, J.E., Ciaccio, E.J., Fabell, M.K.M., Tanik, U.J., Rajinikanth, V., Yeong, C.H., 2019. Automated detection of Alzheimer's disease using brain MRI images—a study with various feature extraction techniques. *J. Med. Syst.* 43 (9), 302.
- Betzal, R.F., Bassett, D.S., 2017. Multi-scale brain networks. *Neuroimage* 160, 73–83.
- Brier, M.R., Mitra, A., McCarthy, J.E., Ances, B.M., Snyder, A.Z., 2015. Partial covariance based functional connectivity computation using Ledoit–Wolf covariance regularization. *NeuroImage* 121, 29–38.
- Chen, X., Zhang, H., Gao, Y., Wee, C.-Y., Li, G., Shen, D., Initiative, A.D.N., 2016. High-order resting-state functional connectivity network for MCI classification. *Hum. Brain Mapp.* 37 (9), 3282–3296.
- Chen, Y., Wu, L., Zaki, M. J., 2019. Deep iterative and adaptive learning for graph neural networks. *arXiv preprint arXiv:1912.07832*
- Daubechies, I., DeVore, R., Fornasier, M., Güntürk, C.S., 2010. Iteratively reweighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math.* 63 (1), 1–38.
- Duchi, J., Shalev-Shwartz, S., Singer, Y., Chandra, T., 2008. Efficient projections onto the l_1 -ball for learning in high dimensions. In: *ICML*, pp. 272–279.
- Eavani, H., Satterthwaite, T.D., Filipovych, R., Gur, R.E., Gur, R.C., Davatzikos, C., 2015. Identifying sparse connectivity patterns in the brain using resting-state fMRI. *Neuroimage* 105, 286–299.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J., 2008. Liblinear: a library for large linear classification. *J. Mach. Learn. Res.* 9, 1871–1874.
- Gan, J., Zhu, X., Hu, R., Zhu, Y., Ma, J., Peng, Z., Wu, G., 2020. Multi-graph fusion for functional neuroimaging biomarker detection. In: *IJCAL*, pp. 580–586.
- Gillan, C.M., Apergis-Schoute, A.M., Morein-Zamir, S., Urcelay, G.P., Sule, A., Fineberg, N.A., Sahakian, B.J., Robbins, T.W., 2015. Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am. J. Psychiatry* 172 (3), 284–293.
- Gillan, C.M., Pappmeyer, M., Morein-Zamir, S., Sahakian, B.J., Fineberg, N.A., Robbins, T.W., de Wit, S., 2011. Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* 168 (7), 718–726.
- Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *PANS* 100 (1), 253–258.
- de Haan, W., Pijnenburg, Y.A.L., Strijers, R.L.M., van der Made, Y., van der Flier, W.M., Scheltens, P., Stam, C.J., 2009. Functional neural network analysis in frontotemporal dementia and Alzheimer's disease using EEG and graph theory. *BMC Neurosci.* 10 (1), 101–112.
- Hu, R., Zhu, X., Zhu, Y., Gan, J., 2020. Robust SVM with adaptive graph learning. *World Wide Web* 23, 1945–1968.
- Huang, H., Liu, X., Jin, Y., Lee, S.-W., Wee, C.-Y., Shen, D., 2019. Enhancing the representation of functional connectivity networks by fusing multi-view information for autism spectrum disorder diagnosis. *Hum. Brain Mapp.* 40 (3), 833–854.
- Karmonik, C., Brandt, A., Elias, S., Townsend, J., Silverman, E., Shi, Z., Frazier, J.T., 2019. Similarity of individual functional brain connectivity patterns formed by music listening quantified with a data-driven approach. *Int. J. Comput. Assist. Radiol. Surg.* 1–11.
- Kong, D., An, B., Zhang, J., Zhu, H., 2020. L2RM: low-rank linear regression models for high-dimensional matrix responses. *J. Am. Stat. Assoc.* 115 (529), 403–424.
- Kong, D., Xue, K., Yao, F., Zhang, H.H., 2016. Partially functional linear regression in high dimensions. *Biometrika* 103 (1), 147–159.
- Liu, F., Guo, W., Fouché, J.-P., Wang, Y., Wang, W., Ding, J., Zeng, L., Qiu, C., Gong, Q., Zhang, W., et al., 2015. Multivariate classification of social anxiety disorder using whole brain functional connectivity. *Brain Struct. Funct.* 220 (1), 101–115.
- Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., Wang, J., 2018. Applications of deep learning to MRI images: asurvey. *Big Data Min. Anal.* 1 (1), 1–18.
- Ma, G., Lu, C.-T., He, L., Philip, S.Y., Ragin, A.B., 2017. Multi-view graph embedding with hub detection for brain network analysis. In: *ICDM*, pp. 967–972.
- Reyes, P., Ortega-Merchan, M.P., Rueda, A., Uriza, F., Santamaria-García, H., Rojas-Serrano, N., Rodriguez-Santos, J., Velasco-Leon, M.C., Rodriguez-Parra, J.D., Mora-Diaz, D.E., et al., 2018. Functional connectivity changes in behavioral, semantic, and nonfluent variants of frontotemporal dementia. *Behav. Neurol.* 1–11.
- Rubbert, C., Mathys, C., Jockwitz, C., Hartmann, C.J., Eickhoff, S.B., Hoffstaedter, F., Caspers, S., Eickhoff, C.R., Sigl, B., Teichert, N.A., et al., 2019. Machine-learning identifies Parkinson's disease patients based on resting-state between-network functional connectivity. *Br. J. Radiol.* 92 (1101), 20180886.
- Scheinost, D., Tokoglu, F., Hampson, M., Hoffman, R., Constable, R.T., 2019. Data-driven analysis of functional connectivity reveals a potential auditory verbal hallucination network. *Schizophr. Bull.* 45 (2), 415–424.
- Schott, J.M., Crutch, S.J., Frost, C., Warrington, E.K., Rossor, M.N., Fox, N.C., 2008. Neuropsychological correlates of whole brain atrophy in Alzheimer's disease. *Neuropsychologia* 46 (6), 1732–1737.
- Seth, A.K., Barrett, A.B., Barnett, L., 2015. Granger causality analysis in neuroscience and neuroimaging. *J. Neurosci.* 35 (8), 3293–3297.
- Shen, H.T., Zhu, X., Zhang, Z., Wang, S.-H., Chen, Y., Xu, X., Shao, J., 2021. Heterogeneous data fusion for predicting mild cognitive impairment conversion. *Inf. Fusion* 66, 54–63.
- Shen, H.T., Zhu, Y., Zheng, W., Zhu, X., 2020. Half-quadratic minimization for unsupervised feature selection on incomplete data. *IEEE Trans. Neural Netw. Learn. Syst.* doi:10.1109/TNNLS.2020.3009632.

- Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., Constable, R.T., 2017. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* 12 (3), 506–518.
- Shu, H., Nan, B., et al., 2019. Estimation of large covariance and precision matrices from temporally dependent observations. *Ann. Stat.* 47 (3), 1321–1350.
- Shu, H., Wang, X., Zhu, H., 2019. D-CCA: a decomposition-based canonical correlation analysis for high-dimensional datasets. *J. Am. Stat. Assoc.* 1–29.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* 58 (1), 267–288.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15 (1), 273–289.
- Wang, H., Nie, F., Huang, H., 2014. Robust distance metric learning via simultaneous l_1 -norm minimization and maximization. In: *ICML*, pp. 1836–1844.
- Wang, M., Huang, J., Liu, M., Zhang, D., 2019. Functional connectivity network analysis with discriminative hub detection for brain disease identification. In: *AAAI*, 33, pp. 1198–1205.
- Wee, C.-Y., Yap, P.-T., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D., 2012. Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. *PLoS One* 7 (5), e37828.
- 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 (3), 2045–2056.
- Weinberger, K.Q., Saul, L.K., 2009. Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* 10 (Feb), 207–244.
- Wu, F., Jing, X.-Y., You, X., Yue, D., Hu, R., Yang, J.-Y., 2016. Multi-view low-rank dictionary learning for image classification. *Pattern Recognit.* 50, 143–154.
- Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., Weinberger, K., 2019. Simplifying graph convolutional networks. In: *ICML*, 97, pp. 6861–6871.
- 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*, pp. 246–253.
- Yao, D., Sui, J., Yang, E., Yap, P.-T., Shen, D., Liu, M., 2020. Temporal-adaptive graph convolutional network for automated identification of major depressive disorder using resting-state fMRI. In: *International Workshop on Machine Learning in Medical Imaging*, pp. 1–10.
- Ye, J., Janardan, R., Li, Q., 2005. Two-dimensional linear discriminant analysis. In: *NIPS*, pp. 1569–1576.
- Zhang, H., Chen, X., Zhang, Y., Shen, D., 2017. Test-retest reliability of α high-order functional connectivity in young healthy adults. *Front. Neurosci.* 11, 439.
- Zhang, S., Dong, Q., Zhang, W., Huang, H., Zhu, D., Liu, T., 2019. Discovering hierarchical common brain networks via multimodal deep belief network. *Med. Image Anal.* 54, 238–252.
- Zhang, S., Li, X., Zong, M., Zhu, X., Cheng, D., 2017. Learning k for kNN classification. *ACM Trans. Intell. Syst. Technol.* 8 (3), 1–19.
- Zhang, S., Li, X., Zong, M., Zhu, X., Wang, R., 2018. Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (5), 1774–1785.
- Zhang, Y., Zhang, H., Chen, X., Liu, M., Zhu, X., Lee, S.-W., Shen, D., 2019. Strength and similarity guided group-level brain functional network construction for MCI diagnosis. *Pattern Recognit.* 88, 421–430.
- Zhang, Y., Zhang, H., Chen, X., Shen, D., 2017. Constructing multi-frequency high-order functional connectivity network for diagnosis of mild cognitive impairment. In: *IWCN*, pp. 9–16.
- Zhu, X., Gan, J., Lu, G., Li, J., Zhang, S., 2020. Spectral clustering via half-quadratic optimization. *World Wide Web* 23, 1969–1988.
- Zhu, X., Song, B., Shi, F., Chen, Y., Hu, R., Gan, J., Zhang, W., Li, M., Wang, L., Gao, Y., et al., 2021. Joint prediction and time estimation of COVID-19 developing severe symptoms using chest CT scan. *Med. Image Anal.* 67, 101824.
- Zhu, X., Yang, J., Zhang, C., Zhang, S., 2019. Efficient utilization of missing data in cost-sensitive learning. *IEEE Trans. Knowl. Data Eng.* doi:10.1109/TKDE.2019.2956530.
- Zhu, X., Zhang, S., Zhu, Y., Zhu, P., Gao, Y., 2020. Unsupervised spectral feature selection with dynamic hyper-graph learning. *IEEE Trans. Knowl. Data Eng.* doi:10.1109/TKDE.2020.3017250.
- Zou, H., Yang, J., 2020. Multiple functional connectivity networks fusion for schizophrenia diagnosis. *Med. Biol. Eng. Comput.*