Medical & Biological Engineering & Computing (2022) 60:2813–2823
https://doi.org/10.1007/s11517-022-02628-7

ORIGINAL ARTICLE

Estimating high-order brain functional networks by correlation-preserving embedding

Hui Su[1] · Limei Zhang[1] · Lishan Qiao[1] · Mingxia Liu[2]

Received: 1 December 2021 / Accepted: 7 June 2022 / Published online: 22 July 2022

Abstract
Brain functional network (FN) has emerged as a potential tool for identifying mental and neurological diseases. Traditional FN estimation methods such as Pearson's correlation (PC) and sparse representation (SR), despite their popularity, can only model low-order relationships between brain regions (i.e., nodes of FN), thus failing to capture more complex interaction in the brain. Recently, researchers proposed to estimate high-order FN (HoFN) and successfully used them in the early diagnosis of neurological diseases. In practice, however, such HoFN is constructed by directly considering the columns (or rows) of the adjacency matrix of low-order FN (LoFN) as node feature vectors that may contain some redundant or noisy information. In addition, it is not really reflected whether the original low-order relationship is maintained during the construction of the HoFN. To address these problems, we propose correlation-preserving embedding (COPE) to re-code the LoFN prior to constructing HoFN. Specifically, we first use SR to construct traditional LoFN. Then, we embed the LoFN via COPE to generate the new node representation for removing the potentially redundant/noisy information in original node feature vectors and simultaneously maintaining the low-order relationship between brain regions. Finally, the expected HoFN is estimated by SR based on the new node representation. To verify the effectiveness of the proposed scheme, we conduct experiments on 137 subjects from the public Alzheimer's Disease Neuroimaging Initiative (ADNI) database to identify subjects with mild cognitive impairment (MCI) from normal controls. Experimental results show that the proposed scheme can achieve better performance than the baseline method.

Keywords High-order correlation · Brain functional network · Sparse representation · Correlation-preserving embedding · Mild cognitive impairment

1 Introduction

Resting state functional magnetic resonance imaging (rs-fMRI), generally by measuring the blood oxygen level dependent (BOLD) signals of the brain, can effectively capture the spontaneous cerebral activity that occurs when the subject remains awake and does not receive any external stimulus [1]. Based on rs-fMRI, one can estimate brain functional networks (FNs) to understand the cerebral working mechanism and provide effective biomarkers for the diagnosis of neurological diseases [2], such as autism spectrum disorder (ASD) [3, 4], major depressive disorder [5, 6], Alzheimer's disease (AD) [7], and its early stage, namely mild cognitive impairment (MCI) [8].

In general, FN is modeled as a graph, where the nodes are defined as the brain regions of interest (ROI) and the edge (generally with weight) denotes the statistical dependence between BOLD signals extracted from the ROI [9, 10]. To date, researchers have proposed many FN estimation methods [11], including Pearson's correlation (PC), partial correlation, regularized partial correlation, sparse low-rank representation [12], and dynamic causal model [13], to name a few. In this paper, we mainly focus on correlation-based methods because they are simple and effective [14].

Among correlation-based methods, PC is the most widely used in practice, but it can only capture the full correlation and obtain dense FN that may contain false connections

✉ Lishan Qiao
qlishan@163.com

✉ Mingxia Liu
mxliu1226@gmail.com

Hui Su
sh19980311@163.com

Limei Zhang
zhanglimei@lcu.edu.cn

[1] School of Mathematics Science, Liaocheng University, Liaocheng 252000, China

[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

caused by confounding effects from other brain regions. To address this problem, a threshold is generally used for sparse processing of FN [15, 16]. As an alternative to PC, partial correlation builds FN by regressing out potential effects from other brain regions. However, the partial correlation method generally requires the calculation of the inverse covariance matrix [17], which may lead to the ill-posed problem. Therefore, an l_1-norm regularization is generally included in the model to obtain more reliable partial correlation, resulting in the sparse representation (SR)–based FN estimation method [18].

Despite their successful applications, the above methods can only model the low-order relationships between brain regions (i.e., nodes of FN), thus failing to capture more complex interaction in the brain. Recently, researchers proposed some high-order FN (HoFN) estimation methods [19–23] towards modelling more complex relationships and used them in the early diagnosis of neurological diseases. Zhao et al. investigated the multi-level HoFN based on PC to diagnose ASD [24]; Chen et al. developed the HoFN estimation method based on sliding window and applied it to the classification of MCI [25]; Zhang et al. studied the method of constructing HoFN based on two consecutive correlation operations for the diagnosis of MCI abnormalities [26]; and Jia et al. constructed the HoFN using two consecutive canonical models for predicting both consciousness level and recovery outcome in acquired brain injury [27]. Due to the fact that these HoFN methods above are all based on two consecutive operations to calculate the high-order correlation, we simply refer to them as correlation's correlation (CC). Note that we can naturally generalize the two consecutive correlation operations to multiple correlations, but this goes beyond our main focus in the current study. Refer to [28] for a detailed investigation about this problem.

The experimental results show that the CC-based HoFN estimation methods are effective for the task of brain disease recognition and classification [26, 27]. However, such HoFN is constructed by directly considering the rows/columns of the LoFN adjacency matrix as node feature vectors that may contain some redundant or noisy information. We will shortly explain this point further in Fig. 1. Additionally, it does not really reflect whether the original low-order relationship is maintained during the construction of the HoFN. To address these problems, we propose correlation-preserving embedding (COPE) to re-code the LoFN prior to HoFN estimation. Specifically, we first use SR to construct traditional LoFN. Then, we embed the LoFN via COPE to generate the new node representation for removing the potentially redundant/noisy information in original node features and simultaneously maintaining the low-order relationship. Finally, the expected HoFN is estimated by SR based on the new node representation. To verify the effectiveness of our proposed scheme, we use the estimated HoFN for MCI recognition on the ADNI dataset. The experimental results illustrate that the proposed method can achieve better performance than the baseline method.

The rest of this paper is organized as follows. In Section 2, we introduce the preparation of data, the related methods for estimating FNs, and our proposed method. In Section 3, we report the experimental setting and evaluate the proposed method by applying it to recognition. In Section 4, we investigate the effect of FN modeling parameters on the experimental results, the selection of discriminative features, the advantages of the proposed method, and discuss the statistical significance. We finally conclude the paper in Section 5.

2 Methods and materials

In this section, we first describe data acquisition and preprocessing. Then, the correlation-based FN construction methods and our proposed scheme are depicted. Finally, we introduce the relationship between CC and our method.

2.1 Data acquisition and preprocessing

In this paper, we use the publicly available ADNI database[1] for evaluating the involved methods. As in the recent study [29, 30], 137 subjects, including 68 MCIs and 69 NCs, are adopted in our experiment. The selected fMRI images were scanned by 3.0 T Philips scanner with the following parameters: slice thickness is 3.3 mm, TE is 30 ms, and TR is $2.2-3.1$ s. The scanning time is 7 min, leading to 140 volumes for each subject. In Table 1, we list the main demographic information of the subjects included in this study.

The original rs-fMRI data is preprocessed via SPM8 toolbox[2] as follows. To remain the signal stabilization, we removed the first three volumes of each subject from the fMRI time course with 140 volumes. For the remaining 137 volumes, we conduct head motion and slice timing correction. Furthermore, nuisance regression is used to reduce the effect of the ventricle and white matter signals based on Friston 24 parameters. After that, the corrected rs-fMRI images are registered to the standard Montreal neurological institute (MNI) space and spatially smoothed by a Gaussian kernel with the full-width-half maximum of 4 mm. Then, band-pass filter ($0.015-0.150$ Hz) is used to remove the extremely low- and high-frequency artifacts. Finally, according to the automated anatomical labeling (AAL) template [31], the brain was partitioned into 116 ROIs, and the mean time series of each ROI were put into a data matrix $X \in R^{137\times116}$.

[1] http://adni.loni.usc.edu/

[2] http://www.fil.ion.ucl.ac.uk/spm/

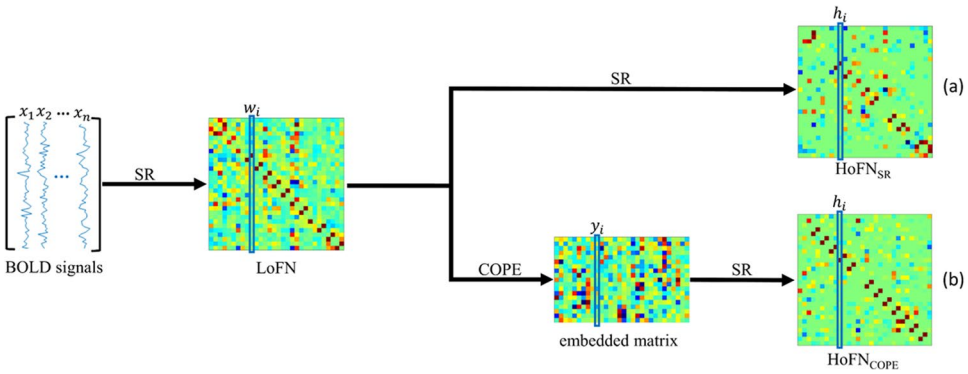

Fig. 1 The pipeline of building HoFN. Branch (**a**) is the CC-based HoFN estimation; Branch (**b**) is the proposed method

2.2 Related methods

As mentioned earlier, researchers have developed many methods to estimate FNs. In this section, we first review two correlation-based LoFN estimation approaches, PC and SR, and then briefly introduce the CC-based HoFN construction method. These methods are all closely related to our study.

2.2.1 Pearson's correlation

It is well known that PC is the simplest and most popular method for estimating FNs. Suppose that the brain has been parcellated into n ROIs based on a certain atlas. Denote $x_i \in R^m (i = 1, \cdots, n)$ as the mean time series extracted from the i th ROI, where m is the number of the time points in each series. Then, the edge weight w_{ij} between the i th and j th ROIs in the PC-based FN can be calculated as follows:

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

where $\bar{x}_i \in R^m$ is the mean vector corresponding to x_i. Without loss of generality, we normalize x_i by $x_i = (x_i - \bar{x}_i)/\sqrt{(x_i - \bar{x}_i)^T (x_i - \bar{x}_i)}$, and thus, the PC can be simply expressed as $w_{ij} = x_i^T x_j$, which corresponds to the optimal solution of the following model:

$$\min_{w_{i,j}} \sum_{i,j=1}^{n} \|x_i - w_{ij} x_j\|^2 \tag{2}$$

According to a previous work [32], Eq. (2) can be further transformed 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 fMRI data matrix, $W = (w_{ij}) \in R^{n \times n}$ is the edge weight matrix of the estimated FN, and $||\bullet||_F$ denotes the Frobenius norm of a matrix.

In general, the constructed FN based on PC is a dense graph (i.e., all vertices are fully-connected by edges), which may contain noisy or uninformative connections. In practice, a threshold is generally used to sparse the PC-based FN by filtering out the weak connections.

2.2.2 Sparse representation

Although PC is empirically effective in building FN, it can only measure the full correlation. By contrast, partial correlation aims to estimate more reliable connections between two ROIs by regressing out the confounding effect from other ROIs. The general approach to calculate partial correlation is based on the estimation of inverse covariance matrix. However, this approach may be ill-posed due to the singularity of the sample covariance matrix. To address this problem, an l_1 regularizer is generally introduced into the traditional model, which leads to SR-based FN estimation as follows:

$$\min_{w_{ij}} \sum_{i=1}^{n} (\|x_i - \sum_{j \neq i} w_{ij} x_j\|^2 + \lambda_1 \sum_{j \neq i} |w_{ij}|) \tag{4}$$

Equivalently, it can be further simplified to the following matrix form:

$$\begin{aligned} &\min_{W} \|X - XW\|_F^2 + \lambda_1 \|W\|_1 \\ &s.t. w_{ii} = 0, \forall i = 1, \cdots n \end{aligned} \tag{5}$$

Table 1 Demographic and clinical information of subjects in the ADNI datasets. Values are reported as mean ± standard deviation. M/F, male/female; MMSE, mini-mental examination

Dataset	Class	Gender (M/F)	Age (years)	MMSE
ADNI	MCI	39/29	76.50 ± 13.50	26.77 ± 1.23
	NC	17/52	71.50 ± 14.50	28.85 ± 1.15

where $||\bullet||_F$ and $||\bullet||_1$ are the F-norm and l_1-norm of a matrix, respectively. The constraint $w_{ii} = 0$ is used here to avoid the trivial solution by removing x_i from X. λ_1 is a regularized parameter that can control the sparsity of the estimated FN and benefit to achieve a stable solution.

2.2.3 HoFN estimation based on correlation's correlation

The two basic FN construction methods described previously can only capture the low-order (second-order) statistics, but ignore more complex interactions between ROIs [26, 27]. In this section, we review a CC-based FN construction method that can model higher-order statistics between ROIs and also closely relates to the proposed method. Due to the fact that SR is used to correlation calculation in the high-order method, we name it HoFN_{SR}, and take it as the baseline in our study.

The HoFN_{SR} method involve two main steps. First, the traditional LoFN $W = (w_{ij})_{n\times n}$ is calculated by SR, as shown in Eq. (5). As a result, for each ROI, we have a node vector (*i.e.*, the column of the FN adjacency matrix) to measure the relationship between this ROI and all other ROIs. In particular, we suppose that $w_i = [w_{i1}, \cdots, w_{ij}, \cdots, w_{in}]$ denotes the feature vector of the node i (corresponding to the i th ROI). Then, based on this LoFN, SR is again carried out on the node feature vector to calculate the HoFN as follows:

$$\min_{h_{ij}} \sum_{i=1}^{n} (\|w_i - \sum_{j\neq i} h_{ij} w_j\|^2 + \lambda_2 \sum_{j\neq i} |h_{ij}|) \tag{6}$$

where h_{ij} is the relationship between the feature vector of i th ROI and j th ROI, and λ_2 is a regularized parameter for controlling the sparsity of the estimated HoFN.

Equation (6) can also be equivalently transformed into the following matrix form:

$$\begin{aligned} &\min_{H} \|W - WH\|_F^2 + \lambda_2 \|H\|_1 \\ &\quad s.t. h_{ii} = 0, \forall i = 1, \cdots n \end{aligned} \tag{7}$$

where $H = (h_{ij}) \in R^{n\times n}$ is the estimated HoFN that is expected to capture more complex interaction between brain regions at a higher-order level.

2.3 Proposed method

As mentioned above, the HoFN is constructed by directly representing the adjacency matrix of LoFN as node features that may contain some redundant or noisy information. Furthermore, the noisy connections inherited from LoFN may in turn lead to unreliable HoFN. More importantly, it is not really reflected whether the original low-order relationship is maintained during the construction of the HoFN. Therefore, in this section, we develop a new scheme COPE to re-code the LoFN prior to construct HoFN. The proposed method aims at not only removing the potentially redundant/noisy information in original node features, but also maintaining the low-order relationship in the embedding space.

Specifically, we first use SR to construct traditional LoFN whose adjacency matrix is denoted by $W = (w_{ij})_{n\times n}$. Then, we embed the LoFN via COPE to generate the new node representation $Y = [y_1, y_2, \cdots, y_n] \in R^{d\times n}$ by minimizing the following objective function:

$$\Phi(Y) = \sum_{i=1}^{n} \| y_i - \sum_{j=1}^{n} w_{ij} y_j \|^2 \tag{8}$$

where y_i is the new node representation of node i in the low-dimensional embedded space, and w_{ij} is the edge weight of LoFN estimated in the previous step. In fact, the weight w_{ij} in LoFN is the low-order information (or rather second-order statistic) that encodes the relationship between the high-dimensional BOLD signals. Then, we use COPE to find a low-dimensional embedding space where the low-order information w_{ij} can be well preserved. This can be done by optimizing Eq. (8) and the new representation $Y = [y_1, y_2, \cdots, y_n] \in R^{d\times n}$, where d controls the dimension of the embedding space. In other words, COPE aims to keep the low-order information contained in the high-dimensional space as much as possible in the low-dimensional vector space after embedding.

In mathematics, Eq. (8) can be transformed into the following matrix form:

$$\begin{aligned} \Phi(Y) &= \sum_{i=1}^{n} \|YI_i - YW_i\|^2 \\ &= \sum_{i=1}^{n} \|Y(I_i - W_i)\|^2 \\ &= \sum_{i=1}^{n} [Y(I_i - W_i)(I_i - W_i)^T Y^T] \\ &= tr(Y(I - W)(I - W)^T Y^T) \end{aligned} \tag{9}$$

For avoiding trivial solution, we further constrain $\sum_{i=1}^{n} tr(y_i y_i^T) = n$, that is, $tr(YY^T) = n$. As a result, we have the final model of COPE as follows:

$$\min_{Y} tr(YMY^T) \ \ s.t. tr(YY^T) = n \tag{10}$$

where $M = (I - W)(I - W)^T$. Next, the Lagrange multiplier method is used to transform the above problem into the following:

$$L(Y) = tr(YMY^T) + tr(\mu(YY^T - n)) \tag{11}$$

By taking the derivative of the above equation, we have the following:

$$MY^T = \mu Y^T \tag{12}$$

This turns COPE into a generalized eigenproblem, and the eigenvectors corresponding to the smallest d non-zero eigenvalues of the matrix M is the optimal solution that

will serve as the new node representation for estimating HoFN as follows:

$$\min_{h_{ij}} \sum_{i=1}^{n} (\|y_i - \sum_{j \neq i} h_{ij} y_j\|^2 + \lambda_2 \sum_{j \neq i} |h_{ij}|) \tag{13}$$

Equation (13) can be further transformed mathematically into the following matrix form:

$$\begin{aligned} &\min_{H} \|Y - YH\|_F^2 + \lambda_2 \|H\|_1 \\ &s.t. h_{ii} = 0, \forall i = 1, \cdots n \end{aligned} \tag{14}$$

where λ_2 is a regularized parameter for controlling the balance of two terms in the objective function. In order to facilitate the subsequent description, the HoFN constructed based on the COPE method is simply named as $\text{HoFN}_{\text{COPE}}$.

The specific pipeline of constructing HoFN is shown in Fig. 1. Branch (a) is the CC-based HoFN estimation: (1) LoFN is first calculated by SR, and the columns of adjacency matrix are taken as node feature vectors (e.g., w_i); (2) then, SR is carried out on the node feature vector to calculate HoFN, where h_i represents the high-order correlation between i th ROI and other ROIs. Branch (b) is the proposed method: (1) With the same step as CC-based method to construct LoFN, (2) LoFN is re-coded via COPE to obtain embedded matrix whose column w_i is taken as new node representation, and (3) HoFN is estimated by SR based on the new node representation. We note that the traditional CC-based methods directly represent the adjacency matrix of LoFN as node features to construct HoFN. In contrast, the proposed method re-code the LoFN via COPE before constructing HoFN. By converting the node information of the LoFN into a vector with smaller dimension, more essential features can be extracted for HoFN construction. Without the COPE operation on the LoFN, the $\text{HoFN}_{\text{COPE}}$ will reduce to the traditional method. In other words, we can consider traditional CC-based methods as a special case of our proposed method.

3 Experiments and results

3.1 Brain functional network estimation

After obtaining the pre-processed fMRI data, we estimate FNs based on different methods. Specifically, four methods, including PC, SR, HoFN_{SR}, and $\text{HoFN}_{\text{COPE}}$, are involved in our experiment. In general, each FN estimation method contains one or more hyper-parameters. For the regularized parameter (i.e., λ_1 in SR, λ_2 in both HoFN_{SR} and $\text{HoFN}_{\text{COPE}}$), we search in the candidate range of $[2^{-10}, 2^{-9}, \cdots, 2^{-1}, 2^{0}]$ to select the optimal parametric value. Note that, for PC-based FN estimation method, no any free parameter is involved in the model itself. However, in order to improve its flexibility and make a fair comparison, we introduce a threshold parameter to sparse the initial estimated FNs by removing some weak connections. To keep in line with the number of parameters subset in other methods, we also use 11 candidate thresholds in the range of $[0, 10\%, \cdots, 90\%, 99\%]$, corresponding to different degrees of sparsity, where the parametric values mean the percentage of weak connections that are discarded. In addition, for $\text{HoFN}_{\text{COPE}}$, the embedding dimension d is selected from $[100, 90, \cdots, 20, 10]$, and the influence of the dimension on classification results is discussed in the following sections.

3.2 Feature selection and classification

With the estimated FNs, we subsequently identify the subjects with MCI from NCs using the popular support vector machine [33] that has been verified to work well in the recent survey [34]. In our experiment, the edge weights of low-order or high-order networks are used as features for the identification task. Since the estimated FN matrix is symmetric, we only consider its upper triangular elements as the input features. In particular, each FN has 116 nodes and thus can produce $116 \times (116 - 1)/2 = 6670$ features (corresponding to 6670 functional connections between 116 ROIs). Compared to the sample size (i.e., the number of subjects), the feature dimension is very high, which often leads to the so-called curse of dimensionality and in turn affects the final classification accuracy. To address this problem, researchers have proposed numbers of approaches for feature selection [35]. In this paper, we only adopt the simplest feature selection method t-test with $p = 0.05$, since our main focus is to evaluate the FN estimation strategies.

The detailed pipeline for feature selection and classification is shown in Fig. 2. Due to the limitation of sample size, we use the leave-one-out cross validation (LOOCV) to estimate the classification performance of the involved methods. Specifically, for the 137 subjects in the dataset, 136 of them are used for training, while the remaining one is used for testing. The final performance can thus be achieved by averaging results of all the runs. The parameters involved in the FN estimation models may affect the structure of the brain network and the final classification result. In order to obtain the optimal parameters of each method, an inner LOOCV is further conducted on the training data by the grid-search strategy as shown in Fig. 2.

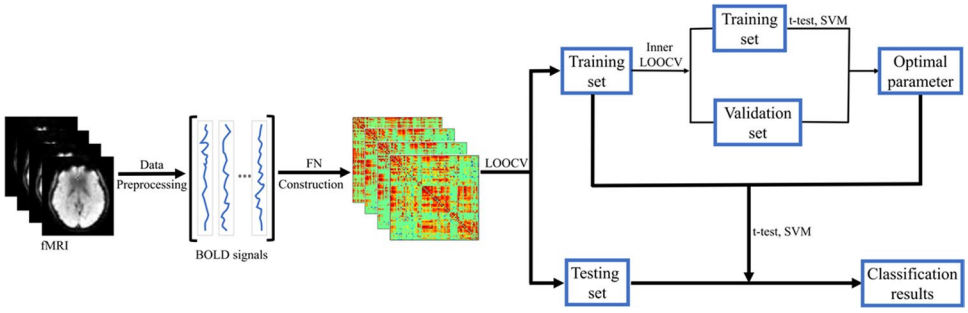

Fig. 2 The MCI identification pipeline based on the estimated FNs in this study, which contains three major modules: (1) data preprocessing, (2) FN construction, and (3) feature selection and classification

3.3 Results

3.3.1 Brain network visualization

For an intuitive comparison among differently estimated FNs, we take one subject from ADNI dataset as an example to visualize the adjacency matrices of the FNs estimated by PC, SR, HoFN_{SR}, and $\text{HoFN}_{\text{COPE}}$ in Fig. 3. It can be observed that the PC based FN is significantly different from those estimated by SR-based methods, since they use different data fitting terms to capture full correlation and partial correlation between ROIs, respectively. In contrast, all the SR-based methods lead to similar topological structures, and the resulted FNs are sparse due to the introduction of l_1-norm regularization. In addition, we select 14 subjects and use the signed modularity maximization algorithm to calculate the modularity scores of HoFNs constructed by two methods under optimal parameter combinations [36, 37]. As shown in Fig. 4, modularity of FN estimated by $\text{HoFN}_{\text{COPE}}$ is generally higher than that by HoFN_{SR}, which indicates that $\text{HoFN}_{\text{COPE}}$ tends to achieve a cleaner FN with clear modularity structure.

3.3.2 Classification performance

In this paper, we adopt five quantitative metrics, including accuracy (*ACC*), sensitivity (*SEN*), specificity (*SPE*), positive predictive value (*PPV*), and negative predictive value (*NPV*) to evaluate the classification performance of different methods. Their mathematical definitions are given as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$SEN = \frac{TP}{TP + FN} \tag{16}$$

$$SPE = \frac{TN}{TN + FP} \tag{17}$$

$$PPV = \frac{TP}{TP + FP} \tag{18}$$

$$NPV = \frac{TN}{TN + FN} \tag{19}$$

where *TP*, *TN*, *FP*, and *FN* indicate the true positive, true negative, false positive and false negative, respectively. Of note, in this work, we treat the subjects with MCI as the positive class while the NCs as the negative class.

In Table 2, we report the classification results for the MCI identification task based on four different FN estimation methods. Especially for the $\text{HoFN}_{\text{COPE}}$, the embedding dimension d is fixed with 80. In the next section, we will discuss in detail the influence of different parametric values on the final classification performance. As can be seen from Table 2, the proposed method significantly outperforms the three baseline methods. More interestingly, we note that the high-order method does not necessarily result in a better performance than the low-order counterpart. This finding is consistent with a recent study [28]. However, with the COPE operation before the high-order network construction, the $\text{HoFN}_{\text{COPE}}$ tends to extract more essential (at least more discriminative) features of the LoFN, thus achieving the best classification performance.

4 Discussion

In this section, we investigate the effect of network modelling parameters on the classification results and show the most discriminative features selected by our method for exploring their relationship with brain disorders. In addition, we further discuss the advantages of the proposed method and the statistical significance of FN estimated by the four methods.

4.1 Sensitivity to network modelling parameters

In general, the parameters involved in the FN estimation methods have an important influence on the ultimate

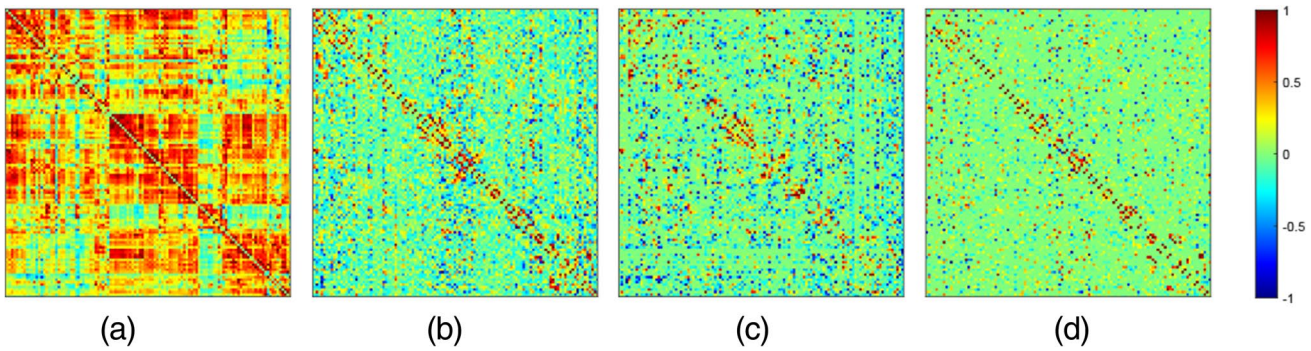

Fig. 3 For one subject, the adjacency matrices of FNs estimated by four different methods, i.e., PC, SR, HoFN_{SR}, and $\text{HoFN}_{\text{COPE}}$. Note that the elements in the adjacency matrices have been normalized into the interval of [−1,1] for the convenience of comparison between the visualized results

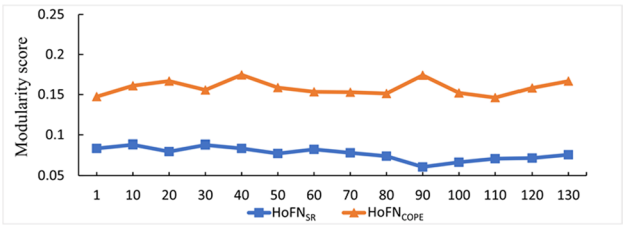

Fig. 4 Modularity scores of HoFN constructed by different methods. Note that the horizontal axis indicates the label of the participant

Table 2 The classification results corresponding to four different FN estimation methods under five performance indices

Method	*ACC*	*SEN*	*SPE*	*PPV*	*NPV*
PC	0.7956	0.7647	0.8261	0.8125	0.7808
SR	0.8029	0.7941	0.8116	0.8060	0.8000
HoFN_{SR}	0.6788	0.6912	0.6667	0.6714	0.6866
$\text{HoFN}_{\text{COPE}}$	**0.8759**	**0.8824**	**0.8696**	**0.8696**	**0.8824**

Bold values indicates the best results

classification performance [38, 39]. In this study, the parameters can be divided into two groups: the embedding dimension d and the regularization parameters in the network estimation models. To investigate the sensitivity of the involved methods to different parametric values, we repeat MCI classification experiments based on different dimensions d and regularization parameter combinations and report the results in Figs. 5 and 6, respectively.

As shown in Fig. 5, the classification performance based on $\text{HoFN}_{\text{COPE}}$ fluctuates with the change of parameter d, where the horizontal line denotes the performance of the baseline method HoFN_{SR} that is independent of the embedding dimension. In particular, a big dimension tends to result in better performance, and with the decrease of the dimension, the classification performance of $\text{HoFN}_{\text{COPE}}$ declines gradually. Especially, when the dimension is reduced to 40, the classification performance of $\text{HoFN}_{\text{COPE}}$ becomes worse than the baseline. This illustrates an over-reduction of the dimension may cause the loss of useful information for discrimination. Therefore, the selection of embedding dimension is crucial to the final classification performance, and we empirically set the dimension d to 80.

In the case of fixed embedding dimension d, the method of constructing HoFN involves two parameters to control the sparsity of the network. Specifically, HoFN_{SR} and $\text{HoFN}_{\text{COPE}}$ both contain two regularization parameters: one (λ_1) is used to control the sparsity of low-order networks, and the other (λ_2) is used to control the sparsity of high-order networks. In Fig. 6, we show the classification accuracy corresponding to different parametric combinations in the proposed method when d is 80. It can be observed that $\lambda_1 = 2^0$ and $\lambda_2 = 2^{-10}$ are the optimal combinations of regularized parameters for HoFN_{SR}, and the accuracy is 77.37%. In contrast, $\text{HoFN}_{\text{COPE}}$ reaches the highest classification accuracy of 90.51% at $\lambda_1 = 2^{-1}$ and $\lambda_2 = 2^{-3}$. Compared with the performance of the two HoFN methods, we found that the classification result of $\text{HoFN}_{\text{COPE}}$ is superior to the traditional HoFN_{SR} on the whole. In addition, under different parameter combinations, the classification performance of HoFN constructed by $\text{HoFN}_{\text{COPE}}$ model is more stable than that of HoFN_{SR}.

The above two groups of experiments discuss the influence of the parameters involved in the HoFN construction methods (i.e., HoFN_{SR} and $\text{HoFN}_{\text{COPE}}$) on the final classification performance. To further study the classification performance of $\text{HoFN}_{\text{COPE}}$, we compare it with LOFN constructed by PC under different thresholds. As shown in Fig. 7, PC-based classification accuracy fluctuates greatly with the change of thresholding values, where the horizontal reference line represents the performance of $\text{HoFN}_{\text{COPE}}$ when the fixed embedding dimension $d = 80$. By observing Fig. 7, we find that the classification accuracy of LoFN, under different thresholds, is lower than $\text{HoFN}_{\text{COPE}}$. This

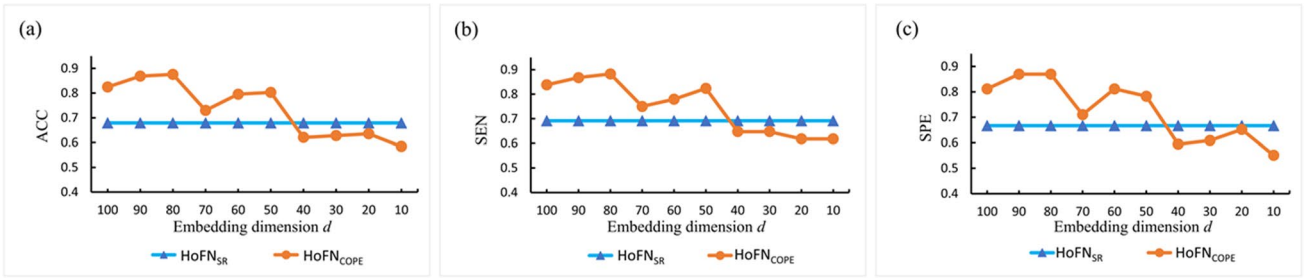

Fig. 5 Influence of embedding dimension d on the performance of $\text{HoFN}_{\text{COPE}}$, including (**a**) *ACC*, (**b**) *SEN*, and (**c**) *SPE*, respectively. Note that we also provide the high-order brain network construction method HoFN_{SR} as a baseline that is independent of the embedding dimensions

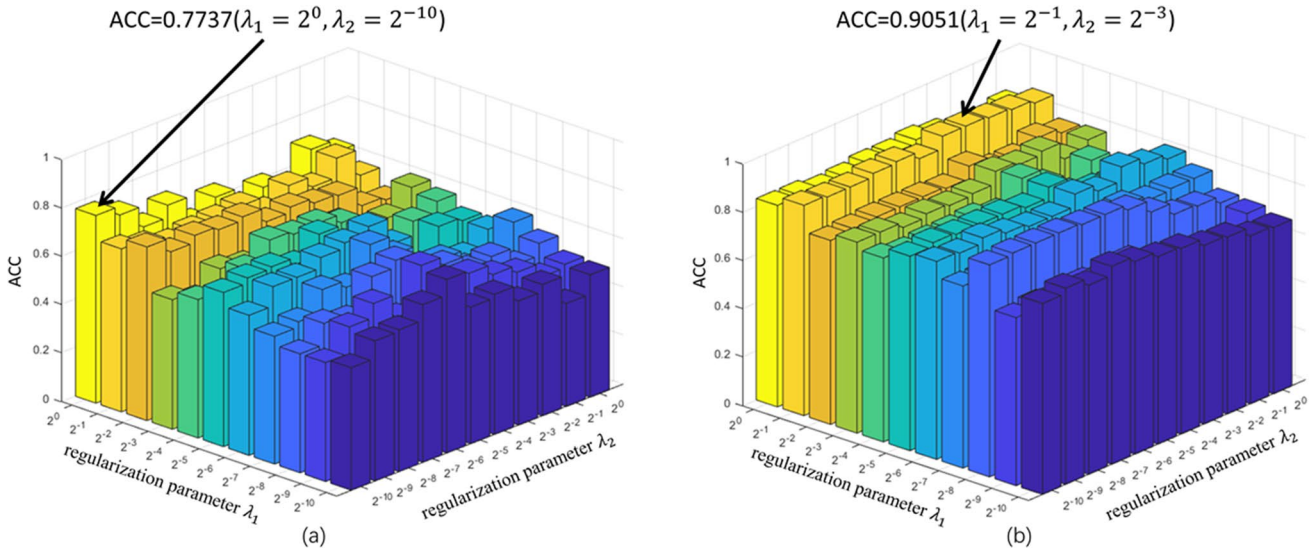

Fig. 6 Classification accuracy corresponding to two high-order brain network construction methods, (**a**) HoFN_{SR} and (**b**) $\text{HoFN}_{\text{COPE}}$, under different regularization parameter combinations. Note that the results of $\text{HoFN}_{\text{COPE}}$ are based on a fixed embedding dimensions $d = 80$

also indicates that HoFN contains more discriminative information to help the classification task.

4.2 Discriminative features

In addition to the classification accuracy itself, an interesting aspect is which features contribute the most to disease identification. As previously mentioned in Sect. 3.2, this paper uses the edge weights of estimated FN as features for classification. Here, after constructing FN by $\text{HoFN}_{\text{COPE}}$ ($d = 80$) model, we apply t-test with p-value of 0.005 to select discriminative features. As a result, we select 61 discriminative connections and visualize them in Fig. 8.

Of note, each arc in Fig. 8 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 p-value. From Fig. 8, we can find that the brain regions associated with top discriminative features include hippocampus, para-hippocampus, and precuneus. According to previous studies [40, 41], these regions are reported as potential biomarkers for AD and MCI.

4.3 The advantages of the proposed method

To date, researchers have proposed many FN estimation methods, including PC, partial correlation, and some directed connection estimation methods such as dynamic causal modelling (DCM) [13], Granger causality [42], structural equation modelling (SEM) [43], and Bayes net methods [44]. In order to evaluate different connectivity estimation methods, Smith et al. conducted a systematical comparative experiment [45] and found that, "correlation-based approaches can be quite successful, where partial correlation methods can give high sensitivity

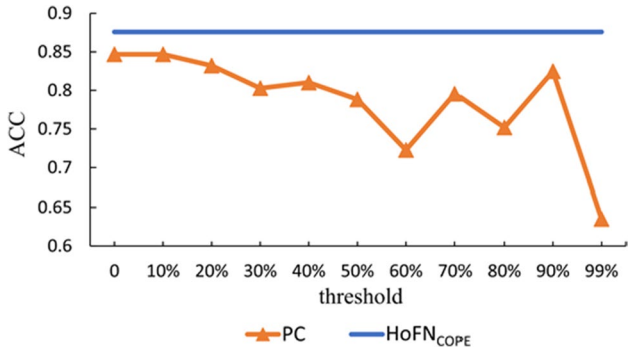

Fig. 7 Impact of different threshold values on the classification accuracy of PC. Note that the results of $\text{HoFN}_{\text{COPE}}$ are based on a fixed embedding dimensions $d = 80$

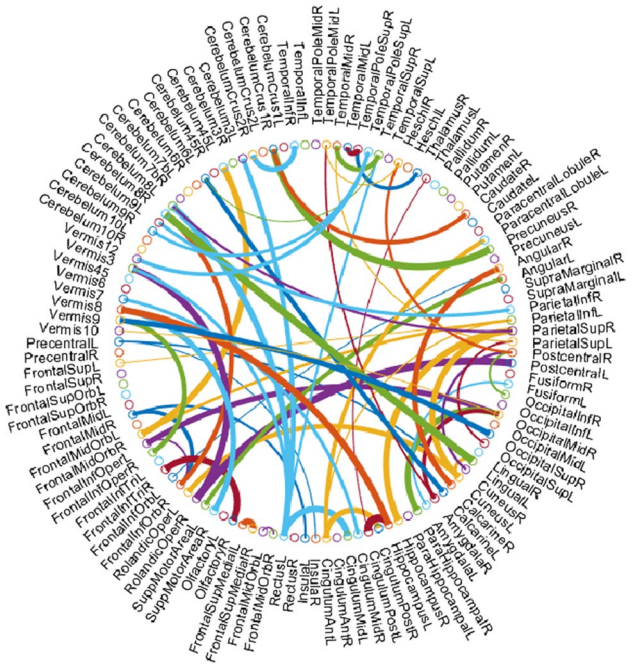

Fig. 8 The most discriminative features (network connections) selected using *t*-test. Note that each arc shows the selected feature between two ROIs, where the color is randomly allocated only for a better visualization, and the thickness of each arc indicates its discriminative power that is inversely proportional to the corresponding *p*-value

to network connection detection on good quality fMRI data." Additionally, they pointed out that "none of the directed methods is very accurate at estimating directionality." Therefore, in this paper, we do not consider the directionality, but use the partial correlation to calculate the connectivity between two brain regions by regressing out the confounding effects from other brain regions.

4.4 Analysis of statistical significance

To evaluate the significance of FNs estimated by PC, SR, HoFN_{SR}, and $\text{HoFN}_{\text{COPE}}$, respectively, we perform statistical significance tests based on surrogate analysis. Specifically, we apply inverse Fourier transform of the original BOLD signals to obtain surrogate datasets, which have the same power spectrum as the original data [46], thus generating three surrogates in this paper. After obtaining the surrogate data, we estimate FNs for the identification task based on four different methods. The classification accuracy is used here as the statistic of the test, and the measure of significance [46, 47] $S = |Q_{\text{orig}} - < Q_{\text{surr}} > |/\sigma_{\text{surr}}$, where Q_{orig} is the statistic for the original data, $< Q_{\text{surr}} >$ is the mean value statistic for the surrogates, and σ_{surr} is the standard deviation of the statistic for the surrogates. We select the significant level $\alpha = 0.05$, and thus, when $S \geq 1.96$, it indicates that the constructed FN is significant. In Table 3, we report significance measures of FNs constructed by four different methods. As can be seen from Table 3, the statistical measures are all greater than 1.96, which indicates that all the differently constructed FNs have statistical significance.

Table 3 The significance measure corresponding to four different FN estimation methods

Method	PC	SR	HoFN_{SR}	$\text{HoFN}_{\text{COPE}}$
S	3.1056	2.9003	2.7282	2.2645

5 Conclusion

In this paper, we propose a new scheme to construct HoFN by embedding the LoFN via COPE to generate the new node representation for removing the potentially redundant/noisy information in original node features and simultaneously maintaining the low-order relationship in the low-dimensional embedding space. To evaluate the effectiveness of the proposed scheme, we conduct experiments on the ADNI dataset to identify subjects with MCI from normal controls. The experimental results demonstrate that the proposed method can achieve better performance than the baseline method. Finally, it is worth pointing out that the proposed COPE is a shallow embedding method that is generally simple and has a better interpretability. However, advanced studies have shown that many deep embedding methods based on graph neural networks can generate more informative node representation [48].

Therefore, we plan to extend COPE to a deep version and use it to estimate HoFN in the future.

Funding This work was partly supported by the National Natural Science Foundation of China (Nos. 61976110, 62176112, 11931008), the Natural Science Foundation of Shandong Province (No. ZR202102270451), and the Open Project of Liaocheng University Animal Husbandry Discipline (No. 319312101–01).

References

1. Liu Y et al (2008) Regional homogeneity, functional connectivity and imaging markers of Alzheimer's disease: a review of resting-state fMRI studies. Neuropsychologia 46(6):1648–1656
2. Azeez AK, Biswal BB (2017) A review of resting-state analysis methods. Neuroimaging Clin 27(4):581–592
3. Geschwind DH, Levitt P (2007) Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiol 17(1):103–111
4. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G (2008) Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. J Am Acad Child Adolesc Psychiatry 47(8):921–929
5. Greicius MD et al (2007) Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. Biol Psychiat 62(5):429–437
6. Cullen KR et al (2014) Abnormal amygdala resting-state functional connectivity in adolescent depression. JAMA Psychiat 71(10):1138–1147
7. Wang L et al (2006) Changes in hippocampal connectivity in the early stages of Alzheimer's disease: evidence from resting state fMRI. Neuroimage 31(2):496–504
8. Rombouts SA, Barkhof F, Goekoop R, Stam CJ, Scheltens P (2005) Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study. Hum Brain Mapp 26(4):231–239
9. Reijneveld JC, Ponten SC, Berendse HW, Stam CJ (2007) The application of graph theoretical analysis to complex networks in the brain. Clin Neurophysiol 118(11):2317–2331
10. Wang J, Zuo X, He Y (2010) Graph-based network analysis of resting-state functional MRI. Front Syst Neurosci 4:16
11. Pervaiz U, Vidaurre D, Woolrich MW, Smith SM (2020) Optimising network modelling methods for fMRI. Neuroimage 211:116604
12. Qiao L, Zhang H, Kim M, Teng S, Zhang L, Shen D (2016) Estimating functional brain networks by incorporating a modularity prior. Neuroimage 141:399–407
13. Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19(4):1273–1302
14. Smith SM et al (2013) Functional connectomics from resting-state fMRI. Trends Cogn Sci 17(12):666–682
15. Schwarz AJ, McGonigle J (2011) Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. Neuroimage 55(3):1132–1146
16. van den Heuvel MP, de Lange SC, Zalesky A, Seguin C, Yeo BT, Schmidt R (2017) Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. Neuroimage 152:437–449
17. Huang S et al (2010) Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. Neuroimage 50(3):935–949
18. Tan Z, Yang P, Nehorai A (2013) Joint-sparse recovery in compressed sensing with dictionary mismatch. 2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) 248–251. https://doi.org/10.1109/CAMSAP.2013.6714054
19. Plis SM et al (2014) High-order interactions observed in multi-task intrinsic networks are dominant indicators of aberrant brain function in schizophrenia. Neuroimage 102:35–48
20. Macke JH, Opper M, Bethge M (2011) Common input explains higher-order correlations and entropy in a simple model of neural population activity. Phys Rev Lett 106(20):208102
21. Guo H, Liu L, Chen J, Xu Y, Jie X (2017) Alzheimer classification using a minimum spanning tree of high-order functional network on fMRI dataset. Front Neurosci 11:639
22. Zhou Y, Qiao L, Li W, Zhang L, Shen D (2018) Simultaneous estimation of low-and high-order functional connectivity for identifying mild cognitive impairment. Front Neuroinform 12:3
23. Chen X, Zhang H, Lee S-W, Shen D (2017) Hierarchical high-order functional connectivity networks and selective feature fusion for MCI classification. Neuroinformatics 15(3):271–284
24. Zhao F, Zhang H, Rekik I, An Z, Shen D (2018) Diagnosis of autism spectrum disorders using multi-level high-order functional networks derived from resting-state functional mri. Front Hum Neurosci 12:184
25. Chen X et al (2016) High-order resting-state functional connectivity network for MCI classification. Hum Brain Mapp 37(9):3282–3296
26. Zhang H et al (2016) Topographical information-based high-order functional connectivity and its application in abnormality detection for mild cognitive impairment. J Alzheimer's Dis 54(3):1095–1112
27. Jia X, Zhang H, Adeli E, Shen D (2017) Consciousness level and recovery outcome prediction using high-order brain functional connectivity network. International Workshop on Connectomics in Neuroimaging 17–24. First online: 02 Sept 2017
28. Guo T, Zhang Y, Xue Y, Qiao L, Shen D (2021) Brain function network: higher order vs. more discrimination. Front Neurosci 1033. https://doi.org/10.3389/fnins.2021.696639
29. Zhou Y, Zhang L, Teng S, Qiao L, Shen D (2018) Improving sparsity and modularity of high-order functional connectivity networks for MCI and ASD identification. Front Neurosci 12:959
30. Sun L, Xue Y, Zhang Y, Qiao L, Zhang L, Liu M (2021) Estimating sparse functional connectivity networks via hyperparameter-free learning model. Artif Intell Med 111:102004
31. Tzourio-Mazoyer N et al (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
32. Li W, Wang Z, Zhang L, Qiao L, Shen D (2017) Remodeling Pearson's correlation for functional brain network estimation and autism spectrum disorder identification. Front Neuroinform 11:55
33. Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):1–27
34. Dadi K et al (2019) Benchmarking functional connectome-based predictive models for resting-state fMRI. Neuroimage 192:115–134
35. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3(Mar):1157–1182

36. Gómez S, Jensen P, Arenas A (2009) Analysis of community structure in networks of correlated data. Phys Rev E 80(1):016114
37. Rubinov M, Sporns O (2011) Weight-conserving characterization of complex functional brain networks. Neuroimage 56(4):2068–2079
38. Jiang X, Zhang L, Qiao L, Shen D (2019) Estimating functional connectivity networks via low-rank tensor approximation with applications to MCI identification. IEEE Trans Biomed Eng 67(7):1912–1920
39. Xue Y, Zhang L, Qiao L, Shen D (2020) Estimating sparse functional brain networks with spatial constraints for MCI identification. PLoS ONE 15(7):e0235039
40. Greicius M (2008) Resting-state functional connectivity in neuropsychiatric disorders. Curr Opin Neurol 21(4):424–430
41. Albert MS et al (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement 7(3):270–279
42 Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica: J Econ Soc 37:424–438
43. Mclntosh A, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Hum Brain Mapp 2(1–2):2–22
44. Ramsey JD, Hanson SJ, Hanson C, Halchenko YO, Poldrack RA, Glymour C (2010) Six problems for causal inference from fMRI. Neuroimage 49(2):1545–1558
45. Smith SM et al (2011) Network modelling methods for FMRI. Neuroimage 54(2):875–891
46. Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D: Nonlinear Phenom 58(1–4):77–94
47 Steuer R, Kurths J, Daub CO, Weise J, Selbig J (2002) The mutual information: detecting and evaluating dependencies between variables. Bioinformatics 18(suppl_2):S231–S240
48. Hamilton WL (2020) Graph representation learning. Synth Lect Artif Intell Mach Learn 14(3):1–159

Hui Su received the B.Sc. degree in Information and Computing Science from Northwest Minzu University in 2019. She is currently a postgraduate student in the School of Mathematics and Science of Liaocheng University. Her research focuses on machine learning and brain functional network analysis.

Limei Zhang received the B.Sc. and M.Sc. degree in Mathematics from Liaocheng University in 2001 and 2007, respectively. In 2012, she received her Ph.D. degree from the Department of Computer Science & Engineering, Nanjing University of Aeronautics and Astronautics. She is currently an associate professor at Liaocheng University. Her research interests focus on pattern recognition and machine learning.

Lishan Qiao received the B.Sc. degree in Mathematics from Liaocheng University in 2001. He received the M.Sc. degree in Applied Mathematics from Chengdu University of Technology in 2004. In 2010, he received the Ph.D. degree in Computer Science, at the Department of Computer Science & Engineering, Nanjing University of Aeronautics and Astronautics. Currently, he is a Professor at the School of Mathematics Science, Liaocheng University, and his research interests pattern recognition, machine learning, and functional brain network analysis.

Mingxia Liu received the B.S. degree and M.S. degree from the School of Communication, Shandong Normal University in 2003 and 2006, and her Ph.D. degree from the Department of Computer Science & Engineering, Nanjing University of Aeronautics and Astronautics in 2015. From May 2014 to January 2017, she worked as a research assistant and postdoctoral research assistant at the University of North Carolina at Chapel Hill. Her research interests include machine learning, pattern recognition and medical image analysis.