Neuroinformatics
https://doi.org/10.1007/s12021-019-09423-0

ORIGINAL ARTICLE

Hierarchical Structured Sparse Learning for Schizophrenia Identification

Mingliang Wang[1,2] · Xiaoke Hao[1] · Jiashuang Huang[1] · Kangcheng Wang[3] · Li Shen[4] · Xijia Xu[5] · Daoqiang Zhang[1] · Mingxia Liu[6]

Abstract

Fractional amplitude of low-frequency fluctuation (fALFF) has been widely used for resting-state functional magnetic resonance imaging (rs-fMRI) based schizophrenia (SZ) diagnosis. However, previous studies usually measure the fALFF within low-frequency fluctuation (from 0.01 to 0.08Hz), which cannot fully cover the complex neural activity pattern in the resting-state brain. In addition, existing studies usually ignore the fact that each specific frequency band can delineate the unique spontaneous fluctuations of neural activities in the brain. Accordingly, in this paper, we propose a novel hierarchical structured sparse learning method to sufficiently utilize the specificity and complementary structure information across four different frequency bands (from 0.01Hz to 0.25Hz) for SZ diagnosis. The proposed method can help preserve the partial group structures among multiple frequency bands and the specific characters in each frequency band. We further develop an efficient optimization algorithm to solve the proposed objective function. We validate the efficacy of our proposed method on a real SZ dataset. Also, to demonstrate the generality of the method, we apply our proposed method on a subset of Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Experimental results on both datasets demonstrate that our proposed method achieves promising performance in brain disease classification, compared with several state-of-the-art methods.

Keywords Schizophrenia · Fractional amplitude of low-frequency fluctuations (fALFF) · Resting-state functional magnetic resonance imaging (rs-fMRI) · Hierarchical feature selection

Introduction

Schizophrenia, associated with symptoms such as hallucination and delusion, is a common chronic and devastating mental disorders that affects 1% of the population worldwide (Bhugra 2005). Until now, the pathological mechanism of schizophrenia is still unclear. Currently, the diagnosis of schizophrenia mainly is based on patient interview and symptom history (Demirci and Calhoun 2009). Hence, identifying valid biomarkers that can characterize the underlying physiological mechanism of SZ has received growing attention in SZ research community.

Currently, resting-state functional magnetic resonance imaging (rs-fMRI) can measure the spontaneous neural activity in blood-oxygenation level-dependent (BOLD) signals (Wang et al. 2010, 2014; Zhang et al. 2018; Jie et al. 2018), and has been increasingly used to find the abnormal neural activity in schizophrenia studies. It has been reported that the spontaneous low frequency oscillations (LFO, from 0.01Hz to 0.08Hz) of the SZ analysis with rs-fMRI are physiologically meaningful. Meanwhile, different frequency bands, i.e., slow-5 (0.01 − 0.027Hz) and slow-4 (0.027 − 0.073Hz) in LFO, have shown distinct oscillators with specific properties and physiological functions (Biswal et al. 1995; Knyazev 2007; Liang et al. 2012; Yu et al. 2014). For instance, Liang et al. (2012) discovered that the global and regional topological effectiveness were higher in the slow-4 than in slow-5. Yu et al. (2014) showed that several brain regions (i.e., *inferior occipital*

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12021-019-09423-0) contains supplementary material, which is available to authorized users.

✉ Xijia Xu
xuxijia@c-nbh.com
✉ Daoqiang Zhang
dqzhang@nuaa.edu.cn

Extended author information available on the last page of the article.

Published online: 23 April 2019

gyrus, *precunus*, and *thalamus*) in the slow-4 and slow-5 exhibit significant interaction between frequency bands. Recently, it has been recognized that the high-frequency bands (i.e., 0.08 − 0.25Hz) also can provide valuable information for SZ diagnosis (Yu et al. 2014; Zuo et al. 2010). For instance, Zuo et al. (2010) revealed that higher amplitude LFO (i.e., slow-3 (0.073 − 0.198Hz) and slow-2 (0.198 − 0.25Hz)) in gray matter were particularly prominent along the midline of the brain. Yu et al. (2014) showed that considering different high-frequency bands helps to measure the intrinsic brain activity of schizophrenia. Although previous studies have shown that both frequency specificity and complementary information play a crucial role in identifying SZ from normal controls, the absence of valid biomarkers of SZ is the major impeding factor in understanding the physiological mechanism that leads to SZ (Huang et al. 2010; Hoptman et al. 2010). Thus, more properly designed feature selection methods that can find the disease-related biomarkers are desired in multi-frequency based schizophrenia identification.

Instead of treating each frequency band as a single-task, multi-task paradigm (Caruana 1997) learns multiple frequency bands jointly, which can better exploit potential information shared by different frequency bands. With the improved performance gain by simultaneously handling several related tasks, multi-task learning has been widely applied in neuroimaging analysis (Wang et al. 2017; Liu et al. 2014, 2017; Jie et al. 2017; Liu and Zhang 2016, 2014). For example, Wang et al. (2017) proposed a multi-modality multi-center classification method for Autism Spectrum Disorder (ASD) diagnosis and selecting predictive biomarkers of ASD. Although typical multi-task methods can achieve superior performance in various applications, they usually assume that all tasks are equally related, in order to select the common information across different tasks (Liu et al. 2009a, 2016). However, if several features are only discriminant to a specific frequency band, using conventional multi-task learning methods will ignore these informative phenotypic biomarkers. Hence, it is desired to develop a unique multi-task learning method to model the structure information of multi-frequency data for automated diagnosis of schizophrenia.

Accordingly, in this paper, we propose a new hierarchical structured sparse (H-TS) learning method (illustrated in Fig. 1) to explicitly model the structure information of multi-frequency data (Buzsáki and Draguhn 2004) (i.e., slow-5 (0.01 − 0.027Hz), slow-4 (0.027 − 0.073Hz), slow-3 (0.073 − 0.198Hz) and slow-2 (0.198 − 0.25Hz)) for diagnosis of schizophrenia. Specifically, we first partition each training image into multiple candidate patches, for extracting patch-based multi-frequency band features from the rs-fMRI data. Afterwards, our hierarchical structured sparse (H-TS) method is used to select the more relevant and discriminative features for disease classification. The basic idea of the hierarchical framework is that we decompose the parameter matrix into a sum of H component matrices, with each component matrix corresponding to one hierarchy. Such hierarchical framework gradually enforces different levels of feature sharing to capture the structure information of multi-frequency data efficiently. At each level, we utilize a novel structure regularization with different parameters to guide the sharing of partial group structures. Also, to reflect the specificity of schizophrenia in each frequency, we introduce the $\ell_{1,1}$-norm regularization that can preserve the unique schizophrenia characteristics. Then, the selected features are used to train support vector machine (SVM) classifiers (Cherkassky 1997; Liu and Zhang 2014) for schizophrenia disease classification based on the training data. Finally, we adopt an ensemble classification strategy, a simple and effective classifier fusion method, to combine

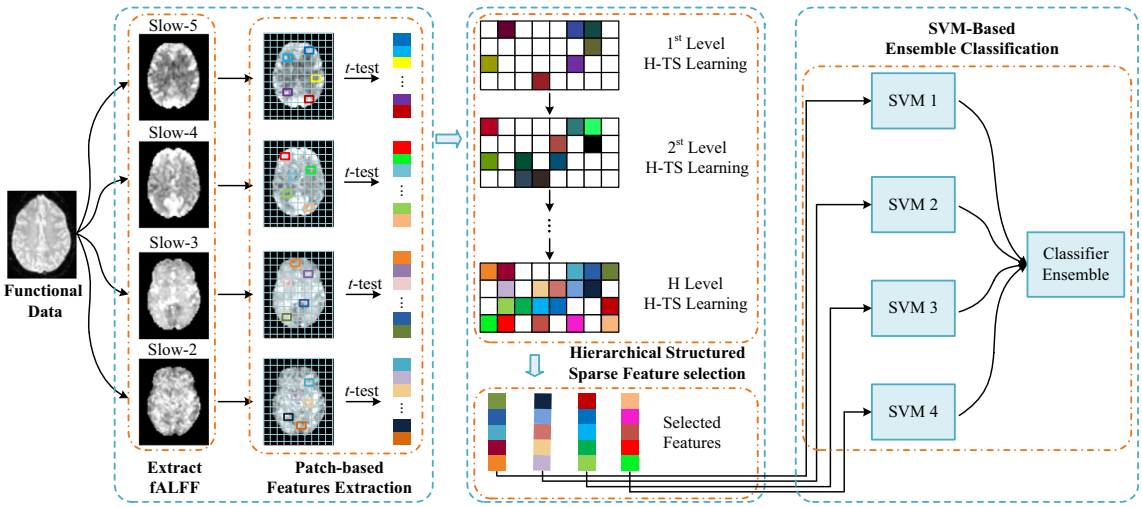

Fig. 1 The flowchart of the proposed classification algorithm

the outputs of different SVM classifiers to make a final prediction.

The remainder of the paper is structured as follows. In "Methodology", we first present the details of the proposed H-TS method. The experiments and results are then described in "Experimental". In "Discussion", we investigate the influence of parameters, present the clinical relevance and discuss the limitations of our method. We finally conclude this paper in "Conclusion".

Methodology

Figure 1 illustrates the proposed classification framework, which consists of three main steps: 1) patch-based multi-frequency band feature extraction, 2) hierarchical structured sparse feature selection, and 3) SVM-based ensemble classification. In the following, we will give the detailed descriptions of these steps. Before that, we will first introduce the subjects used in this paper.

Subjects

Subjects used in this study were obtained from the Department of Psychiatry, Affiliated Nanjing Brain Hospital, Nanjing Medical University (Wang et al. 2017). All patients were identified according to the schizophrenia diagnosis criteria by two qualified psychiatrists. Written informed consent was obtained from all the patients and their guardians, and all of the research procedures and ethical guidelines were followed in accordance with the Institutional Review Board (IRB) of the Affiliated Nanjing Brain Hospital, Nanjing Medical University. Exclusion criteria included the organic brain diseases, infectious diseases, any chronic somatic diseases, and drug abuse. Healthy volunteers with the matching of age and education were recruited via advertisement. The exclusion criteria for controls were similar to those for schizophrenic patients. The detailed demographic and clinical information is summarized in Table 1.

For each participant, their eyes keep open during the rs-fMRI scanning process. In this study, we use 46 participants for performance evaluation, including 23 schizophrenia (SZ) patients and 23 normal controls (NCs). The resting-state function MRI scan was obtained for each participant using a Siemens 3.0T system. Each scan was acquired with the following parameters: TR=3, 000 ms, TE=30 ms, matrix=64 $\times$ 64, slice thickness=3 mm, flip angle=90°, field of view (FOV) =220 $\times$ 220 mm^2. For each participant, 34 trans-axial slices with no gap were acquired to cover the whole brain volume. The scan time of the resting-state fMRI was 6 min.

Table 1 Demographic and clinical information of subjects in the unique schizophrenia dataset. The values are denoted as Mean± Standard Deviation

Diagnosis	Number	Age	Male/Female	PANSS
NC	23	33.5 ± 7.7	8/15	null
SZ	23	32.1 ± 9.1	10/13	98.0 ± 11.1

Image Pre-Processing

For each subject, we first discard the first 10 volumes of its fMRI scan. Subsequently, the remaining volumes are processed by the following procedures sequentially, including 1) slice acquisition adjustment, 2) head-motion correction, 3) normalization (via first linear registration and then nonlinear registration) to the SPM5 Montreal Neurological Institute template (Rosario et al. 2008), and 4) resampling to 3 $\times$ 3 $\times$ 3 mm^3 voxels. Note that any scan with a head motion $>$ 3.0 mm of maximal translation or 1.0° of maximal rotation is excluded. Then, six motion parameters, including the global mean signal, the cerebrospinal fluid (CSF), and white matter signals were used to reduce the negative effects of head motion and non-neuronal BOLD fluctuations as nuisance covariates (Yu et al. 2014). It is worth noting that temporal filtering was not implemented during pre-processing, which enables the different frequency bands can be included in subsequent analysis. After image pre-processing, fractional amplitude of low-frequency oscillations were calculated using REST software (Song et al. 2011).

Patch-Based Multi-frequency Band Feature Extraction

Given an fALFF activation map, all voxels in the brain will be treated as feature vectors in traditional methods. However, voxel-based analysis methods often have very high feature dimensions, which often lead to suboptimal performance (Liu et al. 2016). To address this issue, we use a patch-based feature extraction strategy. Specifically, we first divide each training image into non-overlapping candidate patches, with the patch size of 3 $\times$ 3 $\times$ 3. We then perform patch-wise group comparison to select the most discriminative patches in an image for all four frequency bands based on all training images. That is, centered at each location in the brain image, we compare the patches extracted from SZ patients and those from NC subjects in the training set via the t-test algorithm, where each patch is represented by a 27-dimensional feature vector (with each element denoting the intensity value of a specific voxel within this patch). We then select those patches with the p-value smaller than 0.05 in each frequency band. If one patch is selected in more than two frequency bands, we

treat this patch as the discriminative patch for disease classification. Here, we represent each discriminative patch by the mean intensity value of all 27 voxels within this patch. Finally, we concatenate such patch-based features of all selected discriminative patches to be a feature vector for representing each subject.

Hierarchical Structured Sparse Feature Selection

Based on the patch-based features of each subject, we then apply our proposed H-TS method to learn the feature coefficients jointly. In the following, we will present the H-TS feature selection in detail.

Suppose we have M supervised learning tasks corresponding to the number of frequency bands. Let $\mathbf{X}_m = [\mathbf{x}_{m,1}, \mathbf{x}_{m,2}, \ldots, \mathbf{x}_{m,N}]^T \in \mathbb{R}^{N\times d}$ denotes the training data matrix for the m-th task, where N and d are the number of samples and feature dimensionality, respectively. Similarly, denote $\mathbf{Y} = [y_1, y_2, \ldots, y_N]^T \in \mathbb{R}^N$ as the response vector for training data $\mathbf{X}_m$, where $y_n \in \{-1, 1\}$ is the class label corresponding to the n-th subject. Denote $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_M] \in \mathbb{R}^{d\times M}$ as the coefficient matrix that is learned jointly for all M tasks, where $\mathbf{w}_m \in \mathbb{R}^d$ is the linear discriminant function for the m-th task.

To better capture the complex group structure among different frequency bands, in this paper we propose a hierarchical framework and assume that there are H hierarchies. Then the parameter matrix can be decomposed into H components where each hierarchy can capture different partial group structures. Specifically, the coefficient matrix $\mathbf{W}$ is described as

$$\mathbf{W} = \sum_{h=1}^{H} \mathbf{W}_h \tag{1}$$

where $\mathbf{W}_h = [\mathbf{w}_{h,1}, \ldots, \mathbf{w}_{h,M}] \in \mathbb{R}^{d\times M}$ is the coefficient matrix corresponding to the h-th hierarchy, and $\mathbf{w}_{h,m}$ is the m-th column of $\mathbf{W}_h$ in the h-th hierarchy. Then the objective function of our H-TS model can be defined as follows:

$$min_{\mathbf{W}} \sum_{m=1}^{M} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}_m \sum_{h=1}^{H} \mathbf{w}_{h,m}\|_2^2 + R_t(\mathbf{W}) + R_s(\mathbf{W}) \tag{2}$$

where $R_t(\mathbf{W})$ and $R_s(\mathbf{W})$ are the structure regularization and the $\ell_{1,1}$-norm regularization, respectively. Specifically, the structure regularization is defined as below:

$$R_t(\mathbf{W}) = \sum_{h=1}^{H} \lambda_h \sum_{p<q}^{M} \|\mathbf{w}_{h,p} - \mathbf{w}_{h,q}\|_2 \tag{3}$$

Here, we impose a ℓ_2-norm on the pairwise difference among the column vectors of each coefficient matrix, which induces the structural information across different frequency bands without assuming them to be known in advance. The regularization parameter λ_h controls the strength of feature sharing.

To discover the brain functional activities that are unique to each frequency bands (Yu et al. 2014; Zuo et al. 2010) and to encourage the sparsity of the coefficient matrix $\mathbf{W}$, the $R_s(\mathbf{W})$ is defined as follows:

$$R_s(\mathbf{W}) = \sum_{h=1}^{H} \beta_h \|\mathbf{W}_h\|_{1,1} \tag{4}$$

where $\|\mathbf{W}_h\|_{1,1} = \sum_{i=1}^{d} \|\mathbf{w}_h^i\|_1$ is the sum of ℓ_1-norm of the rows in matrix $\mathbf{W}_h$. The $\ell_{1,1}$-norm is used to encourage many elements of $\mathbf{W}_h$ being zeros as well as select the specific features corresponding to each frequency band (Jalali et al. 2010; Lachowicz and wrzosek 2001). In our study, we will select those features with non-zero elements for subsequent classification tasks. The regularization parameter β_h controls the sparsity of the coefficient matrix.

In this study, we assume a descending order of feature sharing from the H-th hierarchy to the first one, in other words, the subset of sharing partial group structures are gradually descended. Meanwhile, we assume an increasing order for the sparsity of the matrix $\mathbf{W}_h$ from the first hierarchy to the last. Hence, we set $\lambda_h = \lambda_{h-1}/\sigma$ and $\beta_h = \beta_{h-1} \times \sigma$ for $h \geq 2$ with constant $\sigma > 1$. When $\beta_1 = 0$ (i.e., no specific information of respective frequency band is considered), our method will degenerate to a hierarchical task grouping (H-T) method as proposed in Han and Zhang (2015). With $\lambda_1 = 0$ (i.e., no structure regularization item is used), our method will degenerate to a hierarchical lasso method (H-S). In the following section, we will develop an efficient optimization algorithm to solve the objective function defined in Eq. 2.

Optimization

The algorithm, which attempts to solve the objective function in Eq. 2 directly, could lose the generality due to the larger number of parameters in the hierarchical framework. Accordingly, we propose a top-down iterative scheme presented in Algorithm 1, by decomposing Eq. 2 into several sub-problems consistent with hierarchies. The sub-problem of Eq. 2 can be described as follows:

$$\begin{aligned} min_{\mathbf{W}_h} & \sum_{m=1}^{M} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}_m \mathbf{w}_{h,m}\|_2^2 \\ & + \lambda_h \sum_{p<q}^{M} \|\mathbf{w}_{h,p} - \mathbf{w}_{h,q}\|_2 + \beta_h \|\mathbf{W}_h\|_{1,1} \end{aligned} \tag{5}$$

The objective function in Eq. 5 is convex with respect to $\mathbf{W}_h$ but non-smooth. For seeking the optimal value of $\mathbf{W}_h$, we can solve the problem by the smoothing proximal

gradient (SPG) method (Chen et al. 2011), as shown in Algorithm 2. It is worth noting that each hierarchy is optimized respectively utilizing the optimal value from existing hierarchies. The problem solved by the SPG method takes the form

$$min_{\mathbf{W}_h} \widetilde{F}(\mathbf{W}_h) = f(\mathbf{W}_h) + g_\mu(\mathbf{W}_h) + \beta_h \|\mathbf{W}_h\|_{1,1} \tag{6}$$

According to (Chen et al. 2011), the smoothed approximation function of the structure penalty term $g_\mu(\cdot)$ can be rewritten as:

$$\begin{aligned} g_\mu(\mathbf{W}_h) &= \lambda_h \sum_{p<q}^{M} \|\mathbf{w}_{h,p} - \mathbf{w}_{h,q}\|_2 \\ &\cong max_{\mathbf{A}\in \boldsymbol{Q}} \langle \mathbf{C}\mathbf{W}_h^T, \mathbf{A}\rangle - \mu d(\mathbf{A}) \end{aligned} \tag{7}$$

where μ is the positive smoothness parameter and $d(\mathbf{A}) = \frac{1}{2}\|\mathbf{A}\|_F^2$. $\mathbf{A} = (\alpha_1, \ldots, \alpha_{m(m-1)/2})^T$ is the auxiliary matrix variable with a closed and convex set domain $\boldsymbol{Q} \equiv \{\mathbf{A} | \|\alpha_j\|_2 \leq 1, \forall_j \in \mathbb{N}_{m(m-1)/2}\}$. And α_i is a vector of auxiliary variables associated with the i-th row of $\mathbf{C}\mathbf{W}_h^T$. Note that $\mathbf{C} \in \mathbb{R}^{\frac{m(m-1)}{2}\times m}$ is a highly sparse matrix, with each row having only two non-zero elements. Then the gradient of Eq. 7 can be computed as:

$$\nabla g_\mu(\mathbf{W}) = (\mathbf{A}^*)^T \mathbf{C} \tag{8}$$

where $\mathbf{A}^*$ is the optimal solution of Eq. 7. Let $h(\mathbf{W}_h) = f(\mathbf{W}_h) + g_\mu(\mathbf{W}_h)$, the gradient of $h(\mathbf{W}_h)$ is given as:

$$\nabla_{\mathbf{W}_h} h(\mathbf{W}_h) = \nabla_{\mathbf{W}_h} f(\mathbf{W}_h) + (\mathbf{A}^*)^T \mathbf{C} \tag{9}$$

By using the square loss in Eq. 2, the i-th column of $\nabla_{\mathbf{W}_h} f(\mathbf{W}_h)$ can be obtained as $\mathbf{X}_m^T(\mathbf{X}_m \mathbf{w}_{h,m} - \mathbf{Y})$. Moreover, $\nabla_{\mathbf{W}_h} h(\mathbf{W}_h)$ can be proven Lipschitz-continuous easily where the Lipschitz constant L can be determined by numerical approaches (Chen et al. 2011). Therefore, the generalized gradient update step of SPG algorithm is defined as

$$\begin{aligned} \mathbf{W}_h^{t+1} = \; & argmin_{\mathbf{W}_h} \frac{1}{2}\|\mathbf{W}_h - (\widehat{\mathbf{W}}_h^t - \frac{1}{L}\nabla h(\widehat{\mathbf{W}}_h^t))\|_2^2 \\ & + \frac{\beta_h}{L}\|\mathbf{W}_h\|_{1,1} \end{aligned} \tag{10}$$

Let $\mathbf{V}_h = \widehat{\mathbf{W}}_h^t - \frac{1}{L}\nabla h(\widehat{\mathbf{W}}_h^t)$. According to Chen et al. (2011), the closed-form solution for $\mathbf{W}_h^{t+1}$ is given as $\mathbf{W}_h^{t+1,i} = sign(\mathbf{v}^i) max(0, |\mathbf{v}^i| - \frac{\beta_h}{L})$, where $\mathbf{W}_h^{t+1,i}$ and $\mathbf{v}^i$ represent the i-th row of the matrix $\mathbf{W}_h^{t+1}$ and $\mathbf{V}_h$, respectively. In addition, according to Chen et al. (2011), we can directly compute the following formulation to replace the derivation of gradient descent with respect to $\mathbf{W}_h$:

$$\widehat{\mathbf{W}}_h^{t+1} = \mathbf{W}_h^{t+1} + \eta_{t+1}(\mathbf{W}_h^{t+1} - \widehat{\mathbf{W}}_h^t) \tag{11}$$

where $\eta_{t+1} = \frac{(1-\theta_t)\theta_{t+1}}{\theta_t}$ and $\theta_{t+1} = \frac{2}{t+3}$. Let $D = max_{\mathbf{A}\in\boldsymbol{Q}} d(\mathbf{A})$ and $\mathbf{W}_h^*$ be the optimal solution to Eq. 6. If we set ϵ as the desired accuracy and require $|\widetilde{F}(\mathbf{W}_h^t) - \widetilde{F}(\mathbf{W}_h^*)| \leq \epsilon$, the SPG algorithm will convergence in $O(\sqrt{2D}/\epsilon)$ iterations.

Algorithm 1 Hierarchical structured sparse classification algorithm.

Input: $\mathbf{X}$, $\mathbf{Y}$;
Output: $\mathbf{W}$;
Initialize $k = 1, \widehat{\mathbf{W}}_1^0 = 0, \mathbf{W}_1^0 = \cdots = \mathbf{W}_H^0 = \mathbf{0}$;
for $h = 1, 2, \ldots, H$ **do**
  solve Eq. 2;//Algorithm 2
  if $h > 1$ **then**
    $\widehat{\mathbf{W}}_h^0 = \sum_{k<h} \mathbf{W}_k + \mathbf{W}_h^0$;
  endif
  $k = k + 1$;
endfor
$\mathbf{W} = \sum_{h=1}^{H} \mathbf{W}_h$

Algorithm 2 SPG algorithm for solving Eq. 2.

Input: $\mathbf{X}$, $\mathbf{Y}$, $\widehat{\mathbf{W}}_h^0$, μ, h, λ_h, β_h, σ, *step* α, *desired accruacy* ϵ;
Output: $\mathbf{W}_h$;
Initialize $\theta_t = 1, L = 1, t = 0$;
repeat
  Compute the $\nabla h(\widehat{\mathbf{W}}_h^t))$ as in Eq. 11;
  Solve the proximal step:
    $\mathbf{W}_h^{t+1} = argmin_{\mathbf{W}_h} h(\widehat{\mathbf{W}}_h^t) + \langle \mathbf{W}_h - \widehat{\mathbf{W}}_h^t, \nabla h(\widehat{\mathbf{W}}_h^t)\rangle$
      $+\beta_h\|\mathbf{W}_h\|_{1,1} + \frac{L}{2}\|\mathbf{W}_h - \widehat{\mathbf{W}}_h^t\|_2^2$;
  Set $\theta_t = \frac{2}{t+3}$;
  Set $\widehat{\mathbf{W}}_h^{t+1} = \mathbf{W}_h^{t+1} + \frac{1-\theta_t}{\theta_t}\theta_{t+1}(\mathbf{W}_h^{t+1} - \widehat{\mathbf{W}}_h^t)$;
  Set $L = \alpha * L$;
  $t = t + 1$.
until *Some convergence criterion is satisfied*;

SVM-Based Ensemble Classification

Due to the ability to handle high-dimensional data, and the accuracy in classification and prediction, SVM has been widely used in medical aided diagnosis (Liu et al. 2016, 2016, 2016; Zhou et al. 2018). In this paper, to better take advantage of multi-frequency band features, we further propose an SVM-based ensemble classification approach. Based on these selected features, we can construct four classifiers separately, with each classifier corresponding to a specific frequency band. Then we apply a simple and effective classifier fusion method named the majority voting strategy (Lam and Suen 1997) to combine the outputs

of different SVM classifiers. Based on the voting results, we can obtain the predicted class label of a new testing subject. It is worth noting that, under the condition that positive and negative voting results are equal, we consider that the category of the sample is confused, namely, it is a misclassification sample.

Experimental

Methods for Comparison

In the experiments, we compare the H-TS method with two types of methods, namely single-task learning and multi-task learning method. There are 3 single-task learning methods, including 1) *t*-test, 2) Lasso, and 3) Elastic-Net. These methods are very popular in the field of neuroimaging analysis. Also, there are 3 multi-task learning methods, including 1) Multi-Task Feature Learning (MTFL), 2) H-S, and 3) H-T. Following, we summarize the detailed implementation of these methods.

1) ***t*-test** (Guyon et al. 2002).The *t*-test is the most commonly used for hypothesis testing to determine if two sets of data are significantly different from each other. Here, the *t*-test method first selects the discriminative features related to SZ in each frequency band, and then the corresponding SVM classifier is trained using the selected features, respectively.
2) **Lasso** (Hastie et al. 2001). Lasso is a regression analysis method that conducts both feature selection and regularization to enhance the prediction performance. The Lasso method first selects a discriminative subset of features from each frequency band. Then the corresponding SVM classifier is trained using the subset features.
3) **Elastic-Net** (Zou and Hastie 2005). Elastic-Net, being regularized by the linearly combines of the ℓ_1 and ℓ_2-norm, is an extension of Lasso and usually preferred over Lasso method. Different from Lasso, it tends to choose one variable from a group and ignore the others. We also apply it to select the discriminative subset of features in each frequency band, and then feed these features to learn a SVM classifier for classification.
4) **MTFL** (Liu et al. 2009a). MTFL is considered to be the $\ell_{2,1}$-norm regularized model for joint feature selection from multiple tasks. One appealing feature of this method is that it encourages multiple tasks to share a common subset of features. In this paper, we treat each frequency band as a task and jointly select a shared subset of features from multiple frequency bands. We then train multiple SVM classifiers using the selected features corresponding to each frequency band.
5) **H-S**. H-S is a variant of our proposed H-TS method. For H-S method, only the specificity information of each frequency band is considered, and it is denoted as $\lambda_1 = 0$ in Eq. 2. We use H-S to mine the specific patterns of each frequency band, and train the corresponding SVM classifier to diagnose in each frequency band.
6) **H-T** (Han and Zhang 2015). H-T is another variant of our proposed H-TS method. H-T considers the shared structure information among multiple frequency bands, ignoring the specific information of each frequency band, and can be denoted as $\beta_1 = 0$ in Eq. 2. We apply H-T to jointly learn multiple frequency bands to mine the shared structure among different frequency bands, and then train the SVM classifier for each frequency band using the mined patterns.

For each method, we train the corresponding SVM classifier using the subset of features selected from each frequency band. Then we adopt the proposed ensemble-based strategy to combine the outputs of all SVMs for making a final prediction. For the *t*-test method, the *p*-value is chosen from [0.01, 0.05, 0.08, 0.10, 0.12, 0.15]. For Lasso, Elastic-Net and MTFL, the trade-off parameters are chosen from the set $[2^{-4}, 2^{-3}, \cdots, 2^3]$, respectively. The candidate set for both trade-off parameters in H-S (i.e., β_1), H-T (i.e., λ_1) and H-TS (i.e., λ_1 and β_1) is $[2^{-4}, 2^{-3}, \cdots, 2^3]$. The parameter of hierarchy is fixed to 5. Here, we resort to the LIBSVM toolbox (Chang and Lin 2011) for SVM classifier learning and the SLEP toolbox (Liu et al. 2009b) to implement the Lasso, Elastic-Net and MTFL methods. For simplicity, the margin parameter C of linear SVM is fixed as 1 to perform classification for all methods.

Experimental Settings

We adopt a leave-one-out cross-validation (LOOCV) strategy (Besga et al. 2012) for performance evaluation. During LOOCV, each sample is designated as the test sample in turns, while the remaining samples are adopted as the training set. To optimize these parameters for different methods, we further perform an inner LOOCV strategy using each training set, with no testing data used in such process. Unless otherwise specified, we always use LOOCV to select the hyperparameters in all methods. We evaluate the performance of different methods by four classification performance criteria including classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the receiver operating characteristic (ROC) curve (AUC). Here, the accuracy evaluates the proportion of correctly predicted subjects among all subjects, the sensitivity represents the proportion of patients those

are correctly predicted, and the specificity denotes the proportion of normal controls those are correctly classified.

Results

Each Frequency Band Classification Results on SZ Dataset

In the section, we perform schizophrenia disease classification using fMRI data with four frequency bands. Specifically, we compare our method with two categories of methods, 1) single-task learning methods (i.e., *t*-test, Lasso, and Elastic-Net), and 2) multi-task learning methods (i.e., MTFL, H-S, H-T and H-TS). For each competing method, we first perform feature selection and then construct multiple SVM classifiers corresponding to each frequency band. It is worth noting that, the single-task learning methods learn a model for at each frequency band separately, while the multi-task learning methods jointly train models for four frequency bands. We select the brain regions with higher absolute weights at each frequency band for subsequent classification. In Fig. 2, we show the classification results using different methods at each frequency band.

From the results of SZ diagnosis in Fig. 2, we can observe three main points. *First*, the multi-task learning methods generally achieve better than the single-task learning methods in most case. For example, the highest accuracy achieved by Elastic-Net is 69.57%, 65.22%, 67.39% and 63.04% in each frequency band separately, which is noticeably lower than those of multi-task learning methods (i.e., MTFL, H-T, and H-TS). This demonstrates that, the joint learning of four different frequency bands is beneficial to improve the diagnosis of schizophrenia. *Second*, we can observe that the classification accuracy and specificity of all competing methods in slow-2 perform slightly worse compared to other frequency bands. Also, we can see that all methods achieve higher accuracies in slow-3. It is consistent with the previous studies that the different frequency band owns the unique characteristics of schizophrenia (Yu et al. 2014; Zuo et al. 2010). It also illustrates the necessity to model the specific information of each frequency band. *Finally*, the hierarchical methods (i.e., H-S, H-T and H-TS) generally achieve significantly better performance, compared with the single level methods. It is worth noting that, although H-S method has not obtained the better performance in comparison with Elastic-Net and MTFL, it still outperforms the *t*-test and Lasso methods. In contrast to Elastic-Net and MTFL, H-S only considers a single band and ignores the structure information within or across different frequency bands.

Ensemble Classification Results on SZ Dataset

In addition, we report the ensemble experimental results achieved by different methods in Fig. 3 and their receiver operating characteristic (ROC) curves is given in Fig. 4. From Figs. 3 and 4, we can see that the ensemble of different SVM classifiers is helpful to further improve the classification performance. Specifically, H-TS achieves a classification accuracy of 93.48%, a sensitivity of 95.65%, a specificity of 91.30% and an AUC of 97.07%, while the second best (H-T) classification accuracy, sensitivity, specificity and AUC is 91.30%, 91.30%, 91.30%, 96.44%, respectively. H-TS method achieves an improvement of 2.18%, 4.35% compared with the H-T method in

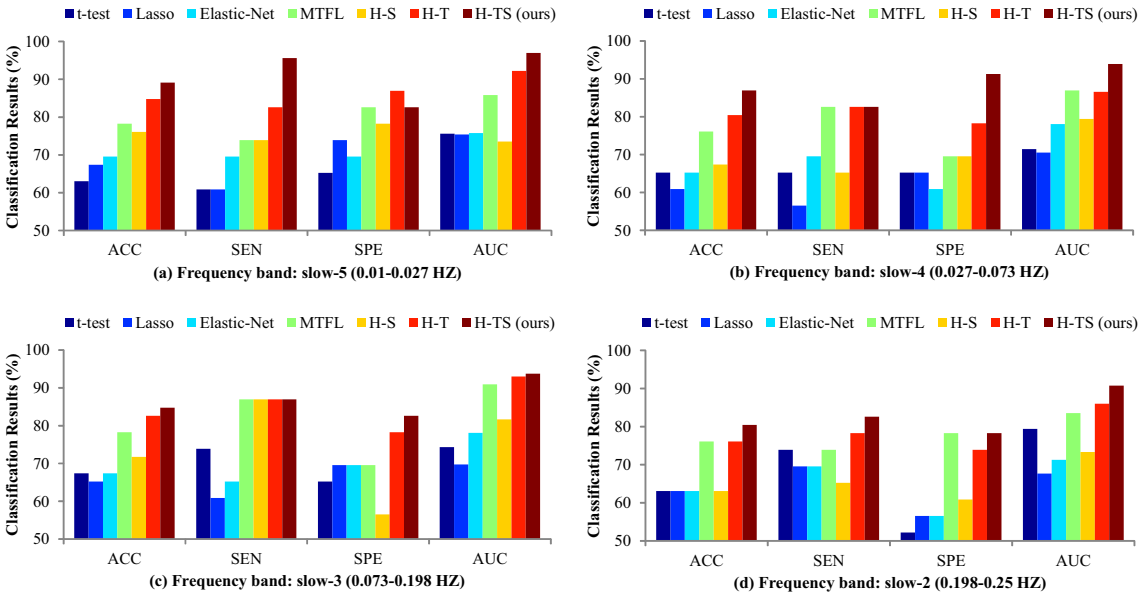

Fig. 2 Classification results achieved by different methods on each frequency band in SZ vs. NC

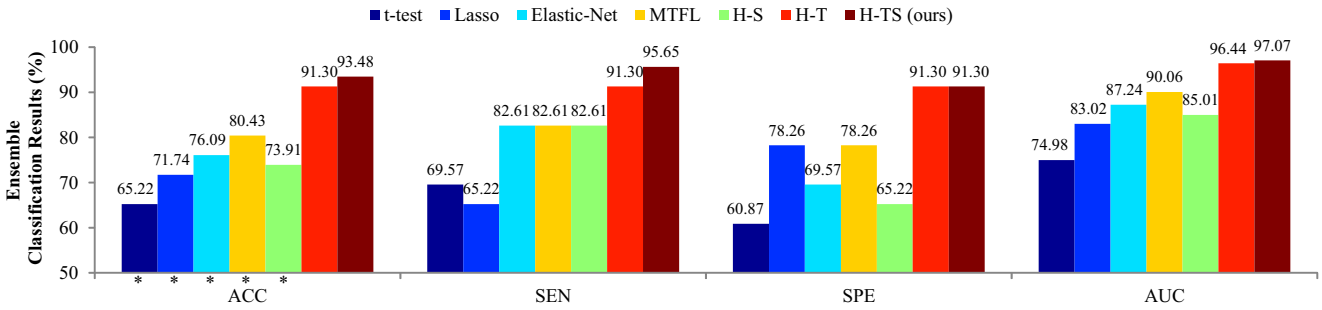

Fig. 3 Classification results achieved by different methods using the ensemble classification strategy in SZ vs. NC. The asterisk (*) denotes that the difference between ACC values of our H-TS method and each competing method is statistically significant ($p < 0.05$)

classification accuracy and sensitivity, respectively. This also implies that specific features corresponding to each frequency band are significant to improve the classification performance. The McNemar test (Dietterich 1998) is used to evaluate the statistical significance of the difference between ensemble classification performances (w.r.t., ACC) of our method with six competing methods. We report the p-values in Fig. 3 and mark statistically significant differences ($p < 0.05$) with the asterisk (*).

Results on AD Dataset

To demonstrate the effectiveness of our proposed method, we also perform experiments on a larger multi-modality dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). One primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). A detailed description of the data can be found on the website (http://www.adni-info.org).

In this study, following (Hao et al. 2016), a total of 558 subjects including 160 AD patients, 187 LMCI (Late Mild Cognitive Impairment) patients and 211 normal controls are used for diagnosis classification. Specifically, we apply our method to the preprocessed multi-modality imaging data, in which each modality with 90 features (i.e., ROIs) is treated as a task, and the number of hierarchies in our H-TS model is empirically set as 3 for two classification tasks. We adopt a 10-fold cross-validation strategy to evaluate the performance of different methods. Table 2, summarizes the averaged classification results of AD versus NC and LMCI versus NC.

From Table 2, we can see that the proposed method consistently achieves better classification performance than the competing methods, which demonstrates the effectiveness of our H-TS method. Specifically, H-TS achieves a classification accuracy of 88.21%, a sensitivity of 90.31%, a specificity of 85.83% and an AUC of 0.9095 for AD versus NC. For LMCI versus NC classification task, it achieves a classification accuracy of 70.00%, a sensitivity of 87.81%, a specificity of 61.80% and an AUC of 0.7091. We further use the McNemar's test (Dietterich 1998) (with the significance level at 0.05) to evaluate whether the difference in classification accuracies between our proposed method and each competing method is significant, with the corresponding p-values reported in Table 2. More results can be found in Section 1 of the Supplementary Materials.

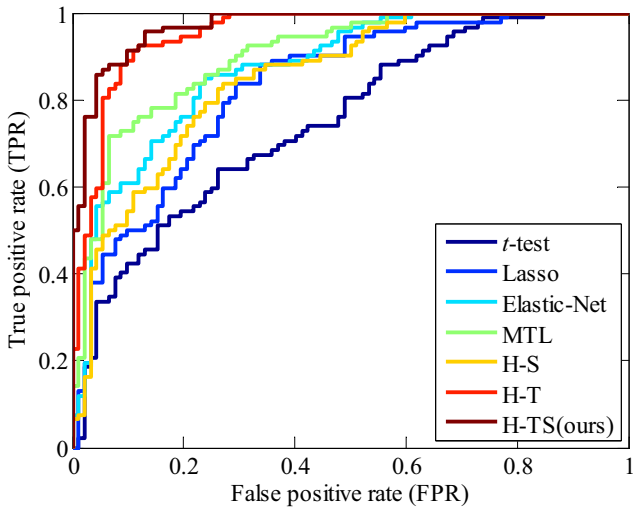

Fig. 4 ROC curves achieved by different methods using the ensemble classification strategy in SZ vs. NC

Most Important Brain Regions

It's meaningful to identify the biomarkers for schizophrenia disease diagnosis (Shen et al. 2010; Takayanagi et al. 2011; Neuhaus et al. 2011; Arribas et al. 2010; Demirci et al. 2008; Jafri and Calhoun 2006; Lian et al. 2019), so we investigate the top ten brain regions identified by our proposed H-TS method in this section. Since we use the LOOCV in our experiments, the selected biomarkers may be different from each cross-validation fold. Hence, we choose the cumulative

Table 2 Performance of AD vs. NC and LMCI vs. NC classification using 10-fold cross-validation on the subset of ADNI database (Hao et al. 2016)

Methods	ACC (%)	SEN (%)	SPE (%)	AUC (%)	*p*-value
AD vs. NC					
t-test	80.00 ± 3.57	86.25 ± 5.89	71.67 ± 4.93	88.03 ± 0.58	< 0.0001
Lasso	82.05 ± 3.66	87.66 ± 5.77	74.58 ± 6.64	88.86 ± 0.70	0.0002
Elastic-Net	81.61 ± 3.04	89.38 ± 4.76	71.25 ± 5.45	88.25 ± 0.59	< 0.0001
MTFL	84.02 ± 3.22	87.97 ± 5.26	78.75 ± 3.90	89.29 ± 0.64	0.0027
H-S	82.32 ± 2.88	84.38 ± 5.80	79.58 ± 5.08	87.27 ± 1.52	< 0.0001
H-T	86.61 ± 2.06	90.00 ± 3.70	81.67 ± 4.59	90.69 ± 0.49	0.0478
H-TS (ours)	**88.21±2.05**	**90.31±2.93**	**85.83±3.78**	**90.95±0.79**	–
LMCI vs. NC					
t-test	62.16 ± 2.80	84.69 ± 6.24	54.52 ± 6.04	68.10 ± 0.85	< 0.0001
Lasso	64.64 ± 4.26	85.47 ± 6.55	54.79 ± 5.60	69.02 ± 0.67	0.0040
Elastic-Net	65.28 ± 3.33	86.88 ± 5.67	60.62 ± 3.70	69.98 ± 0.68	0.0033
MTFL	66.56 ± 4.80	83.91 ± 6.59	60.36 ± 6.61	70.51 ± 0.55	0.0476
H-S	64.88 ± 1.62	69.06 ± 4.70	60.49 ± 5.15	66.59 ± 0.57	0.0001
H-T	68.56 ± 3.60	81.09 ± 6.09	55.41 ± 5.45	70.90 ± 0.42	0.034
H-TS (ours)	**70.00±2.88**	**87.81±5.45**	**61.80±3.94**	**70.91±0.53**	–

The bold symbols mean that the results are superior

absolute weight (Noriaki et al. 2016) as the contribution indicator of each brain region in the classification of SZ and NC. Specifically, we first map the top 500 patch-based features selected by our method into the MNI space, and record the frequency of the brain regions where the patch-based features are located. Then we choose the top 10 high

Table 3 Top 10 selected regions of interest (ROIs) from four different frequency bands

MNI	AAL Area	Cumulative absolute weight	MNI	AAL Area	Cumulative absolute weight
Slow-5			Slow-4		
(−33, −15, 3)	Putamen.L	6.52	(−69, −33, 33)	SupraMarginal.L	6.45
(3, −15, 3)	Thalamus.R	6.46	(−15, −9, 45)	Cingulum.Mid.L	6.10
(39, −57, 45)	Angular.R	5.85	(−9, −63, 21)	Calcarine.L	5.46
(39, 27, 21)	Frontal.Mid.R	5.54	(−45, −45, 3)	Temporal.Mid.L	5.23
(−45, 33, 39)	Frontal.Mid.L	5.50	(−27, −51, 39)	Parietal.Inf.L	5.09
(39, 3, 45)	Precentral.R	5.44	(21, 45, −21)	Frontal.Mid.Orb.R	4.85
(57, 15, 27)	Frontal.Inf.Oper.R	4.99	(−3, −75, 45)	Precuneus.L	4.64
(51, 57, −9)	Temporal.Inf.R	4.68	(45, −69, 27)	Occipital.Mid.R	4.63
(27, −39, −15)	Fusiform.R	4.62	(−51, −57, 45)	Parietal.Inf.L	4.54
(−3, 33, 15)	Cingulum.Ant.L	4.38	(27, 57, 15)	Frontal.Sup.R	4.45
Slow-3			Slow-2		
(27,3,45)	Frontal.Mid.R	7.16	(15,-63,27)	Precuneus.R	7.59
(9,-57,21)	Precuneus.R	6.46	(9,33,3)	Cingulum.Ant.R	6.75
(51,-75,15)	Temporal.Mid.R	6.11	(57,-45,-3)	Temporal.Mid.R	6.61
(15,-57,15)	Precuneus.R	5.80	(-21,-63,27)	Occipital.Sup.L	6.43
(51,-33,27)	SupraMarginal.R	5.41	(27,3,45)	Frontal.Sup.R	6.04
(9,39,-3)	Cingulum.Ant.R	5.32	(57,-45,-3)	Temporal.Mid.R	5.97
(-3,39,-27)	Rectus.L	5.09	(9,33,9)	Cingulum.Ant.R	5.50
(33,3,27)	Frontal.Inf.Oper.R	5.08	(27,15,57)	Frontal.Mid.R	5.45
(-63,-33,15)	Temporal.Sup.L	4.85	(27,-21,57)	Precuneus.R	5.44
(27,9,45)	Frontal.Mid.R	4.69	(39,-57,33)	Angular.R	5.41

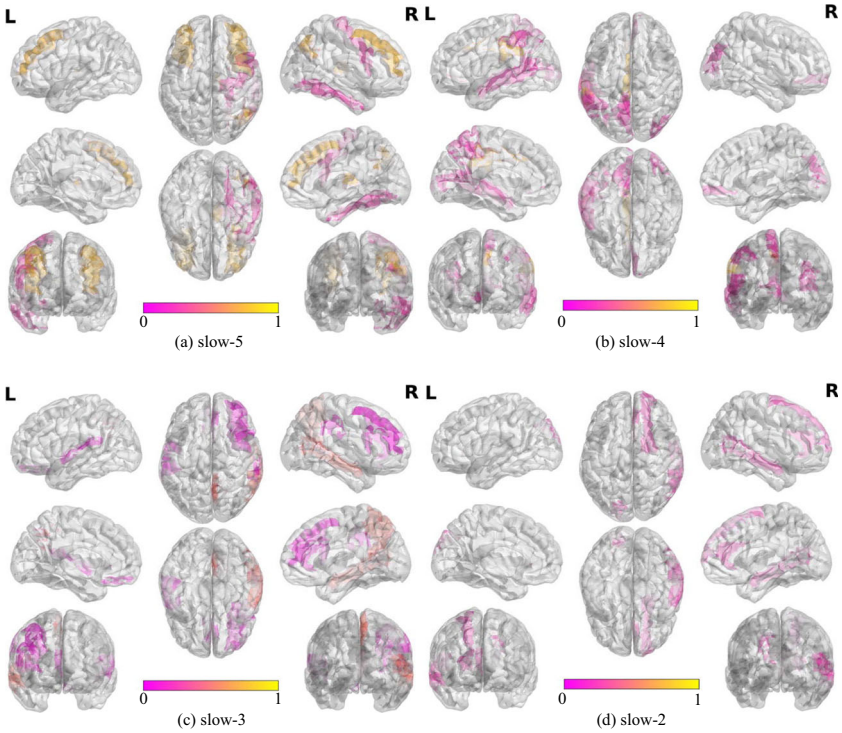

Fig. 5 The biomarkers identified by our H-TS method from four different frequency bands. The magenta color denotes smaller contribution of each brain region for SZ vs. NC classification, while the yellow color denotes larger contribution

frequency brain regions as the identified biomarkers, and calculate the cumulative absolute weight of each patch in the corresponding brain region as the contribution of the brain region for the disease identification. Table 3 lists the top 10 identified biomarkers (i.e., brain regions) and Fig. 5 depicts the biomarkers from the four frequency bands. The color of these biomarkers in Fig. 5 indicates the contribution for identifying the SZ.

The results in Table 3 and Fig. 5 show that key regions of *cognitive control networks*, including *prefrontal cortex* and *anterior cingulate cortex* were found to contribute high weight for identifying schizophrenia in all frequency bands. While *core default mode* network (DMN) regions which include *precuneus*, *angular gyrus*, and *memory* associated *temporal regions*, take a central position in classification of schizophrenia and patients within specific but not all frequency bands. These findings are consistent with the previous reports in the literature (Yu et al. 2014; Zuo et al. 2010; Du et al. 2016; Marsman et al. 2013; Cole and Schneider 2007).

Effectiveness of Selected Biomarkers

The correlation analysis between biomarkers and the PANSS score reveal the effective for diagnosis of SZ, which has been widely used in the SZ research (Morgan et al. 2017). Since the different contributions of LFO amplitudes in the diagnosis of patients and normal controls within four defined frequency bands, we show the relationship between the PANSS score and high frequency biomarkers (i.e., *Frontal_Mid_L*, *Parietal_Inf_L*, *Precuneus_R*, and *Occipital_Sup_L*) in Fig. 6.

From Fig. 6, it is clear to see that the selected biomarkers by the proposed method in different frequency band are highly associated with the PANSS score. Specifically, the values of fALFF in the regions of the selected biomarkers have a significant negative correlation with PANSS scores (i.e., PANSS-positive scale, PANSS-negative scale, and PANSS-general psychopathology). This correlation indicates that high frequency biomarkers are sensitive to the diagnosis of SZ, and have the potential to be used as a diagnostic indicator.

Comparison with State-of-the-Art Methods

We also compare our method with several recent state-of-the-art methods reported in the literature using rs-fMRI data for schizophrenia disease classification. These methods include five FNC-based methods and two multi-frequency methods. For clarity, we list the details of each method in Table 4, including the type of features, classifiers and subjects. From Table 4, we can see that our H-TS method consistent outperforms the comparison methods in SZ versus NC classification. More specifically, H-TS achieves a much higher accuracy (i.e., 93.48%), sensitivity (i.e., 95.65%) and specificity (i.e., 91.30%) compared with other methods. It is worth noting that, our method is the first one to mine and utilize the underlying multi-frequency

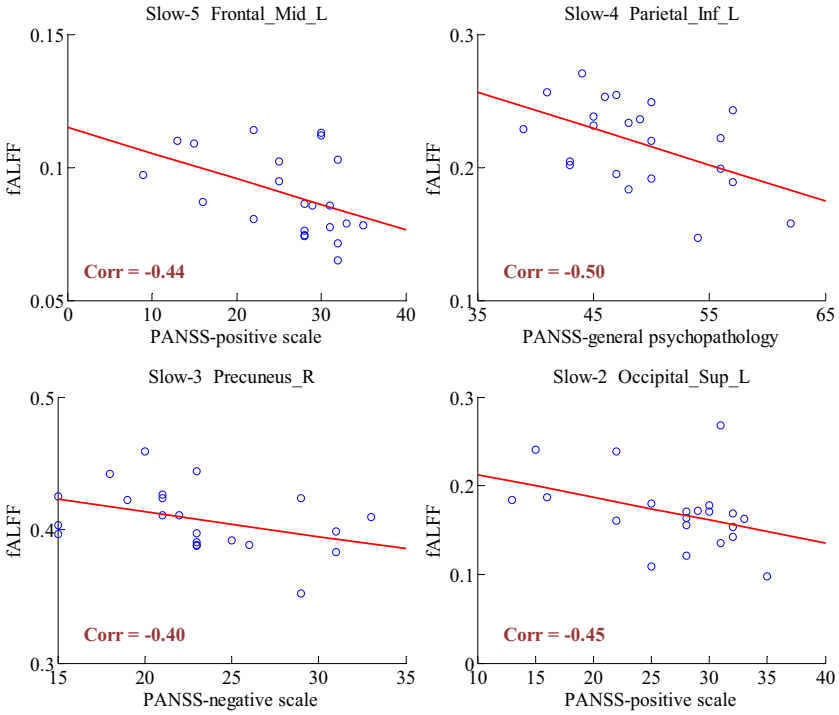

Fig. 6 Correlation between selected biomarker with PANSS

band structure for feature selection in schizophrenia disease diagnosis.

Discussion

In this section, we first investigate the influence of parameters, and then discuss the clinical relevance and limitations of the proposed method.

Parameters Analysis

In our H-TS method, there are three parameters (i.e., λ_h, β_h and H). Here, we discuss the influence of these parameters on the performance of our method using SZ dataset. Specifically, we vary the values of λ_h, β_h and H, and λ_h and β_h are selected from $[2^{-4}, 2^{-3}, \cdots, 2^3]$, while H is chosen from $[1, 2, \cdots, 10]$. We first conduct experiments to illustrate the effectiveness of hierarchical structure, where the ACC and AUC are used to measure the performance of the hierarchical methods. Note that the H-T method proposed in the paper (Han and Zhang 2015) is a variant of our H-TS method, so we do experiments on these two hierarchical methods. Fig. 7 shows the changes in ACC and AUC results for H-TS and H-T methods on the training dataset by varying hierarchies. From Fig. 7, we can observe that at the beginning with the increases of hierarchies, the classification performance improves in terms of ACC and AUC. When hierarchy reaches five, both methods achieve the best performance. Then, for larger hierarchies, the performance becomes slightly worse while still outperforms the first few hierarchies. Therefore, we just set the number of hierarchies to five in our experiments.

We also evaluate the influence of these regularization parameters (i.e., λ_h and β_h) on the performance of our H-TS method. The classification accuracies achieved by H-TS using different parameters are shown in Fig. 8, where we fix H into 1, 3, and 5 and vary the values of the other two parameters. From Fig. 8, we can see that the fluctuations of performance will slightly increase with the varying of hierarchy. The potential mechanism is our method can gradually enforce different levels of information sharing, when H reaches to five it can capture more structural information. This also demonstrates that our proposed H-TS method is not particularly sensitive to the parameter values.

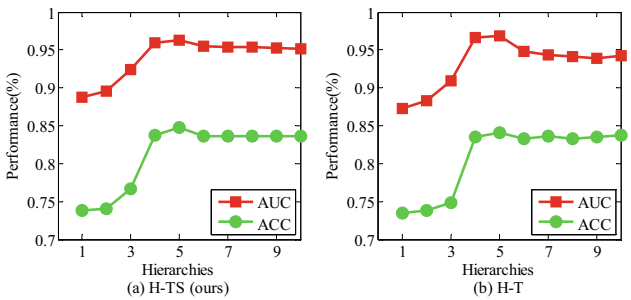

Fig. 7 Results with respect to different hierarchies in SZ vs. NC achieved by different methods on training dataset. (a) H-TS (ours); (b) H-T

Table 4 Comparison with existing studies using rs-fMRI data for schizophrenia identification. FNC: Functional Network Connectivity; DNN: deep neural network; R-LDA: Regularized linear discriminant analysis; ALFF: Amplitude of Low Frequency Fluctuations; fALFF: fractional Amplitude of Low Frequency Fluctuations; V-ELM: ensembles of Extreme Learning Machine; ITG: inferior temporal gyrus

Methods	Feature Type	Classifier	Subjects	ACC (%)	SEN (%)	SPE (%)
Bassett et al. (2012)	FNC	SVM	29 HC + 29 SZ	75.00	85.00	64
Su et al. (2013)	FNC	SVM	32 HC + 23 SZ	81.20	81.20	81.20
Cheng et al. (2015)	FNC	SVM	29 HC + 19 SZ	79.00	83.00	74.00
Kim et al. (2015)	FNC	DNN	50 HC + 50 SZ	85.80	–	–
Kaufmann et al. (2015)	FNC	R-LDA	196 HC + 71 SZ	77.90	50.70	87.80
Chyzhyk et al. (2015)	ALFF	V-ELM	74 HC + 72 SZ	87.67	90.00	85.00
Chyzhyk et al. (2015)	fALFF	V-ELM	74 HC + 72 SZ	82.19	–	–
Guo et al. (2014)	fALFF of the left ITG	SVM	46 HC + 46 SZ	75.00	71.74	78.26
H-TS (ours)	fALFF	SVM ensemble	23 HC + 23 SZ	**93.48**	**95.65**	**91.30**

Clinical Relevance

This study proposes a novel H-TS classification method for SZ diagnosis. Different from previous low-frequency band analysis, we have combined low and high-frequency bands for the diagnosis of schizophrenia. Experimental results on the unique SZ database demonstrate that our method can consistently and substantially outperform the existing classification methods. Specifically, our method can achieve a high accuracy of 93.04% for SZ versus NC classification. It is worth noting that our method is general for other disease diagnosis where multi-modality features can be obtained. For example, the results on the subset of ADNI dataset (Hao et al. 2016) show the efficacy of our method.

In recent studies, several works have shown that different frequency bands within the same neuronal network reveal different patterns of spontaneous fluctuations in physiological (Liang et al. 2012; Yu et al. 2014). However, most existing studies on schizophrenia simply use low-frequency bands, thus ignoring complementary information in high-frequency bands. For example, researchers in Chyzhyk et al. (2015) proposed an extreme learning machine based feature selection method using the fALFF features within the low-frequency, without considering the valuable information conveyed in high-frequency bands. To this end, we propose a H-TS feature selection method, where the underlying specificity and structure information of multi-frequency band data can be well utilized. As can be seen from Fig. 2, our method achieves superior performance over the compared methods.

In this study, we also propose an ensemble classification method by combining the results of multiple classifiers corresponding to multiple frequency bands. As can be seen from Fig. 3, the methods using ensemble strategy often perform better than methods without ensemble, demonstrating the effectiveness of the ensemble strategy in boosting the classification results based on multiple frequency band data. In Table 2, we also show the ensemble classification results on the subset of the ADNI database. It can be seen that our proposed method achieves better performance than the compared methods in both AD versus NC and LMCI versus NC classification tasks.

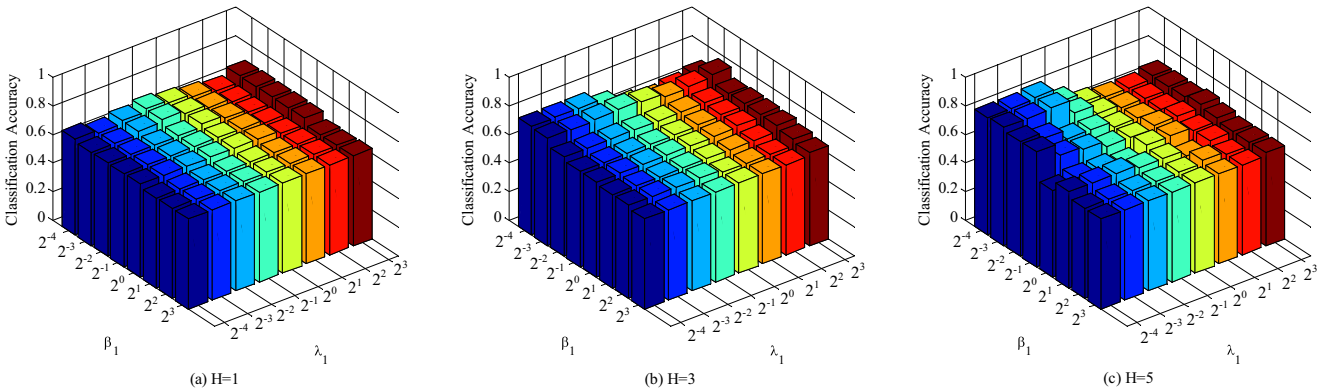

Fig. 8 Accuracies of SZ vs. NC by our H-TS method with different hierarchies on training dataset. **a** $H = 1$, **b** $H = 3$, **c** $H = 5$

Furthermore, in Table 4, we compare our method with the recent state-of-the-art methods for fMRI-based SZ vs. NC classification. It is clear that our method achieves the best classification performance. To investigate the influence of parameters on the performance of our method, we further perform experiments by varying the parameter values (i.e., λ_h, β_h and H) and record their corresponding results, as shown in Figs. 7 and 8. We can see that our method is not very sensitive to the selection of parameter values.

Limitations

There are still several limitations to be considered in this study. First, in the current study, we investigate only the classification between one status of SZ (either treatment-resistant SZ or non-treatment-resistant SZ) and normal controls, and do not test the ability of the classifier to simultaneously discriminate multiple status of SZ, i.e., multi-class classification of treatment-resistant SZ, non-treatment-resistant SZ, and normal controls. However, there may be some problem to straightforwardly treat the multi-class classification as simply multiple binary-class (Duda et al. 2001; Liu and Zhang 2016), and this will be an interesting topic. In addition, in the practical diagnosis of SZ, besides diagnostic label, multiple clinical variables are generally acquired, e.g., PANSS-positive scale and PANSS-negative scale, etc. Since the clinical variables are helpful to reflect the status of SZ with objective assessments, the brain ROI predictors will be detected by regression, which can eliminate the noise from category labels. Therefore, continuous values can be used to discover the robust disease-relevant biomarkers in the future.

Conclusion

In this paper, we proposed a hierarchical structured sparse learning method to effectively utilize the underlying specificity and complex structure information of multiple frequency bands in rs-fMRI data for schizophrenia diagnosis. Specifically, we first extracted the multi-frequency band based features from rs-fMRI data, based on the most discriminative patches selected in each frequency band. We then construct a hierarchical structured sparse learning model to capture the specificity and partial group structures among multiple frequency band data. Finally, we train the corresponding SVM classifier using selected features from each frequency band and combine the outputs of multiple SVMs together by a majority voting strategy. Experimental results on the schizophrenia dataset demonstrate that our H-TS method helps improve the diagnostic accuracy of schizophrenia based on rs-fMRI data with multi-frequency bands.

Information Sharing Statement

The code for hierarchical structured sparse learning will be shared upon individual inquiry to the corresponding author, Dr. Daoqiang Zhang (dqzhang@nuaa.edu.cn). The Alzheimer's disease data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, which can be downloaded at the website http://adni.loni.usc.edu/. According to the data protection policy of the collaborating institution, direct general public access is not available for the schizophrenia dataset.

Acknowledgements This work was supported in part by the National Key Research and Development Program of China (Nos. 2016YFC1306900, 2018YFC2001602), the National Natural Science Foundation of China (Nos. 81771444, 61876082, 61861130366, 61703301), the Royal Society-Academy of Medical Sciences Newton Advanced Fellowship (No. NAF\R1\180371), and the Fundamental Research Funds for the Central Universities (No. NP2018104).

Compliance with Ethical Standards

Conflict of interests The authors declare no conflict of interest.

References

Arribas, J.I., Calhoun, V.D., Adalı, T. (2010). Automatic bayesian classification of healthy controls, bipolar disorder and schizophrenia using intrinsic connectivity maps from fMRI data. *IEEE Transactions on Biomedical Engineering*, *57*(12), 2850–2860.

Bassett, D.S., Nelson, B.G., Mueller, B.A., Camchong, J., Lim, K.O. (2012). Altered resting state complexity in schizophrenia. *NeuroImage*, *59*(3), 2196–2207.

Besga, A., Termenon, M., Graña, M., Echeveste, J., Pérez, J.M., Gonzalezpinto, A. (2012). Discovering Alzheimer's disease and bipolar disorder white matter effects building computer aided diagnostic systems on brain diffusion tensor imaging features. *Neuroscience Letters*, *520*(1), 71–76.

Bhugra, D. (2005). The global prevalence of schizophrenia. *Plos Medicine*, *2*(5), 372–373.

Biswal, B., Zerrin Yetkin, F., Haughton, V.M., Hyde, J.S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926–1929.

Caruana, R. (1997). Multitask learning. *Machine Learning*, *28*(1), 41–75.

Chang, C.C., & Lin, C.J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, *2*(3), 389–396.

Chen, X., Lin, Q., Kim, S., Carbonell, J.G., Xing, E.P. (2011). Smoothing proximal gradient method for general structured sparse learning. In *Twenty-seventh conference on uncertainty in artificial intelligence* (pp. 105–114): ACM.

Cheng, H., Newman, S., Goñi, J., Kent, J.S., Howell, J., Bolbecker, A., Puce, A., O'Donnell, B.F., Hetrick, W.P. (2015). Nodal centrality of functional network in the differentiation of schizophrenia. *Schizophrenia Research*, *168*(2), 345–352.

Cherkassky, V. (1997). The nature of statistical learning theory. *IEEE Transactions on Neural Networks*, *8*(6), 1564–1564.

Chyzhyk, D., Savio, A., Graña, M. (2015). Computer aided diagnosis of schizophrenia on resting state fMRI data by ensembles of ELM. *Neural Networks*, *68*, 23–33.

Cole, M.W., & Schneider, W. (2007). The cognitive control network: integrated cortical regions with dissociable functions. *NeuroImage*, *37*(1), 343–360.

Demirci, O., & Calhoun, V.D. (2009). Functional magnetic resonance imaging-implications for detection of schizophrenia. *European Neurological Review*, *4*(2), 103–106.

Demirci, O., Clark, V.P., Magnotta, V.A., Andreasen, N.C., Lauriello, J., Kiehl, K.A., Pearlson, G.D., Calhoun, V.D. (2008). A review of challenges in the use of fMRI for disease classification/characterization and a projection pursuit application from multi-site fMRI schizophrenia study. *Brain Imaging and Behavior*, *2*(3), 207–226.

Dietterich, T.G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*(7), 1895–1923.

Du, Y., Pearlson, G.D., Yu, Q., He, H., Lin, D., Jing, S., Wu, L., Calhoun, V.D. (2016). Interaction among subsystems within default mode network diminished in schizophrenia patients: a dynamic connectivity approach. *Schizophrenia Research*, *170*(1), 55–65.

Duda, R.O., Hart, P.E., Stork, D.G. (2001). *Pattern classification*. USA: Wiley.

Guo, W., Su, Q., Yao, D., Jiang, J., Zhang, J., Zhang, Z., Yu, L., Zhai, J., Xiao, C. (2014). Decreased regional activity of default-mode network in unaffected siblings of schizophrenia patients at rest. *European Neuropsychopharmacology*, *24*(4), 545–552.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1), 389–422.

Han, L., & Zhang, Y. (2015). Learning multi-level task groups in multi-task learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 2638–2644): AAAI.

Hao, X., Yao, X., Yan, J., Risacher, S.L., Saykin, A.J., Zhang, D., Shen, L. (2016). Identifying multimodal intermediate phenotypes between genetic risk factors and disease status in Alzheimer's disease. *Neuroinformatics*, *14*(4), 439–452.

Hastie, T., Tibshirani, R., Friedman, J. (2001). *The elements of statistical learning*. New York: Springer.

Hoptman, M.J., Zuo, X.N., Butler, P.D., Javitt, D.C., D'Angelo, D., Mauro, C.J., Milham, M.P. (2010). Amplitude of low-frequency oscillations in schizophrenia: a resting state fMRI study. *Schizophrenia Research*, *117*(1), 13–20.

Huang, X.Q., Lui, S., Deng, W., Chan, R.C.K., Wu, Q.Z., Jiang, L.J., Zhang, J.R., Jia, Z.Y., Li, X.L., Li, F. (2010). Localization of cerebral functional deficits in treatment-naive, first-episode schizophrenia using resting-state fMRI. *NeuroImage*, *49*(4), 2901–2906.

Jafri, M.J., & Calhoun, V.D. (2006). Functional classification of schizophrenia using feed forward neural networks. In *2006 International conference of the IEEE engineering in medicine and biology society* (pp. 6631–6634).

Jalali, A., Sanghavi, S., Ruan, C., Ravikumar, P.K. (2010). A dirty model for multi-task learning. In *Advances in neural information processing systems 23* (pp. 964-972). Curran Associates, Inc.

Jie, B., Liu, M., Liu, J., Zhang, D., Shen, D. (2017). Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, *64*(1), 238–249.

Jie, B., Liu, M., Zhang, D., Shen, D. (2018). Sub-network kernels for measuring similarity of brain connectivity networks in disease diagnosis. *IEEE Transactions on Image Processing*, *27*(5), 2340–2353.

Kaufmann, T., Skåtun, K.C., Alnæs, D., Doan, N.T., Duff, E.P., Tønnesen, S., Roussos, E., Ueland, T., Aminoff, S.R., Lagerberg, T.V. (2015). Disintegration of sensorimotor brain networks in schizophrenia. *Schizophrenia Bulletin*, *41*(6), 1326–1335.

Kim, J., Calhoun, V.D., Shim, E., Lee, J.H. (2015). Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *NeuroImage*, *124*, 1326–1335.

Knyazev, G.G. (2007). Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neuroscience and Biobehavioral Reviews*, *31*(3), 377–95.

Lachowicz, M., & wrzosek, D. (2001). Nonlocal bilinear equations: equilibrium solutions and diffusive limit. *Mathematical Models and Methods in Applied Sciences*, *11*(08), 1393–1409.

Lam, L., & Suen, S.Y. (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans*, *27*(5), 553–568.

Lian, C., Liu, M., Zhang, J., Shen, D. (2019). Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2018.2889096.

Liang, X., Wang, J., Yan, C., Shu, N., Xu, K., Gong, G., He, Y. (2012). Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: a resting-state functional MRI study. *Plos One*, *e32*(3), 766.

Liu, F., Wee, C.Y., Chen, H., Shen, D. (2014). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification. *NeuroImage*, *84*, 466–475.

Liu, J., Ji, S., Ye, J. (2009). Multi-task feature learning via efficient l2,1-norm minimization. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (pp. 339–348): AUAI Press.

Liu, J., Ji, S., Ye, J. (2009). SLEP: sparse learning with efficient projections. Arizona State University.

Liu, M., & Zhang, D. (2014). Sparsity score: a novel graph-preserving feature selection method. *International Journal of Pattern Recognition and Artificial Intelligence*, *1450*(04), 009.

Liu, M., & Zhang, D. (2016). Feature selection with effective distance. *Neurocomputing*, *215*, 100–109.

Liu, M., Zhang, D., Adeli, E., Shen, D. (2016). Inherent structure based multi-view learning with multi-template feature representation for Alzheimer's disease diagnosis. *IEEE Transactions on Biomedical Engineering*, *63*(7), 1473–1482.

Liu, M., Zhang, D., Chen, S., Xue, H. (2016). Joint binary classifier learning for ECOC-based multi-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(11), 2335–2341.

Liu, M., Zhang, D., Shen, D. (2016). Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. *IEEE Transactions on Medical Imaging*, *35*(6), 1463–1474.

Liu, M., Zhang, J., Adeli, E., Shen, D. (2017). Deep multi-task multi-channel learning for joint classification and regression of brain status. In *International conference on medical image computing and computer-assisted intervention* (pp. 3–11): Springer.

Marsman, A., Mandl, R.C., Mp, V.D.H., Boer, V.O., Wijnen, J.P., Klomp, D.W., Luijten, P.R., Hilleke, E. (2013). H.P.: Glutamate changes in healthy young adulthood. *European Neuropsychopharmacology*, *23*(11), 1484–1490.

Morgan, A.R., Touchard, S., O' Hagan, C., Sims, R., Majounie, E., Escott-Price, V., Jones, L., Williams, J., Morgan, B.P. (2017). The correlation between inflammatory biomarkers and polygenic risk score in Alzheimer's disease. *Journal of Alzheimer's Disease*, *56*(1), 25–36.

Neuhaus, A.H., Popescu, F.C., Grozea, C., Hahn, E., Hahn, C., Opgenrhein, C., Urbanek, C., Dettling, M. (2011). Single-subject classification of schizophrenia by event-related potentials during selective attention. *NeuroImage*, *55*(2), 514–521.

Noriaki, Y., Jun, M., Ryuichiro, H., Giuseppe, L., Kazuhisa, S., Yuki, K., Hitoshi, K., Miho, K., Takashi, Y., Fukuda, M. (2016). A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nature Communications*, *7*(11254), 1–12.

Rosario, B.L., Ziolko, S.K., Weissfeld, L.A., Price, J.C. (2008). Assessment of parameter settings for SPM5 spatial normalization of structural MRI data: application to type 2 diabetes. *NeuroImage*, *41*(2), 363–370.

Shen, H., Wang, L., Liu, Y., Hu, D. (2010). Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *NeuroImage*, *49*(4), 3110–3121.

Song, X.W., Dong, Z.Y., Long, X.Y., Li, S.F., Zuo, X.N., Zhu, C.Z., He, Y., Yan, C.G., Zang, Y.F. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *Plos One*, *6*(9), 1–12.

Su, L., Wang, L., Shen, H., Feng, G., Hu, D. (2013). Discriminative analysis of non-linear brain connectivity in schizophrenia: an fMRI study. *Frontiers in Human Neuroscience*, *7*(702), 1–12.

Takayanagi, Y., Takahashi, T., Orikabe, L., Mozue, Y., Kawasaki, Y., Nakamura, K., Sato, Y., Itokawa, M., Yamasue, H., Kasai, K., Kurachi, M., Okazaki, Y., Suzuki, M. (2011). Classification of first-episode schizophrenia patients and healthy subjects by automated mri measures of regional brain volume and cortical thickness. *PLOS ONE*, *6*(6), 1–10.

Wang, J., Wang, Q., Peng, J., Nie, D., Zhao, F., Kim, M., Zhang, H., Wee, C.Y., Wang, S., Shen, D. (2017). Multi-task diagnosis for autism spectrum disorders using multi-modality features: a multi-center study. *Human Brain Mapping*, *38*(6), 3081–3097.

Wang, J., Zuo, X., He, Y. (2010). Graph-based network analysis of resting-state functional MRI. *Frontiers in Systems Neuroscience*, *4*(16), 1–14.

Wang, M., Hao, X., Huang, J., Wang, K., Xu, X., Zhang, D. (2017). Multi-level multi-task structured sparse learning for diagnosis of schizophrenia disease. In *International conference on medical image computing and computer-assisted intervention* (pp. 46–54): Springer.

Wang, Z., Zhang, Z., Liao, W., Xu, Q., Zhang, J., Lu, W., Jiao, Q., Chen, G., Feng, J., Lu, G. (2014). Frequency-dependent amplitude alterations of resting-state spontaneous fluctuations in idiopathic generalized epilepsy. *Epilepsy Research*, *108*(5), 853–860.

Yu, R., Chien, Y.L., Wang, H.L.S., Liu, C.M., Liu, C.C., Hwang, T.J., Ming, H.H., Hwu, H.G., Tseng, W.Y.I. (2014). Frequency-specific alternations in the amplitude of low-frequency fluctuations in schizophrenia. *Human Brain Mapping*, *35*(2), 627–637.

Zhang, D., Huang, J., Jie, B., Du, J., Tu, L., Liu, M. (2018). Ordinal pattern: a new descriptor for brain connectivity networks. *IEEE Transactions on Medical Imaging*, *37*(7), 1711–1722.

Zhou, Z., Wang, J.B., Zang, Y.F., Pan, G. (2018). PAIR comparison between two within-group conditions of resting-state fMRI improves classification accuracy. *Frontiers in Neuroscience*, *11*, 740.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *67*(2), 301–320.

Zuo, X.N., Di, M.A., Kelly, C., Shehzad, Z.E., Gee, D.G., Klein, D.F., Castellanos, F.X., Biswal, B.B., Milham, M.P. (2010). The oscillating brain: complex and reliable. *NeuroImage*, *49*(2), 1432–1445.

Affiliations

Mingliang Wang[1,2] · Xiaoke Hao[1] · Jiashuang Huang[1] · Kangcheng Wang[3] · Li Shen[4] · Xijia Xu[5] · Daoqiang Zhang[1] · Mingxia Liu[6]

Mingliang Wang
wml489@nuaa.edu.cn

[1] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing, China

[2] The State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, Shaanxi, China

[3] Department of Psychology, Southwest University, Chongqing, China

[4] Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA USA

[5] Department of Psychiatry, Affiliated Nanjing Brain Hospital, Nanjing Medical University, Nanjing, China

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