

Functional Brain Network Estimation with Time Series Self-scrubbing

Weikai Li, Lishan Qiao*, Zhengxia Wang, and Dinggang Shen

Abstract—Functional brain network (FBN) has been becoming an increasingly important measurement for exploring the cerebral working mechanism and mining informative biomarkers for assisting diagnosis of some neurodegenerative disorders. Despite its potential performance in discovering the valuable patterns hidden in the brains, the estimated FBNs are often heavily influenced by the quality of the observed data (e.g., BOLD signal series). In practice, a preprocessing pipeline is usually employed for improving the data quality prior to the FBN estimation; but, even so, some data points in the time series are still not clean enough, possibly including original artifacts (e.g., micro head motion), non-resting functional disturbing (e.g., mind-wandering), and new “noises” caused by the preprocessing pipeline per se. Therefore, not all data points in the time series can contribute to the subsequent FBN estimation. To address this issue, in this paper, we propose a novel FBN estimation method by introducing a latent variable as an indicator of the data quality, and develop an alternating optimization algorithm for scrubbing the data and estimating FBN simultaneously in a single framework. As a result, we can obtain more accurate FBNs with the self-scrubbing data. To illustrate the effectiveness of the proposed method, we conduct experiments on two publicly available datasets to identify mild cognitive impairment (MCI) patients from normal control (NC) subjects based on the estimated FBNs. Experimental results show that the proposed FBN modelling method can achieve higher classification accuracy, significantly outperforming the baseline methods.

Index Terms—Functional Brain Network; Resting-state Functional Magnetic Resonance Imaging (fMRI); Mild Cognitive Impairment (MCI)

I. INTRODUCTION

FUNCTIONAL brain network (FBN) provides an increasingly important way to explore the brain integration mechanism [1], discover potential brain activities [2], and mine sensitive biomarkers for neural disease diagnosis [3,4] (such as autism spectrum disorder [5-7], Alzheimer’s disease [8,9] and Parkinson’s disease [10]). All these rely heavily on the quality of the final FBNs, and thus it becomes essential to estimate an accurate FBN [11].

* Corresponding author.

This work was partly supported by National Natural Science Foundation of China (61300154, 61402215), Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ1500501), Chongqing Graduate Student Research Innovation Project (CYS16183), and NIH grants (EB022880, AG049371, AG042599).

Weikai Li is with the School of Mathematics Science, Liaocheng University, Liaocheng 252000, China and with College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China (e-mail: leweikai@outlook.com).

Currently, researchers have proposed many different methods towards better FBN estimation, and most of them can be explained under a regularized framework that requires an accurate FBN estimation model to *not only* fit the data, but also effectively encode the priors of brain organization [12]. In practice, the commonly-used priors include sparsity [13-16], group-sparsity [17,18], low-rank [19], and modularity [12], which can be transformed into their corresponding regularization terms in the FBN estimation models and can often improve the performance of the obtained FBNs.

Besides the regularizers, the data-fitting terms (see Eq. (5) and related comments for details) also have a high influence on FBN estimation. However, the artifacts or noises involved in the observed data (or time series) usually lead to a poor fitting result. Therefore, in practice, a preprocessing pipeline, including motion correction, spatial smoothing and temporal filtering, is generally employed to improve the quality of the data before FBN estimation [20]. Even so, it is hard to eliminate all the artifacts/noises in the data due to the weak fMRI signals and complex disturbing sources. Moreover, some preprocessing steps (e.g., spatial normalization) in the pipeline may also cause new “dirty” points in the time series.

As a recently developed preprocessing step, the scrubbing operation has been investigated to further clean the data by removing some potentially “dirty” time points from the fMRI series [21]. Despite its seeming appeal, there are also some debates on such a scheme [21-25]. Moreover, this scheme mainly focuses on specific physical artifacts (e.g., micro head motion) that can only cover a limited range of problematic time points. In fact, a large number of factors may introduce “dirty” data points into the fMRI series. For example, the “resting-state” fMRI data tend to involve many different functional processes, e.g., mind-wandering [26], thus resulting in non-resting-state time points in the fMRI data that cannot be detected by the traditional scrubbing technique. More notably, the scrubbing operation is independent of the FBN estimation model used, and thus cannot guarantee that the preserved data points can necessarily benefit for the subsequent FBN estimation, while the removed data points are not necessarily helpful. In addition,

Lishan Qiao is with the School of Mathematics Science, Liaocheng University, Liaocheng 252000, China (e-mail: qiaolishan@lcu.edu.cn).

Zhengxia Wang is with the College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China (e-mail: zxiawang@163.com).

Dinggang Shen is with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA and with Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea. (e-mail: dgshen@med.unc.edu).

it is hard for the common scrubbing method to control the length of the remaining time series [21], and thus the final series sometimes cannot contain enough time points (samples) for estimating a reliable FBN [12].

To address these issues, in this paper, we propose a novel FBN estimation strategy by introducing a latent indicator variable into the FBN optimization model. The latent variable indicates the quality or state of each time point in the fMRI series, and can be learned automatically with the FBN together from the data. Based on the latent variable model [27,28], we then develop an alternating optimization algorithm to simultaneously estimate the indicator variable and FBN in a single framework. Consequently, our proposed method will be able to automatically identify and remove the “dirty” time points from the fMRI series; that is, it can make a self-scrubbing operation at the FBN estimation procedure. In the following, the contributions and the main advantages of our proposed framework are summarized.

1) Different from traditional methods that often conduct the data scrubbing and FBN estimation in two separate sequential steps, our proposed framework combines data scrubbing and FBN estimation into a single model. By joint optimization, we can obtain FBNs with potentially higher accuracy and efficiency.

2) Technically, we introduce a latent variable into the FBN estimation model as an indicator of the data quality, and design an alternating approach to estimate the optimal indicator, by which we can scrub the time points that are potentially not helpful for FBN estimation. Moreover, compared with the traditional scrubbing operation that often removes too many data points, our proposed framework can control the size of scrubbed data by a hyper-parameter.

3) Finally, it is worth emphasizing that our proposed framework is not competitive to the original scrubbing strategy, since it can also work on the data that have been already scrubbed by the traditional methods. Therefore, the proposed method is more flexible and adaptive for the FBN estimation than the traditional scrubbing scheme.

To verify the effectiveness of the proposed method, we apply it to estimate FBNs based on the resting-state functional magnetic resonance imaging (R-fMRI), and then identify mild cognitive impairment (MCI) patients from normal control (NC) subjects via the estimated FBNs. Experiments are conducted on two publicly available datasets, and the experimental results illustrate that our proposed method works well on both scrubbed and non-scrubbed R-fMRI data. For facilitating efforts to replicate our results, we have also shared both pre-processed data and codes in <https://github.com/Cavin-Lee/self-scrubbing/>.

The rest of this paper is organized as follows. In Section II, we review the main-stream FBN estimation methods. In Section III, we propose a novel FBN estimation framework, including motivation, modeling, and algorithm. In Section IV, we conduct experiments on a simulated data and two real-world datasets. In Section V, we conclude the paper with a brief discussion.

II. RELATED WORKS

In this paper, we mainly focus on the correlation-based methods, which are currently the most popular ways of FBN estimation and have been empirically demonstrated to be more sensitive than the complex or higher-order methods [11]. In the following, we review several representative correlation-based methods that, in fact, provides a platform for developing our model.

A. Pearson’s Correlation

According to a recent review [1], Pearson’s Correlation (PC) is the simplest method for estimating FBNs. We suppose that each brain has been parcellated into N regions of interest (ROIs) based on a certain atlas, and the fMRI time series associated with the i th ROI is represented by $\mathbf{x}_i \in R^T$, $i = 1, \dots, N$, where T is the number of time points in each series. Then, the edge weights of the FBN based on PC can be calculated as follows:

$$W_{ij} = \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T (\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T (\mathbf{x}_i - \bar{\mathbf{x}}_i)} \sqrt{(\mathbf{x}_j - \bar{\mathbf{x}}_j)^T (\mathbf{x}_j - \bar{\mathbf{x}}_j)}} \quad (1)$$

By defining a new $\mathbf{x}_i \triangleq (\mathbf{x}_i - \bar{\mathbf{x}}_i) / \sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T (\mathbf{x}_i - \bar{\mathbf{x}}_i)}$, a centralized and normalized counterpart of the original \mathbf{x}_i , we can simplify the PC as $W_{ij} = \mathbf{x}_i^T \mathbf{x}_j$, which corresponds to the solution of the following optimization problem:

$$\min_{\mathbf{W}} \|\mathbf{W} - \mathbf{X}^T \mathbf{X}\|_F^2 \quad (2)$$

Here, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in R^{T \times N}$ is the data matrix, \mathbf{W} is the edge weight matrix, and $\|\cdot\|_F$ denotes the F-norm of a matrix. Since the BOLD signals commonly contain noises, the original PC tends to result in a FBN with dense connections [29]. In practice, a threshold scheme is generally used to sparsify the PC-based FBN by filtering out the noisy or weak connections. For detailed discussion on the thresholding strategy, please refer to Section 3.2.1 in [29].

B. Partial Correlation

Partial correlation is used in FBN estimation for treating the confounding problem involved in the full correlation methods such as PC [30]. A general approach to calculate partial correlation is based on the estimation of inverse covariance matrix [31]. However, this approach may be ill-posed due to the singularity of the estimated sample covariance matrix $\mathbf{\Sigma} = \mathbf{X}^T \mathbf{X}$. To alleviate this issue, a regularization term $R(\mathbf{W})$ is generally introduced into the FBN estimation model as follows.

$$\min_{\mathbf{W}} \sum_{i,j=1}^n \|\mathbf{x}_i - W_{ij} \mathbf{x}_j\|^2 + \lambda R(\mathbf{W}) \quad (3)$$

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

$$\min_{\mathbf{W}} \|\mathbf{X} - \mathbf{XW}\|_F^2 + \lambda R(\mathbf{W}) \quad (4)$$

In Eq. (3) or (4), the first term implies to invert the covariance matrix $\mathbf{\Sigma}$ [31], and λ is a regularized parameter to control the balance between the first (data-fitting) term and the second (regularization) term. The most popular regularizer is l_1 -norm, i.e., $R(\mathbf{W}) = \|\mathbf{W}\|_1 = \sum_{i,j=1}^n |W_{ij}|$, which corresponds to the sparsity prior of FBN, and leads to the LASSO or sparse representation (SR) model [15,32,33]. As discussed previously,

some other regularizers have been investigated in recent years to encode different priors, which, however, is beyond our main focus of this paper.

C. A Regularized FBN Estimation Framework

According to a recent study [12], a large family of FBN estimation models can be summarized by the following regularized framework:

$$\min_{\mathbf{W}} f(\mathbf{X}, \mathbf{W}) + \lambda R(\mathbf{W}), \text{ s. t. } \mathbf{W} \in \Delta \quad (5)$$

where $f(\mathbf{X}, \mathbf{W})$ is a data-fitting term, aiming to capture some statistics of the data (e.g., the covariance or inverse covariance structure), and $R(\mathbf{W})$ is the regularization term, aiming to encode the biological/physical priors of the FBN. Sometimes, specific constraints (e.g., symmetry or positive semi-definite) are included in Δ for shrinking the search space of \mathbf{W} towards better FBNs.

Such a regularized framework can *not only* stabilize the statistical estimation, *but also*, more importantly, provide a general platform for designing the new FBN estimation method in this paper.

III. THE PROPOSED METHOD

As mentioned earlier, both prior information and data quality have a significant influence on the performance of the estimated FBNs. The priors can be generally encoded by the regularization terms. In contrast, however, the quality of time points in the fMRI series cannot be easily measured without the guidance of a specific learning/estimation task. Therefore, in this section, we propose a novel solution for identifying the quality of time points with the FBN estimation (task) together, by which we expect to get cleaner data, and, in turn, more accurate FBNs.

A. Motivation

Before presenting the novel FBN estimation model, we first introduce the motivation and the basic idea behind it by a toy example. As shown in Fig. 1, we have a set of fMRI series from $N=5$ ROIs, each of which includes $T=30$ time points. According to the previous discussion, even a sophisticated preprocessing pipeline cannot guarantee all the T time points in the fMRI series are clean. To illustrate this more clearly, some data points in Fig. 1 are labelled as “clean”, while the others are labelled as “dirty”. Equivalently, a binary indicator can also be used to denote the quality of the data, with 0 corresponding to “dirty”, and 1 to “clean”.

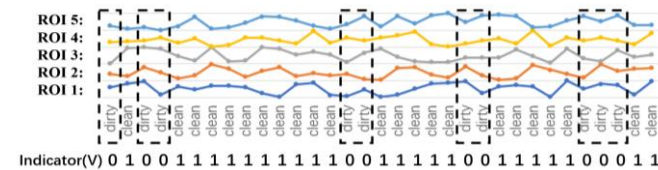


Fig. 1. A toy example for illustrating the motivation of the proposed method.

However, the indicator cannot be observed directly in the real-world applications. Traditionally, a data-scrubbing operation is used to identify the quality of the data points (or, equivalently, determine the value of the indicator) by detecting the micro head motion [21]. Despite its potential effectiveness (also with some debates [23-25]), such a scheme 1) cannot remove all the “dirty” data points caused by different factors; 2) tends to scrub too many time points due to the lack of a control mechanism [12,21]; 3) is independent on the specific task, and thus will not necessarily benefit the ensuing FBN estimation.

B. Model

To address the above problems, in this section, we propose a task-dependent data scrubbing scheme for FBN estimation. In particular, we consider the binary indicator as a variable v_t for denoting the quality of the t th time point ($v_t = 0$ for “dirty”, and $v_t = 1$ for “clean”), and thus the regularized FBN estimation framework in Eq. (5) can be extended to the following form with an indicator variable.

$$\min_{\mathbf{W}, v_t \in \{0,1\}} \sum_{i=1}^N \sum_{t=1}^T v_t f(\mathbf{X}^{(t)}, \mathbf{W}) + \lambda R(\mathbf{W}), \quad (6)$$

where $\mathbf{X}^{(t)}$ is the t th row of the data matrix \mathbf{X} . Note that, when the indicator variable $v_t = 0$, the t th time point in the fMRI series will be removed, meaning that it has no contribution to the FBN estimation; when $v_t = 1$ for all $t = 1, 2, \dots, T$, Eq. (6) will reduce to the original FBN estimation framework given in Eq. (5).

However, in practice, Eq. (6) will always get a trivial solution (i.e., $v_t = 0$ for all $t = 1, 2, \dots, T$), since it can really minimize the objective function. Therefore, we introduce a negative l_1 -norm regularization term, $-\gamma \sum_{t=1}^T |v_t|$, into Eq. (6). Additionally, we relax¹ the binary indicator variable $v_t \in \{0,1\}$ to a real range $[0,1]$, for simplifying the solving of the optimization problem. Then, we get the following model:

$$\min_{\mathbf{W}, v_t \in [0,1]} \sum_{i=1}^N \sum_{t=1}^T v_t f(\mathbf{X}^{(t)}, \mathbf{W}) + \lambda R(\mathbf{W}) - \gamma \sum_{t=1}^T |v_t|, \quad (7)$$

where λ and γ are two regularized parameters for controlling the balance among the three terms in the objective function. Especially for γ , it plays a role to determine the number of removed time points from the whole series. When γ has a small value close to 0, most of the time points will be removed. On the other hand, when γ has a large value, all the time points in the series will be kept for FBN estimation. In other words, the proposed model can scrub the data adaptively in the process of the FBN estimation, by controlling the hyper-parameter γ and learning the indicator variable v_t from the data. Therefore, we name our proposed scheme as FBN estimation with self-scrubbing (SS model for short).

In principle, any data-fitting term can be used to realize the SS model, but, in this paper, we adopt the partial correlation strategy shown in Eq. (3) or (4), since it overcomes the confounding effect and is empirically verified to be more effective [11] than the full correlation. Consequently, the partial correlation based self-scrubbing FBN estimation model is given

¹ In Section C, we will find that such a relaxation is tight, thus resulting in a binary solution for v_t .

as follows:

$$\min_{\mathbf{W}, v_t \in [0,1]} \sum_{i=1}^N \sum_{t=1}^T v_t (x_i^{(t)} - \sum_{j \neq i} W_{ij} x_j^{(t)})^2 + \lambda R(\mathbf{W}) - \gamma \sum_{t=1}^T |v_t|, \quad (8)$$

where $x_i^{(t)}$ is the t th time point of the series associated with the i th node (ROI). For simplicity, we rewrite Eq. (8) into the following matrix form:

$$\min_{\mathbf{W}, \mathbf{V}} \|\mathbf{V}\mathbf{X} - \mathbf{V}\mathbf{X}\mathbf{W}\|^2 + \lambda R(\mathbf{W}) - \gamma \|\mathbf{V}\|_1 \quad (9)$$

s. t. $W_{ii} = 0, 0 \leq v_t \leq 1, \forall i, t$

where $\mathbf{V} = \text{diag}(v_1, v_2, \dots, v_T) \in R^{T \times T}$ is a diagonal matrix containing the indicator variables on its principal diagonal. The constraint, $W_{ii} = 0$, is employed only to avoid the trivial solution that leads to \mathbf{W} being an identity matrix. For $R(\mathbf{W})$, we can, in principle, use any off-the-shelf regularizers, such as l_1 -norm, $l_{2,1}$ -norm [18,36], trace norm and their combination. However, this problem goes beyond our main focus in this paper. Therefore, we only attempt the l_1 -norm (sparsity prior) due to its simplicity and effectiveness [32], and get the specific FBN estimation model (named **SR+SS**) as follows.

$$\min_{\mathbf{W}, \mathbf{V}} \|\mathbf{V}\mathbf{X} - \mathbf{V}\mathbf{X}\mathbf{W}\|^2 + \lambda \|\mathbf{W}\|_1 - \gamma \|\mathbf{V}\|_1 \quad (10)$$

s. t. $W_{ii} = 0, 0 \leq v_t \leq 1, \forall i, t$

C. Optimization Algorithm

Considering that there are two variables \mathbf{V} and \mathbf{W} involved in Eq. (10), in this paper we employ the alternative convex search (ACS) [37] method to solve them alternately by the two following steps. Before that, we first initialize the indicator \mathbf{V} as an identity matrix, meaning that we select all time points in the first iteration.

Step 1: With a fixed \mathbf{V} , Eq. (10) reduces to a traditional FBN estimation problem that can be solved by many convex optimization methods. Here, we use the proximal method [38] due to its efficiency and simplicity. Two main steps are involved in the proximal method, including gradient descent and proximal operation. First, for the data-fitting term $f(\mathbf{X}, \mathbf{W}) = \|\mathbf{V}\mathbf{X} - \mathbf{V}\mathbf{X}\mathbf{W}\|_F^2$ whose gradient w.r.t \mathbf{W} is $\nabla_{\mathbf{W}} f(\mathbf{X}, \mathbf{W}) = 2\mathbf{X}^T \mathbf{V}^T \mathbf{V}\mathbf{X}\mathbf{W} - \mathbf{X}^T \mathbf{V}^T \mathbf{V}\mathbf{X}$, we have the following update formula according to the gradient descent criterion:

$$\mathbf{W}_k = \mathbf{W}_{k-1} - \alpha_k \nabla_{\mathbf{W}} f(\mathbf{X}, \mathbf{W}_{k-1}), \quad (11)$$

where α_k denotes the step size of the gradient descent. Then, the proximal operator is imposed on the current \mathbf{W} . For the sparsity regularizer $\lambda \|\mathbf{W}\|_1$, the proximal operator is defined [38] as follows.

$$\text{prox}_{\lambda \|\cdot\|_1}(\mathbf{W}) = [\text{sgn}(W_{ij}) \times \max(\text{abs}(W_{ij}) - \lambda, 0)]_{N \times N} \quad (12)$$

where $\text{sgn}(W_{ij})$ and $\text{abs}(W_{ij})$ return the sign and absolute value of W_{ij} , respectively.

Step 2: With a fixed \mathbf{W} , we update \mathbf{V} . Now, Eq. (10) reduces to the following optimization problem,

$$\min_{\mathbf{V}} \|\mathbf{V}\mathbf{X} - \mathbf{V}\mathbf{X}\mathbf{W}\|^2 - \gamma \|\mathbf{V}\|_1, \quad \text{s. t. } 0 \leq v_t \leq 1, \forall t \quad (13)$$

or

$$\min_{v_t \in [0,1]} \sum_{i=1}^N \sum_{t=1}^T v_t (x_i^{(t)} - \sum_{j \neq i} W_{ij} x_j^{(t)})^2 - \gamma \sum_{t=1}^T |v_t| \quad (14)$$

Eq. (14) can be further simplified to the following problem.

$$\min_{v_t \in [0,1]} \sum_{t=1}^T (c_t - \gamma) v_t, \quad (15)$$

where $c_t = \sum_{i=1}^N (x_i^{(t)} - \sum_{j \neq i} W_{ij} x_j^{(t)})^2 = \|\mathbf{X}^{(t)} - \mathbf{X}^{(t)}\mathbf{W}\|^2$ is a constant. We note that Eq. (15) is a linear programming problem, and thus can easily get its optimal solution as follows.

$$v_t = \begin{cases} 1, & c_t = \|\mathbf{X}^{(t)} - \mathbf{X}^{(t)}\mathbf{W}\|^2 < \gamma \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

This means that 1) if $\|\mathbf{X}^{(t)} - \mathbf{X}^{(t)}\mathbf{W}\|^2 > \gamma$, the t th time point is more likely to be removed by labelling it with 0; on the contrary, 2) if $\|\mathbf{X}^{(t)} - \mathbf{X}^{(t)}\mathbf{W}\|^2 < \gamma$, the t th time point will be kept by labelling it with 1. Such a formula for updating v_t coincides well with the intuition that the ‘‘dirty’’ time points (labelled as 0) can cause a poor fitting with high bias/residual (i.e., $\|\mathbf{X}^{(t)} - \mathbf{X}^{(t)}\mathbf{W}\|^2 > \gamma$). In fact, in the experimental section, we will further illustrate this problem based on a simulated dataset, and empirically verify that the proposed method can automatically detect the ‘‘dirty’’ time points. Finally, we summarize the algorithm for solving Eq. (10) in ALGORITHM I.

ALGORITHM I ESTIMATING FBN WITH SELF-SCRUBBING

Input: $\mathbf{X}, \lambda, \gamma$
Output: \mathbf{W}, \mathbf{V}

Initialize \mathbf{V} ;
while *not converged*
 while *not converged*
 $\mathbf{W}_{k+1} = \mathbf{W}_k - 2\alpha(\mathbf{X}^T \mathbf{V}^T \mathbf{V}\mathbf{X}\mathbf{W}_k - \mathbf{X}^T \mathbf{V}^T \mathbf{V}\mathbf{X})$;
 $\mathbf{W}_{k+1} = \text{prox}(\mathbf{W}_{k+1})$; // based on Eq. (12).
 end
 update \mathbf{V} by Eq. (16);
end
return \mathbf{W}, \mathbf{V} ;

IV. EXPERIMENTS

In this section, we first illustrate how the proposed method works by an experiment on simulated data in Section A, and then verify its effectiveness by identifying Mild Cognitive Impairment (MCI) from Normal Control (NC) based on the estimated FBNs in Section B.

A. An Illustrative Experiment on Simulated Data

For the convenience of interpretation and visualization, in this example, we only consider the simplest case that $N=2$. Thus, we have two time series, x_1 and x_2 , associated with ROI₁ and ROI₂, respectively. Without loss of generality, we suppose that there is a strong connection between ROI₁ and ROI₂. For simulating the strong connection (i.e., correlation), we generate the data by an approximately linear relationship that $x_1 \approx 5 \times x_2$, as shown in Fig. 2(a) and (b). Based on the generated data, we can easily calculate the correlation coefficient between ROI₁ and ROI₂, and the result is 0.985.

However, as discussed previously, the correlation can be significantly affected by the data quality. To illustrate this, we change several time points randomly in the generated data for simulating the artifacts or noises. In particular, as shown in Fig. 2(c) and (d), we 1) introduce a large-amplitude shake in the first

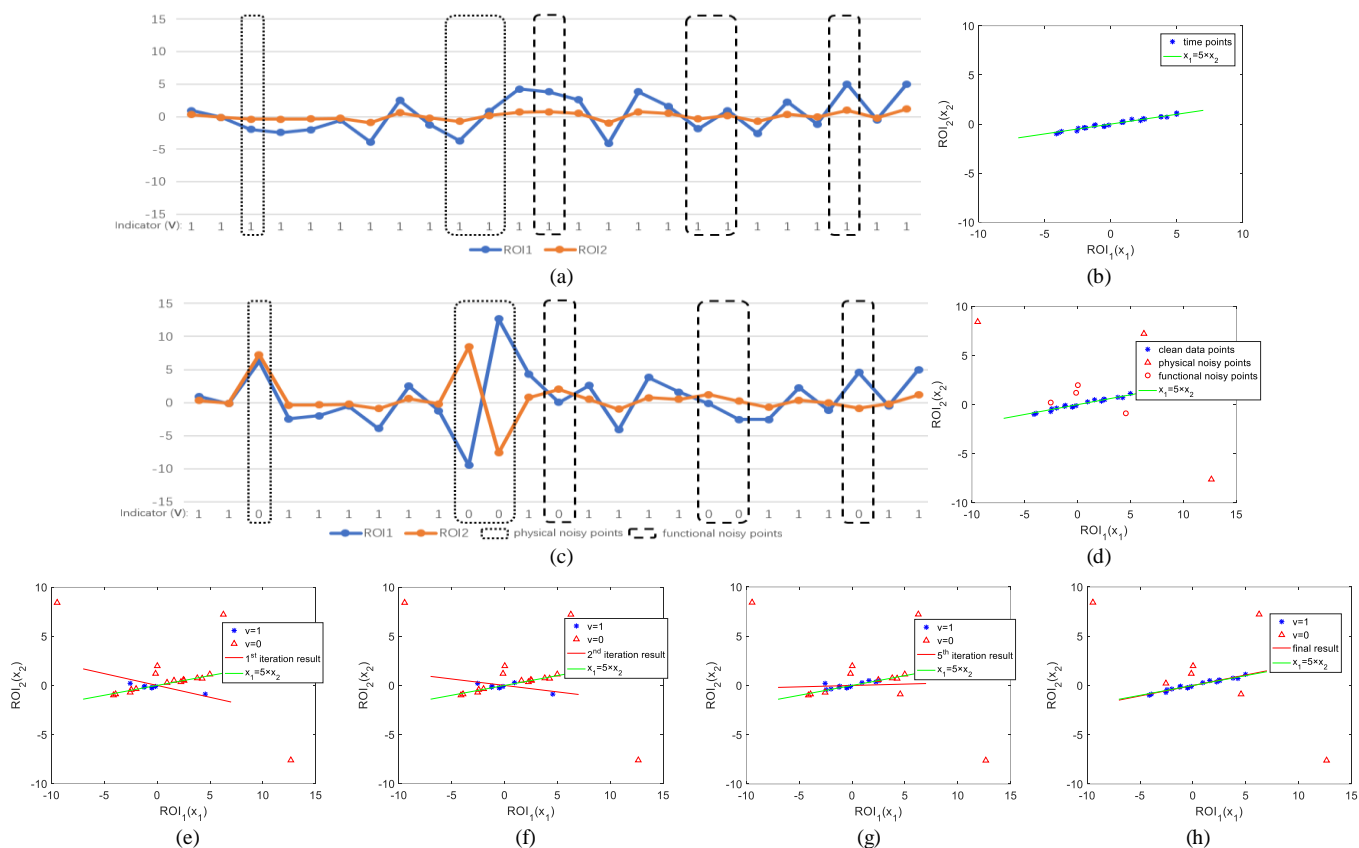


Fig. 2. A toy example for illustrating how the proposed method works. (a) Original ideal signals simulated without noises; (b) scatter plot corresponding to (a); (c) observed signals simulated by introducing 7 noisy points into the original signals in (a); (d) scatter plot corresponding to (b); (e)-(h) the iterative results of the proposed algorithm. Note that the proposed method can effectively remove the noisy points gradually with the iterations.

three “dirty” points to simulate the physical noises, since the physical noises (e.g., motion artifacts) have a high relation with the large-amplitude changes in signals [34,35]; 2) introduce four “dirty” points to simulate the possible functional noises, by setting them off the main direction (i.e., $x_1 = 5 \times x_2$). Here, we simply consider the functional noises as non-resting-state signals that lie in the normal range, but deviate from the resting-state signals, according to a recent study [39].

Now, the correlation between ROI₁ and ROI₂ is -0.399 , meaning that even limited change of data points can have a big influence on estimation of the functional connection. Based on the simulated data with “dirty” time points, we run the proposed algorithm, and find that it can remove the “dirty” points gradually with the iterations, as shown in Fig. 2(e) - (h).

B. MCI Identification

1) Data Acquisition and Preprocessing

In this study, we validate the proposed method by MCI and NC classification on two publicly available datasets. One is from Alzheimer’s Disease Neuroimaging Initiative (ADNI)², and the other is from Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC)³ shared by a recent study

[12].

For ADNI dataset, 110 participants, including 51 MCIs and 59 NCs, are adopted in this experiment. The fMRIs are obtained by 3.0T Philips scanners with the following parameters: TR/TE = 3000/30mm, flip angle = 80, imaging matrix=64×64, 48 slices, 140 volumes, and voxel thickness = 3.3mm. SPM8 toolbox⁴ and DPARSFA (version 2.2) [40] are used to preprocess the fMRI data according to the well accepted pipeline. The first 10 R-fMRI volumes of each subject are discarded to avoid signal shaking. The remaining images are first corrected for different slice acquisition timing and head motion [41]. Then, regression of ventricular and WM signals as well as six head-motion profiles are conducted to further reduce the effects of nuisance signals. Mean fMRI series of each ROI is band-pass filtered (0.01-0.08Hz). Depending on the automated anatomical labeling (AAL) atlas [42], the pre-processed BOLD time series signals are partitioned into 116 ROIs. At last, we put these time series into a data matrix $\mathbf{X} \in R^{137 \times 116}$.

For NITRC dataset, it includes 46 MCIs and 45 NCs that come from the participants recruited via advertisements in local newspapers and media. A similar preprocessing pipeline [12] are employed as in the ADNI dataset, except that a scrubbing operation is applied, for NITRC dataset, to remove the time

² <http://adni.loni.ucla.edu>

³ <http://www.nitrc.org/projects/modularbrain/>

⁴ <http://www.fil.ion.ucl.ac.uk/spm>

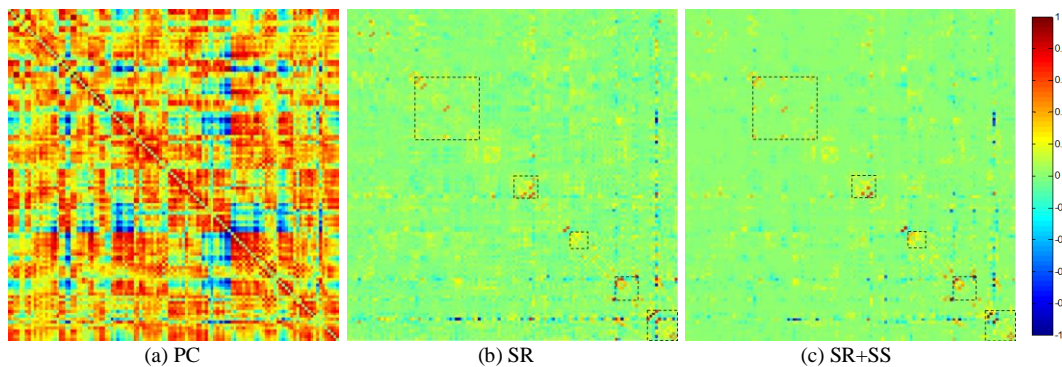


Fig. 3. The FBN adjacency matrices of a subject estimated by 3 different methods. The rectangles in (b) and (c) help for distinguishing the differences between the two figures.

points with frame-wise displacement larger than 0.5 [21]. We mainly use these two datasets for verifying that our proposed method can work well, whether the fMRI series are scrubbed or not.

2) FBN estimation

After obtaining the preprocessed fMRI data, we estimate FBNs based on three different methods, PC, SR, and the proposed SR+SS. In Fig. 3, we visualize the adjacency matrices⁵ of the FBN estimated by these methods. For SR, we simply set the regularized parameter $\lambda = 1$, and for SR+SS, $\lambda = 1$ and $\gamma = 0.5$.

As shown in Fig. 3, the FBN estimated by PC has a topology highly different from that of the partial correlation-based methods (i.e., SR and SR+SS), since they use different data fitting term. In contrast, SR and SR+SS lead to a similar FBN structure by using the same kind of data fitting term. The differences of the FBNs estimated by SR and SR+SS methods lie mainly in several specific brain regions. In order to examine which brain regions have changed the connections following the self-scrubbing, we simply sum up the connection weights of all the subjects for each region. The result is shown in Fig. 4, where the height of each bar represents the sum of connection weights changing across different brain regions.

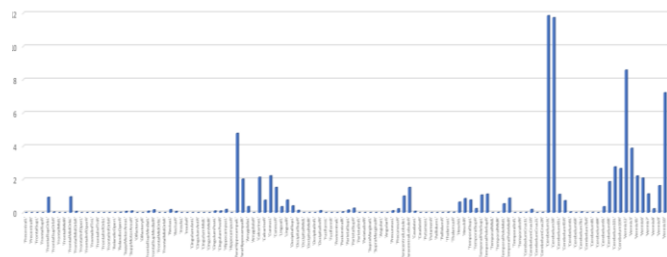


Fig. 4. The connection weight change of FBN on each node (brain region) between SR and SR+SS. We first calculate the absolute value of the difference between the connections estimated by SR and SR+SS respectively. Then, we sum up the different weight values of the subjects for each brain region.

As we can see in Fig. 4, by scrubbing the “noisy” time points, a large number of changes happen in cerebellum regions, which

is possibly related to the head motion or some other tiny movements. Also, the changes take place in hippocampus and frontal brain regions, which may be related to the psychological phenomena such as mind-wandering.

3) Feature selection and Classification

Once we obtain the FBNs of all subjects, the subsequent task is to classify the MCI and NC based on the estimated FBNs. Then, the problem turns to determine which features and classifiers should be used for classification. Considering the big influence of different steps in the classification pipeline on the final accuracy [18], it is difficult to conclude whether the FBN estimation methods or the ensuing feature selection and classification methods contribute to the ultimate result. Therefore, we only adopt the simplest feature selection method (t-test with $p < 0.01$) and the most popular SVM [43] classifier (linear SVM with default parameter $C=1$) in our experiment.

Due to limited samples, we test the involved methods using the leave-one-out cross validation (LOOCV) strategy, in which only one subject is left for testing while the others are used to train the models and get the optimal parameters. For the choice of the optimal parameters, an inner LOOCV is further conducted on the training data by a grid-search. For the regularized parameter λ , the candidate values range in $[2^{-5}, 2^{-4}, \dots, 2^0, \dots, 2^4, 2^5]$; for the regularized parameter γ , the candidate values range in $[0.1, 0.2, \dots, 0.9, 1]$; for the threshold in PC, we use 20 sparsity levels from $[5\%, 10\%, \dots, 95\%, 100\%]$, where, for example, 90% means that 10% of the weak edges are filtered out from the FBN.

A set of quantitative measurements, including accuracy, sensitivity and specificity, are used to evaluate the classification performance of different methods. The classification results corresponding to these methods on ADNI and NITRC datasets are given in Fig. 5. Based on the results, we observe that the proposed SR+SS method achieves the best performance in accuracy, sensitivity and specificity on both ADNI and NITRC datasets. This illustrates that the self-scrubbing strategy can

⁵ The elements of the adjacency matrix indicate the connection strengths of the node pairs in the network. Here, for the convenience of comparison among different methods, all the weights are normalized to the interval $[-1, 1]$.

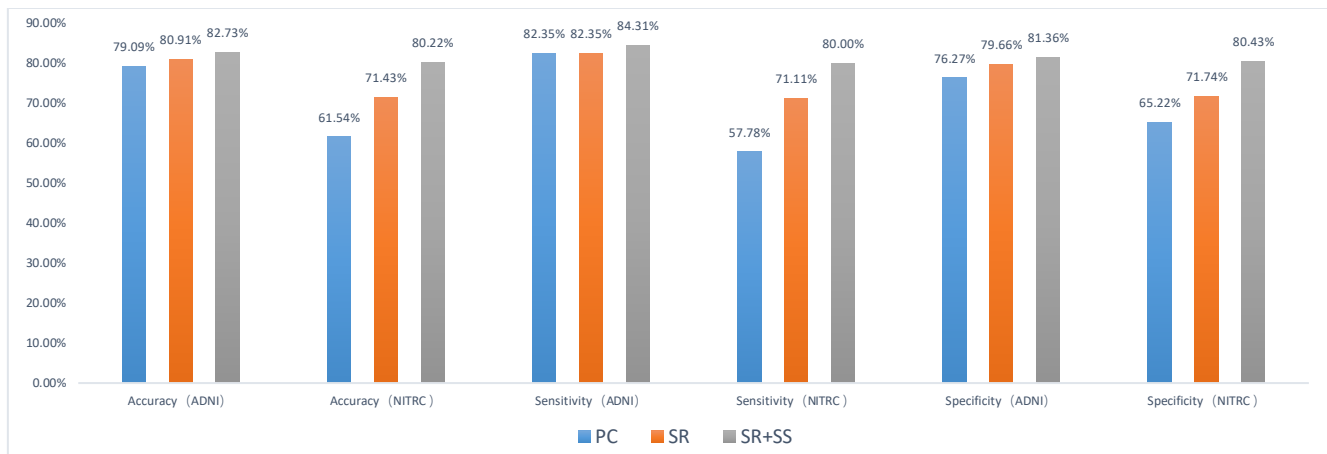


Fig. 5. Comparison of classification results based on three different methods. The results obtained by LOOCV test show that the proposed SR+SS achieves the best performance.

improve the performance of SR method in both scrubbed and non-scrubbed datasets.

V. CONCLUSION

The observed fMRI time series commonly contain various artifacts or noises, thus leading to a poor estimation of the FBN. In this paper, we propose a novel FBN estimation method by incorporating a latent variable into the regularized FBN estimation model as an indicator of the data quality. Then, we design an alternating optimization algorithm to solve the new model. As a consequence, the proposed method can estimate FBN and scrub the fMRI series adaptively in a single framework. In particular, we adopt the SR as a simple test platform in this paper for developing our method that is then validated on simulated and real-world (both scrubbed and non-scrubbed) datasets. The experimental results show that our proposed method significantly outperforms the baseline methods. Also, in order to examine how FBN changes following the self-scrubbing, we compare the FBNs estimated by SR and SR+SS methods. The result shows that changes mainly happen in the high-order brain regions (e.g., hippocampus and frontal brain region) and cerebellum region, which illustrates that “dirty” points may be highly related to motion artifacts and functional processes.

The proposed method still has several shortcomings that need to be improved. For example, we select all time points in the first iteration, which may lead to a bad regression result (local minimum) when a large number of “dirty” points exists. In addition, the possibly useful information from the removed points can be lost, since the proposed method simply discards them based on the hard indicator. In the future, we plan to develop a soft (or probabilistic) version of the proposed method for higher flexibility. It is also worth pointing out that the proposed self-scrubbing strategy can be easily transferred into the other FBN estimation models with any combination of data-fitting and regularization terms. Therefore, we will try more combinations towards better accuracy in the future.

REFERENCES

- [1] S. M. Smith, D. Vidaurre, C. F. Beckmann, M. F. Glasser, M. Jenkinson, K. L. Miller, T. E. Nichols, E. C. Robinson, G. Salimi-Khorshidi, and M. W. Woolrich, “Functional connectomics from resting-state fMRI,” *Trends in Cognitive Sciences*, vol. 17, no. 12, pp. 666-682, 2013.
- [2] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, “Functional connectivity in the resting brain: a network analysis of the default mode hypothesis,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 1, pp. 253-258, 2003.
- [3] A. Fomito, A. Zalesky, and M. Breakspear, “The connectomics of brain disorders,” *Nature Reviews Neuroscience*, vol. 16, no. 3, pp. 159-72, 2015.
- [4] C. J. Stam, “Modern network science of neurological disorders,” *Nature Reviews Neuroscience*, vol. 15, no. 10, pp. 683-695, 2014.
- [5] C. Y. Wee, P. T. Yap, and D. Shen, “Diagnosis of Autism Spectrum Disorders Using Temporally Distinct Resting-State Functional Connectivity Networks,” *Cns Neuroscience & Therapeutics*, vol. 22, no. 3, 2016.
- [6] C. Y. Wee, L. Wang, F. Shi, P. T. Yap, and D. Shen, “Diagnosis of autism spectrum disorders using regional and interregional morphological features,” *Human Brain Mapping*, vol. 35, no. 7, pp. 3414-30, 2014.
- [7] S. Delmonte, L. Gallagher, E. O’Hanlon, J. Mcgrath, and J. H. Balsters, “Functional and structural connectivity of frontostriatal circuitry in Autism Spectrum Disorder,” *Frontiers in Human Neuroscience*, vol. 7, no. 7, pp. 233-242, 2012.
- [8] S. Huang, J. Li, L. Sun, J. Liu, T. Wu, K. Chen, A. Fleisher, E. Reiman, and J. Ye, “Learning Brain Connectivity of Alzheimer’s Disease from Neuroimaging Data.” pp. 808-816.
- [9] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius, “Network Analysis of Intrinsic Functional Brain Connectivity in Alzheimer’s Disease,” *Plos Computational Biology*, vol. 4, no. 6, pp. 1-11, 2008.
- [10] C. Duval, J. F. Daneault, W. D. Hutchison, and A. F. Sadikot, “A brain network model explaining tremor in Parkinson’s disease,” *Neurobiology of Disease*, vol. 85, 2015.
- [11] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, “Network modelling methods for FMRI,” *Neuroimage*, vol. 54, no. 2, pp. 875-91, 2011.
- [12] L. Qiao, Z. Han, M. Kim, S. Teng, L. Zhang, and D. Shen, “Estimating functional brain networks by incorporating a modularity prior,” *Neuroimage*, vol. 141, pp. 399-407, 2016.
- [13] L. Zhou, L. Wang, and P. Ogunbona, “Discriminative Sparse Inverse Covariance Matrix: Application in Brain Functional Network Classification.” pp. 3097-3104.
- [14] M. J. Rosa, L. Portugal, T. Hahn, A. J. Fallgatter, M. I. Garrido, J. Shawe-Taylor, and J. Mourao-Miranda, “Sparse network-based

- models for patient classification using fMRI," *Neuroimage*, vol. 105, no. 3, pp. 493-506, 2015.
- [15] J. Peng, P. Wang, N. Zhou, and J. Zhu, "Partial Correlation Estimation by Joint Sparse Regression Models," *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 735-746, 2009.
- [16] O. Yamashita, M. A. Sato, T. Yoshioka, F. Tong, and Y. Kamitani, "Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns," *Neuroimage*, vol. 42, no. 4, pp. 1414-1429, 2008.
- [17] R. Yu, H. Zhang, L. An, X. Chen, Z. Wei, and D. Shen, *Correlation-Weighted Sparse Group Representation for Brain Network Construction in MCI Classification*: Springer International Publishing, 2016.
- [18] C. Y. Wee, P. T. Yap, D. Zhang, L. Wang, and D. Shen, "Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification," *Brain Structure & Function*, vol. 219, no. 2, pp. 641-656, 2014.
- [19] A. Benichoux, and T. Blumensath, "A spatially constrained low-rank matrix factorization for the functional parcellation of the brain." pp. 1-5.
- [20] R. A. Poldrack, J. A. Mumford, and T. E. Nichols, *Handbook of functional MRI data analysis*: Cambridge University Press, 2011.
- [21] C. G. Yan, B. Cheung, C. Kelly, S. Colcombe, R. C. Craddock, A. D. Martino, Q. Li, X. N. Zuo, F. X. Castellanos, and M. P. Milham, "A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics," *Neuroimage*, vol. 76, no. 1, pp. 183-201, 2013.
- [22] K. Murphy, R. M. Birn, and P. A. Bandettini, "Resting-state fMRI confounds and cleanup," *Neuroimage*, vol. 80, no. 1, pp. 349-359, 2013.
- [23] K. R. Van Dijk, M. R. Sabuncu, and R. L. Buckner, "The influence of head motion on intrinsic functional connectivity MRI," *Neuroimage*, vol. 59, no. 1, pp. 431-438, 2012.
- [24] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion," *Neuroimage*, vol. 59, no. 3, pp. 2142, 2012.
- [25] S. Yang, T. J. Ross, Y. Zhang, E. A. Stein, and Y. Yang, "Head motion suppression using real-time feedback of motion information and its effects on task performance in fMRI," *Neuroimage*, vol. 27, no. 1, pp. 153, 2005.
- [26] M. F. Mason, M. I. Norton, J. D. V. Horn, D. M. Wegner, S. T. Grafton, and C. N. Macrae, "Wandering minds: the default network and stimulus-independent thought," *Science*, vol. 317, no. 5834, pp. 43; author reply 43, 2007.
- [27] L. Jiang, D. Meng, Q. Zhao, S. Shan, and A. Hauptmann, "Self-paced Curriculum Learning."
- [28] M. P. Kumar, B. Packer, and D. Koller, "Self-Paced Learning for Latent Variable Models." pp. 1189-1197.
- [29] A. Fomito, A. Zalesky, and E. Bullmore, *Fundamentals of brain network analysis*: Academic Press, 2016.
- [30] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrianiissac, S. Lehericy, J. Doyon, and H. Benali, "Partial correlation for functional brain interactivity investigation in functional MRI," *Neuroimage*, vol. 32, no. 1, pp. 228-37, 2006.
- [31] K. V. Mardia, J. T. Kent, and J. M. Bibby, "Multivariate analysis," *Mathematical Gazette*, vol. 37, no. 1, pp. 123-131, 1979.
- [32] H. Lee, D. S. Lee, H. Kang, B. N. Kim, and M. K. Chung, "Sparse brain network recovery under compressed sensing," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1154-65, 2011.
- [33] S. Ryali, K. Supekar, D. A. Abrams, and V. Menon, "Sparse logistic regression for whole-brain classification of fMRI data," *Neuroimage*, vol. 51, no. 2, pp. 752-764, 2010.
- [34] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion," *Neuroimage*, vol. 59, no. 3, pp. 2142-54, 2012.
- [35] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Methods to detect, characterize, and remove motion artifact in resting state fMRI," *Neuroimage*, vol. 84, no. 1, pp. 320, 2014.
- [36] A. Gilad, "Brain covariance selection: better individual functional connectivity models using population prior," *Quantitative Biology*, vol. 1, no. 6, pp. 665-672, 2010.
- [37] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd Edition: Wiley, 2013.
- [38] P. L. Combettes, and J. C. Pesquet, "Proximal Splitting Methods in Signal Processing," *Heinz H Bauschke*, vol. 49, pp. págs. 185-212, 2015.
- [39] Z. Jie, C. Wei, Z. Liu, Z. Kai, L. Xu, Y. Ye, B. Becker, Y. Liu, K. M. Kendrick, and G. Lu, "Neural, electrophysiological and anatomical basis of brain-network variability and its characteristic changes in mental disorders," *Brain*, vol. 139, no. 8, 2016.
- [40] Y. Chao-Gan, and Z. Yu-Feng, "DPARFSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI," *Frontiers in Systems Neuroscience*, vol. 4, no. 13, pp. 13, 2010.
- [41] K. J. Friston, S. Williams, R. Howard, R. S. J. Frackowiak, and R. Turner, "Movement-Related effects in fMRI time-series," *Magnetic Resonance in Medicine*, vol. 35, no. 3, pp. 346-355, 1996.
- [42] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," *Neuroimage*, vol. 15, no. 1, pp. 273-289, 2002.
- [43] C. C. Chang, and C. J. Lin, "LIBSVM: A library for support vector machines," *Acm Transactions on Intelligent Systems & Technology*, vol. 2, no. 3, article 27, pp. 389-396, 2007.