Fusion of ULS Group Constrained High- and Low-Order Sparse Functional Connectivity Networks for MCI Classification

Yang Li[1] · Jingyu Liu[1] · Ziwen Peng[2,3] · Can Sheng[4,5] · Minjeong Kim[6] · Pew-Thian Yap[7] · Chong-Yaw Wee[8] · Dinggang Shen[7]

Abstract

Functional connectivity networks, derived from resting-state fMRI data, have been found as effective biomarkers for identifying mild cognitive impairment (MCI) from healthy elderly. However, the traditional functional connectivity network is essentially a low-order network with the assumption that the brain activity is static over the entire scanning period, ignoring temporal variations among the correlations derived from brain region pairs. To overcome this limitation, we proposed a new type of sparse functional connectivity network to precisely describe the relationship of temporal correlations among brain regions. Specifically, instead of using the simple pairwise Pearson's correlation coefficient as connectivity, we first estimate the temporal low-order functional connectivity for each region pair based on an ULS Group constrained-UOLS regression algorithm, where a combination of ultra-least squares (ULS) criterion with a Group constrained topology structure detection algorithm is applied to detect the topology of functional connectivity networks, aided by an Ultra-Orthogonal Least Squares (UOLS) algorithm to estimate connectivity strength. Compared to the classical least squares criterion which only measures the discrepancy between the observed signals and the model prediction function, the ULS criterion takes into consideration the discrepancy between the weak derivatives of the observed signals and the model prediction function and thus avoids the overfitting problem. By using a similar approach, we then estimate the high-order functional connectivity from the low-order connectivity to characterize signal flows among the brain regions. We finally fuse the low-order and the high-order networks using two decision trees for MCI classification. Experimental results demonstrate the effectiveness of the proposed method on MCI classification.

Keywords High-order network · Low-order network · Mild cognitive impairment · Ultra-least squares · Computer-aided detection and diagnosis

---

Yang Li, Jingyu Liu and Ziwen Peng contributed equally to this project.

✉ Jingyu Liu
liujingyu@buaa.edu.cn

✉ Dinggang Shen
dgshen@med.unc.edu

[1] School of Automation Sciences and Electrical Engineering, Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing Advanced Innovation Center for Big Date-based Precision Medicine, Beihang University, 37 XueYuan Road, HaiDian District, Beijing, China

[2] Shenzhen Kangning Hospital, Shenzhen University School of Medicine, Shenzhen, China

[3] College of Psychology and Sociology, Shenzhen University, Shenzhen, China

[4] Department of Neurology, XuanWu Hospital of Capital Medical University, Beijing, China

[5] Center of Alzheimer's Disease, Beijing Institute for Brain Disorders, Beijing, China

[6] Department of Computer Science, University of North Carolina at Greensboro, Greensboro, NC 27402-6170, USA

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

[8] Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore

Introduction

Alzheimer's disease (AD), characterized by the progressive decline in episodic memory, reasoning and other cognitive domains (Alzheimer's Association 2015), is considered as the most common form of dementia. It has been reported that AD incidence rate increases exponentially with aging (Ziegler-Graham et al. 2008) and 1 in 85 persons worldwide will suffer from the disease by 2050 (Brookmeyer et al. 2007). The latest statistics show that AD is the fifth leading cause of death for people over 65 years old and the sixth leading cause of death in the United States (Alzheimer's Association 2015). Recently, however, there is no effective treatment to stop the damage to neurons leading to the clinical symptoms of AD (Alzheimer's Association 2015), achieving early diagnosis at the stage of MCI is of great values. Mild cognitive impairment (MCI) is considered as the transitional state between normal senility and AD (Petersen et al. 2001; Gauthier et al. 2006; Li et al. 2019b). Approximately 10% to 15% of MCI patients deteriorate to AD every year, and more than half of them develop AD within 5 years (Petersen et al. 2001; Gauthier et al. 2006). Due to this high conversion rate, it is thus crucial to accurately identify MCI so that appropriate actions can be taken to slow down the progress of the disease. However, MCI is difficult to diagnose because of its relatively subtle symptoms (Eshkoor et al. 2015). In this case, many researchers dedicated to the diagnosis of AD and MCI with the aid of neuroimaging techniques (Davatzikos et al. 2011; Suk et al. 2014; Huang et al. 2010; Liu et al. 2014).

Neuroimaging techniques, such as Magnetic Resonance Imaging (MRI) (Fan et al. 2008; Hu et al. 2016; Cuingnet et al. 2011) and Magnetoencephalography (MEG) (Amezquita-Sanchez et al. 2016), are considered to be a powerful tool for classification of neurodegenerative diseases, especially for MCI and AD (Josef Golubic et al. 2017; Sandanalakshmi and Sardius 2016). On the other hand, resting-state functional Magnetic Resonance Imaging (rs-fMRI), which employs the blood-oxygenation-level-dependent (BOLD) signal as neurophysiological index, has been used recently for the early diagnosis of MCI before the appearance of clinical symptoms (McKenna et al. 2016; Wee et al. 2016; Khazaee et al. 2017). The correlation between BOLD signals of two brain regions is regarded as the functional connectivity between that region pair (Greicius 2008). The functional connectivity or temporal correlations among all regions within the brain (Van Den Heuvel and Pol 2010) are often characterized through functional connectivity networks using the graph theory (Fornito et al. 2010; Lee et al. 2017; Li et al. 2013). The differences between normal and damaged functional connectivity networks caused by pathological attacks can be considered as the biomarkers to study the pathological underpinnings of MCI (Jie et al. 2014; Li et al. 2014; Qi et al. 2010; Chand et al. 2017; Wee et al. 2014).

The most classical functional connectivity network modeling approach is based on pairwise Pearson's correlation coefficient (Wee et al. 2012a; Power et al. 2011). The correlation-based methods are easy to understand and have less computational complexity. However, in fact, brain regions are connected with only a limited number of regions, not all brain regions (Liao et al. 2017). Thus, the correlation-based networks are inconformity with the sparse nature of actual brain networks (Lee et al. 2011; Li et al. 2014). To overcome this limitation, some sparse modeling approaches have been proposed to construct sparse functional connectivity network based on whole brain connectivity information, i.e., identifying a small number of connections from the whole brain dense connections. For example, Lee et al. (2011) constructed sparse brain networks based on a l_1-norm regularized linear regression model, and the difference between the modular structures was derived from the sparse brain networks of autism spectrum disorders and pediatric control subjects, respectively. Rosa et al. (2015) proposed a sparse functional connectivity network modeling method based on l_1-norm regularized maximum likelihood estimation and Gaussian graphical models, and discovered the discriminative changes in brain networks between major depressive disorders and normal controls (NCs). These sparse brain network modeling methods are considered more sensitive to reject the spurious connections than the correlation-based methods (Lee et al. 2011; Rosa et al. 2015). In addition, to minimize the influence of inter-subject variability, some recent approaches have also adopted $l_{2,1}$-norm penalization to ensure the consistency of non-zero connections across subjects (Li et al. 2018d, c; Wee et al. 2014). For example, a sparse causality model aid by $l_{2,1}$-norm penalization was proposed to detect causal interactions in multivariate time series (Haufe et al. 2008). The group constrained causality model yields a better performance than some classical methods including Granger Causality, Ridge Regression, and Lasso in the causal structure detection. Ryali et al. (2012) adopted a sparse partial correlation method which combines l_1- and l_2-norm penalization to estimate the functional connectivity between brain regions in fMRI data, where l_1-norm penalty provides sparse interpretable solutions and l_2-norm penalty improves the sensitivity of the model. Thus, this method can provide a more accurate brain connection estimation result. Furthermore, in order to explore study the directional causal interactions among brain regions, a cross-spectral density connectivity estimation method was proposed to estimate the effective connection from the whole brain fMRI data (Lennartz et al. 2018). With a reduced dependence on hemodynamic variability, this method may produce more reliable connectivity estimation results. In addition, a group constrained effective connectivity inference method was further proposed for MCI identification (Li et al. 2018d). This approach combines the $l_{2,1}$-norm penalization with the time-dependent effective connectivity estimation, and thus can generate effective brain connectivity networks with the consistent topology among subjects. In all the

aforementioned modeling methods, the vertices of brain networks correspond to the brain regions and the edges correspond to the correlation among brain regions, thus producing the so-called low-order network (Chen et al. 2016). It is notable that the low-order correlation network is normally calculated over the whole time series, without considering potential temporal variations among the correlations (Chen et al. 2016). However, the brain activities are indeed not static across the entire scanning period and the correlations among brain regions vary across time (Allen et al. 2014; Hutchison et al. 2013; Liu et al. 2017). Therefore, the conventional methods, which ignore temporal variations among correlations, may fail to diagnose MCI accurately (Chen et al. 2017). Recently, a high-order network modeling method has been proposed to preserve the dynamic correlation information neglected in the conventional methods (Chen et al. 2016), where the vertices of high-order network correspond to the brain region pairs, and the edges correspond to the correlation between the brain region pairs. The high-order network modeling approach takes the temporal variations among correlations into account, and discovers discriminative dynamic correlation information for MCI classification. However, the existing high-order network modeling approach (Chen et al. 2016) is derived using the pairwise Pearson's correlation-based low-order functional connectivity, which is inconsistent with the sparse nature and small-world characteristics of most biological networks (Supekar et al. 2008).

To overcome this deficiency, we propose a novel high-order network modeling method that utilizes a specially designed unified sparse regression framework. Specifically, we construct the high-order functional connectivity networks by using a novel ULS Group constrained topology structure detection algorithm which is accompanied with an Ultra-Orthogonal Least Squares (UOLS) algorithm. The former, which consists of an ultra-least squares (ULS) criterion and a Group constrained topology structure detection algorithm, is applied to detect the topology of functional connectivity networks. The latter, which consists of the ULS criterion and an Orthogonal Least Squares (OLS) algorithm (Li et al. 2019a), is employed to estimate the strength of functional connectivity. The rationale of using the ULS criterion in our proposed method is that, besides extracting the classical dependent relation between the fMRI time series of a region pair, it further extracts the dependent relation of the associated weak derivatives, and thus avoids the overfitting problem which is common in the conventional least squares criterion (Guo et al. 2016). The weak derivative, which can be calculated for all integrable functions, is a measure describing interconnections among the data points, where the definition of the weak derivative can be found at Appendix. In other words, the classical least squares (LS) criterion based method does not take into consideration the continuity between fMRI time series. The

absence of the connection information between data points may lead to an inaccurate model structure. To overcome this limitation, we integrate the ULS criterion, which can describe the relationship among data points, into our proposed framework. Additionally, different from the traditional sparse regression algorithm with a l_1-norm penalization which leads to different network structures at individual level (Lee et al. 2011), the Group constrained topology structure detection algorithm in our modeling method utilizes a $l_{2,1}$-norm penalization to encourage an identical network topology among subjects. Identical network topology ensures an easier comparison between subjects, thus achieving a better generalization performance in brain disease classification (Wee et al. 2014; Zhu et al. 2014).

The high-order network is able to encode the temporal variations of correlation between brain regions, while it is unable to characterize the holistic correlation calculated based on the whole time series as in the low-order correlation network. Therefore, in order to incorporate both the low-order correlation and the temporally dynamic information encoded in the high-order correlation for better classification performance, we first construct a decision tree (DCT) for each type of correlations and then fused their classification scores together to provide the final classification decision. The fused DCT model takes into account not only the correlation derived based on the whole time series, but also the temporal variations between correlations. We have compared our proposed framework (i.e., fusion of high-order and low-order functional connectivity networks) with the state-of-the-art methods on the same dataset, and the experimental results demonstrate the superiority of the proposed framework for MCI classification.

In summary, the main contributions of our proposed framework are three-fold:

- Taking into consideration the discrepancy between the weak derivatives of the observed signals and the model prediction function during functional connectivity network estimation;
- Derive the functional networks using sparse regression framework to preserve the sparse nature of brain networks while enforcing identical network topology among all subjects to ease the between-subject comparison;
- Simultaneously considering the dynamic correlation information and the holistic correlation information for MCI classification by fusing the high-order and low-order networks.

The rest of the paper is organized as follows. "Materials and Methods" section furnishes information on the data

acquisition and post-processing, followed by the proposed framework for the construction and fusion of low- and high-order functional connectivity networks for MCI classification. Then, we evaluate and discuss the performance of the proposed framework in "Results and Discussions" section. Finally, we conclude this paper in "Conclusion" section.

Materials and Methods

Proposed Framework

In Fig. 1, we provide the proposed MCI classification framework, based on the fusion of high-order and low-order sparse functional connectivity networks. "Data Acquisition and Preprocessing" section provides the details of data acquisition and preprocessing; In "Low-Order Functional Connectivity Networks" section and "High-Order Functional Connectivity Networks" section, we construct the low- and high-order functional connectivity networks, respectively. "Feature Extraction, Selection, and Classification" section gives the process of the feature extraction, selection, and fusion classification. Finally, we summarize the methodology in "Summary of Methodology" section.

Data Acquisition and Preprocessing

This study is approved by the local ethical committee and all of participants voluntarily participated in this study with informed consents. The participants are recruited via advertisements in local newspapers and media. All the participants carry on regular neuropsychological assessment, including the mini-mental state examination (MMSE) (Van Patten et al. 2018), the hospital anxiety and depression Scale (HAD) (McKenzie et al. 2018), and the lawton's instrumental activities of daily living (IADL) (Mao et al. 2018). All the participants are also evaluated using the Clinical Dementia Rating scale (CDR) (Das et al. 2018). MCI patients are diagnosed according to the criteria proposed by Petersen (2004), which are: (1) memory loss complaint corroborated by an informant; (2) objective cognitive impairment in single or multiple domains, adjusted for age and education; (3) preserved general cognitive function; (4) failure to meet the criteria for dementia; (5) the Clinical Dementia Rating (CDR) score is 0.5. The inclusion criteria of NCs were: (1) no complaint of memory loss; (2) CDR score is 0; (3) no severe visual or auditory impairment.

Twenty-eight MCI patients and thirty-three demographically matched NCs are selected from the participants. All subjects are scanned using a standard echo-planar imaging (EPI) sequence on a 3 Tesla Siemens TRIO scanner with the following parameters: TR = 3000 ms, TE = 30 ms, acquisition matrix = 74 × 74, 45

slices, and voxel thickness = 3 mm. One-hundred and thirty resting-state fMRI volumes are acquired. Standard preprocessing pipeline of the fMRI images is performed using Statistical Parametric Mapping 8 (SPM8) software package, which includes removal of first 10 fMRI volumes, slice timing correction, head-motion correction, regression of nuisance signals (ventricle, white matter, global signal, and head-motion with Friston's 24-parameter model (Wee et al. 2016)), signal de-trending. Next, the brain space was parcellated into 90 region-of-interests (ROIs) based on the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al. 2002). Finally, we adopt a band-pass filter (0.01–0.08 Hz) to eliminate the effects of low- and high-frequency noise.

Low-Order Functional Connectivity Networks

ULS Group Constrained Topology Structure Detection

Suppose that there are N subjects ($N = 61$), each subject includes the total of M ROIs ($M = 90$), y_m^n denotes the ROI time series of the m-th ROI from the n-th subject. Compared to the widely-used Lasso algorithm (Lee et al. 2011) with a l_1-norm penalization, which may generate different network topologies for different subjects, the Group constrained topology structure detection algorithm with a $l_{2,1}$-norm penalization minimize this inter-subject variability by encouraging an identical network topology across subjects (Wee et al. 2014). The objective function of the Group constrained topology structure detection algorithm is given by

$$f(\boldsymbol{\Theta}_m) = \sum_{n=1}^{N} \left\| \boldsymbol{y}_m^n - \boldsymbol{A}_m^n \boldsymbol{\theta}_m^n \right\|_2^2 + \lambda \|\boldsymbol{\Theta}_m\|_{2,1} \qquad (1)$$

where $\boldsymbol{A}_m^n = \left[\boldsymbol{y}_1^n, \ldots, \boldsymbol{y}_{m-1}^n, \boldsymbol{y}_{m+1}^n, \ldots, \boldsymbol{y}_M^n \right]$ is a matrix which includes all ROIs time series except the m-th ROI, $\lambda > 0$ is the regularization parameter that controls the sparsity level of the regression model, $\boldsymbol{\theta}_m^n = \left[\theta_1^n, \ldots, \theta_{m-1}^n, \theta_{m+1}^n, \ldots, \theta_M^n \right]^T$ is the weight vector that describes the relationship between the m-th ROI and the other eighty-nine ROIs for n-th subject, and $\boldsymbol{\Theta}_m = \left[\boldsymbol{\theta}_m^1, \boldsymbol{\theta}_m^2, \ldots, \boldsymbol{\theta}_m^N \right]$ is the weight matrix for all subject. It should be noted that each row of the matrix $\boldsymbol{\Theta}_m$ represents the coefficient vector of one ROI for all subjects, while each column of the matrix $\boldsymbol{\Theta}_m$ represents the coefficient vector of all ROIs for one subject (i.e. $\boldsymbol{\theta}_m^n$). $\|\boldsymbol{\Theta}_m\|_{2,1}$ is the summation of the l_2-norm of each row in the matrix $\boldsymbol{\Theta}_m$, not the summation of the l_2-norm of each column in $\boldsymbol{\Theta}_m$ (i.e. $\sum_{n=1}^{N} \left\| \boldsymbol{\theta}_m^n \right\|_2$). Therefore, Eq. (1) encourages consistent non-zero elements for the given ROI across all subjects and cannot be rewritten as $f(\boldsymbol{\theta}_m^n) = \left\| \boldsymbol{y}_m^n - \boldsymbol{A}_m^n \boldsymbol{\theta}_m^n \right\|_2^2 + \lambda \left\| \boldsymbol{\theta}_m^n \right\|_2$ ($n = 1, 2, \ldots, N$). Constructing the sparse network structure can be considered as an optimization problem, i.e., minimizing the above objective function.
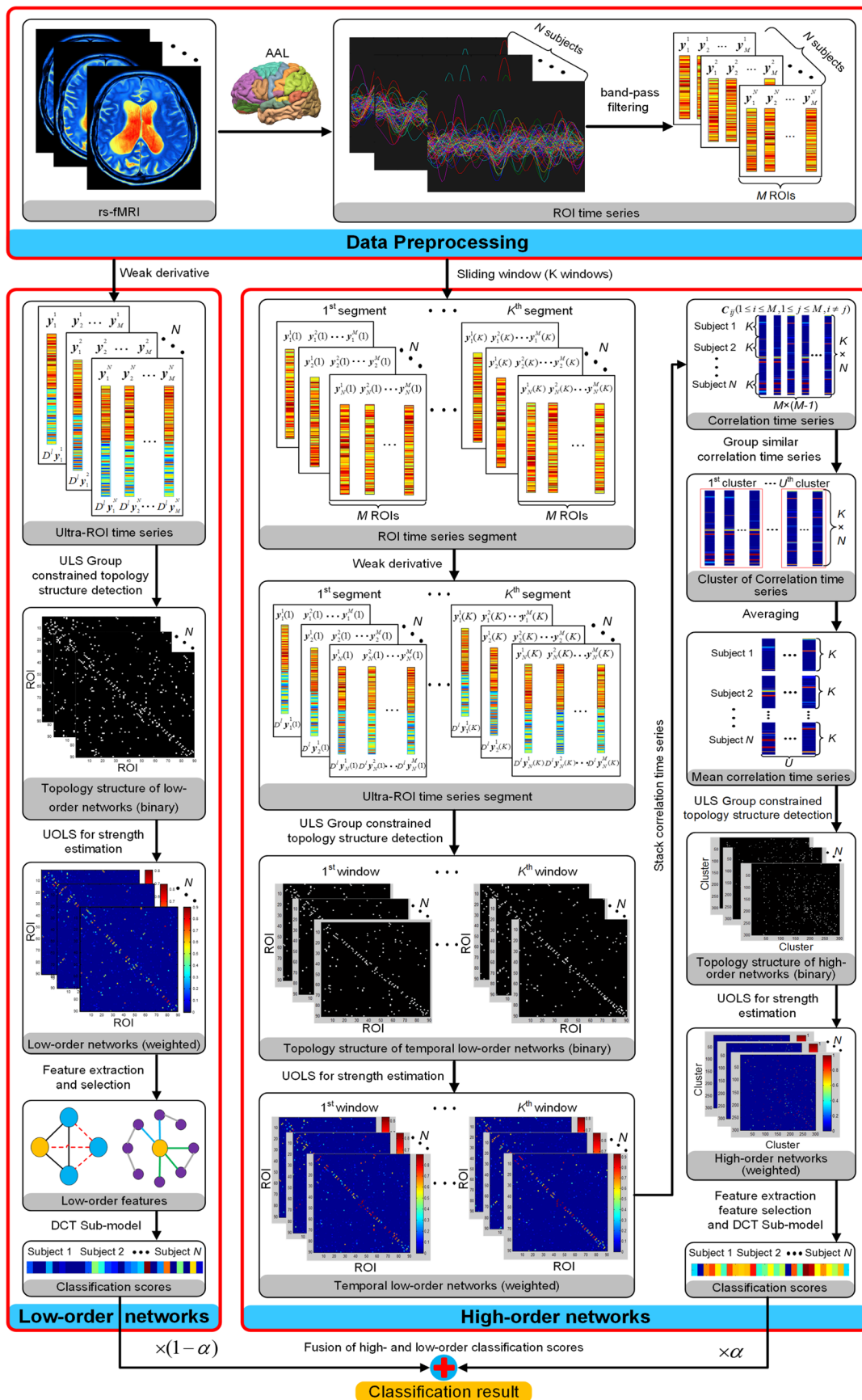
Fig. 1 The schematic diagram of the proposed framework

In order to obtain a more accurate evaluation standard for the model fitness, an ULS criterion is integrated into the Group constrained topology structure detection algorithm to generate an ULS Group constrained topology structure detection algorithm. The ULS criterion, which considers the discrepancy between the weak derivatives of the observed signals and the model prediction function and thus avoids the overfitting problem, is defined by

$$J_{ULS} = \left\| \tilde{\boldsymbol{y}}_m^n - \tilde{\boldsymbol{A}}_m^n \boldsymbol{\theta}_m^n \right\|_2^2, \qquad (2)$$

where $\tilde{\boldsymbol{y}}_m^n = \left[\left(\boldsymbol{y}_m^n\right)^T, \left(D^1 \boldsymbol{y}_m^n\right)^T, \left(D^2 \boldsymbol{y}_m^n\right)^T, ..., \left(D^L \boldsymbol{y}_m^n\right)^T \right]^T$, generated by connecting the original ROI time series $\boldsymbol{y}_m^n$ with its weak derivatives $D^l \boldsymbol{y}_m^n (l = 1, 2, ..., L)$, is the ultra-ROI time series of the m-th ROI from the n-th subject, and $\tilde{\boldsymbol{A}}_m^n = \left[\tilde{\boldsymbol{y}}_1^n, ..., \tilde{\boldsymbol{y}}_{m-1}^n, \tilde{\boldsymbol{y}}_{m+1}^n, ..., \tilde{\boldsymbol{y}}_M^n\right]$ is a matrix consisting of all ultra-ROI time series except the m-th self ROI (the details of the ULS criterion and weak derivatives are provided in Appendix). Specifically, as discussed in Appendix, the ULS criterion can be incorporated into the Group constrained topology structure detection algorithm by replacing the original ROI time series $\boldsymbol{y}_m^n$ with the ultra-ROI time series $\tilde{\boldsymbol{y}}_m^n$ in Eq. (1). Therefore, the ULS Group constrained topology structure detection algorithm is adopted to detect a more accurate network topology via the following objective function:

$$f(\boldsymbol{\Theta}_m) = \sum_{n=1}^N \left\| \tilde{\boldsymbol{y}}_m^n - \tilde{\boldsymbol{A}}_m^n \boldsymbol{\theta}_m^n \right\|_2^2 + \lambda \|\boldsymbol{\Theta}_m\|_{2,1}. \qquad (3)$$

By separating the weak derivative part $D^l \boldsymbol{y}_m^n (l = 1, 2, ..., L)$ from the original ROI time series part $\boldsymbol{y}_m^n$, Eq. (3) can be rewritten as

$$f(\boldsymbol{\Theta}_m) = \sum_{n=1}^N \|\boldsymbol{y}_m^n - \boldsymbol{A}_m^n \boldsymbol{\theta}_m^n\|_2^2 + \sum_{n=1}^N \sum_{l=1}^L \|D^l \boldsymbol{y}_m^n - D^l \boldsymbol{A}_m^n \boldsymbol{\theta}_m^n\|_2^2 + \lambda \|\boldsymbol{\Theta}_m\|_{2,1} \qquad (4)$$

It should be noted that the estimated coefficients matrix $\boldsymbol{\Theta}_m$ based on Eq. (4) cannot be regarded as the functional connectivity strengths, because they are biased as the result of group-constrained sparse penalization (Li et al. 2018d). Particularly, some of the coefficients are even negative, leading to difficulty in interpreting and analyzing the functional connectivity network. Therefore, the estimated coefficients matrix $\boldsymbol{\Theta}_m$ based on Eq. (4) are only treated as the network topology indicator. The ROIs with non-zero elements in the coefficient matrix $\boldsymbol{\Theta}_m$ are considered to have functional connections with the target m-th ROI, while the zero element in $\boldsymbol{\Theta}_m$ indicates the non-connection between the corresponding ROI with the target m-th ROI.

Strength Estimation of Sparse Functional Connectivity Networks Via UOLS

Supposing that P ROIs have been found correlated to the target m-th ROI based on the ULS Group constrained

topology structure detection algorithm in "ULS Group Constrained Topology Structure Detection" section, the associated the ultra-ROI time series of these ROIs

$$\boldsymbol{y}_{m_p}^n = \left[\left(\boldsymbol{y}_{m_p}^n\right)^T, \left(D^1 \boldsymbol{y}_{m_p}^n\right)^T, \left(D^2 \boldsymbol{y}_{m_p}^n\right)^T, ..., \left(D^L \boldsymbol{y}_{m_p}^n\right)^T \right]^T (p = 1, 2, ..., P)$$

are selected as the candidate time series to be used for the functional connectivity strength estimation, while the ultra-ROI time series of other ROIs are discarded. Then, we utilize an UOLS algorithm to estimate the functional connectivity strengths between these P ROIs with the target m-th ROI (Li et al. 2018d). The UOLS algorithm is the combination of the ULS criterion (Eq. (2)) and the OLS algorithm (Guo et al. 2016). According to Appendix, the ULS criterion can be integrated into the OLS algorithm by incorporating the weak derivatives into the original ROI time series. Therefore, we can obtain the UOLS algorithm by replacing the original ROI time series with the ultra-ROI time series (containing the original ROI time series and its weak derivatives) in the OLS algorithm. The detailed procedure of the UOLS algorithm can be founded in Table 1. In the UOLS algorithm, the functional connectivity strength is estimated in a stepwise orthogonal forward procedure. The value of $MaxUerr\left(\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_{m_p}^n\right)$ is regarded as the functional connectivity strength between the target m-th ROI and candidate p-th ROI.

In summary, we first repeat the "ULS Group Constrained Topology Structure Detection" section M times to detect the network topology for all subjects (each time a different ROI for all subjects will be selected to be the target ROI in Eq. (4)). Then, the network connectivity strengths for all subjects are estimated by repeating the "Strength Estimation of Sparse Functional Connectivity Networks via UOLS" section $M \times N$ times. Each time a different ROI for one subject will be regarded as the target ROI. In this way, we can obtain a low-order ULS Group constrained-UOLS network for each subject.

High-Order Functional Connectivity Networks

Ultra-ROI Time Series Segment

The first step in the high-order network construction is to employ a sliding window to partition each ROI time series into multiple overlapping segments. For the one ROI time series containing Z temporal image volumes, $\boldsymbol{y}_m^n$, the total number of time series segments by using the sliding window can be computed as $K = [(Z - S)/r] + 1$, where S is the size of the sliding window and r denotes the step size between adjacent windows. Letting $\boldsymbol{y}_m^n(k)$ be the k-th segment generated from $\boldsymbol{y}_m^n$, for the n-th subject, the k-th segments of all M ROIs can be represented in a matrix form as $\boldsymbol{Y}^n(k) = \left[\boldsymbol{y}_1^n(k), \boldsymbol{y}_2^n(k), ..., \boldsymbol{y}_M^n(k)\right] \epsilon R^{S \times M}$. Further, the set of the k-th segments for all subjects and all ROIs can be

Table 1 Strength estimation of functional connectivity via UOLS algorithm

---

Input: Ultra-ROI time series of m-th ROI $\tilde{\boldsymbol{y}}_m^n$ and those of its correlated ROIs $\left\{\tilde{\boldsymbol{y}}_{m_p}^n | p = 1, 2, ..., P\right\}$

Output: $MaxUerr(\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_{m_p}^n)$, $p = 1, 2, ..., P$ $(MaxUerr(\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_{m_p}^n)$ is the connectivity strength between the m-th ROI and candidate p-th ROI.)

01: **initialization:** $index = \{1, 2, \cdots, P\}$ ($index$ denotes the candidate ROIs.)

02: **while** $index \neq [\ ]$ **do**

03:     **for all** $p \in index$ **do**

04:         $Uerr(\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_{m_p}^n) = \frac{<\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_p^n>^2}{<\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_m^n><\tilde{\boldsymbol{y}}_p^n, \tilde{\boldsymbol{y}}_p^n>}$ $(<\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_p^n>$ is the inner product of vectors $\tilde{\boldsymbol{y}}_m^n$ and $\tilde{\boldsymbol{y}}_p^n$)

05:     **end for**

06:     $b = \text{argmax}_{p \in index} \left\{ Uerr(\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_{m_p}^n) | p \in index \right\}$

07:     $MaxUerr(\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_{m_b}^n) = Uerr(\tilde{\boldsymbol{y}}_m^n, \tilde{\boldsymbol{y}}_{m_b}^n)$

08:     $index = index \setminus \{b\}$

09:     **for all** $p \in index$ **do**

10:         $\tilde{\boldsymbol{y}}_{m_p}^n = \tilde{\boldsymbol{y}}_{m_p}^n - \frac{<\tilde{\boldsymbol{y}}_{m_p}^n, \tilde{\boldsymbol{y}}_{m_b}^n>}{<\tilde{\boldsymbol{y}}_{m_b}^n, \tilde{\boldsymbol{y}}_{m_b}^n>} \tilde{\boldsymbol{y}}_{m_b}^n$

11:     **end for**

12: **end while**

---

represented by $Y(k) \epsilon R^{S \times M \times N}$. Thus, by applying a sliding window to all fMRI time series, we can obtain K sets of time series segments $Y(k)$. Then, we generate the ultra-ROI time series segment $\tilde{Y}(k) = \left[(Y(k))^T, (D^1 Y(k))^T, (D^2 Y(k))^T, ..., (D^L Y(k))^T \right]^T$ by connecting the original time series segment $Y(k)$ with its weak derivatives $D^l Y(k) (l = 1, 2, ..., L)$, where '$T$' in the $(Y(k))^T$ denotes the transpose of interchanging the first and second dimensions for the three-dimensional tensor. The details of the weak derivative and the ULS criterion are provided in Appendix.

Construction of Temporal Low-Order Functional Connectivity Networks

For each ultra-ROI time series segment $\tilde{Y}(k)$, we apply the ULS Group constrained topology structure detection algorithm across subjects to detect the functional connectivity network structure and then apply an UOLS at the individual level to re-estimate the connectivity strength of each derived connection, producing a total of $K \times N$ temporal low-order functional connectivity networks. Taking each ROI with $\left\{ \boldsymbol{y}_m^n(k) \right\}$ as the vertex and $\left\{ C_{m_1, m_2}^n(k) \right\} (1 \leq m_1 \leq M, 1 \leq m_2 \leq M, m_1 \neq m_2)$ as the connectivity strength for each pair of vertices, the temporal low-order functional connectivity network can be expressed

as $\boldsymbol{G}^n(k) = \left(\{ \boldsymbol{y}_m^n(k) \}, C_{m_1, m_2}^n(k) \right) (k = 1, 2, ..., K)$, where $\left\{ C_{m_1, m_2}^n(k) \right\}$ is the correlation between the m_1-th ROI and m_2-th ROI of the n-th subject in the k-th window. A larger value of $\left\{ C_{m_1, m_2}^n(k) \right\}$ indicates a stronger connection between the m_1-th ROI and m_2-th ROI in the k-th window.

The next step is to extract the correlation time series from the temporal low-order functional connectivity networks. For the n-th subject, the correlation time series of each ROI-pair (m_1, m_2), $\boldsymbol{C}_{m_1, m_2}^n = \left[C_{m_1, m_2}^n(1), C_{m_1, m_2}^n(2), ..., C_{m_1, m_2}^n(K) \right]^T \epsilon R^K$, can be obtained by concatenating all $C_{m_1, m_2}^n(k) (k = 1, 2, ..., K)$. Different from $\boldsymbol{y}_m^n$ that represents the time series of a ROI, $\boldsymbol{C}_{m_1, m_2}^n$ characterizes the variations of the correlation of ROI-pair (m_1, m_2) across time (Chen et al. 2016). Considering the low-order network is asymmetric, the total number of correlation time series $\left\{ \boldsymbol{C}_{m_1, m_2}^n | 1 \leq m_1 \leq M, 1 \leq m_2 \leq M, m_1 \neq m_2 \right\}$ for each subject is $M(M - 1)$.

Construction of High-Order Functional Connectivity Networks

The final step is to construct high-order functional connectivity networks based on the correlation time series. For the set of all correlation time series $\left\{ \boldsymbol{C}_{m_1, m_2}^n | 1 \leq m_1 \leq M, 1 \leq m_2 \leq M, m_1 \neq m_2, 1 \leq n \leq N \right\}$,

we employ the ULS Group constrained topology structure detection algorithm to detect the high-order network structure and then employ the UOLS to estimate the high-order connectivity strength, establishing a total of N high-order networks with identical network structure. Specifically, for each pair of correlation time series ($\left\{ \boldsymbol{C}^n_{m_1,m_2} \right\}$, $\left\{ \boldsymbol{C}^n_{m_3,m_4} \right\}$), $\left\{ E^n_{(m_1,m_2),(m_3,m_4)} \right\}$ is used to represent the high-order correlation between them. In other words, $\left\{ E^n_{(m_1,m_2),(m_3,m_4)} \right\}$ is the high-order connectivity strength between the ROI-pairs (m_1, m_2) and (m_3, m_4). By considering ROI-pair (m_1, m_2) with $\left\{ \boldsymbol{C}^n_{m_1,m_2} \right\}$ as vertex and $\left\{ E^n_{(m_1,m_2),(m_3,m_4)} \right\}$ as the weights of edges, the high-order network can be expressed as $\boldsymbol{G}^n = \left(\left\{ \boldsymbol{C}^n_{m_1,m_2} \right\}, \left\{ E^n_{(m_1,m_2),(m_3,m_4)} \right\} \right)$. Therefore, a high-order network is devoted to describe the relationship of temporal correlations among brain regions.

However, there is an obvious limitation about the high-order network, i.e., the scale of the high-order network is too large. As mentioned above, the number of vertices $\left\{ \boldsymbol{C}^n_{m_1,m_2} \right\}$ is $M(M-1)$, thus the number of edges is proportional to M^4. This large scale of a high-order network leads to a large amount of computation complexity and the poor generalization performance. To overcome this limitation, the Ward's hierarchical grouping (Chen et al. 2014) is applied to group the correlation time series into different clusters. Specifically, when grouping the correlation time series for all subjects into different clusters, we need to ensure the consistency of the clustering results between different subjects. Thus, the correlation time series $\left\{ \boldsymbol{C}^n_{m_1,m_2} \right\}$ for all subjects ($n = 1, 2, \ldots, N$) are first connected together $\boldsymbol{C}_{m_1,m_2} = \left[\left(\boldsymbol{C}^1_{m_1,m_2} \right)^T, \left(\boldsymbol{C}^2_{m_1,m_2} \right)^T, \ldots, \left(\boldsymbol{C}^N_{m_1,m_2} \right)^T \right]^T \epsilon R^{KN \times 1}$. Then, we apply the Ward's hierarchical grouping (Chen et al. 2014) to group $\left\{ \boldsymbol{C}_{m_1,m_2} \right\}$ into different U clusters and obtain the grouping results $\{ \Omega_1, \Omega_2, \ldots, \Omega_U \}$, where ROI-pair is $(m_1, m_2) \in \Omega_u$ if $\boldsymbol{C}_{m_1,m_2}$ belongs to the u-th cluster. The correlation time series that belongs to the same cluster have the similar time variation. Then, the mean correlation time series of the u-th cluster for the n-th subject can be calculated as follows:

$$\overline{\boldsymbol{C}}^n_u = \frac{\sum_{(m_1,m_2) \in \Omega_u} \boldsymbol{C}^n_{m_1,m_2}}{|\Omega_u|}, \tag{5}$$

where $|\Omega_u|$ is the number of elements in Ω_u. Finally, the mean correlation time series of each cluster $\left\{ \overline{\boldsymbol{C}}^n_u \right\}$ is used as the vertices of the high-order network, instead of the correlation time series $\left\{ \boldsymbol{C}^n_{m_1,m_2} \right\}$. As a result, the scale of the high-order network can be significantly reduced and the generalization capability of the high-order network

can be significantly improved (Chen et al. 2016).

Feature Extraction, Selection, and Classification

Feature Extraction

The ULS Group constrained-UOLS sparse regression is an asymmetric network model. Thus, the low-order and high-order ULS Group constrained-UOLS networks are asymmetric networks. In this work, we extract four typical asymmetric network features from the low-order and high-order ULS Group constrained-UOLS networks, respectively, including the weighted clustering coefficient, betweenness centrality, in-degree, and out-degree (Rubinov and Sporns 2010; NeuroimageFagiolo 2007). Specifically, the weighted-clustering coefficient can quantify the probability that the neighbors of one ROI are connected to each other (Chen et al. 2016), and reflects the prevalence of clustered connectivity around this ROI (Rubinov and Sporns 2010). Betweenness centrality reflects the idea that the central nodes, which take part in many shortest paths from one ROI to another, consequently act as the important controls of information flow in the brain network (Rubinov and Sporns 2010). In-degree and out-degree are the most fundamental measures of the asymmetric network and most other measures are based on the degree of nodes (Bullmore and Sporns 2009). Therefore, we finally extract these four features from the low-order and high-order networks, which are widely used to describe the characteristics of brain networks (Chen et al. 2018, 2016; Khazaee et al. 2015, 2016). For a node i of an asymmetric network, the weighted clustering coefficient is defined by

$$WC_i = \frac{\sum_{j \neq i} \sum_{h \neq (i,j)} \left(c_{i,j}^{1/3} + c_{j,i}^{1/3} \right) \left(c_{i,h}^{1/3} + c_{h,i}^{1/3} \right) \left(c_{j,h}^{1/3} + c_{h,j}^{1/3} \right)}{2 \left[d_i(d_i - 1) - 2d_i^{\leftrightarrow} \right]}, \tag{6}$$

where $c_{i,j}$ is the connectivity strength from a node i to j, d_i represents the number of adjacent points of a node i, and $d_i^{\leftrightarrow}$ denotes the number of bilateral edges between i and its adjacent nodes. For a node i, the betweenness centrality is defined by

$$BC_i = \sum_{j \neq i, k \neq (i,j)} \frac{\rho_{jk}(i)}{\rho_{jk}}, \tag{7}$$

where ρ_{jk} is the number the of shortest paths from k to j, and $\rho_{jk}(i)$ denotes the number of the shortest paths from k and j that pass through i. The in-degree and out-degree of a node i are defined as the number of connections from its adjacent nodes to i, and the number of connections from i to its adjacent nodes, respectively.

In this section, four types of features were extracted from each ROI (in low-order network) or cluster (in high-order

network). Hence, we can obtain $4 \times M$ low-order features and $4 \times U$ high-order features from each subject ($M = 90$ and the parameter optimization of U is provided in "Impact of the Number of Clusters U" section). Then, we arrange these features into a low-order feature vector (with $4 \times M$ features) and a high-order feature vector (with $4 \times U$ features), respectively.

Feature Selection

It is inevitable that the features extracted from functional connectivity networks contain some irrelevant or redundant features for MCI classification. Those irrelevant or redundant features increase the running time of the learning algorithm and deteriorate the generalization performance of the classifier (Jain and Singh 2018). Thus, we input the low-order and high-order feature vectors obtained in "Feature Extraction" section into a feature selection filter to eliminate these irrelevant and redundant features. Specifically, for all features, the correlation coefficients between the features and the class labels of training samples are first calculated. The feature with a larger correlation coefficient is considered as a more discriminative feature for MCI classification. The feature with its coefficient larger than a predefined threshold ε_1 will be selected to form a feature set γ_1. Then, we apply the Relief algorithm (Zhang et al. 2016) to compute the weights for the features that do not belong to γ_1. The features with their weights larger than a predefined threshold ε_2 will be selected to form a feature set γ_2. Next, by treating the feature values as the corresponding classification scores for the training samples, we can compute the area under curve (AUC) for the features that do not belong to γ_1 and γ_2. The features whose AUC are larger than a predefined threshold ε_3 will be selected (denoted as γ_3). Finally, $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3$ will be the final feature vector for MCI classification.

Classification

Due to the limited sample size, in this work, a bagging-based nested 10-fold cross-validation scheme is adopted to determine the optimal parameter and evaluate the classification performance. The nested 10-fold cross-validation scheme contains two cross-validation loops: the inner cross-validation loop based on the training set (containing 9 fold subjects) is used to determine the optimal parameters including feature selection parameters ε_1 ([0.1, 0.15,···, 0.5]), ε_2 ([0.01, 0.02, ···, 0.1]) and ε_3 ([0.1, 0.15,···, 0.5]), and the outer cross-validation loop based on the test set (containing 1 fold subjects) is applied to evaluate the generalization performance of DCT model.

The bagging approach generates a strong classifier by combining a series of weak classifiers which are constructed based on the bootstrapped samples from the training dataset (Shah et al. 2015; Breiman 1996). It should be noted that the weak classifiers need to be relatively unstable, such that the classification results of them can change considerably even with a slightly different training dataset. Therefore, we adopt a typically unstable classifier (i.e. DCT) to construct the weak classifiers. Then, we combine the weak classifiers via the bagging approach to generate a strong classifier.

For each fold of the outer cross-validation loop, two DCT sub-models are constructed based on the low-order feature vector and the high-order feature vector selected in "Feature Selection" section, respectively. For the test data set, each DCT sub-model can provide a classification score for each subject, and we get the final classification score for each subject by combining the classification scores from two DCT sub-models with a certain fusion weight α.

To characterize performance comparison between different methods, we use four frequently used performance measures including accuracy (ACC), sensitivity (SEN), specificity (SPC), and AUC (Li et al. 2018b; Wang et al. 2017). The accuracy, sensitivity, and specificity are defined as follows (Li et al. 2017):

$$\mathrm{ACC} = \frac{TP + TN}{TP + FN + TN + FP} \tag{8}$$

$$\mathrm{SEN} = \frac{TP}{TP + FN} \tag{9}$$

$$\mathrm{SPC} = \frac{TN}{TN + FP} \tag{10}$$

where TP, FP, TN, and FN denote the true positive, false positive, true negative, and false negative, respectively. In order to obtain reliable evaluation results, we repeat the 10-fold cross-validation for 20 times and report our classification results as the average of these 20 repetitions.

Summary of Methodology

The proposed fusion of high- and low-order functional connectivity networks framework is summarized and listed as follows:

(1) **Data Acquisition and Preprocessing:**
  1) Scan the fMRI images of twenty-eight MCI patients and thirty-three NCs using a standard echo-planar imaging (EPI) sequence on a 3 Tesla Siemens TRIO scanner;
  2) Preprocess the fMRI images is performed using the SPM8 software package, including removal of first 10 fMRI volumes, slice timing correction, head-motion correction, regression of nuisance signals, signal detrending, and band-pass filtering.
(2) **The construction of the low-order networks:**
  1) Connect the ROIs time series with its weak derivatives to generate the ultra-ROI time series;

2) Detect the topology of low-order networks via the ULS Group constrained topology structure detection algorithm;

3) Apply an UOLS algorithm to estimate the functional connectivity strength of the low-order networks.

(3) **The construction of the high-order networks:**

1) Apply a sliding window to partition the original ROIs time series into multiple overlapping segments of subsequence;

2) Generate the ultra-ROI time series segment by connecting the each ROIs time series segment with its weak derivatives;

3) Similar to the construction of low-order networks, construct the temporal low-order networks based on the ultra-ROI time series segments via the ULS Group constrained topology structure detection algorithm and UOLS algorithm;

4) Stack all temporal low-order networks of all subjects together (Leonardi et al. 2013) to generate a set of correlation time series, with each correlation time series corresponding to the strength of the edge in the same coordinate of stacked networks;

5) Apply Ward's hierarchical grouping (Chen et al. 2014) to group the correlation time series into different clusters. Calculate the mean correlation time series in each cluster, and take these mean correlation time series as the new vertices of the high-order network;

6) Different from the construction of low-order networks based on the ROI time series, the high-order networks are constructed based on the mean correlation time series via the ULS Group constrained topology structure detection algorithm and UOLS algorithm, where, for each subject, we can obtain a high-order functional connectivity network by this step.

(4) **Feature Extraction, Selection, and Classification:**

1) Extract low- and high-order features from low- and high-order networks, respectively. Select the most discriminative features from the low- and high-order features, respectively;

2) Construct two DCT sub-models (Akhoondzadeh 2016) based on the selected low- and high-order features, respectively, and obtain the final classification scores for each subject by combining the classification scores from the two DCT sub-models with a certain fusion weight α.

Results and Discussions

The Impact of Parameters Optimization

In this section, we investigate the influence of different parameters on the classification performance, including the highest order of the weak derivatives L, the window size S, the step size r, the sliding window function, the number of clusters U, and the fusion weight α, respectively.

Effect of the Highest Order of the Weak Derivatives L

In order to extract the information of relations among fMRI data points, we incorporate the weak derivatives $D^l y_m^n$ $(l = 1, 2, ..., L)$ into the original ROI time series y_m^n, where L is the highest order of the weak derivatives and determines the number of weak derivatives used in the low- and high-order network construction. We seek to explore how the classification performance of the low-order, high-order, and fusion frameworks are influenced by the changes of L. Specifically, we varied L from 0 to 4 with a step of 1, and reported the classification accuracy and AUC values of the low-order, high-order and fused methods in Fig. 2. When $L = 0$, these classification frameworks are constructed only based on the original ROI time series without the weak derivatives. It can be noticed that the classification performance at $L = 0$ is inferior to that at $0 < L \leq 4$, confirming the importance of incorporating the weak derivatives into the original ROI time series to improve the classification performance. Furthermore, the maximum accuracy and AUC values are achieved at $L = 2$, indicating that the combination of the

Fig. 2 Classification accuracy (**a**) and AUC (**b**) with respect to L
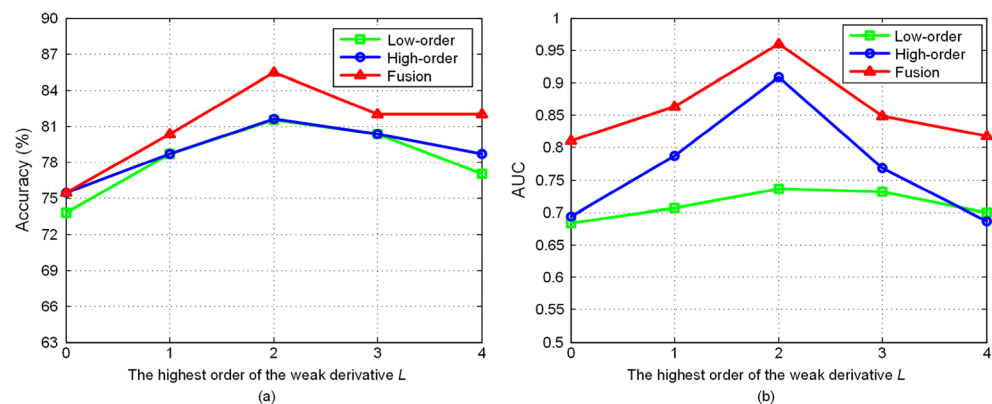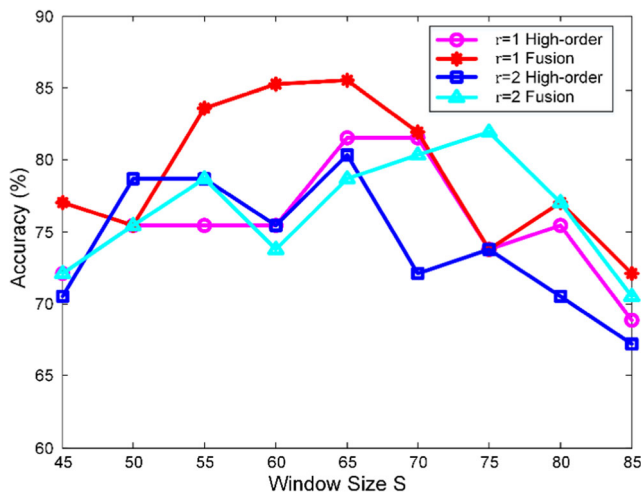
Fig. 3 Classification accuracy of high-order and fused methods against different S and r

first-order derivative and the second-order derivative is capable of representing the relations among fMRI data points. This phenomenon can be interpreted as that the effect of weak-noise may be amplified in high-order weak derivatives ($L > 2$).

Influence of Window Size, Step Size and Window Function on Classification Performance

One important step of the proposed method is using a sliding window to partition each ROI time series into multiple overlapping segments. Determining the optimal window size (S), step size (r) and function of the sliding window is essential to the proposed classification framework. In this section, we seek to investigate how the window size, step size, and function of the sliding window affect the classification performance of the proposed high-order and fused methods. Specifically, we varied S from 45 to 85 in steps of 5, and r from 1 to 2. With the window function maintained as the rectangular function, the classification accuracy of the high-order and fused methods with respect to S and r is shown in Fig. 3.

Table 2 Mathematical representation of the tested sliding window functions

Windows functions	$\omega(s)$
Rectangular	1
Triangular	$1-\left\|\frac{s-(S-1)/2}{S/2}\right\|$
Hann	$0.5\left(1-\cos\left(\frac{2\pi s}{S-1}\right)\right)$
Hamming	$0.54-0.46\cos\left(\frac{2\pi s}{S-1}\right)$
Sine	$\sin\left(\frac{\pi s}{S-1}\right)$
Gaussian	$e^{-\frac{1}{2}\left(\frac{s-(S-1)/2}{0.4(S-1)/2}\right)^2}$
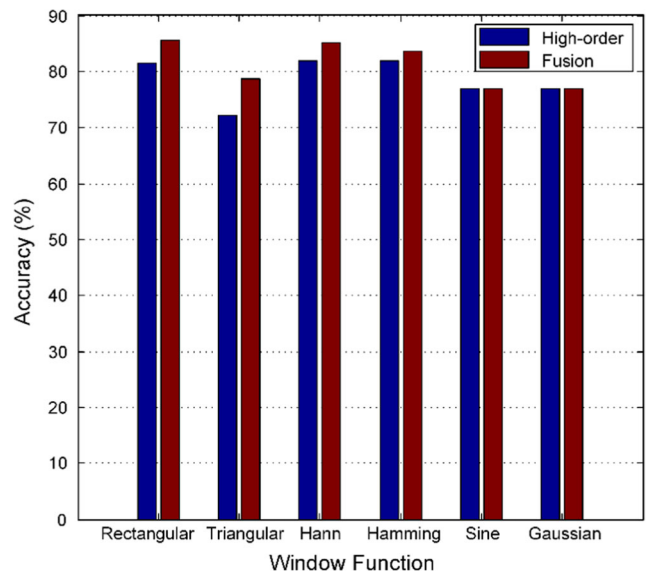
Fig. 4 Classification accuracy of high-order and fused methods against different sliding window functions

Compared to $r = 2$, the high-order and fused methods yield a relatively preferable performance when using $r = 1$. This phenomenon can be interpreted as that a larger step size leads to a less number of time series segments, while the number of time series segments is equal to the length of correlation time series in the high-order functional connectivity. A relatively short correlation time series may cause the insufficient information for representing the temporal variations among the correlations. It also can be understood as some temporal variations among correlations may be ignored while a relatively large step size is adopted.

By fixing $r = 1$, a relatively high and robust accuracy for the fused method can be achieved, when S varies between 55 and 70, and the highest accuracies for the high-order method are achieved within the range of $65 \leq S \leq 70$. However, when S becomes too small or too large, the classification accuracy decreases gradually. This can be interpreted from two aspects. First, when S is too small, the temporal low-order networks are
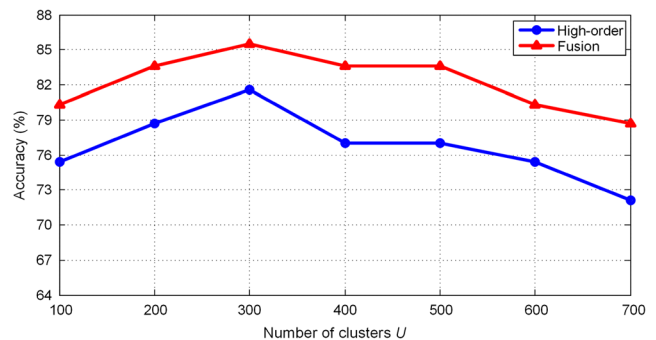
Fig. 5 Classification accuracy of high-order and fused methods with different the number of clusters U

Fig. 6 Classification accuracy (**a**) and AUC (**b**) of the proposed fused method with respect to α
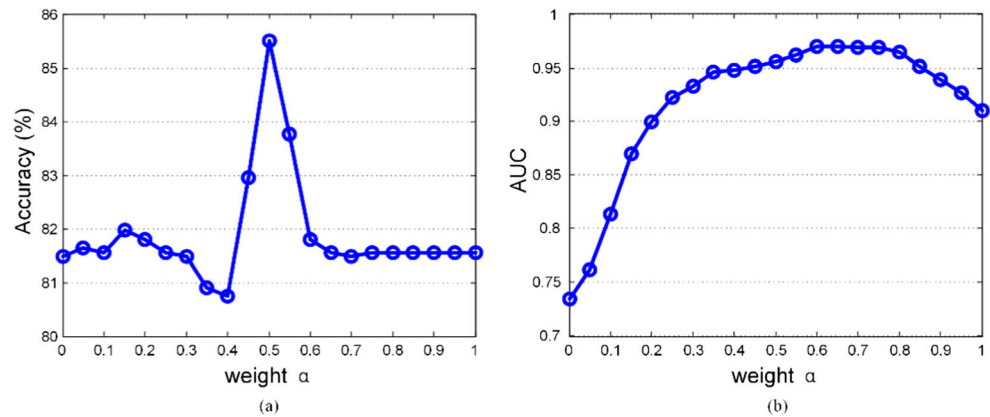
constructed based on small time series segments which can seriously reduce the reliability of correlation in the temporal low-order networks. Second, when S is too large, the number of time series segments is insufficient in computing the correlation time series, causing the high-order correlation, which is calculated based on the low-order correlation time series, unreliable.

By fixing $r = 1$ and $S = 65$, we explored the influence of sliding window function using six commonly used window functions, including Rectangular, Triangular, Hann, Hamming, Sine, and Gaussian. The mathematical representation of these windows is shown in Table 2. The classification accuracies of high-order and fused methods are shown in Fig. 4. The Triangular window achieved the lowest classification accuracy while the Rectangular, Hann, Hamming window achieved relatively high classification accuracy.

Impact of the Number of Clusters U

In order to reduce the scale of the high-order networks, we perform a Ward's hierarchical grouping method (Chen et al. 2014) to group the correlation time series into different clusters. We perform a series of experiments to explore the influence of the different number of clusters U on the classification performance. Specifically, we varied U from

100 to 700 in steps of 100 and report the ACCs of the proposed high-order and fused methods in Fig. 5. It can be observed that the high-order and fused methods yield a relatively high and robust classification accuracy. The highest ACCs of the high-order and fused methods are achieved at $U = 300$. However, the classification accuracy decreases gradually when U becomes too small or too large. This is reasonable since U determines the number of clusters and further controls the scale of the high-order networks. When U is too small, the correlation time series with different temporal variations are grouped into the same cluster, thus reducing the purity of clusters and the reliability of the mean correlation time series of each cluster. It further leads to the unreliability of high-order networks, which is constructed based on the mean correlation time series of each cluster. Meanwhile, when U is too large, the correlation time series with the similar temporal variations are partitioned into different clusters. It may cause the scale of the high-order network to be too large and produce redundant high-order features, thus deteriorating the performance of the classifiers.

Effect of the Fusion Weight α

The fusion weight α determines the contribution of the low-order and high-order networks in the final fusion

Table 3 The number of extracted features and selected features for each method

Method	Number of extracted features	Number of selected features
Low-order correlation-based	180	30
High-order correlation-based	600	82
Fused correlation-based	780	112
Low-order ULS Group constrained-UOLS	360	32
High-order ULS Group constrained-UOLS	1200	299
Fused ULS Group constrained-UOLS (Proposed)	1560	331

Table 4 Performance comparison between correlation-based and ULS Group constrained-UOLS networks

Method	ACC(%)	SEN(%)	SPC(%)	AUC	*p*-values
Low-order correlation-based	62.4	51.4	71.7	0.626	2.7×10^{-17}
High-order correlation-based	69.3	63.9	73.9	0.772	3.8×10^{-13}
Fused correlation-based	69.7	63.0	75.3	0.800	1.8×10^{-12}
Low-order ULS Group constrained-UOLS	81.5	74.3	87.6	0.736	2.0×10^{-13}
High-order ULS Group constrained-UOLS	81.6	72.9	**88.9**	0.908	5.6×10^{-11}
Fused ULS Group constrained-UOLS (Proposed)	**85.5**	**86.6**	84.6	**0.960**	–

The numbers marked bold indicate the best performance

classification framework. A larger α value indicates the larger contribution of the high-order networks and smaller contribution of the low-order networks to the classification. When $\alpha = 0$, the fusion classification framework is reduced to the low-order sub-model, and when $\alpha = 1$, the fusion classification framework is reduced to the high-order sub-model. In this experiment, we seek to investigate how the classification performance of the fusion framework is influenced by the changes of α. Specifically, we changed the value of α from 0 to 1 with a step of 0.05, and reported the classification accuracy and AUC values of the fusion classification framework in Fig. 6. Relatively high accuracy values was achieved with the range of $0.45 \leq \alpha \leq 0.55$, indicating the relatively equal contribution from the low- and high-order connectivity. On the other hand, AUC of more than 0.95 can be achieved over a relatively wide range of $0.45 \leq \alpha \leq 0.85$, indicating relatively robust and consistent classification performance with respect to α. Furthermore, higher

classification accuracy and AUC were achieved when the fusion strategy is used ($0 < \alpha < 1$), indicating the dominance of fusing the dynamic correlation information and the holistic correlation information.

Classification Accuracy

In this work, we compare the MCI classification performance of the proposed fused ULS Group constrained-UOLS method with other methods, which include the conventional low-order correlation-based method, the high-order correlation-based method (Chen et al. 2016), the fused correlation-based method (Chen et al. 2016), the low-order ULS Group constrained-UOLS method, and the high-order ULS Group constrained-UOLS method on the same dataset. As mentioned in "Feature Extraction" section, we extract four types of features from the low-order and high-order ULS Group constrained-UOLS networks, including the weighted clustering coefficient, betweenness centrality, in-degree, and out-degree. Therefore, $4 \times M$ low-order features and $4 \times U$ high-order features are extracted from low-order and high-order ULS Group constrained-UOLS networks, respectively ($M = 90$ and $U = 300$). Because low-order and high-order correlation-based networks are non-sparse symmetric matrices, only the weighted clustering coefficient and degree can be extracted from the correlation-based networks. Hence, we extract $2 \times M$ low-order features and $2 \times U$ high-order features from low-order and high-order correlation-based networks, respectively ($M = 90$ and $U = 300$). For each competing method, we use the nested 10-fold cross-validation scheme described in "Classification" section to optimize the feature selection parameters ε_1, ε_2 and ε_3, and evaluate the classification performance. The number of selected features fed into the classifier of each method is provided in Table 3.

The performance measures are computed according to the average results of the 10-fold cross-validation that is repeated 20 times and shown in Table 4. The proposed fused ULS Group constrained-UOLS method yields the best accuracy of 85.5%, the best sensitivity of 86.6%, and the best AUC of
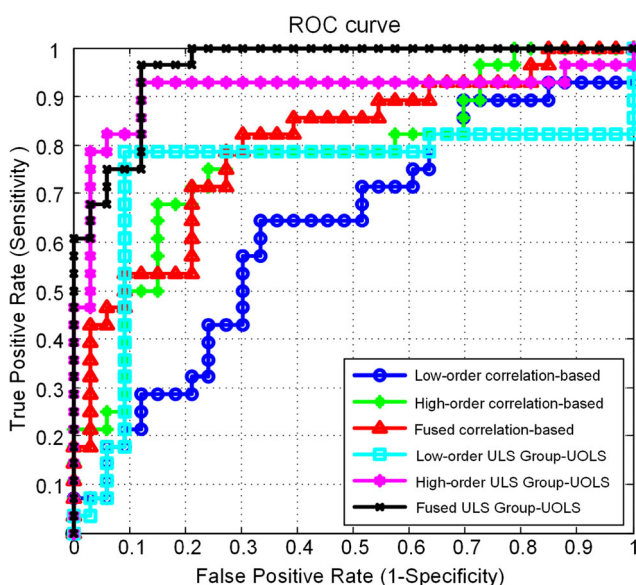
**Fig. 7** ROC curves for all comparison methods
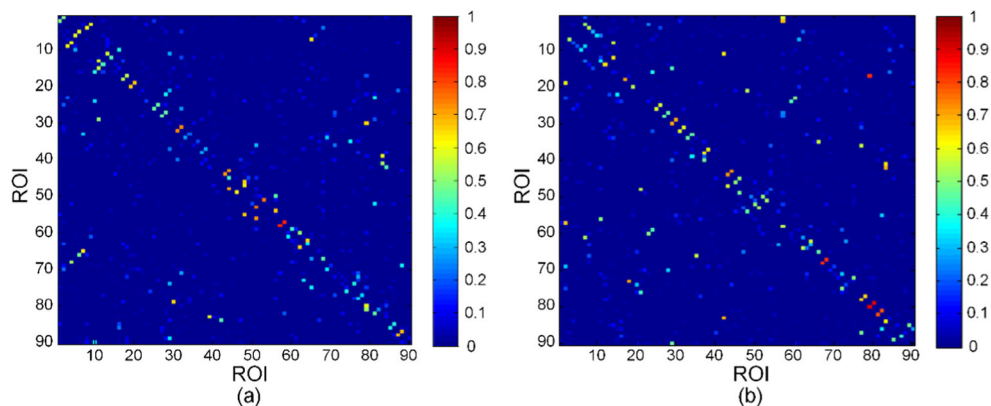
0.960, respectively. Although the best specificity of 88.9% is obtained by the high-order ULS Group constrained-UOLS method, it also indicates that the dynamic correlation information is important and should be incorporated into MCI classification. By comparing these results, we can demonstrate the superiority of fused ULS Group constrained-UOLS method from three aspects. First, the sparse regression-based models (ULS Group constrained-UOLS) outperform the correlation-based models, indicating that the sparse network structure and the information of weak derivatives are beneficial to MCI classification. Second, the high-order networks perform better than the low-order networks, confirming the importance of dynamic correlation information for MCI classification. Third, the fusion networks have better classification performance than the corresponding low-order and high-order networks, proving the necessity of simultaneously considering the dynamic correlation information and the holistic correlation information. Furthermore, our proposed method performs significantly better than all the competing methods in terms of ACC, SPC, and AUC based on the two-sample t-test results

on 20 repetitions. Table 4 also shows the p-values for the comparison of AUC between the proposed method and other methods. Figure 7 shows the ROC curves of all compared methods. It can be observed that the proposed method achieves the largest area under ROC curves, indicating its excellent diagnostic ability with respect to the changes of the discrimination threshold.

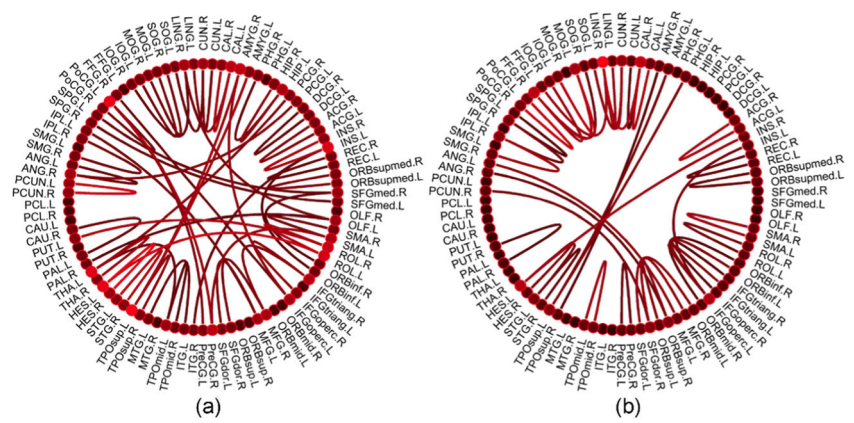## Low-Order Functional Connectivity Network

Figure 8a, b show two examples of the low-order functional connectivity networks for MCI and NC, respectively. The discrepancy network obtained by subtracting the MCI network from the NC network is shown in Fig. 9. The two functional connectivity networks show the relatively different patterns in terms of network topology and strength. For instance, the connectivity between bilateral precuneus is disrupted in MCI subjects, compared to the NC subjects who show stronger connectivity strength between them, which is consistent with the previous study (Haussmann et al. 2017; Mi et al. 2017) that the atrophy in precuneus is related to AD/MCI. Figure 10a, b graphically show the topology of the aforementioned networks for MCI and NC subjects, respectively. With the same threshold of 0.4, totally 51 connections are survived in the NC network while only 42 connections are survived in the MCI networks. This observation is consistent with the previous study (Zhou et al. 2013), indicating that some connections of the MCI network have been disrupted by pathological attacks.

Moreover, to demonstrate the temporal variation of correlations, we compare a series of temporal low-order functional connectivity networks. Specifically, Fig. 11a shows the low-order functional connectivity network of one MCI subject, which characterizes the holistic correlation between different ROIs calculated on the whole time series. The low-order correlation between a pair of ROIs shows no temporal variation. In order to extract
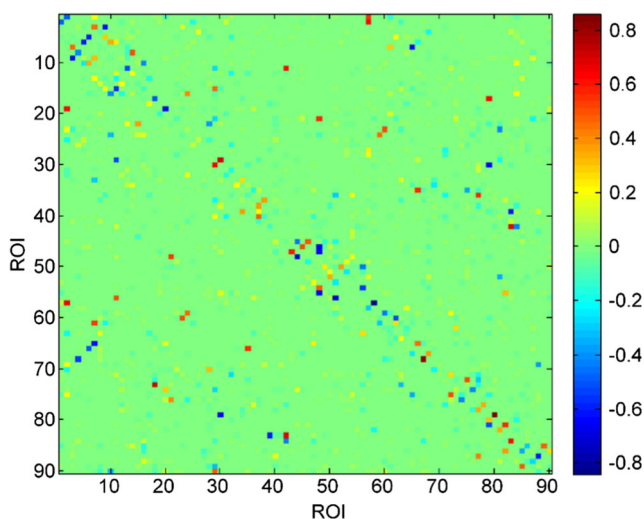

Fig. 9 The discrepancy between the low-order networks for MCI and NC

Fig. 10 The structure of two aforementioned networks for MCI (**a**) and NC (**b**)


(a)   (b)

temporal variation information contained in the fMRI time series, the sliding window approach is employed to construct a collection of temporal low-order functional connectivity networks, with each of them computed based on a segment of fMRI subseries that is equivalent to the length of sliding window. Some of the temporal networks are provided in Fig. 11b-f. It is clearly observed that these temporal low-order networks, which are computed using subsequence fMRI subseries, consistently change over the scanning period. Therefore, we can conclude that the conventional low-order network, which is

computed using the entire fMRI time series, is unable to characterize the temporal variations of correlations between different ROIs over the entire scanning period, and thus the other type of network should be constructed to effectively characterize the dynamic correlation between the ROIs. The existing methods, which use the dynamic networks (temporal low-order networks) for diagnosing brain diseases (Wee et al. 2016; Zhu et al. 2016), extract the network features from the temporal low-order networks directly to construct the brain disease classifiers. These methods do not take into account the
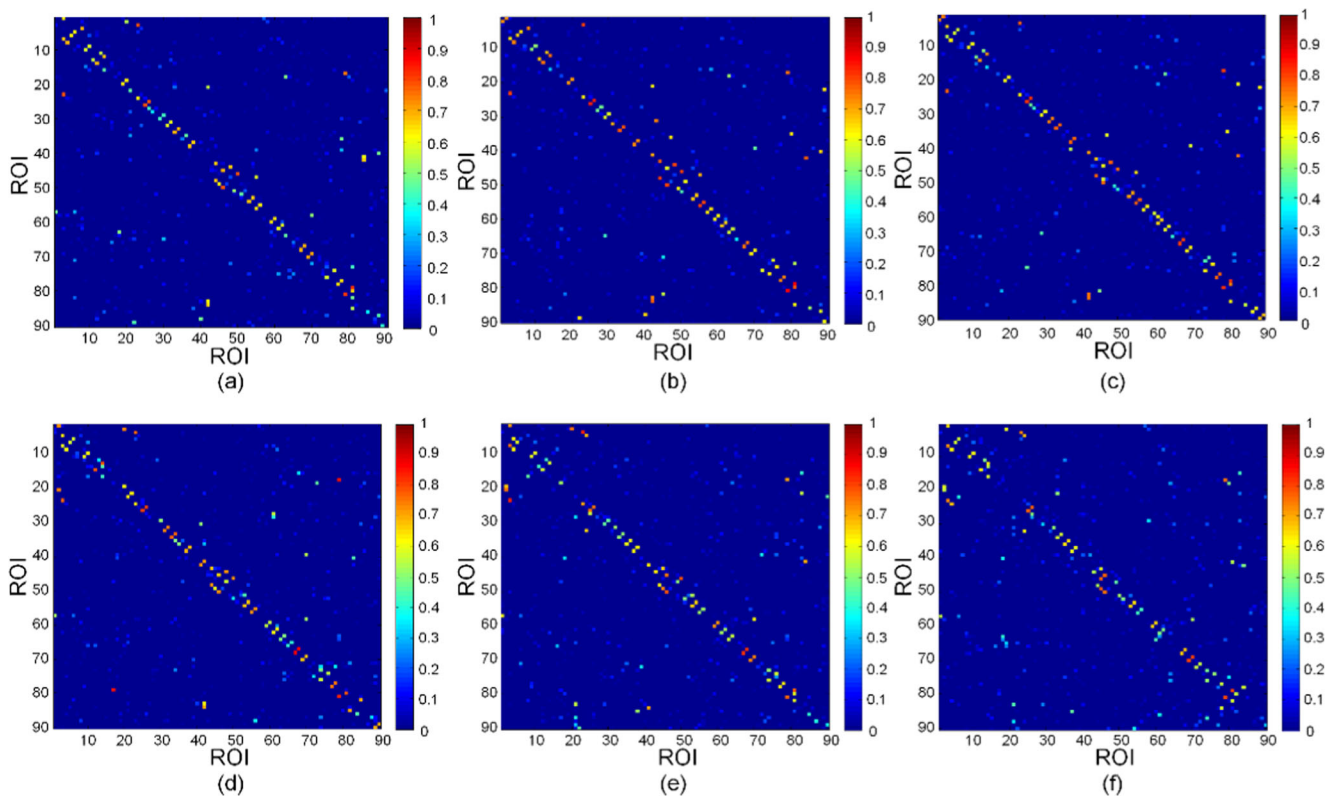
Fig. 11 Conventional and temporal low-order functional connectivity networks for MCI. (**a**) Conventional network; (**b–f**) Temporal networks
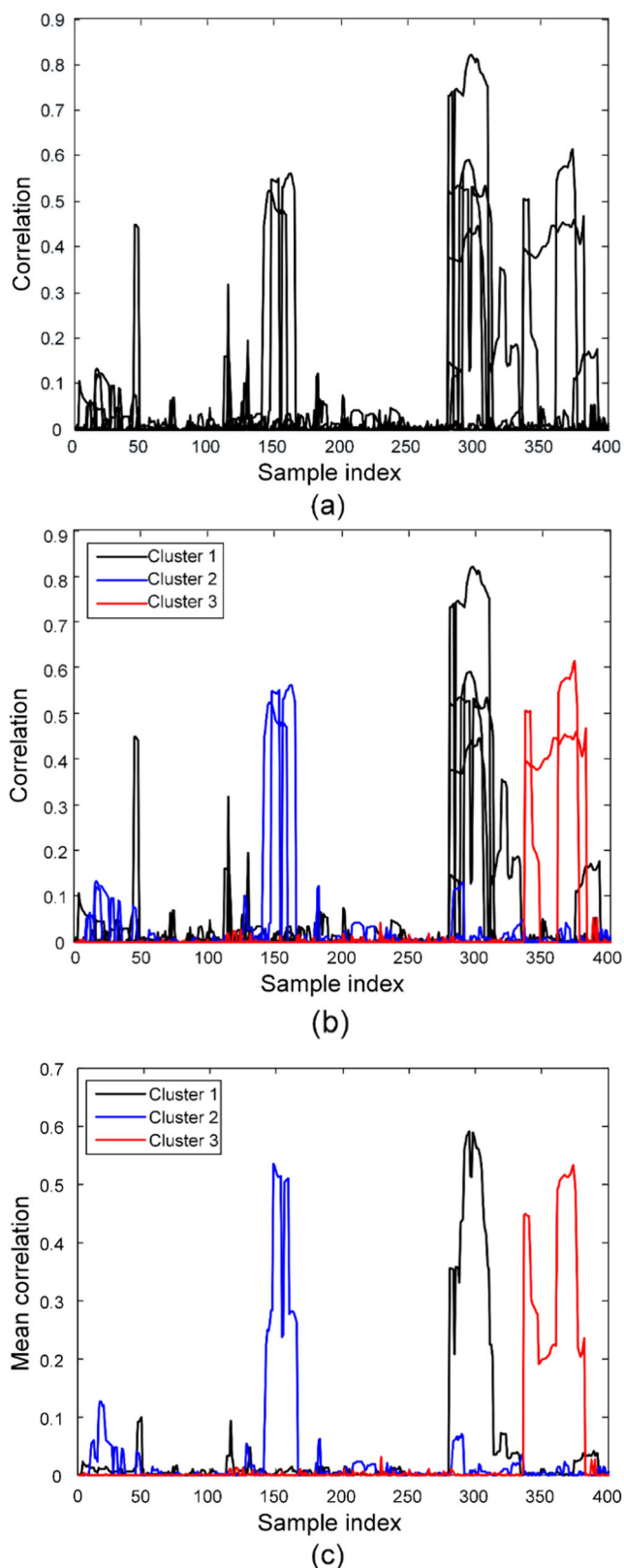
Fig. 12 Clustering results of the correlation time series. **a** Original correlation time series; **b** Three clusters of the correlation time series based on Ward's algorithm (correlation time series from the same cluster are depicted with the same color); **c** The mean correlation time series of each cluster

temporal continuity of correlations between ROIs. In other words, these methods consider each temporal low-order network independently and omit the information of relative positions of the temporal low-order networks. The absence of this information may lead to a suboptimal classification model. Therefore, we further construct the high-order networks to overcome the limitation of the temporal low-order network methods.

The Clustering in the Construction of High-Order Functional Connectivity Network

As mentioned earlier, Ward's hierarchical grouping (Chen et al. 2014) is applied to group the correlation time series into different clusters. In this clustering method, parameter initialization is not required, reducing the dependence of the method on the hyper-parameters and thus improving the robustness of clustering results. Some correlation time series are displayed in Fig. 12a. It can be clearly observed that the correlations between ROIs undergo large variation over the scanning period. Furthermore, the significant correlation between ROIs is only occurred for a short period of time and is almost zero in the rest of the time series. Figure 12b shows the clustering results for the correlation time series, with those belonging to the same cluster being depicted in the same color. By comparing Figs. 12a and b, we find that the correlation time series with similar temporal variations are grouped into the same cluster, while the time series with dissimilar temporal variation are assigned to different clusters. In Fig. 12c, we provide the mean correlation time series for each cluster. By using the mean correlation time series of each cluster as the new vertices of a high-order network, we construct a collection of small-scale high-order functional connectivity networks, without largely losing the important dynamic correlation information.

The Most Discriminative Regions and Clusters

The most discriminative brain regions and clusters are defined as the ones with the highest frequency of selection in 20 repetitions of 10-fold cross-validation, corresponding to the low-order and the high-order network, respectively. As mentioned above, we use clusters to replace the original correlation time series as the new vertices of high-order networks for reducing the scale of networks. Hence, for high-order networks, we select the most discriminative clusters as the biomarker for MCI identification.

The Most Discriminative Brain Regions and Low-Order Connections

The most discriminative brain regions selected from the low-order networks are listed in Table 5 and displayed in Fig. 13, which include the right olfactory cortex
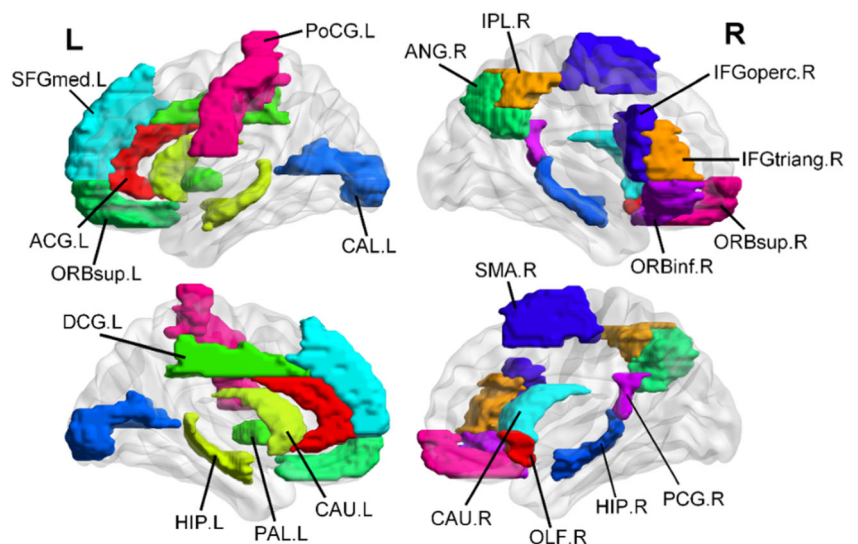
Table 5 The top 20 most discriminative ROIs selected from the low-order network

| No.	ROI index	ROI abbr.	References	Location
1	22	OLF.R	(Sun et al. 2012; Vasavada et al. 2015)	Frontal lobe
2	14	IFGtriang.R	(Salvatore et al. 2015; Chen et al. 2016; Xu et al. 2016)	Frontal lobe
3	37	HIP.L	(Salvatore et al. 2015)	Limbic lobe
4	33	DCG.L	(Matsuda 2013)	Limbic lobe
5	5	ORBsup.L	(Xu et al. 2016)	Frontal lobe
6	23	SFGmed.L	(Xu et al. 2016)	Frontal lobe
7	43	CAL.L	(Xu et al. 2016)	Occipital lobe
8	20	SMA.R	(Rose et al. 2006)	Frontal lobe
9	16	ORBinf.R	(Salvatore et al. 2015)	Frontal lobe
10	6	ORBsup.R	(Xu et al. 2016; Wee et al. 2012b)	Frontal lobe
11	72	CAU.R	(Salvatore et al. 2015)	Sub cortical gray nuclei
12	62	IPL.R	(Salvatore et al. 2015)	Parietal lobe
13	71	CAU.L	(Salvatore et al. 2015)	Sub cortical gray nuclei
14	75	PAL.L	(Matsuda 2013)	Sub cortical gray nuclei
15	66	ANG.R	(Xu et al. 2016)	Parietal lobe
16	31	ACG.L	(Salvatore et al. 2015; Wee et al. 2012b)	Limbic lobe
17	38	HIP.R	(Xu et al. 2016)	Limbic lobe
18	12	IFGoperc.R	(Chen et al. 2016)	Frontal lobe
19	36	PCG.R	(Wee et al. 2012b)	Limbic lobe
20	57	PoCG.L	(Xu et al. 2016)	Central region

(OLF.R), right inferior frontal gyrus triangular part (IFGtriang.R), left hippocampus (HIP.L), left median cingulate and paracingulate gyri (DCG.L), and left superior frontal gyrus orbital part (ORBsup.L). These regions are frequently reported as highly associated with AD/MCI pathology (Chen et al. 2016; Matsuda 2013; Salvatore et al. 2015; Vasavada et al. 2015; Xu et al. 2016). Particularly, olfactory deficits are prevalent in AD/MCI patients (Vasavada et al. 2015). The symptoms of olfactory deficits normally appear before the clinical cognitive deficits and memory deficits, which is consistent with the fact that AD attacks the central olfactory structures preferentially (Vasavada et al. 2015). Hence, the decline in olfactory cortex (OLF), can be regarded as an earlier and effective biomarker for identifying MCI from healthy elderly. Previous works also reported that hippocampus (HIP) is a key brain region for memory (Jaroudi et al. 2017; Du et al. 2001; Sze et al. 1997).

Fig. 13 The top 20 most discriminative ROIs

Fig. 14 Connectivity strengths among the top 20 most discriminative ROIs identified from the low-order network
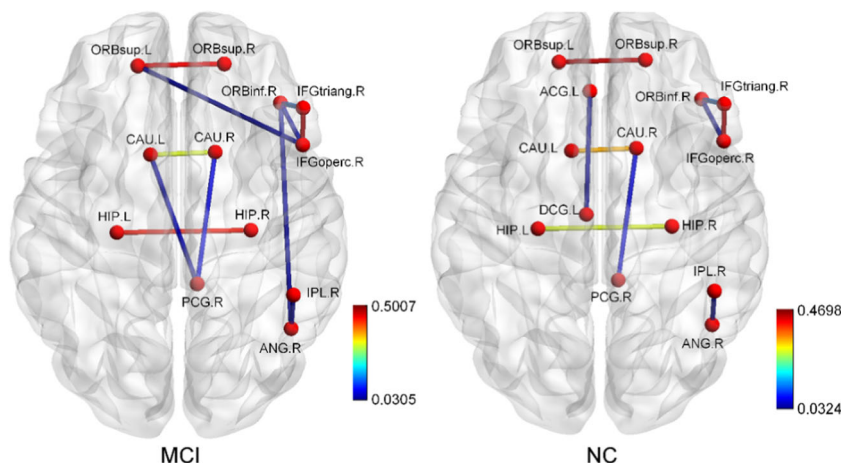

The atrophy in HIP is believed to be an early biomarker of AD/MCI. On the other hand, the most discriminative brain regions are located mainly in the frontal lobe and limbic lobe. In the previous studies, the frontal lobe, which plays an important role in the integration of non-task based memories stored, has been reported as the important brain area associated with AD/MCI pathology (Ribeiro and Busatto Filho 2016). The limbic lobe is believed to be highly related with language and memory, in which AD patients showed an obvious reduction in regional cerebral blood flow (Takahashi et al. 2017).

A multiple linear regression between CDR score and feature values for the selected top 20 most discriminative ROIs is performed to examine the relationship between the CDR score and the selected ROIs. Then, an Analysis of Variance (ANOVA) is adopted to check the significance of the correlation between the CDR score and the selected ROIs. The ANOVA F-test value is 2.003 (p-value

is 0.03), which is larger than the critical value of F-test under 95% confidence interval, indicating that the CDR scores show a significant linear correlation with feature values of the selected ROIs. Thus, the most discriminative ROIs selected in this work are highly correlated to the CDR score, in line with the diagnosis criteria of MCI subjects used in this study.

In addition, we calculate the low-order connections among the top 20 most discriminative ROIs by averaging the low-order connectivity strengths of all MCI and NC subjects, respectively. The low-order connections are provided in Fig. 14. The discriminative power of each connection between MCIs and NCs are evaluated by using the two-sample t-test, and p-values of all connections are provided in Fig. 15a. Figure 15b provides the connections that are significantly different ($p < 0.05$) between NC and MCI subjects. According to statistical analysis results, we observe a significant difference in functional connection
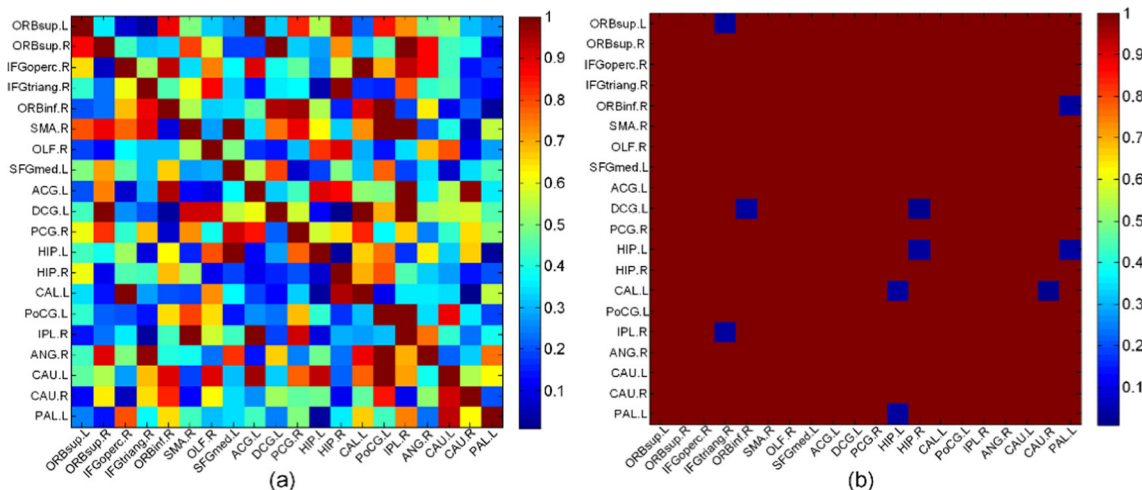
Fig. 15 p-values of connections among the top 20 most discriminative ROIs (**a**), and p-values less than 0.05 (**b**)

Table 6 The top 20 most discriminative clusters selected from the high-order network

Clusters	ROI-pairs	Clusters	ROI-pairs
Cluster 1	CUN.L−CUN.R	Cluster 11	IOG.L−MOG.L
Cluster 2	SMG.R−IPL.R	Cluster 12	SMG.R−SMG.L
Cluster 3	INS.R−HES.R	Cluster 13	MOG.L−IOG.L
Cluster 4	STG.L−ROL.L	Cluster 14	MFG.R−IFGtriang.R; IFGtriang.R−MFG.R
Cluster 5	SMA.R−SMA.L	Cluster 15	TPOsup.R−AMYG.L
Cluster 6	SFGmed.L−SFGdor.L	Cluster 16	ROL.R−INS.L; INS.L−ROL.R
Cluster 7	HES.R−INS.R	Cluster 17	INS.R−INS.L
Cluster 8	SOG.R−MOG.L	Cluster 18	AMYG.L−TPOsup.L
Cluster 9	MOG.L−MOG.R	Cluster 19	HES.L−STG.L
Cluster 10	INS.L−INS.R	Cluster 20	MOG.R−MOG.L

Note that the brain networks constructed in this paper are asymmetric networks, thus the ROI-pair (m_1, m_2) is different from the ROI-pair (m_2, m_1)

between left and right hippocampi (HIP.L and HIP.R) for MCI and NC subjects. The connectivity strength between HIP.L and HIP.R was significantly larger in MCI subjects compared to NC subjects. This phenomenon can be interpreted as that the MCI subjects may require stronger interactions between some brain regions than NC subjects for brain compensation due to cognitive impairment (Salvatore et al. 2015), which is consistent with the brain compensatory phenomenon that has been reported in the literature (Jie et al. 2016; Qi et al. 2010).

The Most Discriminative Clusters and High-Order Connections

As for the high-order networks, the top 20 most discriminative clusters and the ROI-pairs contained in these clusters are listed in Table 6. Most of these clusters contain only one ROI-pair. The ROI-pairs in the most discriminative clusters include the left middle occipital gyri−right middle occipital gyri (MOG.L−MOG.R), left Heschl gyrus−left superior temporal gyrus (HES.L−STG.L), left middle occipital gyri −left inferior occipital gyrus (MOG.L−IOG.L), suggesting

that the time-varying characteristics of the correlation time series corresponding to these ROI-pairs play an important role in MCI classification.

Additionally, Fig. 16 shows the difference of high-order connections among the top 20 most discriminative clusters between MCI and NC subjects. Different from the low-order connection represents the functional connectivity strength of between two ROIs, the high-order connection, which is calculated based on the low-order correlation time series, describes the relation between two clusters (i.e. the relation between the low-order connections). The significant difference in high-order connection is observed between the cluster 13 (containing ROI-pair MOG.L−IOG.L) and cluster 9 (containing ROI-pair MOG.L−MOG.R). This high-order connection (row 13, column 9) in NC subjects is stronger than that in MCI subjects, which implies that the temporal variations of the low-order connectivity strength between MOG.L and IOG.L is highly correlated with the temporal variations of the low-order connectivity strength between MOG.L and MOG.R in NC subjects, but not in MCI subjects. This

Fig. 16 Comparison of correlation difference among the top 20 most discriminative clusters identified based on the high-order connectivity networks, where elements with zero indicate no correlation between two clusters
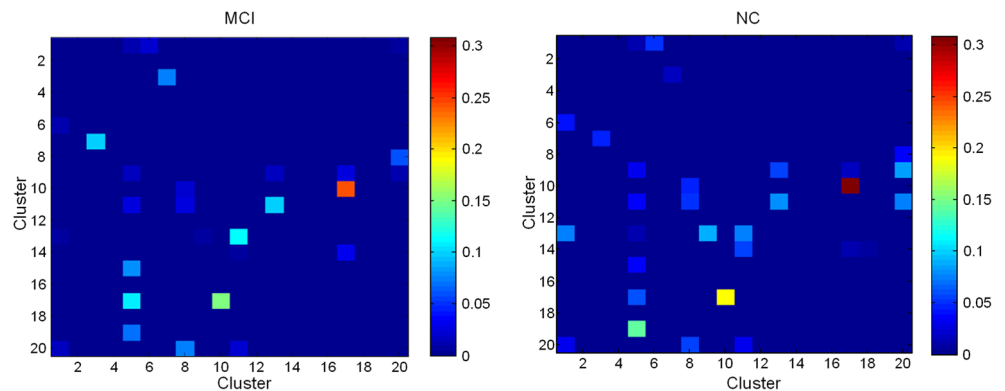
difference between MCI and NC can be used as an effective biomarker for identifying MCI from NC.

Conclusion

In this paper, we proposed a novel fusion approach to infer the high- and low-order functional connectivity networks for MCI classification. By using the ULS Group constrained topology structure detection algorithm and UOLS algorithm, our approach successfully extracts the temporal information of connections among fMRI data, which can be further used to construct the functional connection for the low- and high-order networks. Furthermore, hierarchical grouping (based on Ward's algorithm) is applied to reduce the scale and computational complexity of the high-order network. By fusing the DCT sub-models that are trained using the low- and high-order networks separately, our proposed approach integrates both the information of holistic correlations and the temporal variations at the disease identification stage. Promising results obtained demonstrate the superiority of our proposed method and also the importance of integrating both the conventional holistic correlation and the dynamic temporal correlation information for improving the performance of MCI classification.

Information Sharing Statement

The dataset used in this paper are from the Alzheimer's Disease Neuroimaging Initiative (ADNI, RRID:SCR_003007) which are available at http://adni.loni.usc.edu/. The source code developed in this paper are available via GitHub website (https://github.com/792866506/High-and-Low-order-Networks).

Compliance with Ethical Standards

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the local ethical committee.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest The authors declare that they have no conflicts of interest.

Appendix

The weak derivative and Ultra-Least Squares Criterion

A linear system with k inputs and one output can be described by a linear regression model below:

$$y(t) = \sum_{i=1}^{k} \theta_i x_i(t) + e, \tag{11}$$

where $y(t)$ and $x_i(t)$ denote the system output and input variables, θ_i is the system parameter, and e is the system noise. For this system, the ordinary least squares regression problem can be solved via the least squares criterion as follows:

$$J_{LS} = \left\| y(t) - \sum_{i=1}^{k} \theta_i x_i(t) \right\|_2^2, \tag{12}$$

where $t \in [0, T]$, $y(t)$ and $x_i(t)$ are time dependent signals with finite amplitude on the interval $[0, T]$, and thus $y(t)$ and $x_i(t)$ are L^2 integrable functions belong to the Lebesgue space $L^2([0, T])$, where $L^2([0, T]) = \{x(t) | \int_{[0, T]} |x(t)|^2 dt < +\infty\}$. Supposing $\hat{y}(t)$ is prediction function of $y(t)$, it is obvious that the least squares criterion only measures the discrepancy between $y(t)$ and $\hat{y}(t)$ on the whole interval $[0, T]$, ignoring how the discrepancy distributes at every individual time point. Therefore, the least squares criterion cannot accurately describe the similarity of function shapes and discards the information of correlations among data points, leading to a common overfitting problem for identification of the dynamic system (Li et al. 2018a; Guo et al. 2016).

In order to overcome this limitation, we integrate a weak derivative part into the least squares criterion to construct an ULS criterion:

$$J_{ULS} = \left\| y(t) - \sum_{i=1}^{k} \theta_i x_i(t) \right\|_2^2 + \sum_{l=1}^{L} \left\| D^l y(t) - \sum_{i=1}^{k} \theta_i D^l x_i(t) \right\|_2^2 \tag{13}$$

where D^l is the l-th order weak derivative ($l = 1, 2, \ldots, L$). The weak derivative, which measures interconnections among the data points, is a generalization of the derivative that is in the usual sense. Different from the derivatives which can be calculated only for the differentiable functions, the weak derivatives can be calculated for all integrable functions. Supposing that $x(t)$ belong to the Lebesgue space $L^2([0, T])$, the l-th order weak derivative of $x(t)$ is defined as the function $D^l x(t) \in L^2([0, T])$ which satisfies

$$\int_{[0,T]} x(t) D^l \varphi(t) dt = (-1)^l \int_{[0,T]} \varphi(t) D^l x(t) dt, \tag{14}$$

for all infinitely differentiable functions $\varphi(t)$ with $\varphi(0) = \varphi(T) = 0$. As discussed in (Guo et al. 2016), the regression model fitted by weak derivatives takes into account the relationship among data points and is therefore more effective and accurate. Given discrete observations of the system

signals, $\{y(j)\}$, $\{x_i(j)\}$, $j = 1, 2, \ldots, J$, the l-th order weak derivative can be calculated as

$$D^l y(h) = \Sigma_{j=h}^{h+J_0} y(j)\varphi^{(l)}(j-h) \ (h = 1, 2, \ldots, J-J_0) \quad (15)$$

$$D^l x_i(h) = \Sigma_{j=h}^{h+J_0} x_i(j)\varphi^{(l)}(j-h) \ (h = 1, 2, \ldots, J-J_0) \quad (16)$$

where $\varphi(t)$ ($t \in [0, J_0]$) is the test function, which is l-th order derivable on the interval $[0, J_0]$, $\varphi^{(l)}(t)$ denotes the l-th order derivative of the $\varphi(t)$. Due to the l-th order weak derivative ($l = 1, 2, \ldots, L$) of original signals is used in this work, the test function is required to have L-th order derivative. Therefore, the $(L+1)$-th order B-spline basis function which satisfies the above condition is adopted as the test function in this paper. More details of B-spline basis function and weak derivative can be found at Guo et al. (2016).

The Lebesgue space $L^2([0, T]) = \{x(t)| \int_{[0, T]}|x(t)|^2 dt < +\infty\}$ is a function space, in which the functions are L^2 integrable (i.e. the l_2-norm of the function is finite). Meanwhile, the Sobolev space $H^L([0, T]) = \{x(t)| \ x(t) \in L^2([0, T]), D^l x \in L^2([0, T]), l = 1, 2, \cdots, L\}$ is a subspace of $L^2([0, T])$, in which not only the functions but also the l-th order weak derivatives of the functions ($l = 1, 2, \ldots, L$) are L^2 integrable (i.e. belong to $L^2([0, T])$). The definition of Sobolev space $H^L([0, T])$ can also be written as

$$H^L([0, T]) \quad (17)$$
$$= \left\{x(t)| \int_{[0,T]}|x(t)|^2 dt < +\infty, \int_{[0,T]}|D^l x|^2 dt < +\infty, l = 1, 2, \cdots, L\right\}.$$

The least squares criterion only needs to calculate l_2-norm of the discrepancy between the observed signal $y(t)$ and the model prediction function $\sum_{i=1}^k \theta_i x_i(t)$, and thus is defined in the Lebesgue space $L^2([0, T])$. However, the ULS criterion calculate not only the l_2-norm of the discrepancy between $y(t)$ and $\sum_{i=1}^k \theta_i x_i(t)$, but also the l_2-norm of the discrepancy between the weak derivatives of $y(t)$ and $\sum_{i=1}^k \theta_i x_i(t)$. Therefore, the weak derivatives of $y(t)$ and $\sum_{i=1}^k \theta_i x_i(t)$ are required to belong to $L^2([0, T])$, and further the functions $y(t)$ and $x_i(t)$ of the ULS criterion are required to belong to $H^L([0, T])$. For these reasons, the ULS criterion is defined in the Sobolev space $H^L([0, T])$.

The fMRI time series is a low-frequency signal with finite energy. Thus, the fMRI time series and its weak derivatives are L^2 integrable functions (i.e. belong to $L^2([0, T])$). The fMRI time series can be further considered as the discrete observations of the signals belonging to $H^L([0, T])$. Therefore, the ULS criterion is applicable to the study of fMRI time series.

The new criterion considers *not only* the discrepancy between the observed signal and the model prediction function, *but also* the discrepancy between their weak derivatives. Thus, the ULS criterion is a more accurate evaluation standard for the model fitness. Essentially, the ULS criterion is the combination of the least squares criterion with the weak

derivative of the original signals. By connecting the original signals $y(t)$ and $x_i(t)$ with their weak derivatives $D^l y(t)$ and $D^l x_i(t)$ ($l = 1, 2, \ldots, L$), we generate the corresponding ultra-signals $\tilde{y}(t) = \left[(y(t))^T, (D^1 y(t))^T, (D^2 y(t))^T, \ldots, (D^L y(t))^T\right]^T$ and $\tilde{x}_i(t) = \left[(x_i(t))^T, (D^1 x_i(t))^T, (D^2 x_i(t))^T, \ldots, (D^L x_i(t))^T\right]^T$, and Eq. (13) can be rewritten as

$$J_{ULS} = \|\tilde{y}(t) - \Sigma_{i=1}^k \theta_i \tilde{x}_i(t)\|_2^2. \quad (18)$$

Therefore, we can integrate the ULS criterion into our proposed framework by incorporating the weak derivatives into the original time series.

References

Akhoondzadeh, M. (2016). Decision tree, bagging and random forest methods detect TEC seismo-ionospheric anomalies around the time of the Chile, (M-w=8.8) earthquake of 27 February 2010. *Advances in Space Research, 57*(12), 2464–2469. https://doi.org/10.1016/j.asr.2016.03.035.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex, 24*(3), 663–676. https://doi.org/10.1093/cercor/bhs352.

Alzheimer's Association. (2015). 2015 Alzheimer's disease facts and figures. *Alzheimers & Dementia, 11*(3), 332–384. https://doi.org/10.1016/j.jalz.2015.02.003.

Amezquita-Sanchez, J. P., Adeli, A., & Adeli, H. (2016). A new methodology for automated diagnosis of mild cognitive impairment (MCI) using magnetoencephalography (MEG). *Behavioural Brain Research, 305*, 174–180. https://doi.org/10.1016/j.bbr.2016.02.035.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140. https://doi.org/10.1023/A:1018054314350.

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., & Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimers & Dementia, 3*(3), 186–191. https://doi.org/10.1016/j.jalz.2007.04.381.

Bullmore, E. T., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186–198. https://doi.org/10.1038/nrn2575.

Chand, G., Wu, J., Hajjar, I., & Qiu, D. (2017). Interactions of the salience network and its subsystems with the default-mode and the central-executive networks in normal aging and mild cognitive impairment. *Brain Connectivity, 7*, 401–412. https://doi.org/10.1089/brain.2017.0509.

Chen, X. B., Xiao, Y., Cai, Y. F., & Chen, L. (2014). Structural max-margin discriminant analysis for feature extraction. *Knowledge-Based Systems, 70*, 154–166. https://doi.org/10.1016/j.knosys.2014.06.020.

Chen, X. B., Zhang, H., Gao, Y., Wee, C. Y., Li, G., & Shen, D. G. (2016). High-order resting-state functional connectivity network for MCI classification. *Human Brain Mapping, 37*(9), 3282–3296. https://doi.org/10.1002/hbm.23240.

Chen, X. B., Zhang, H., Zhang, L. C., Shen, C., Lee, S. W., & Shen, D. G. (2017). Extraction of dynamic functional connectivity from brain Grey matter and white matter for MCI classification. *Human Brain Mapping, 38*(10), 5019–5034. https://doi.org/10.1002/hbm.23711.

Chen, L., Zhang, H., Lu, J., Thung, K., Aibaidula, A., Liu, L., Chen, S., Jin, L., Wu, J., Wang, Q., Zhou, L., & Shen, D. (2018). Multi-label nonlinear matrix completion with Transductive multi-task feature selection for joint MGMT and IDH1 status prediction of patient with high-grade gliomas. *IEEE Transactions on Medical Imaging, 37*(8), 1775–1787. https://doi.org/10.1109/tmi.2018.2807590.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M. O., et al. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *Neuroimage, 56*(2), 766–781. https://doi.org/10.1016/j.neuroimage.2010.06.013.

Das, K., Rana, S., & Roy, S. (2018). Evaluation of Alzheimer's disease progression based on clinical dementia rating scale with missing responses and covariates. *Journal of Biopharmaceutical Statistics, 28*(5), 893–908. https://doi.org/10.1080/10543406.2017.1402780.

Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., & Trojanowski, J. Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiology of Aging, 32*(12), 2322.e19–2322.e27. https://doi.org/10.1016/j.neurobiolaging.2010.05.023.

Du, A. T., Schuff, N., Amend, D., Laakso, M. P., Hsu, Y. Y., Jagust, W. J., et al. (2001). Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *Journal of Neurology Neurosurgery and Psychiatry, 71*(4), 441–447. https://doi.org/10.1136/jnnp.71.4.441.

Eshkoor, S. A., Hamid, T. A., Mun, C. Y., & Ng, C. K. (2015). Mild cognitive impairment and its management in older people. *Clinical Interventions in Aging, 10*, 687. https://doi.org/10.2147/CIA.S73922.

Fan, Y., Gur, R. E., Gur, R. C., Wu, X. Y., Shen, D. G., Calkins, M. E., & Davatzikos, C. (2008). Unaffected family members and schizophrenia patients share brain structure patterns: A high-dimensional pattern classification study. *Biological Psychiatry, 63*(1), 118-124. https://doi.org/10.1016/j.biopsych.2007.03.015.

Fornito, A., Zalesky, A., & Bullmore, E. T. (2010). Network scaling effects in graph analytic studies of human resting-state FMRI data. *Frontiers in Systems Neuroscience, 4*, 22. https://doi.org/10.3389/fnsys.2010.00022.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J. L., de Leon, M., Feldman, H., Ganguli, M., Hampel, H., Scheltens, P., Tierney, M. C., Whitehouse, P., & Winblad, B. (2006). Mild cognitive impairment. *Lancet, 367*(9518), 1262–1270. https://doi.org/10.1016/S0140-6736(06)68542-5.

Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Current Opinion in Neurology, 21*(4), 424–430. https://doi.org/10.1097/WCO.0b013e328306f2c5.

Guo, Y. Z., Guo, L. Z., Billings, S. A., & Wei, H. L. (2016). Ultra-orthogonal forward regression algorithms for the identification of non-linear dynamic systems. *Neurocomputing, 173*, 715–723. https://doi.org/10.1016/j.neucom.2015.08.022.

Haufe, S., Nolte, G., Mueller, K. R., & Kraemer, N. (2008). Sparse causal discovery in multivariate time series. In *NIPS workshop on causality, 6*, 97–106.

Haussmann, R., Werner, A., Gruschwitz, A., Osterrath, A., Lange, J., Donix, K. L., Linn, J., & Donix, M. (2017). Precuneus structure changes in amnestic mild cognitive impairment. *American Journal of Alzheimers Disease and Other Dementias, 32*(1), 22–26. https://doi.org/10.1177/1533317516678087.

Hu, K., Wang, Y. J., Chen, K. W., Hou, L. K., & Zhang, X. Q. (2016). Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis. *Neurocomputing, 175*, 132–145. https://doi.org/10.1016/j.neucom.2015.10.043.

Huang, S. A., Li, J., Sun, L., Ye, J. P., Fleisher, A., Wu, T., et al. (2010). Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *Neuroimage, 50*(3), 935–949. https://doi.org/10.1016/j.neuroimage.2009.12.120.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., Handwerker, D. A., Keilholz, S., Kiviniemi, V., Leopold, D. A., de Pasquale, F., Sporns, O., Walter, M., & Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *Neuroimage, 80*, 360–378. https://doi.org/10.1016/j.neuroimage.2013.05.079.

Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal, 19*(3), 179–189. https://doi.org/10.1016/j.eij.2018.03.002.

Jaroudi, W., Garami, J., Garrido, S., Hornberger, M., Keri, S., & Moustafa, A. A. (2017). Factors underlying cognitive decline in old age and Alzheimer's disease: The role of the hippocampus. *Reviews in the Neurosciences, 28*, 705–714. https://doi.org/10.1515/revneuro-2016-0086.

Jie, B., Shen, D. G., & Zhang, D. Q. (2014). Brain connectivity hyper-network for MCI classification. In *International conference on medical image computing and computer-assisted intervention, 8674*, 724–732.

Jie, B., Wee, C. Y., Shen, D., & Zhang, D. Q. (2016). Hyper-connectivity of functional networks for brain disease diagnosis. *Medical Image Analysis, 32*, 84–100. https://doi.org/10.1016/j.media.2016.03.003.

Josef Golubic, S., Aine, C. J., Stephen, J. M., Adair, J. C., Knoefel, J. E., & Supek, S. (2017). MEG biomarker of Alzheimer's disease: Absence of a prefrontal generator during auditory sensory gating. *Human Brain Mapping, 38*, 5180–5194. https://doi.org/10.1002/hbm.23724.

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2015). Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. *Clinical Neurophysiology, 126*(11), 2132–2141. https://doi.org/10.1016/j.clinph.2015.02.060.

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2016). Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. *Brain Imaging and Behavior, 10*(3), 799–817. https://doi.org/10.1007/s11682-015-9448-7.

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2017). Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI. *Behavioural Brain Research, 322*, 339–350. https://doi.org/10.1016/j.bbr.2016.06.043.

Lee, H., Lee, D. S., Kang, H., Kim, B. N., & Chung, M. K. (2011). Sparse brain network recovery under compressed sensing. *IEEE Transactions on Medical Imaging, 30*(5), 1154–1165. https://doi.org/10.1109/Tmi.2011.2140380.

Lee, W. H., Bullmore, E., & Frangou, S. (2017). Quantitative evaluation of simulated functional brain networks in graph theoretical analysis. *Neuroimage, 146*, 724–733. https://doi.org/10.1016/j.neuroimage.2016.08.050.

Lennartz, C., Schiefer, J., Rotter, S., Hennig, J., & LeVan, P. (2018). Sparse estimation of resting-state effective connectivity from fMRI cross-spectra. *Frontiers in Neuroscience, 12*, 19. https://doi.org/10.3389/fnins.2018.00287.

Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J. M., Schluep, M., Vuilleumier, P., & van de Ville, D. (2013). Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. *Neuroimage, 83*, 937–950. https://doi.org/10.1016/j.neuroimage.2013.07.019.

Li, Y., Jewells, V., Kim, M., Chen, Y. S., Moon, A., Armao, D., et al. (2013). Diffusion tensor imaging based network analysis detects alterations of Neuroconnectivity in patients with clinically early relapsing-remitting multiple sclerosis. *Human Brain Mapping, 34*(12), 3376–3391. https://doi.org/10.1002/hbm.22158.

Li, Y., Wee, C. Y., Jie, B., Peng, Z. W., & Shen, D. G. (2014). Sparse multivariate autoregressive modeling for mild cognitive impairment classification. *Neuroinformatics, 12*(3), 455–469. https://doi.org/10.1007/s12021-014-9221-x.

Li, Y., Wang, X., Luo, L., Li, K., Yang, X., & Guo, Q. (2017). Epileptic seizure classification of eegs using time-frequency analysis based multiscale radial basis functions. *IEEE Journal of Biomedical and Health Informatics, 22*(2), 386–397. https://doi.org/10.1109/JBHI.2017.2654479.

Li, Y., Cui, W. G., Guo, Y. Z., Huang, T., Yang, X. F., & Wei, H. L. (2018a). Time-varying system identification using an ultra-orthogonal forward regression and multiwavelet basis functions with applications to EEG. *IEEE Transactions on Neural Networks & Learning Systems, 29*(7), 2960–2972. https://doi.org/10.1109/TNNLS.2017.2709910.

Li, Y., Cui, W. G., Luo, M. L., Li, K., & Wang, L. N. (2018b). Epileptic seizure detection based on time-frequency images of EEG signals using gaussian mixture model and gray level co-occurrence matrix features. *International Journal of Neural Systems, 28*(7), 1850003. https://doi.org/10.1142/S012906571850003X.

Li, Y., Liu, J., Huang, J., Li, Z., & Liang, P. (2018c). Learning brain connectivity sub-networks by group- constrained sparse inverse covariance estimation for Alzheimer's disease classification. *Frontiers in Neuroinformatics, 12*. https://doi.org/10.3389/fninf.2018.00058.

Li, Y., Yang, H., Lei, B., Liu, J., & Wee, C.-Y. (2018d). Novel effective connectivity inference using ultra-group constrained orthogonal forward regression and elastic multilayer perceptron classifier for MCI identification. *IEEE Transactions on Medical Imaging*, 1. https://doi.org/10.1109/tmi.2018.2882189.

Li, Y., Cui, W. G., Huang, H., Guo, Y. Z., Li, K., & Tan, T. (2019a). Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the Fisher vector approach. *Knowledge-Based Systems, 164*(15), 96–106. https://doi.org/10.1016/j.knosys.2018.10.029.

Li, Y., Liu, J., Gao, X., Jie, B., Kim, M., Yap, P.-T., Wee, C. Y., & Shen, D. (2019b). Multimodal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification. *Medical Image Analysis, 52*, 80–96. https://doi.org/10.1016/j.media.2018.11.006.

Liao, X. H., Vasilakos, A. V., & He, Y. (2017). Small-world human brain networks: Perspectives and challenges. *Neuroscience and Biobehavioral Reviews, 77*, 286–300. https://doi.org/10.1016/j.neubiorev.2017.03.018.

Liu, F., Wee, C. Y., Chen, H. F., & Shen, D. G. (2014). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification. *Neuroimage, 84*, 466–475. https://doi.org/10.1016/j.neuroimage.2013.09.015.

Liu, F., Wang, Y. F., Li, M. L., Wang, W. Q., Li, R., Zhang, Z. Q., et al. (2017). Dynamic functional network connectivity in idiopathic generalized epilepsy with generalized tonic-clonic seizure. *Human Brain Mapping, 38*(2), 957–973. https://doi.org/10.1002/hbm.23430.

Mao, H. F., Chang, L. H., Tsai, A. Y. J., Huang, W. N. W., Tang, L. Y., Lee, H. J., Sun, Y., Chen, T. F., Lin, K. N., Wang, P. N., Shyu, Y. I. L., & Chiu, M. J. (2018). Diagnostic accuracy of instrumental activities of daily living for dementia in community-dwelling older adults. *Age and Ageing, 47*(4), 551–557. https://doi.org/10.1093/ageing/afy021.

Matsuda, H. (2013). Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease. *Aging and Disease, 4*(1), 29–37.

McKenna, F., Koo, B. B., & Killiany, R. (2016). Comparison of ApoE-related brain connectivity differences in early MCI and normal aging populations: An fMRI study. *Brain Imaging and Behavior, 10*(4), 970–983. https://doi.org/10.1007/s11682-015-9451-z.

McKenzie, D. P., Downing, M. G., & Ponsford, J. L. (2018). Key Hospital Anxiety and Depression Scale (HADS) items associated with DSM-IV depressive and anxiety disorder 12-months post traumatic brain injury. *Journal of Affective Disorders, 236*, 164–171. https://doi.org/10.1016/j.jad.2018.04.092.

Mi, Z. P., Abrahamson, E. E., Ryu, A. Y., Fish, K. N., Sweet, R. A., Mufson, E. J., et al. (2017). Loss of precuneus dendritic spines immunopositive for spinophilin is related to cognitive impairment in early Alzheimer's disease. *Neurobiology of Aging, 55*, 159–166. https://doi.org/10.1016/j.neurobiolaging.2017.01.022.

NeuroimageFagiolo, G. (2007). Clustering in complex directed networks. *Physical Review E, 76*(2), 026107. https://doi.org/10.1103/PhysRevE.76.026107.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine, 256*(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x.

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., Ritchie, K., Rossor, M., Thal, L., & Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology, 58*(12), 1985–1992. https://doi.org/10.1001/archneur.58.12.1985.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron, 72*(4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006.

Qi, Z. G., Wu, X., Wang, Z. Q., Zhang, N., Dong, H. Q., Yao, L., et al. (2010). Impairment and compensation coexist in amnestic MCI default mode network. *Neuroimage, 50*(1), 48–55. https://doi.org/10.1016/j.neuroimage.2009.12.025.

Ribeiro, L. G., & Busatto Filho, G. (2016). Voxel-based morphometry in Alzheimers disease and mild cognitive impairment: Systematic review of studies addressing the frontal lobe. *Dementia & Neuropsychologia, 10*(2), 104–112. https://doi.org/10.1590/s1980-5764-2016dn1002006.

Rosa, M. J., Portugal, L., Hahn, T., Fallgatter, A. J., Garrido, M. I., Shawe-Taylor, J., & Mourao-Miranda, J. (2015). Sparse network-based models for patient classification using fMRI. *Neuroimage, 105*, 493–506. https://doi.org/10.1016/j.neuroimage.2014.11.021.

Rose, S. E., Mcmahon, K. L., Janke, A. L., O'Dowd, B., De, Z. G., Strudwick, M. W., et al. (2006). Diffusion indices on magnetic resonance imaging and neuropsychological performance in amnesic mild cognitive impairment. *Journal of Neurology Neurosurgery & Psychiatry, 77*(10), 1122–1128. https://doi.org/10.1136/jnnp.2005.074336.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage, 52*(3), 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003.

Ryali, S., Chen, T. W., Supekar, K., & Menon, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage, 59*(4), 3852–3861. https://doi.org/10.1016/j.neuroimage.2011.11.054.

Salvatore, C., Cerasa, A., Battista, P., Gilardi, M. C., Quattrone, A., & Castiglioni, I. (2015). Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach. *Frontiers in Neuroscience, 9*. https://doi.org/10.3389/fnins.2015.00307.

Sandanalakshmi, R., & Sardius, V. (2016). Selected saliency based analysis for the diagnosis of Alzheimer's disease using structural magnetic resonance image. *Journal of Medical Imaging and Health Informatics, 6*(1), 177–184. https://doi.org/10.1166/jmihi.2016.1610.

Shah, S. A. A., Aziz, W., Arif, M., & Nadeem, M. S. A. Decision Trees based Classification of Cardiotocograms using Bagging Approach. In *13th International Conference on Frontiers of Information*

Technology, New York, 2015 (pp. 12–17): IEEE. https://doi.org/10.1109/fit.2015.14.

Suk, H. I., Lee, S. W., & Shen, D. G. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *Neuroimage, 101,* 569–582. https://doi.org/10.1016/j.neuroimage.2014.06.077.

Sun, G. H., Raji, C. A., MacEachern, M. P., & Burke, J. F. (2012). Olfactory identification testing as a predictor of the development of Alzheimer's dementia: A systematic review. *Laryngoscope, 122*(7), 1455–1462. https://doi.org/10.1002/lary.23365.

Supekar, K., Menon, V., Rubin, D., Musen, M., & Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Computational Biology, 4*(6), e1000100. https://doi.org/10.1371/journal.pcbi.1000100.

Sze, C. I., Troncoso, J. C., Kawas, C., Mouton, P., Price, D. L., & Martin, L. J. (1997). Loss of the presynaptic vesicle protein synaptophysin in hippocampus correlates with cognitive decline in Alzheimer disease. *Journal of Neuropathology and Experimental Neurology, 56*(8), 933–944. https://doi.org/10.1097/00005072-199708000-00011.

Takahashi, M., Oda, Y., Okubo, T., & Shirayama, Y. (2017). Relationships between cognitive impairment on ADAS-cog and regional cerebral blood flow using SPECT in late-onset Alzheimer's disease. *Journal of Neural Transmission, 124*(9), 1109–1121. https://doi.org/10.1007/s00702-017-1734-7.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978.

Van Den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology, 20*(8), 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008.

Van Patten, R., Britton, K., & Tremont, G. (2018). Comparing the minimental state examination and the modified mini-mental state examination in the detection of mild cognitive impairment in older adults. *International Psychogeriatrics, 1,* 1–9. https://doi.org/10.1017/s1041610218001023.

Vasavada, M. M., Wang, J. L., Eslinger, P. J., Gill, D. J., Sun, X. Y., Karunanayaka, P., et al. (2015). Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment. *Journal of Alzheimers Disease, 45*(3), 947–958. https://doi.org/10.3233/Jad-141947.

Wang, L., Xue, W., Li, Y., Luo, M., Huang, J., Cui, W., & Huang, C. (2017). Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis. *Entropy, 19*(6). https://doi.org/10.3390/e19060222.

Wee, C. Y., Yap, P. T., Denny, K., Browndyke, J. N., Potter, G. G., Welsh-Bohmer, K. A., Wang, L., & Shen, D. (2012a). Resting-state multi-Spectrum functional connectivity networks for identification of MCI patients. *PLoS One, 7*(5), 11. https://doi.org/10.1371/journal.pone.0037828.

Wee, C. Y., Yap, P. T., Zhang, D. Q., Denny, K., Browndyke, J. N., Potter, G. G., et al. (2012b). Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage, 59*(3), 2045–2056. https://doi.org/10.1016/j.neuroimage.2011.10.015.

Wee, C. Y., Yap, P. T., Zhang, D. Q., Wang, L. H., & Shen, D. G. (2014). Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Structure & Function, 219*(2), 641–656. https://doi.org/10.1007/s00429-013-0524-8.

Wee, C. Y., Yang, S., Yap, P. T., & Shen, D. G. (2016). Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. *Brain Imaging and Behavior, 10*(2), 342–356. https://doi.org/10.1007/s11682-015-9408-2.

Xu, L. L., Wu, X., Li, R., Chen, K. W., Long, Z. Y., Zhang, J. C., et al. (2016). Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers. *Journal of Alzheimers Disease, 51*(4), 1045–1056. https://doi.org/10.3233/Jad-151010.

Zhang, J. H., Chen, M., Zhao, S. K., Hu, S. Q., Shi, Z. G., & Cao, Y. (2016). ReliefF-based EEG sensor selection methods for emotion recognition. *Sensors, 16*(10). https://doi.org/10.3390/s16101558.

Zhou, L. P., Wang, L., Liu, L. Q., Ogunbona, P., & Shen, D. G. (2013). Discriminative brain effective connectivity analysis for Alzheimer's disease: A kernel learning approach upon sparse Gaussian Bayesian network. *2013 IEEE conference on computer vision and pattern recognition*, 2243–2250, https://doi.org/10.1109/Cvpr.2013.291, 2013.

Zhu, X. F., Suk, H. I., & Shen, D. G. (2014). A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *Neuroimage, 100*, 91–105. https://doi.org/10.1016/j.neuroimage.2014.05.078.

Zhu, Y. Y., Zhu, X. F., Zhang, H., Gao, W., Shen, D. G., & Wu, G. R. (2016). Reveal consistent spatial-temporal patterns from dynamic functional connectivity for autism Spectrum disorder identification. *International conference on medical image computing and computer-assisted intervention*, 9900, 106–114, https://doi.org/10.1007/978-3-319-46720-7_13.

Ziegler-Graham, K., Brookmeyer, R., Johnson, E., & Arrighi, H. M. (2008). Worldwide variation in the doubling time of Alzheimer's disease incidence rates. *Alzheimers & Dementia, 4*(5), 316–323. https://doi.org/10.1016/j.jalz.2008.05.2479.