Artificial Intelligence In Medicine 111 (2021) 102004

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Estimating sparse functional connectivity networks via hyperparameter-free learning model

Lei Sun [a,1], Yanfang Xue [a,1], Yining Zhang [a], Lishan Qiao [a], Limei Zhang [a,*], Mingxia Liu [b,*]

[a] School of Mathematics Science, Liaocheng University, Liaocheng 252000, China
[b] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

ARTICLE INFO

Keywords:
Functional connectivity network
Pearson's correlation
Sparse representation
Thresholding
Mild cognitive impairment
Autism spectrum disorder

ABSTRACT

Functional connectivity networks (FCNs) provide a potential way for understanding the brain organizational patterns and diagnosing neurological diseases. Currently, researchers have proposed many methods for FCN construction, among which the most classic example is Pearson's correlation (PC). Despite its simplicity and popularity, PC always results in dense FCNs, and thus a thresholding strategy is usually needed in practice to sparsify the estimated FCNs prior to the network analysis, which undoubtedly causes the problem of threshold parameter selection. As an alternative to PC, sparse representation (SR) can directly generate sparse FCNs due to the l_1 regularizer in the estimation model. However, similar to the thresholding scheme used in PC, it is also challenging to determine suitable values for the regularization parameter in SR. To circumvent the difficulty of parameter selection involved in these traditional methods, we propose a hyperparameter-free method for FCN construction based on the global representation among fMRI time courses. Interestingly, the proposed method can automatically generate sparse FCNs, without any thresholding or regularization parameters. To verify the effectiveness of the proposed method, we conduct experiments to identify subjects with mild cognitive impairment (MCI) and Autism spectrum disorder (ASD) from normal controls (NCs) based on the estimated FCNs. Experimental results on two benchmark databases demonstrate that the achieved classification performance of our proposed scheme is comparable to four conventional methods.

1. Introduction

Resting state functional magnetic resonance imaging (rs-fMRI), which measures the bold-oxygen-level-dependent (BOLD) signals of subjects without performing any explicit task, has been widely used in the field of neuroimaging analysis [1]. In practice, the BOLD fMRI is not quantitative (without unit), and thus it is meaningless to simply compare the signal intensity between different subjects [2]. In contrast, functional connectivity network (FCN), which tends to capture the dependency between BOLD signals of brain regions, provides an effective way of comparing different subjects, and has been used to discover potential neuroimaging biomarkers for diagnosing neurological disorders [3,4], such as Alzheimer's disease and its prodromal state (*i.e.*, mild cognitive impairment, MCI) [5,6], Autism spectrum disorder (ASD) [7,8], and Parkinson's disease [9], etc.

Mathematically, FCN can be expressed by a graph, where each node corresponds to a specific region-of-interest (ROI) in the brain and each edge delineates the dependency between the BOLD signals associated with a pair of ROIs [10,11]. In order to estimate the edges in FCN, researchers have proposed many different schemes, such as Pearson's correlation (PC) based method [12], sparse representation (SR) based method [13], dynamic time warping distance (DTW) [14] and dynamic causal model (DCM) [15]. In this paper, we mainly focus on methods based on the second-order statistics, including PC and SR, because a recent study [16] has empirically validated that they are more effective than many complex higher-order methods.

Despite its simplicity and popularity, PC always yields dense FCN in which all nodes are connected to each other. In fact, the brain is widely accepted to organize in a form of sparse network, meaning that the PC-based FCN may include a substantial percentage of spurious connections. To remove these uninformative/noisy connections, a thresholding operation is generally involved as a post-processing step to sparse the initially constructed FCN [17,18]. However, the thresholding strategy has two major disadvantages. *First*, it is often challenging to determine

* Corresponding authors.
E-mail addresses: zhanglimeilcu@163.com (L. Zhang), mxliu1226@gmail.com (M. Liu).
[1] L. Sun and Y. Xue contributed equally to this work..

https://doi.org/10.1016/j.artmed.2020.102004
Received 13 April 2020; Received in revised form 14 October 2020; Accepted 15 December 2020
Available online 18 December 2020

the optimal value for thresholding FCNs in a specific problem. *Second*, it usually leads to different connection densities across subjects, which in turn results in a statistical bias between patient and normal group. Although several schemes have been improved towards addressing the above problems, the thresholding selection in PC-based methods currently remains a conundrum [19].

As an alternative to PC, partial correction can explain more complex interactions among multiple ROIs. However, the estimation of partial correction has an effect of amplifying noise, and is usually ill-posed due to the singularity of the sample covariance matrix [20]. To overcome these problems, an l_1-norm regularizer is generally incorporated into the partial correction model, which results in the popular SR-based method for FCN construction. Unfortunately, similar to the thresholding scheme used in PC, it is also difficult for SR to determine the optimal value for the regularization parameter when estimating FCNs.

To circumvent the difficulty of parametric selection, in this paper, we propose a novel FCN estimation scheme based on a hyperparameter-free graph learning model, as illustrated in Fig. 1. Different from PC-based and SR-based approaches that need to select optimal thresholds or the regularization parameters for sparsifying FCNs, the proposed method does not contain any free parameter, but, interestingly, can automatically generate sparse FCNs. To validate the effectiveness of our proposed method, we first apply it to estimate FCNs, and then use the estimated FCNs for classification. Two classification tasks are considered: (1) MCI identification (*i.e.*, MCI vs. NC classification) on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, and (2) ASD identification (*i.e.*, ASD vs. NC classification) on the Autism Brain Imaging Data Exchange (ABIDE) dataset. Experimental results on these two benchmark databases suggest that the performance of our proposed method is comparable to competing methods for FCN construction. For the convenience of replicating our work, we share the pre-processed data and source code online[2] .

The rest of this paper is organized as follows. In Section 2, we first introduce the preprocessed data, and review the most relevant studies, and then propose the novel FCN estimation scheme, including its model and algorithm. In Section 3, we describe the experimental setting and report the experimental results. In Section 4, we investigate the effect of brain parcellation, data preprocessing step, feature selection and FCN modelling parameters on the final results, and discuss the selection of baseline methods and the discriminative features (potential biomarkers) for identifying brain disorders. We finally conclude this paper in Section 5.

2. Materials and methods

In this section, we describe the data preparation (including data acquisition and preprocessing), and FCN construction methods (including the baseline and the proposed methods).

2.1. Data preparation

Two benchmark databases are used in this study to verify the effectiveness of the proposed method, including the public ADNI and ABIDE databases with rs-fMRI data. For the ADNI database, 137 subjects (68 MCIs and 69 NCs) were selected and preprocessed as in a recent study [21]. In Table 1, we represent the demographic and clinical information of subjects used in this study. Note that, the information of these subjects shown in Table 1 conforms to the general inclusion/exclusion criteria for samples in the ADNI dataset, and the criteria available are briefly listed below: (1) NC subjects: Mini-Mental State Examination (MMSE) scores between 20 and 30 (inclusive), a Clinical Dementia Rating (CDR) of 0, nondepressed, non-MCI, and non-demented; (2) MCI subjects: MMSE scores between 24 and 30 (inclusive), impaired memory, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia.

For the preprocessing pipeline of each subject in ADNI dataset, the scanning time was 7 *min*, corresponding to 140 volumes. To remain the signal stabilization, the first three volumes of each subject were removed from the fMRI time course. Then, the remaining volumes were processed by a well-accepted pipeline based on the Data Processing Assistant for Resting-State fMRI (DPARSF) toolbox. More specifically, the first preprocessing step was head motion correction. The frame-wise displacement (FD) was calculated based on head motion parameters, and the subjects with more than 2.5 *min* of FD larger than 0.5 *mm* were excluded from the dataset. Then, nuisance regression was used to reduce the influence of the ventricle and white matter signals, as well as the high-order effect of head motion based on Friston 24-parameters model. After that, the corrected images were registered to the standard Montreal Neurological Institute (MNI) space, followed by spatially smoothing with the full-width-half-maximum of 4 *mm* and temporal band-pass filtering ($0.015 - -0.150\,Hz$). Note that, we did not perform scrubbing operation for the obtained fMRI data, since it may break the data autocorrection structure [22] and produce additional artifacts [23]. Finally, according to the automated anatomical labeling (AAL) atlas [24], the brain was partition into 116 ROIs, and the mean time series of each ROI were placed in order into a data matrix $X \in R^{137\times116}$. It is worth noting that we use AAL atlas to parcellate ROIs mainly due to its popularity and simplicity. In the discussion section, we will give more details of the atlas selection problem and its possible effects on the subsequent results.

For the ABIDE dataset, 184 subjects (including 79 ASDs and 105 NCs) from the largest site (*i.e.*, NYU) were used in our study. The demographic information of these subjects was shown in Table 1. All fMRI data were acquired based on a standard echo-planar imaging sequence on a clinical routine 3.0 Tesla Allegra scanner with the following imaging parameters: TR/TE is 2, 000/15 *ms* with 180 volumes, the number of slices is 33, and the slice thickness is 4.0 *mm*. All the involved fMRI data are provided by the Preprocessed Connectome Project initiative, and further preprocessed by DPARSF [25]. Specifically, the preprocessed pipeline can be mainly divided into four steps: (1) volume slices and head motion correction, (2) nuisance signals regression (ventricle, white matter signals and the high-order effect of head motion described by Friston 24-parameters model), (3) registration to MNI space, and (4) temporal filtering ($0.01 - -0.10\,Hz$). Subsequently, according to AAL atlas, the brain was partitioned into 116 ROIs, and the extracted mean time series from all these ROIs were put into a data matrix $X \in R^{175\times116}$.

2.2. Baseline methods for FCN construction

2.2.1. Pearson's correlation

PC is the simplest and most widely used method for estimating FCNs. Suppose that our brain has been divided into n ROIs based on a certain atlas. Denote $x_i \in R^m$ as the mean time series extracted from the ith ROI, where m is the number of the time points in each series. Let $W = (w_{ij}) \in R^{n\times n}$ be the adjacency matrix of the estimated FCN. Then, PC-based FCN can be formulated as

$$w_{ij} = \frac{(x_i - \bar{x}_i)^T (x_i - \bar{x}_i)}{\sqrt{(x_i - \bar{x}_i)^T (x_i - \bar{x}_i)}\sqrt{(x_j - \bar{x}_j)^T (x_j - \bar{x}_j)}}. \quad (1)$$

Without loss of generality, we redefine $x_i = (x_i - \bar{x}_i)/\sqrt{(x_i - \bar{x}_i)^T (x_i - \bar{x}_i)}$. That is, the new x_i has been centralized by $x_i - \bar{x}_i$ and normalized by $\sqrt{(x_i - \bar{x}_i)^T (x_i - \bar{x}_i)}$. As a result, Eq. (1) can be simplified into $w_{ij} = x_i^T x_j$, which corresponds to the optimal solution of the following model:

[2] https:github.com/Leisun981/LeisunFCN

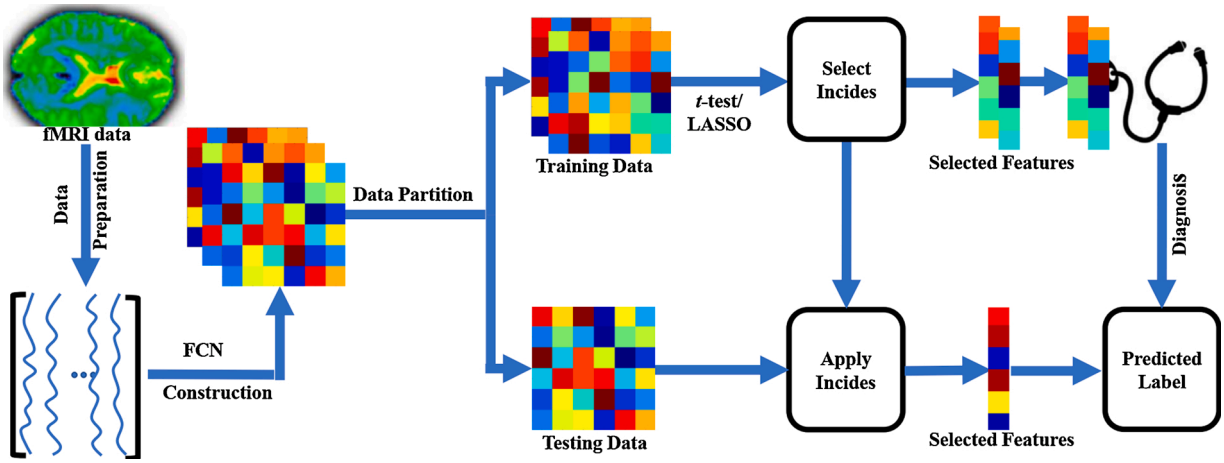


Fig. 1. The main pipeline of disorder identification used in this study, which contains three major modules: (1) data preparation, (2) FCN construction, and (3) classification with feature selection, where (1) and (2) will be discussed in Section 2, and (3) will be discussed in Section 3.

Table 1
Demographic and clinical information of subjects in the ADNI and ABIDE datasets. Values are reported as mean±standardard deviation. M/F: Male/Female; MMSE: Mini-Mental Examination; GCDR: Global Clinical Dementia Rating; FIQ: Full-Scale Intelligence Quotient; VIQ: Verbal Intelligence Quotient; PIQ: Performance Intelligence Quotient.

Datasets	Class	Gender (M/F)	Age (Years)	MMSE	GCDR	FIQ	VIQ	PIQ
ADNI	MCI	39/29	76.50 ± 13.50	28.01 ± 1.64	0.48 ± 0.02	–	–	–
	NC	17/52	71.50 ± 14.50	29.25 ± 1.07	0	–	–	–
ABIDE	ASD	68/11	18.58 ± 11.45	–	–	107.92 ± 3.15	105.81 ± 1.23	108.81 ± 2.10
	NC	79/26	19.13 ± 11.85	–	–	113.15 ± 2.45	113.13 ± 1.15	115.07 ± 2.08

$$min_{w_{ij}} \sum_{ij}^{n} \| x_i - w_{ij}x_j \|^2. \tag{2}$$

According to a previous work [26], Eq. (2) can be further transformed mathematically into the following matrix form:

$$min_W \ \| W - X^T X \|_F^2, \tag{3}$$

where $X = [x_1, x_2, \cdots, x_n] \in R^{m \times n}$ is the BOLD data matrix, and $\| \cdot \|_F$ represents the Frobenius-norm of a matrix [27].

In general, the constructed FCN based on PC is a dense graph (*i.e.*, all vertices are fully-connected by edges), which may contain noisy or uninformative information. To remove those noisy connections, the thresholding strategy is generally used to sparsify FCNs, requiring additional thresholding parameters to be tuned.

2.2.2. Sparse representation

SR is one of the commonly-used methods for calculating the partial correction. As an alternative to PC, SR aims to estimate more reliable connections between two ROIs by regressing out the confounding effect from other ROIs. The mathematical model of SR is expressed as follows:

$$\begin{aligned} min_{w_{ij}} \quad & \sum_{i=1}^{n} (\| x_i - \sum_{j \neq i} w_{ij}x_j \|^2 + \lambda \sum_{j \neq i} |w_{ij}|_1, \\ s.t. \quad & w_{ii} = 0, \forall i = 1, \cdots, n \end{aligned} \tag{4}$$

which can be further rewritten by the following matrix form:

$$\begin{aligned} min_W \quad & \| X - XW \|_F^2 + \lambda \| W \|_1, \\ s.t. \quad & w_{ii} = 0, \forall i = 1, \cdots, n \end{aligned} \tag{5}$$

where $\| X - XW \|_F^2$ is a data fitting term for capturing the partial correction information, $\| W \|_1$ is an l_1-regularized term for obtaining sparse solutions of W, and λ is a regularization parameter for controlling the balance between these two terms. Note that the constraint $w_{ii} = 0$ is used here to avoid the trivial solution (*i.e.*, $W = I$, the identity matrix) by implicitly removing x_i from X.

It can be seen from Eq. (5) that, to generate sparse connections in an FCN, we need to determine the regularization parameter λ, which is usually a challenging task in practice. To address the limitations of PC- and SR-based methods, in this work, we propose a hyperparameter-free method for constructing sparse FCNs. More details can be found in the following section.

2.2.3. FCN construction based on proposed method

As mentioned above, the selection of an appropriate threshold or regularization parameter for sparsifying FCN is currently still an open problem. To estimate sparse FCNs and simultaneously remove the difficulty of parameter selection, we propose a hyperparameter-free FCN learning model as

$$\begin{aligned} min_{w_{ij}} \quad & \sum_{i=1}^{n} \| \sum_{j \neq i} w_{ij}x_i - \sum_{j \neq i} w_{ij}x_j \|^2, \\ s.t. \quad & \sum_{j \neq i} w_{ij} \geq 1, \forall i = 1, \cdots, n \\ & w_{ij} = w_{ji} \geq 0, \forall i, j = 1, \cdots, n \end{aligned} \tag{6}$$

In Eq. (6), we use a similar data-fitting term to that in SR (*i.e.*, Eq. (4) or Eq. (5)) for capturing the partial correlation structure in data, but remove the l_1 regularizer from SR for avoiding the difficulty of selecting the optimal regularization parameter. In particular, to prevent trivial solution (*i.e.*, $w_{ij} = 0, \forall i,j = 1, \cdots, n$), we include two constraints in Eq. (6) to encourage that the degree of each node is no less than 1 and the edge weight in the estimated FCN is nonnegative, respectively. Note that, the assumption that edge weight is nonnegative is supported by the structural equilibrium theory [28,29], in which the estimated network is

structurally balanced if and only if the sign product of all the edges is positive. Also, this assumption can simplify the subsequent FCN analysis, while many functional connectivity measures, such as mutual information [30], are also nonnegative. In addition, the FCN is constrained to be symmetric, which avoids the post-processing step (*i.e.*, symmetrization) involved in traditional methods such as SR.

Due to the symmetry of the estimated FCNs, we only consider its upper triangular elements, and concatenate these elements in a column vector $w = [w_{12}, w_{13}, \cdots, w_{1n}, w_{23}, w_{24}, \cdots, w_{n-1,n}]^T \in R^{n(n-1)}$. As a result, Eq. (6) can be equivalently transformed into the following quadratic programming (QP) problem:

$$\begin{array}{ll} min_w & w^T \mathrm{S} w, \\ s.t. & \mathrm{A} w \geq b \end{array} \tag{7}$$

where S, A and b are only dependent of the fMRI data matrix X. Their definitions and the related mathematical formulations are provided in Appendix. Finally, we solve the QP problem in Eq. (7) by CVX toolbox [31], due to its simplicity and popularity.

From Eqs. (6)–(7), we can see that the proposed model requires no parameters for constructing FCNs, which is particularly useful in real-world applications. Interestingly, even no free parameters are involved in the proposed method, we empirically find that it can generate sparse FCNs, as shown in the following experiments.

2.3. Relationship between regularization term and constraints

In particular, l_1-norm used in SR is the sum of the absolute values of all edge weight w_{ij} that can be rewritten $\sum_{i,j=1}^{n} w_{ij}$, if the non-negative constraint requires that $w_{ij} \geq 0$. However, we cannot prove the relationship between the constraints and sparsity rigorously in mathematics, but can only provide empirical results of sparsity from a perspective of experiments (Please see Section 3 for details).

3. Experiments and results

3.1. Experimental setting

After obtaining the pre-processed fMRI data, we estimate FCNs based on different methods. In the experiments, we first compare our method with two typical methods for FCN construction, *i.e.*, PC and SR. Note that the proposed method constrains edge weights of the estimated FCNs to be nonnegative. Therefore, besides the comparison with the original PC and SR methods, we also compare our method with two additional methods, PC+ and SR+, which only keep the positive edges in PC and SR, respectively. Since the threshold or regularization parameter is included in the baseline methods, we select the threshold corresponding to the different sparsity in the set of [0, 10%, ···, 90%, 99%] for PC- and PC+-based methods, where the percentage indicates the proportion of the edges that are removed, and search the value of regularization parameters in the range of $[2^{-5}, 2^{-4}, \cdots, 2^4, 2^5]$ for SR- and SR+-based methods. In contrast, the proposed FCN estimation method does not contain any parameters.

Once we have obtained FCNs of all subjects, the subsequent task is to identify the subjects with MCI (or ASD) from NCs. Now, the problem turns to determine which features and classifier should be chosen for MCI (or ASD) identification. Since the adjacency matrix of the estimated FCN is symmetric, we only consider its upper triangular elements as features. In our experiment, each FCN has 116 nodes, and thus can produce 6, 670 features (corresponding to 6, 670 functional connections between 116 ROIs). Compared to the sample size (less than two hundred), the feature dimension is very high, which not only brings expensive computation, but also may affect the classification accuracy, due to the so-called curse of dimensionality [32]. To alleviate this problem, researchers have proposed numbers of approaches for feature selection, such as *t*-test, least absolute shrinkage and selection operator (LASSO) [33], genetic algorithm (GA) [34] and so on. In our paper, we only adopt the simple feature selection method (*t*-test with four accepted *p*-values and LASSO) and linear SVM classifier [35] with default parameter $C = 1$ in our experiment, since it, due to the large marginal theory [36], is relatively immune to high dimensions.

Further, in order to make the best of limited samples (*i.e.*, the estimated FCNs of different subjects) in the training stage, we choose leave-one-out cross validation (LOOCV) and 5-fold cross validation (5-fold CV) to evaluate the performance of involved methods [37]. Of note, the proposed FCN estimation method does not contain any hyper-parameters, and thus will result in a unique classification accuracy. In contrast, the threshold or regularization parameter used in the baseline methods may affect the final accuracy. Therefore, an inner LOOCV and 5-fold CV are utilized in the training stage to select the optimal parametric values for these methods.

To evaluate the classification performance of different methods, seven performance metrics, including accuracy (ACC), sensitivity (SEN), specificity (SPE), balanced accuracy (BAC), positive predictive value (PPV), negative predictive value (NPV) and the area under the receiver operating characteristic (ROC) curve (AUC) [38], were utilized in this paper. These metrics are defined as follows: BAC=(SEN+SPE)/2, PPV=TP/(TP+FP) and NPV=TN/(TN+FN), where TP, TN, FP and FN indicate true positive, true negative, false positive and false negative respectively.

3.2. Results

3.2.1. Results of FCN estimation

In this section, we take one of subject from ADNI dataset as an example to visualize the FCNs estimated by five different methods (*i.e.*, PC, PC+, SR, SR+ and Ours). Note that the proposed method constrains edge weights of the estimated FCN to be nonnegative. Therefore, besides the original PC and SR, we also include PC+ and SR+ in our experiments for a more comprehensive comparison in the subsequent MCI (or ASD) identification task, where PC+ and SR+ only keep the positive edges and turn the negative edges into zero in PC and SR, respectively. The results are shown in Fig. 2, where the thresholds or regularization parameters involved in PC- and SR-based methods are determined according to their best classification accuracy. In particular, the thresholds used in PC and PC+ are 50% and 40%, while the values of the regularization parameter λ used in SR and SR+ are 2^2 and 2^5, respectively.

Based on Fig. 2, we have several observations as follows. *First*, the topological structure of PC- and PC+-based FCNs is significantly different from that of SR- and SR+-based FCNs, mainly due to the fact that they capture full and partial corrections, respectively, using different data-fitting terms. *Second*, compared with SR, PC generates denser FCN, since it is sensitive to both direct and indirect relationships between ROIs. Although PC+ improves the sparsity of the estimated FCN by removing the negative edges, it is still denser than SR-based FCN, because the original PC has already contained false connections (*e.g.*, those connections from indirect relationships between ROIs). *Third*, FCNs estimated by SR and SR+ are sparse, not only due to the introduction of the l_1 regularization term, but also due to their data-fitting term that can remove the indirect relationships by regressing out the confounding effect from other ROIs. *Finally*, compared with the baseline methods, the FCN estimated by the proposed method looks very clean, and, more interestingly, its topological structure is similar to that of SR- and SR+-based FCNs, even though no any hyperparameters are involved.

To further discuss the sparsity structure of the FCN estimated by the proposed method, we calculate the degree distribution and the number of edges of the FCN shown in Fig. 2 (*i.e.*, Fig. 2-Ours), and the result shown in Fig. 3. Based on Fig. 3, we can observe that the maximum and minimum degree of node (*i.e.*, ROI) are 19 and 4 respectively, which is significantly less than the value of node degree in the case of full-connected FCN (*i.e.*, 115), and all the total number of edges of this

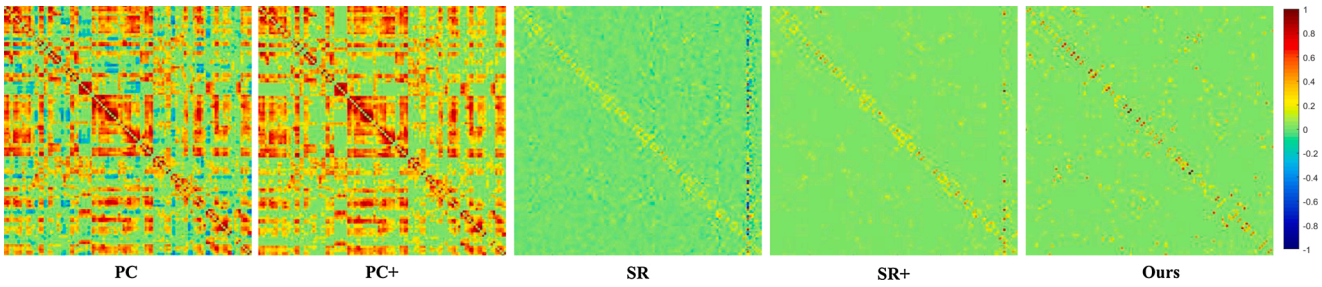


Fig. 2. The adjacency matrices of FCNs constructed by five different methods (*i.e.*, PC, PC+, SR, SR+ and Ours). Note that the elements in the adjacency matrices have been normalized into interval of [−1,1] for the convenience of comparison. PC+ and SR+ only keep the positive edges in PC and SR by turning their negative edges into zero, respectively.

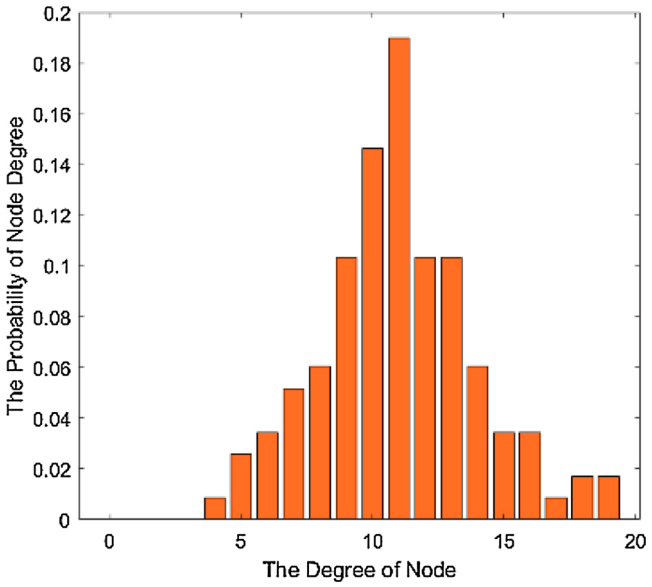


Fig. 3. The degree distribution of the FCNs estimated by the proposed method. Of note, the maximum and minimum degree of node (*i.e.*, ROI) are 19 and 4 respectively, which is significantly less than the value of node degree in the case of full-connected FCN (*i.e.*, 115), and all the total number of edges of this network is 657 (*i.e.*, the density of the network less than 10%).

network is 657 (*i.e.*, the density of the network less than 10%), meaning that the FCNs estimated by the proposed method is sparse. This finding is consistent with the visualized results in Fig. 2.

3.2.2. Results of MCI identification

In this section, we report the MCI classification performance of different methods under seven evaluation metrics based on LOOCV and 5-fold CV (repeated 1000 times) in Fig. 4 respectively.

From Fig. 4, we have the following observations for MCI classification task. *First*, on the basis of the average ACC values, we note that the methods that only use the positive edge weights (*e.g.*, PC+ and SR+) can generally result in a higher classification accuracy than their original counterparts in most cases. For example, for LOOCV, PC+ and SR+ achieve the best classification accuracy of 78.10% and 87.59% with *p*-values of 0.005 and 0.01 respectively; for 5-fold CV, PC+ and SR+ also obtain the highest accuracy of 68.49% and 78.00% with *p*-values of 0.005 and 0.01, respectively. This is a counterintuitive finding, which drives us to think that the positive dependency between ROIs may contain more discriminative information for MCI identification. On the other hand, the removal of negative edge weights from the estimated FCN can reduce the number of features, thus alleviating the curse of dimensionality to some extent. *Second*, even without free parameters for FCN estimation, the proposed method is comparable to four competing methods in terms of seven evolution metrics based on LOOCV and 5-fold CV. Especially with the small *p*-value of 0.001, for LOOCV, our method achieves the average results of ACC=81.75%, SEN=86.76%, BAC=81.79%, NPV= 85.48%, and AUC=89.41%; for 5-fold CV, our method obtains the average results of ACC=76.00%, SEN=76.10%, SPE=75.90%, BAC=75.32%, NPV=74.79%, and AUC=88.26%. They are significantly better than four competing methods.

3.2.3. Results of ASD identification

For ASD identification, we report the results of ASD vs. NC classification achieved by both methods based on LOOCV and 5-fold CV respectively in Fig. 5. It can be seen from Fig. 5 that the proposed method outperforms the competing methods in terms of the average ACC value under the case that *p*-values are 0.001 and 0.005 based on LOOCV. In particular, when *p*-value is 0.005, the proposed method achieves the ACC of 71.74%, SPE of 64.71%, BAC of 70.76%, PPV of 73.33% and AUC of 79.89%, most of which are higher than the baseline methods. For the remaining *p*-values (*i.e.*, 0.01, 0.05), the proposed method (even with unsatisfactory performance) is also comparable to the competing methods in the sense of classification accuracy.

4. Discussion

In this section, we investigate the effect of brain parcellation, data preprocessing steps, feature selection and network modelling parameters on the classification results. Then, we discuss the selection of baseline methods, and further show the most discriminative features selected by our method for exploring their relationship with brain disorders.

4.1. Brain parcellation

To our best knowledge, many brain parcellation schemes have been developed in the past decades and they can be roughly divided into two categories, *i.e.*, atlas-based and data-driven approaches. In atlas-based methods, the voxels within the same ROI are supposed to share the similar structure or function. Typical atlases include (but not limited to) AAL (a structural atlas with 116 ROIs by anatomy of a reference subject) [24], Harvard Oxford (HO, a probabilistic atlas of anatomical structures with 118 ROIs) [39] and functional connectivity multivariate pattern analysis (fcMVPA, a functional atlas with 160 ROIs) [40]. In the second category, data-driven schemes can directly work on the obtained fMRI data, mainly including clustering method [41] and group independent component analysis methods (GICA) [42].

To investigate the influence of brain parcellation on the final classification results, we use the proposed method to estimate FCNs based on three different schemes (*i.e.*, AAL, fcMVPA and clustering) for ASD classification task, and then show the experimental results in Table 2. It can be observed that: when using LOOCV under the *p*-value of 0.005, our method based on AAL atlas can obtain the best classification accuracy, while the *p*-value of 0.01, the performance of our method with fcMVPA

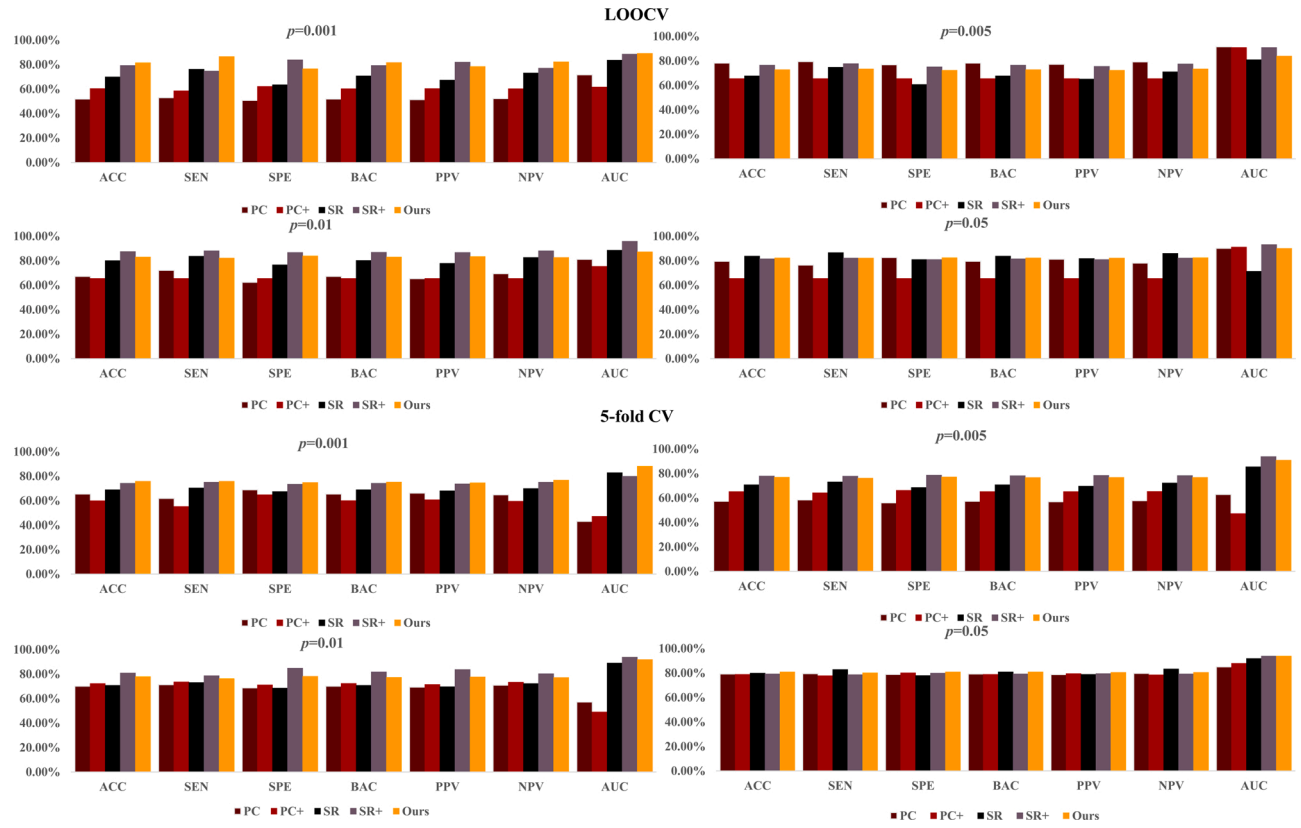


Fig. 4. The MCI classification results of five methods (*i.e.*, PC, PC+, SR, SR+ and Ours) based on seven performance metrics (*i.e.*, ACC, SEN, SPE, BAC, PPV, NPV and AUC) using LOOCV and 5-fold CV. In particular, each subplot represents the classification results under seven performance metrics based on different *p*-values involved in *t*-test, and the horizontal axis of each subplot shows the average classification results.

is optimal; for the remaining *p*-values using LOOCV and 5-fold CV, our method utilizing clustering-based for brain parcellation achieves the highest classification performance. In summary, the performance of different approaches (or atlases) may vary under different conditions (*e.g.*, *p*-value or verification mode). There is no atlas that is always optimal for ROI definition and classification task.

4.2. Influence of data preprocessing step

As mentioned above, many steps are involved in the data preprocessing pipeline, such as volume slices correction, nuisance regression, registration, temporal filtering, and so on. However, in this section, we only select the nuisance regression (of head motion) as an example to investigate the influence of data preprocessing step on the final classification results, since head motion has significant, systematic influence on FCN measures, even as small as 0.1 *mm* [43,44]. In particular, we conduct ASD vs. NC classification experiment based on the ABIDE dataset in two cases. In the first case, we process the data without the regression of head motion, and then use the FCNs estimated by the proposed method to conduct the classification task based on LOOCV and 5-fold CV. By contrast, in the second case, we process the data by including the nuisance regression step, and then perform the subsequent tasks. The experimental results in Table 3 show that the proposed method based on the preprocessed data can obtain a better classification results in most cases, meaning that the data preprocessing steps (at least regression of head motion) have a significant effect on the final identification performance

4.3. Selection of baseline methods

As mentioned above, researchers have proposed numbers of approaches for FCN construction, including PC, SR, DTW and DCM. In our paper, we only select PC and SR-based methods as the baseline for comparison, since one of our main focuses is the simplicity in estimating sparse FCNs by circumventing the challenge of parameter selection. Recent reviews [45,46] for exploring network modelling methods also employed these two methods as baseline methods. In contrast, DTW and DCM are not utilized as baseline methods, because they all assume that functional connectivity exhibits dynamic changes within time scales, while the proposed method in this paper works on the assumption of temporal stationarity.

4.4. Parameter analysis

To our best knowledge, feature selection as a dimensionality reduction technique helps eliminate redundant/noisy features, and is thus widely used in fMRI data [47]. At present, there are amounts of feature selection methods have been developed, including *t*-test (or equivalently, Fisher score [48]), LASSO method [33] and GA [34]. To investigate the influence of different feature selection approaches on the final classification results, we take the simple *t*-test based on different *p*-values and LASSO as examples to perform a control analysis [49–53]. Further, we report the optimal results based on*t*-test under four different *p*-values (*i.e.*, 0.001, 0.005, 0.01, 0.05) and LASSO under different parameters (*i.e.*, 2^{-5}, 2^{-4},⋯, 2^{4}, 2^{5}) in Table 4.

Based on the results in Table 4, we note that several methods using LASSO tend to perform well under seven performance indices in some cases, indicating that LASSO method takes into account the importance of combined features, and thus generally obtain better performance compared with *t*-test. In other words, an appropriate feature selection approach can contribute to achieve better classification results.

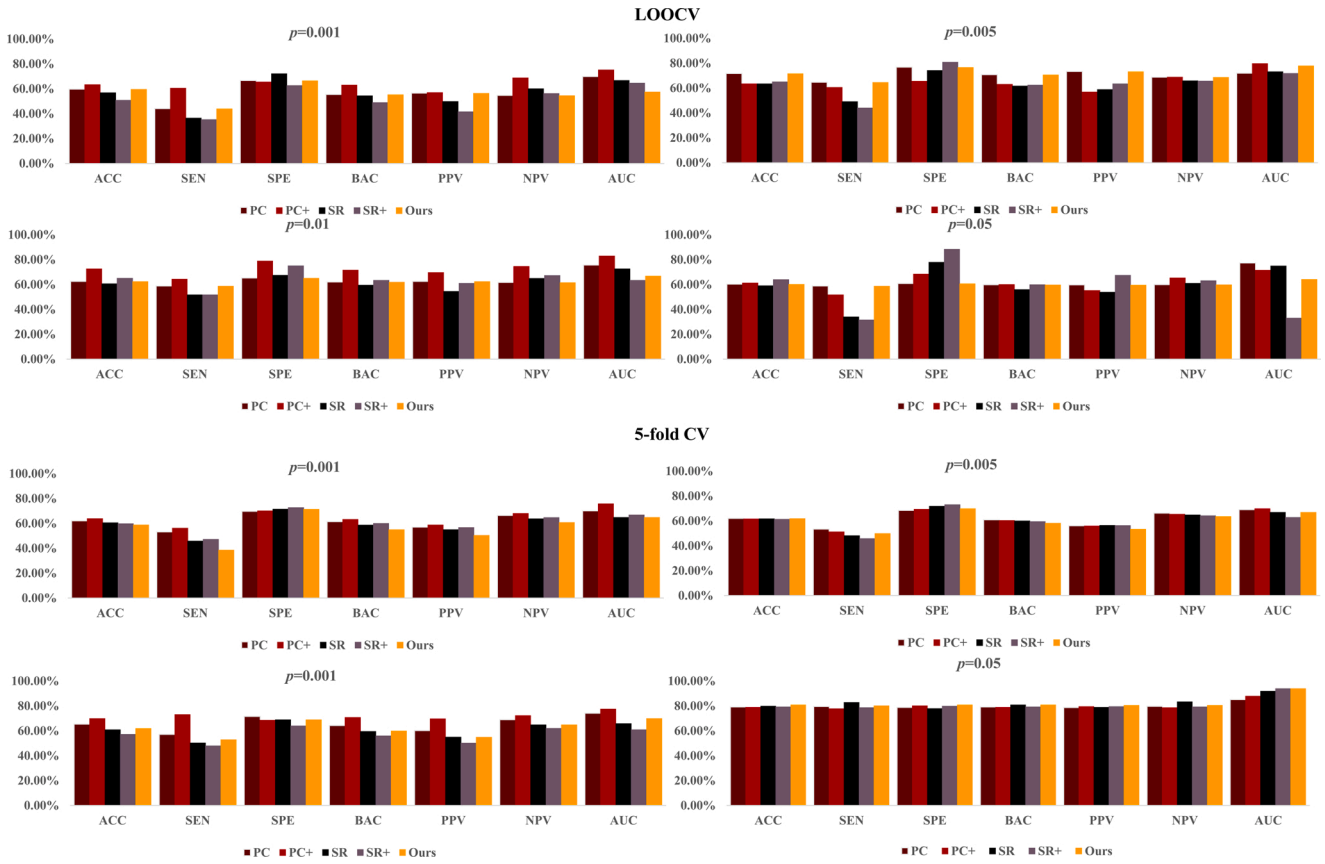


Fig. 5. The ASD classification results of five methods (*i.e.*, PC, PC+, SR, SR+ and Ours) based on seven performance metrics (*i.e.*, ACC, SEN, SPE, BAC, PPV, NPV and AUC) using LOOCV and 5-fold CV. In particular, each subplot represents the classification results under seven performance metrics based on different *p*-values involved in *t*-test, and the horizontal axis of each subplot shows the average classification results.

Table 2
The classification results based on the proposed method using different atlases for brain parcellation. Specifically, the first, second and third row of each *p*-value represents the results based on the AAL, fcMVPA [40] and clustering [41] for defining ROIs. CV: cross validation.

CV	*p*-value	Atlas	ACC	SEN	SPE	BAC	PPV	NPV	ROC
LOOCV	*p*=0.001	AAL	59.78%	44.12%	66.67%	55.39%	56.60%	54.76%	57.69%
		fcMVPA	58.70%	41.77%	71.43%	56.60%	52.38%	61.98%	48.51%
		Clustering	**68.48%**	**58.23%**	**76.19%**	**67.21%**	**64.79%**	**70.80%**	**74.83%**
	p=0.005	AAL	**71.74%**	64.71%	**76.81%**	**70.76%**	**73.33%**	68.83%	**78.06%**
		fcMVPA	64.67%	46.84%	78.10%	62.47%	61.67%	66.13%	69.73%
		Clustering	70.11%	69.62%	70.48%	70.05%	63.95%	75.51%	75.53%
	p=0.01	AAL	62.50%	58.82%	65.22%	62.02%	62.50%	61.64%	67.00%
		fcMVPA	**66.85%**	**63.29%**	**69.52%**	**66.41%**	**60.98%**	**71.57%**	**70.40%**
		Clustering	59.24%	49.37%	66.67%	58.02%	52.07%	63.64%	61.77%
	p=0.05	AAL	60.33%	58.82%	60.87%	59.85%	59.70%	60.06%	64.21%
		fcMVPA	54.35%	39.24%	65.71%	52.48%	46.27%	58.97%	50.45%
		Clustering	**71.74%**	**69.62%**	**73.33%**	**71.48%**	**66.27%**	**76.24%**	**79.16%**
5-fold CV	*p*=0.001	AAL	59.00%	38.75%	**71.56%**	55.16%	50.61%	60.87%	65.00%
		fcMVPA	59.30%	51.20%	65.83%	58.52%	59.17%	64.30%	68.54%
		Clustering	**59.31%**	**52.82%**	65.80%	58.52%	51.57%	64.30%	**69.46%**
	p=0.005	AAL	62.10%	54.10%	70.10%	58.33%	53.60%	63.77%	67.00%
		fcMVPA	58.44%	51.73%	64.07%	57.90%	52.13%	63.94%	63.64%
		Clustering	**65.65%**	**58.46%**	**71.47%**	**64.97%**	**60.85%**	**69.66%**	**76.30%**
	p=0.01	AAL	62.11%	53.23%	69.47%	60.13%	55.24%	65.00%	70.19%
		fcMVPA	58.88%	52.71%	64.00%	58.35%	52.57%	64.39%	65.14%
		Clustering	**66.86%**	**59.45%**	**72.81%**	**66.13%**	**62.35%**	**70.58%**	**77.65%**
	p=0.05	AAL	57.60%	**57.44%**	**57.72%**	57.58%	50.59%	64.36%	74.18%
		fcMVPA	59.77%	51.36%	66.81%	59.08%	53.85%	64.69%	67.39%
		Clustering	**66.74%**	56.54%	**74.95%**	**65.75%**	**63.10%**	**69.70%**	**77.43%**

Table 3
The classification results of the proposed method based on different *p*-values using LOOCV and 5-fold CV. In particular, each *p*-value corresponds to two different classification results, where the NE label 0 and 1 indicates the results based on the preprocessed data without NE (NE=0) and with NE (NE=1), respectively. NE: Nuisance Regression.

CV	*p*-value	NE	ACC	SEN	SPE	BAC	PPV	NPV	ROC
LOOCV	p=0.001	0	60.87%	31.65%	82.86%	57.25%	58.14%	61.70%	46.67%
		1	59.78%	44.30%	71.43%	57.87%	53.85%	63.03%	57.69%
	p=0.005	0	51.09%	37.97%	60.95%	49.46%	42.25%	56.64%	48.43%
		1	**71.74%**	**65.82%**	**76.19%**	**71.01%**	**67.53%**	**74.77%**	**78.06%**
	p=0.01	0	61.41%	48.10%	71.43%	59.76%	55.88%	64.66%	63.61%
		1	**62.50%**	**59.49%**	**64.76%**	**62.13%**	**55.95%**	**68.00%**	**66.99%**
	p=0.05	0	54.89%	43.04%	63.81%	53.42%	47.22%	59.82%	53.69%
		1	**60.33%**	**59.49%**	**60.95%**	**60.22%**	**53.41%**	**66.67%**	**64.21%**
5-fold CV	p=0.001	0	56.28%	27.48%	77.96%	52.72%	48.64%	58.82%	51.04%
		1	**57.61%**	**38.61%**	**71.91%**	**55.26%**	**50.87%**	**60.91%**	**62.45%**
	p=0.005	0	53.95%	40.41%	64.15%	52.28%	45.93%	58.86%	55.46%
		1	**59.70%**	**48.40%**	**68.19%**	**58.30%**	**53.43%**	**63.75%**	**67.26%**
	p=0.01	0	56.03%	46.53%	63.17%	54.85%	48.76%	61.12%	62.00%
		1	**59.80%**	**51.16%**	**66.30%**	**58.73%**	**53.38%**	**64.37%**	**68.80%**
	p=0.05	0	57.96%	50.06%	63.90%	56.98%	51.12%	62.99%	66.33%
		1	57.92%	58.03%	57.94%	57.94%	50.91%	64.73%	67.56%

Table 4
The classification results under seven performance indices of five different methods. Specifically, the first and second row of each method represents the optimal results on *t*-test under four different *p*-values (0.001, 0.005, 0.01, 0.05) and LASSO under different parameters (2^{-5}, 2^{-4}, $\cdots$, 2^{4}, 2^{5}). CV: cross validation.

Dataset	CV	Method	Feature selection	ACC	SEN	SPE	BAC	PPV	NPV	ROC
ADNI	LOOCV	PC	*t*-test	78.10%	80.88%	75.36%	78.12%	76.39%	80.00%	90.15%
			LASSO	**83.21%**	**85.29%**	**81.16%**	**83.23%**	**81.69%**	**84.85%**	89.41%
		PC+	*t*-test	83.94%	88.24%	79.71%	83.97%	81.08%	87.30%	91.37%
			LASSO	**84.67%**	82.83%	**85.51%**	**84.67%**	**85.07%**	84.29%	90.37%
		SR	*t*-test	81.02%	85.29%	76.81%	81.05%	78.38%	84.13%	91.47%
			LASSO	**86.13%**	85.29%	**86.96%**	**86.13%**	**86.57%**	**85.71%**	**93.22%**
		SR+	*t*-test	86.86%	88.24%	85.51%	86.87%	85.71%	88.06%	93.35%
			LASSO	**90.51%**	**91.18%**	**89.86%**	**90.52%**	**89.86%**	**91.18%**	**95.65%**
		Ours	*t*-test	83.21%	82.35%	84.06%	83.21%	83.58%	82.86%	88.94%
			LASSO	**87.14%**	**85.31%**	**88.13%**	**86.72%**	**86.49%**	**87.53%**	**92.77%**
	5-fold CV	PC	*t*-test	79.09%	79.45%	78.75%	79.10%	78.70%	79.59%	89.26%
			LASSO	76.88%	79.08%	77.03%	78.06%	76.37%	78.43%	86.54%
		PC+	*t*-test	79.06%	77.92%	80.20%	79.06%	79.62%	78.71%	88.10%
			LASSO	77.62%	**78.62%**	79.34%	78.98%	76.77%	78.61%	87.34%
		SR	*t*-test	80.55%	83.00%	78.10%	81.00%	79.00%	83.44%	92.20%
			LASSO	**82.96%**	81.34%	**81.76%**	**81.55%**	**80.37%**	**84.78%**	**92.56%**
		SR+	*t*-test	79.37%	78.74%	80.00%	79.37%	79.64%	79.35%	94.00%
			LASSO	**86.35%**	**86.29%**	**86.84%**	**86.56%**	**86.88%**	**86.23%**	**95.20%**
		Ours	*t*-test	81.10%	80.20%	82.00%	81.05%	80.64%	80.64%	94.10%
			LASSO	**84.24%**	**87.99%**	80.60%	**84.29%**	**82.53%**	**87.19%**	90.13%
ABIDE	LOOCV	PC	*t*-test	70.11%	59.49%	78.10%	68.79%	67.14%	71.93%	77.38%
			LASSO	68.48%	53.16%	**80.00%**	66.58%	66.67%	69.42%	77.87%
		PC+	*t*-test	73.37%	69.92%	76.19%	72.91%	68.75%	76.92%	83.19%
			LASSO	65.69%	61.76%	69.57%	65.66%	66.67%	64.86%	73.80%
		SR	*t*-test	68.48%	62.03%	73.33%	67.68%	63.64%	71.96%	75.07%
			LASSO	64.96%	60.29%	69.57%	64.93%	**66.13%**	64.00%	71.73%
		SR+	*t*-test	69.57%	59.49%	77.14%	68.32%	66.20%	71.68%	71.96%
			LASSO	54.01%	33.82%	73.91%	54.01%	56.10%	53.14%	41.42%
		Ours	*t*-test	71.74%	65.82%	76.19%	71.01%	67.53%	74.77%	78.06%
			LASSO	67.15%	64.71%	69.57%	67.14%	**67.69%**	66.67%	68.44%
	5-fold CV	PC	*t*-test	65.94%	57.80%	72.07%	64.94%	60.96%	69.44%	74.09%
			LASSO	61.64%	52.78%	69.38%	61.08%	57.78%	65.66%	62.71%
		PC+	*t*-test	70.93%	73.23%	68.63%	70.93%	69.81%	73.24%	77.60%
			LASSO	67.14%	60.00%	74.24%	67.12%	63.69%	71.34%	74.74%
		SR	*t*-test	65.05%	28.77%	92.34%	60.56%	73.96%	63.28%	40.37%
			LASSO	62.89%	49.99%	73.33%	61.66%	58.07%	66.12%	65.08%
		SR+	*t*-test	64.07%	27.53%	91.57%	59.55%	71.05%	62.69%	39.94%
			LASSO	59.21%	46.82%	68.66%	57.74%	55.89%	62.92%	61.40%
		Ours	*t*-test	62.00%	53.00%	69.00%	60.20%	55.00%	65.10%	70.00%
			LASSO	**65.60%**	**59.79%**	**70.78%**	**65.29%**	**60.64%**	**70.07%**	**71.95%**

4.5. Sensitivity to network modelling parameters

In practice, the final classification accuracies are not only affected by *p*-values used in the stage of feature selection, but also by different network modelling parameters (*e.g.*, thresholds and regularization parameters). In order to investigate their influence on the classification accuracy, we conduct MCI (and ASD) classification experiments, and report their results under different values of thresholds or regularization parameters in Fig. 6.

Based on Fig. 6, we discuss the effect of the FCN modelling parameters on the MCI (and ASD) classification task. *First*, the accuracies of the baseline methods for both MCI and and ASD classification fluctuate heavily with the change of the parametric values, indicating that they are sensitive to the FCN modelling parameters. In contrast, the proposed method achieves a unique solution, since no free parameter is included in the FCN estimation model. *Second*, the baseline methods can generally obtain high accuracy under some specific thresholds or regularization parameters, but the high accuracy cannot be reached in their final results shown in Fig. 4 (or Fig. 5) by the inner loop of cross validation. For example, SR achieves the accuracy of 81.02% under the regularization parameter of 2^5 with the *p*-value of 0.005 for MCI identification, while its final accuracy with parametric values selected by cross validation is only 67.88%. *Third*, it can be observed that, SR and SR+ work well for MCI identification, but perform poorly for ASD classification, even though the regularization parameters are included in their model, which indicates that no method is always effective for all tasks or databases (as described by the famous no-free-lunch theorem [54]). *Finally*, although it achieves a comparable performance in many cases, we also note that, for ASD identification task, the proposed method performs poorly (especially under some *p*-values), suggesting that our method is not flexible enough to estimate FCNs, since no free parameter can be adjusted in practice. From this point of view, the selection of the hyperparameter is in fact a double-edged sword. We need to choose suitable FCN modelling methods according to specific tasks.

4.6. Discriminative features

Besides the MCI and ASD classification accuracy itself, an interesting problem is which features (*i.e.*, functional connections or corresponding ROIs in FCN) contribute most to specific disease identification tasks. In this group of experiments, we first select the 75 (for *p*-value of 0.01) and 19 (for *p*-value of 0.005) most discriminative features corresponding to the best accuracy of MCI and ASD classification tasks, and further visualize them in Fig. 7(A)-(B), respectively.

Specifically, each arc in Fig. 7 shows the selected feature between two ROIs, where the color is randomly allocated only for a better visualization. The thickness of each arc indicates its discriminative power that is inversely proportional to the corresponding the *p*-value.

For MCI classification, we rank the brain regions based on the selected discriminative features (functional connections), and find the top 3 brain regions are the left cerebellum6, the right angular and the right inferior temporal gyrus. According to previous studies [55,56], the two of the three brain regions (the first and third) are reported as potential biomarkers for MCI identification.

Similarly, for ASD classification problem, we find the top 3 most discriminative brain regions are the bilateral thalamus, the right

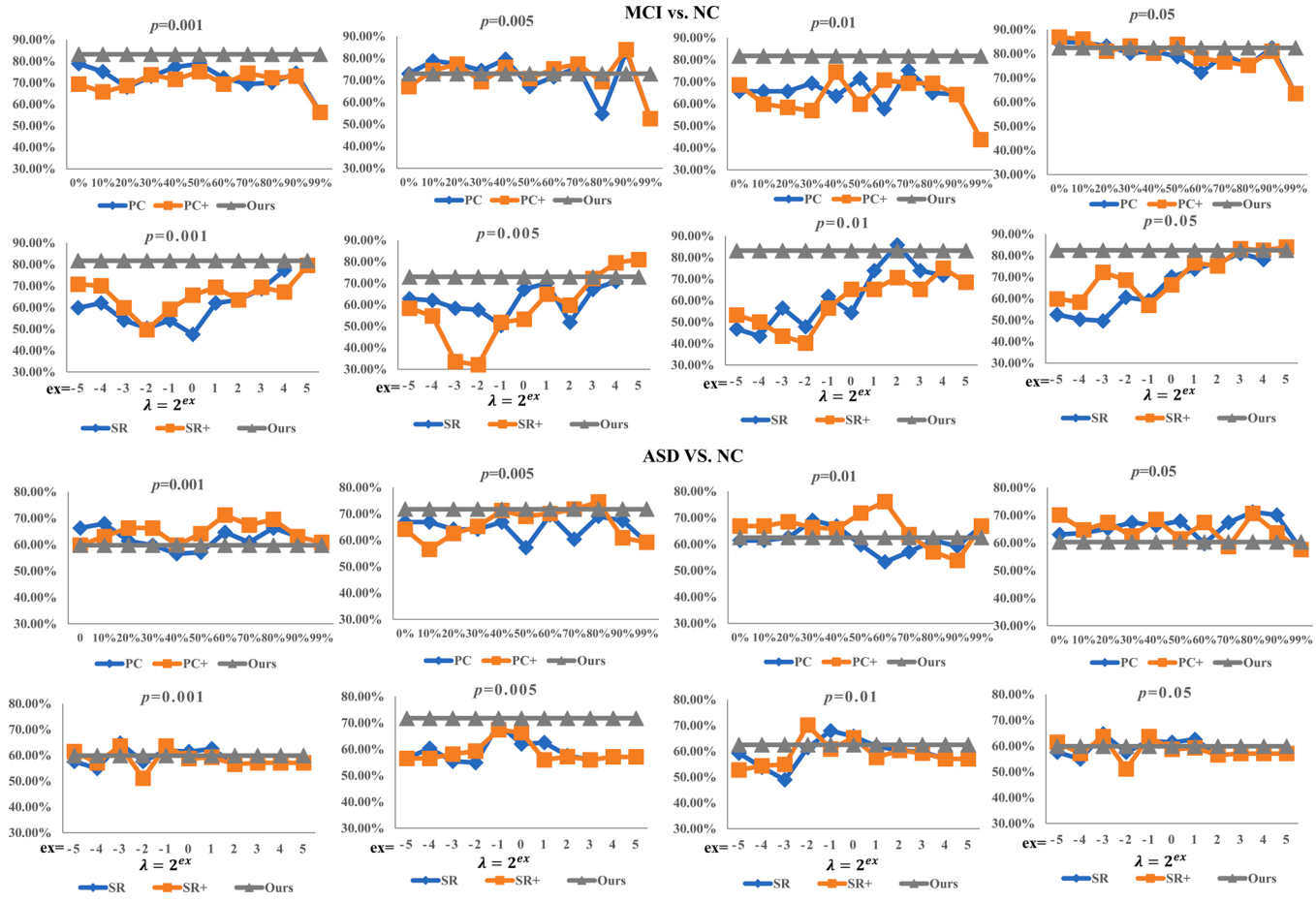
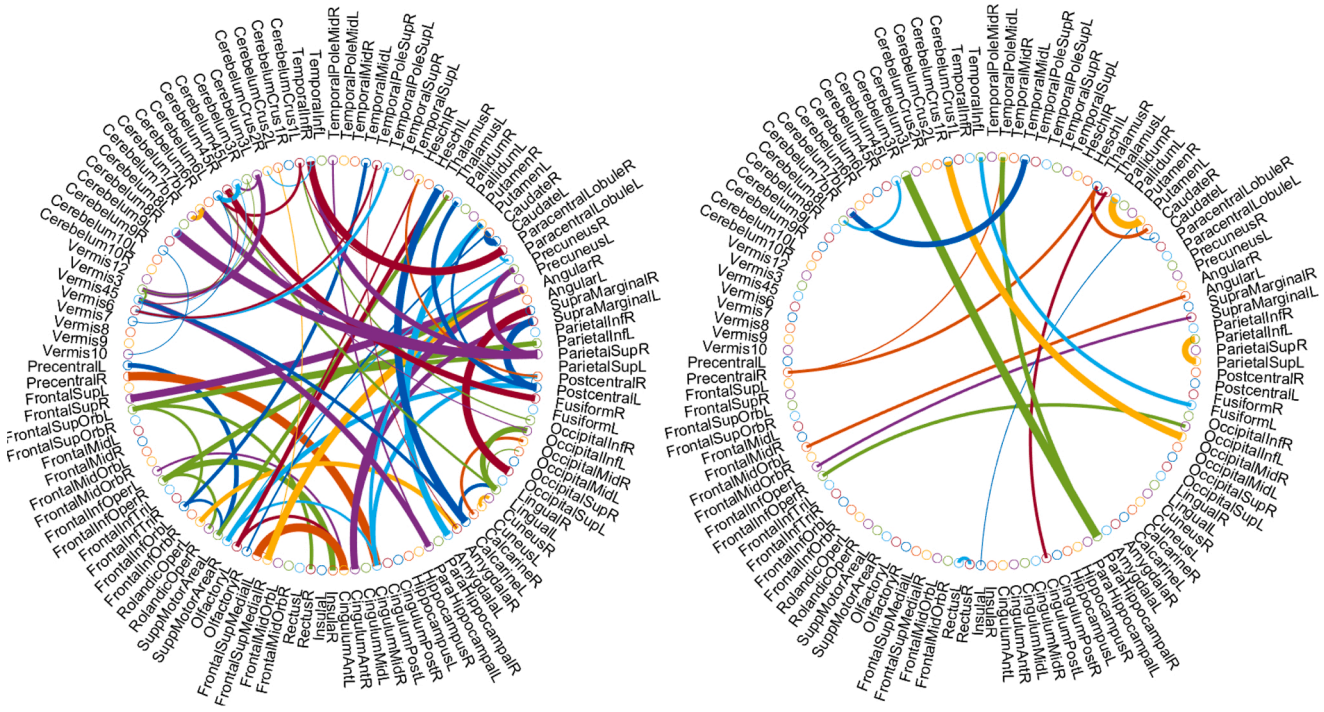


Fig. 6. The MCI and ASD average classification accuracies based on five FCN estimation methods (*i.e.*, PC, PC+, SR, SR+ and Ours) with different network modelling parametric values are shown in the top and bottom part of this figure respectively. Note that, the horizontal axis for PC-based methods represents different thresholds, and for SR-based methods show the regularization parameters, while the vertical axis for five methods represents the average classification accuracies.



(A) MCI VS. NC Classification (B) ASD VS. NC Classification

Fig. 7. The most discriminative features selected by the best accuracy of MCI identification (*i.e.*, MCI vs. NC classification) and ASD identification (*i.e.*, ASD vs. NC classification) tasks based on AAL template respectively. Note that, each arc shows the selected feature between two ROIs, where the color is randomly allocated only for a better visualization, and the thickness of each arc indicates its discriminative power that is inversely proportional to the corresponding the p-value. In addition, this figure is created by a Matlab function (*i.e.*, circularGraph) shared by Paul Kassebaum (http://www.mathworks.com/matlabcentral/fileexchange/48576-circulargraph).

parahippocampal gurus. According to previous studies [57,58], all brain regions may be biologically associated with ASD. These results further validate that our method is potentially useful in discovering fMRI biomarkers for MCI and ASD identification.

5. Conclusion

In this paper, we propose a hyperparameter-free FCN estimation model to circumvent the parametric selection problem in previous PC-based and SR-based methods. Then, we transform this model into a QP problem that can be efficiently solved by an off-the-shelf toolbox. To evaluate the effectiveness of the proposed method, we conduct experiments on two benchmark databases with rs-fMRI data for both MCI and ASD identification. The experimental results demonstrate that the proposed method can achieve comparable performance even though no parameter is included in our model. However, the selection of the hyperparameter is a double-edged sword. It avoids the difficulty of parameter selection, but, on the other hand, reduces the flexibility of the FCN estimation model. Additionally, the proposed method can only encode the positive interaction between ROIs. Therefore, we plan to develop more powerful hyperparameter-free FCN modelling methods that can capture both positive and negative dependency between ROIs in the future.

Conflicts of interest

The authors have no conflict of interest to disclose.

Acknowledgments

L. Sun, Y. Xue, Y. Zhang, L. Qiao and L. Zhang were partly supported by National Natural Science Foundation of China (Nos. 61976110, 11931008) and Natural Science Foundation of Shandong Province (No. ZR2018MF020, ZR2019YQ27).

Appendix

As shown in Section 2, the proposed model is formulated as

$$\begin{aligned} \min_{w_{ij}} \quad & f(W) = \sum_{i=1}^{n} \| \sum_{j \neq i} w_{ij} x_i - \sum_{j \neq i} w_{ij} x_j \|^2. \\ s.t. \quad & \sum_{j \neq i} w_{ij} \geq 1, \forall i = 1, \cdots, n \\ & w_{ij} = w_{ji} \geq 0, \forall j = 1, \cdots, n \end{aligned} \tag{8}$$

In this appendix, we give mathematical derivations for transforming Eq. (8) into a QP problem as follows

$$\begin{aligned} min_w \quad & w^T \mathrm{S} w. \\ s.t. \quad & \mathrm{A} w \geq b \end{aligned} \tag{9}$$

In particular, the mathematical derivations include the following two steps.

Step I: Proof of the equivalence between the objective functions in Eqs. (6) and (7) under the symmetric constraint of $w_{ij} = w_{ji} \geq 0, \forall j = 1, \cdots, n$.

Concretely, let $d_i = \sum_{j \neq i} w_{ij}$ be the degree of the ith node, e_i be a unit column vector in which the ith element is 1 and the other elements are 0, and $W_i = [w_{i1}, w_{i2}, \cdots, w_{in}]^T$ be the ith column of W. Then $f(W)$ can be rewritten as follows

$$\begin{aligned} f(W) &= \sum_{i=1}^{n} \| \sum_{j \neq i} w_{ij} x_i - \sum_{j \neq i} w_{ij} x_j \|^2 \\ &= \sum_{i=1}^{n} \| \mathrm{X} e_i d_i - \mathrm{X} \mathrm{W}_i \|^2 \\ &= \mathrm{tr}(\sum_{i=1}^{n} (\mathrm{X} e_i d_i - \mathrm{X} \mathrm{W}_i)(\mathrm{X} e_i d_i - \mathrm{X} \mathrm{W}_i)^T) \\ &= \mathrm{tr}(X \sum_{i=1}^{n} (e_i d_i - W_i)(e_i d_i - W_i)^T X^T) \\ &= \mathrm{tr}(X(\mathrm{D}\mathrm{D}^T - \mathrm{D}\mathrm{W}^T - \mathrm{W}\mathrm{D} + \mathrm{W}\mathrm{W}^T) X^T) \\ &= \mathrm{tr}(X(D - W)^T (D - W) X^T) \\ &= \| (D - W) X^T \|_F^2, \end{aligned} \tag{10}$$

where $tr(\cdot)$ indicates the trace of a matrix, and D is a diagonal matrix whose diagonal elements are given by d_i.

Furthermore, we transform $f(W) = \| (D - W) X^T \|_F^2$ into a standard quadratic form according to the trick used in. First, let $x^{(k)}$ be the kth column of X^T, $z = n(n-1)/2$ be the number of all possible edges, and $w = [w_{12}, w_{13}, \cdots, w_{1n}, w_{23}, \cdots, w_{n-1,n}]^T \in R^z$ be a column vector concatenated by the upper triangular elements in the matrix W. Then, we define G to be an $n \times z$ matrix whose nonzero elements are given as follows:

$$\begin{aligned} G_{i, \mathrm{index}(w_{a,b})} &= 1, \forall i = a = 1, \cdots, n-1 \\ G_{j, \mathrm{index}(w_{a,b})} &= -1, \forall j = a = 2, \cdots, n, \end{aligned} \tag{11}$$

where $\mathrm{index}(w_{ab}) \in 1, 2, \cdots, z$ is the index of w_{ab} in w. In addition, we define P as a diagonal matrix with diagonal elements given by w. As a result, $D - W$ can be formulated as GPG^T. Therefore, we have the following expression:

$$\begin{aligned} f(W) &= \| (D - W) X^T \|_F^2 \\ &= \sum_{k=1}^{m} \| (D - W) x^{(k)} \|^2 \\ &= \sum_{k=1}^{m} \| \mathrm{GPG}^T x^{(k)} \|^2 \\ &= \sum_{k=1}^{m} \| \mathrm{GPh}^{(k)} \|^2 \\ &= \sum_{k=1}^{m} \| \mathrm{GH}^{(k)w} \|^2 \\ &= w^T \mathrm{S} w, \end{aligned} \tag{12}$$

where $h^{(k)}$ is $G^T x^k$, $H^{(k)}$is a diagonal matrix whose diagonal elements are given by $h^{(k)}$, and $S = \sum_{k=1}^{m} (\mathrm{GH}(k))^T (\mathrm{GH}^k)$.

Step II: Proof of the equivalence between the constraints in Eqs. (6) and (7).

Concretely, the constraint $\sum_{j \neq i} w_{ij} \geq 1, \forall i = 1, \cdots, n$ can be equally converted into $Cw \geq e$, where $C = (abs(G_{il})) \in R^{n \times z}, \forall i = 1, \cdots, n, l = 1, \cdots, z$, and $e \in R^n$ is the all 1 column vector. Since w only contains the upper triangular elements of the matrix W, the constraint $w_{ij} = w_{ji} \geq 0, \forall i, j = 1, \cdots, n$ can be changed to $Iw \geq 0$, where I is the unit matrix, and 0 is the all zeros column vector. To further simplify the above forms, the final constraints in Eqs. (6) can be written as $Aw \geq b$, where $A = [C^T I]^T$, $b = [e^T 0^T]^T$.

References

[1] Huettel SA, Song AW, McCarthy G, et al. Functional magnetic resonance imaging. Sunderland, MA: Sinauer Associates; 2004. vol. 1.

[2] Fornito A, Zalesky A, Bullmore E. Fundamentals of brain network analysis. Academic Press; 2016.

[3] Van Den Heuvel MP, Pol HEH. Exploring the brain network: a review on resting-state fMRI functional connectivity. Eur Neuropsychopharmacol 2010;20(8):519–34.

[4] Fornito A, Zalesky A, Breakspear M. The connectomics of brain disorders. Nat Rev Neurosci 2015;16(3):159–72.

[5] Selkoe DJ. Alzheimer's disease: genes, proteins, and therapy. Physiol Rev 2001:12–3.

[6] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. Neurology 1984;34(7):939.

[7] Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiol 2007;17(1):103–11.

[8] Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. J Am Acad Child Adolesc Psychiatry 2008;47(8):921–9.

[9] Chaudhuri KR, Schapira AH. Non-motor symptoms of parkinson's disease: dopaminergic pathophysiology and treatment. Lancet Neurol 2009;8(5):464–74.

[10] Reijneveld JC, Ponten SC, Berendse HW, Stam CJ. The application of graph theoretical analysis to complex networks in the brain. Clin Neurophysiol 2007;118(11):2317–31.

[11] Wang J, Zuo X, He Y. Graph-based network analysis of resting-state functional MRI. Front Syst Neurosci 2010;4:16.

[12] Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 1995;34(4):537–41.

[13] Lee H, Lee DS, Kang H, Kim B-N, Chung MK. Sparse brain network recovery under compressed sensing. IEEE Trans Med Imaging 2011;30(5):1154–65.

[14] Meszlényi RJ, Hermann P, Buza K, Gál V, Vidnyánszky Z. Resting state fMRI functional connectivity analysis using dynamic time warping. Front Neurosci 2017; 11:75.
[15] Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage 2003;19 (4):1273–302.
[16] Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW. Network modelling methods for fMRI. Neuroimage 2011;54(2):875–91.
[17] Schwarz AJ, McGonigle J. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. Neuroimage 2011;55 (3):1132–46.
[18] van den Heuvel MP, de Lange SC, Zalesky A, Seguin C, Yeo BT, Schmidt R. Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: issues and recommendations. Neuroimage 2017;152:437–49.
[19] Van Wijk BC, Stam CJ, Daffertshofer A. Comparing brain networks of different size and connectivity density using graph theory. PloS One 2010;5(10):e13701.
[20] Huang S, Li J, Sun L, Ye J, Fleisher A, Wu T, Chen K, Reiman E, Initiative ADN, et al. Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. Neuroimage 2010;50(3):935–49.
[21] Zhou Y, Zhang L, Teng S, Qiao L, Shen D. Improving sparsity and modularity of high-order functional connectivity networks for MCI and ASD identification. Front Neurosci 2018;12:959.
[22] Pruim RH, Mennes M, van Rooij D, Llera A, Buitelaar JK, Beckmann CF. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. Neuroimage 2015;112:267–77.
[23] Wang M, Lian C, Yao D, Zhang D, Liu M, Shen D. Spatial-temporal dependency modeling and network hub detection for functional MRI analysis via convolutional-recurrent network. IEEE Trans Biomed Eng 2019.
[24] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 2002;15(1):273–89.
[25] Yan C-G, Wang X-D, Zuo X-N, Zang Y-F. DPABI: data processing & analysis for (resting-state) brain imaging. Neuroinformatics 2016;14(3):339–51.
[26] Li W, Wang Z, Zhang L, Qiao L, Shen D. Remodeling Pearson's correlation for functional brain network estimation and Autism spectrum disorder identification. Front Neuroinform 2017;11:55.
[27] Qiao L, Zhang L, Chen S, Shen D. Data-driven graph construction and graph learning: a review. Neurocomputing 2018;312:336–51.
[28] Heider F. Attitudes and cognitive organization. J Psychol 1946;21(1):107–12.
[29] Cartwright D, Harary F. Structural balance: a generalization of heider's theory. Psychol Rev 1956;63(5):277.
[30] Salvador R, Martinez A, Pomarol-Clotet E, Sarró S, Suckling J, Bullmore E. Frequency based mutual information measures between clusters of brain regions in functional magnetic resonance imaging. Neuroimage 2007;35(1):83–8.
[31] Grant M, Boyd S, Ye Y. CVX: matlab software for disciplined convex programming. 2009.
[32] Qiao L, Zhang H, Kim M, Teng S, Zhang L, Shen D. Estimating functional brain networks by incorporating a modularity prior. Neuroimage 2016;141:399–407.
[33] Tibshirani R. Regression shrinkage and selection via the LASSO. J R Stat Soc Series B 1996;58(1):267–88.
[34] Szenkovits A, Meszlényi R, Buza K, Gaskó N, Lung RI, Suciu M. Feature selection with a genetic algorithm for classification of brain imaging data. Advances in feature selection for data and pattern recognition. Springer; 2018. p. 185–202.
[35] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2(3):1–27.
[36] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.
[37] Qiu S, Joshi PS, Miller MI, Xue C, Zhou X, Karjadi C, Chang GH, Joshi AS, Dwyer B, Zhu S, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain 2020.
[38] Fletcher GS. Clinical epidemiology: the essentials. Lippincott Williams & Wilkins; 2019.
[39] Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31(3):968–80.
[40] Dosenbach NU, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC, Lessov-Schlaggar CN, et al. Prediction of individual brain maturity using fMRI. Science 2010;329(5997):1358–61.
[41] Craddock RC, James GA, Holtzheimer III PE, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum Brain Mapp 2012;33(8):1914–28.
[42] Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. Hum Brain Mapp 2001;14(3):140–51.
[43] Yan C-G, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, Li Q, Zuo X-N, Castellanos FX, Milham MP. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. Neuroimage 2013;76:183–201.
[44] Van Dijk KR, Sabuncu MR, Buckner RL. The influence of head motion on intrinsic functional connectivity MRI. Neuroimage 2012;59(1):431–8.
[45] Pervaiz U, Vidaurre D, Woolrich MW, Smith SM. Optimising network modelling methods for fMRI. Neuroimage 2020;211:116604.
[46] Dadi K, Rahim M, Abraham A, Chyzhyk D, Milham M, Thirion B, Varoquaux G, Initiative ADN, et al. Benchmarking functional connectome-based predictive models for resting-state fMRI. Neuroimage 2019;192:115–34.
[47] Wang M, Jie B, Bian W, Ding X, Zhou W, Wang Z, Liu M. Graph-kernel based structured feature selection for brain disease classification using functional connectivity networks. IEEE Access 2019;7:35001–11.
[48] Yu L, Liu H. Feature selection for high-dimensional data: a fast correlation-based filter solution. Proceedings of the 20th international conference on machine learning (ICML-03) 2003:856–63.
[49] Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, et al. A multi-modal parcellation of human cerebral cortex. Nature 2016;536(7615):171–8.
[50] Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, Yang Z, Chu C, Xie S, Laird AR, et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. Cereb Cortex 2016;26(8):3508–26.
[51] Altmann A, Ng B, Landau SM, Jagust WJ, Greicius MD. Regional brain hypometabolism is unrelated to regional amyloid plaque burden. Brain 2015;138 (12):3734–46.
[52] Dresler M, Shirer WR, Konrad BN, Müller NC, Wagner IC, Fernández G, Czisch M, Greicius MD. Mnemonic training reshapes brain networks to support superior memory. Neuron 2017;93(5):1227–35.
[53] Jin D, Wang P, Zalesky A, Liu B, Song C, Wang D, Xu K, Yang H, Zhang Z, Yao H, et al. Grab-AD: generalizability and reproducibility of altered brain activity and diagnostic classification in alzheimer's disease. Hum Brain Mapp 2020.
[54] Duda RO, Hart PE, Stork DG. Pattern classification. John Wiley & Sons; 2012.
[55] Suk H-I, Wee C-Y, Lee S-W, Shen D. Supervised discriminative group sparse representation for mild cognitive impairment diagnosis. Neuroinformatics 2015;13 (3):277–95.
[56] Kesler SR. Default mode network as a potential biomarker of chemotherapy-related brain injury. Neurobiol Aging 2014;35:S11–9.
[57] Sussman D, Leung R, Vogan V, Lee W, Trelle S, Lin S, Cassel D, Chakravarty M, Lerch J, Anagnostou E, et al. The autism puzzle: diffuse but not pervasive neuroanatomical abnormalities in children with ASD. NeuroImage Clin 2015;8: 170–9.
[58] Chen X, Zhang H, Lee S-W, Shen D, Initiative ADN, et al. Hierarchical high-order functional connectivity networks and selective feature fusion for MCI classification. Neuroinformatics 2017;15(3):271–84.