High-Accuracy Classification of Attention Deficit Hyperactivity Disorder With $l_{2,1}$-Norm Linear Discriminant Analysis and Binary Hypothesis Testing

YIBIN TANG [1], **XUFEI LI**[1], **YING CHEN**[2], **YUAN ZHONG**[3], **AIMIN JIANG**[1], **AND CHUN WANG**[4]

[1]College of Internet of Things Engineering, Hohai University, Changzhou 213022, China
[2]School of Information Science and Engineering, Southeast University, Nanjing 210096, China
[3]School of Psychology, Nanjing Normal University, Nanjing 210024, China
[4]Department of Mood Disorders, Nanjing Medical University, Nanjing 210024, China

Corresponding authors: Yibin Tang (tangyb@hhuc.edu.cn) and Chun Wang (fm51109@163.com)

ABSTRACT Attention Deficit Hyperactivity Disorder (ADHD) is a high incidence of neurobehavioral disease in school-age children. Its neurobiological diagnosis (or classification) is meaningful for clinicians to give proper treatment for ADHD patients. The existing ADHD classification methods suffer from two problems, i.e., insufficient data and noise disturbance. In this paper, a high-accuracy classification method is proposed by using brain Functional Connectivity (FC) as ADHD features, where an $l_{2,1}$-norm Linear Discriminant Analysis (LDA) model and a binary hypothesis testing framework are effectively employed. In detail, we introduce a binary hypothesis testing framework to cope with insufficient data of ADHD database. The FCs of test data (without seeing its label) are used for training and thus affect the subspace learning of training data under binary hypotheses. On other hand, the $l_{2,1}$-norm LDA model generates a subspace to represent ADHD features, aiming to overcome noise disturbance. By robustly learning ADHD features, the subspace energy difference between binary hypotheses becomes more discriminative. Thereby, the true hypothesis can be rightly estimated with its larger subspace energy, which provides reliable evidence to predict the label of test data. By the test on ADHD-200 database, it shows that our method outperforms other state-of-the-art methods with the significant average accuracy of 97.6%. Moreover, the corresponding result analysis with ADHD symptom score and the explanation of discriminative FCs between ADHD and healthy control groups are given, which further verifies the validity of our classification method.

INDEX TERMS ADHD classification, binary hypothesis, feature learning, LDA, subspace clustering.

I. INTRODUCTION

Attention Deficit Hyperactivity Disorder (ADHD) is a high incidence of neurobehavioral disease in children, characterized by difficulty paying attention, excessive activity and other related behavioral disorders. About 5-7% of school-age children suffer from ADHD, and 30-50% of them keep

The associate editor coordinating the review of this manuscript and approving it for publication was Charith Abhayaratne.

ADHD symptom in their adulthood [1]. It is of vital importance to diagnose this disease as accurate as possible such that the treatment can be in time provided for children patients. Current clinical diagnosis depends on subjective scoring via various Hamilton scales, where ADHD patients are directly observed to identify their symptomatological features [2]. But it needs experienced clinicians and has limited ability to discover the potential ADHD bioinformation. Thus, more neurobiological diagnostic methods

are proposed, wherein ADHD classifications with machine learning (ML) and deep learing (DL) are rapidly developed in recent years.

Unlike symptomatological diagnosis using subjective scoring, neurobiological diagnosis relies on the physics data of brain collected from objective observations, e.g., EEG, PET, sMRI and fMRI [3], [4]. With some data procedures, various biological signals are generated to give a full description of brain status, and provide useful bioinformation for the final ADHD judgment. The frequently used biosignals include cortical thickness, gray matter probability, Regional Homogeneity (ReHo) and Amplitude of Low Frequency Fluctuation (ALFF) [5]. Among these biosignals, Functional Connectivity (FC) has been proven to reveal the difference between ADHD and healthy control individuals with great success [6], [7]. Therefore, given a set of resting-state fMRI data, we here focus our classification work on the FC analysis.

Generally speaking, ADHD classification with machine learning is mainly divided into three phases, i.e., feature selection, feature extraction and label decision via classifiers. In feature selection, some discriminative biosignals are sampled from a huge number of biosignals as their features. Several successful feature selection approaches have been adopted in ML-based classification, such as Support Vector Machines Recursive Feature Elimination (SVM-RFE) [8], Least Absolute Shrinkage and Selection Operator (LASSO) [9] and Elastic Net [10]. Besides, more advanced strategies are in pursuit of better selection performance. For example, a Reliable Relief (R-Relief) method is recently presented and gives a set of feature weights to fractional ALFFs during feature selection [11]. As for FCs, a graph-kernel regularized LASSO is performed on the FC network, which preserves the local structure among the selected FCs [12]. An ensemble learning strategy is also employed for the FC selection by combining various Boosting approaches [13].

Feature extraction contributes to find a learned space to effectively represent the selected features. In this learned space, either dimensionality reduction is performed to strengthen the reliable components of features, or a better clustering is achieved to minimize the inter-class disturbance and centralize the same labeled features. To this end, some plain data analysis models such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) [14] and Independent Components Analysis (ICA) [10], [15] are pioneered to tackle the selected features. Recently, more remarkable signal models are cast in feature extraction. In [16], a dynamic sparse coding algorithm is presented on EEG signal to detect ADHD biomarkers in a 'dictionary' space composed of learned atoms. A sparse representation model is further incorporated on FCs to identify ADHD individuals with the grouped dictionary learning [17]. Subspace projection algorithms are also well developed in feature extraction. Several subspaces are effectively designed in consideration of intra- and inter-class relationship of subjects to improve the classification accuracy [18], [19]. Moreover,

since FCs can describe a topographic map of brain, some special graph-based methods are performed for ADHD classification. Various graph measures are carried out to explore the FC network features [20], [21]. Among these methods, a fusion fMRI method [22] attracts more attention due to its recognition of reliable FCs by using the affinity propagation clustering on FC network. In [23], the discriminative FC sub-networks are further discussed such that a robust FC pattern is found for the ADHD feature extraction.

To the best of our knowledge, a training and testing scheme is sophisticatedly adopted in label decision. Here, a target classifier is learned by the features of labeled training data, and then the predicted label of test data is given via this classifier. Therefore, the design of classifier becomes a key point. Various well-developed classifiers have been applied in ADHD classification, including logistic regression, bagging, random forest and decision tree [11], [24]. Especially, the classifiers with SVM [20], [25] and Extreme Learning Machine (ELM) [26] are more welcomed due to their fitness for the ADHD databases of small size. In an L_1BioSVM method [27], the existing l_1-norm SVM model is redesigned with a bi-objective optimization problem, and thus it achieves attractive classification accuracy. Moreover, an ELM combined with a classifier-selecting genetic approach is given to identify ADHD subjects, which only needs to detect three brain regions with their sMRI data [28].

On the other hand, deep learning has shown its potential power in ADHD classification. DL-based classification utterly removes the barriers among the above three phases. Its multilayered and nonlinear learning strategy provides more flexible ability in feature extraction and label decision. In the early work, a Convolutional Neural Network (CNN), named as FCNet [29], is proposed to describe the FCs of ADHD individuals, which efficiently uses a fully connected network to compute the similarity among the extracted features. As an advanced version of FCNet, Deep fMRI [30] sequentially introduces a self-learned subnetwork as its classifier to generate a whole deep learning framework from feature extraction to label prediction. Later, a three-dimension CNN (3D-CNN) model is given to investigate the local spatial patterns of ADHD from various fMRI datasets [31]. Very recently, a four-dimension CNN is presented for the better spatio-temporal feature learning [32].

Though ADHD classification with machining learning and deep learning has obtained some impressive achievements, its accuracy performance still has a gap to meet the requirement of clinical diagnosis. In fact, the accuracies of most state-of-the-art methods fluctuate between 65% and 87% [11], [18], [22], [27], [30]–[32]. The reason comes from two aspects. The lack of data is the first and major cause to hinder the improvement of accuracy, since ADHD databases are with their limited subjects. For ML-based methods, an effective training and testing scheme is founded on the assumption, that is, the features of test data are contained in the feature space of training data. But the features of training data in practice cannot well represent those of test data due to the small

size of ADHD databases. Thus the classification accuracy is inevitable deteriorated. A dearth of data also has significant impact on DL-based methods. DL is always eager for a huge amount of data, since thousands or millions of parameters need be trained in deep neural network. It naturally conflicts with the limited ADHD data. For the sake of inadequate training, the accuracy of DL-based classification is sometimes even worse than those of ML-based methods. The second cause is the noise disturbance in ADHD data. These noises are not only from the biosignal collection procedure, such as head movement and uncorrected image aligment in MRI scans, but also from other pathological factors. For example, ADHD children often have other associate diseases, e.g., anxiety and learning disorders. It greatly increases the difficulty in ADHD feature learning, since the feature noise of other diseases should be pre-excluded in principle. Unfortunately, most of existing methods do not well treat the noise. They pay less attention on the robust feature learning to remove these noises. As a result, the accuracy is decreased.

Motivated by the recent process, we propose an ADHD classification method by using $l_{2,1}$-norm LDA and binary hypothesis testing. Here, a binary hypothesis testing framework is introduced for ADHD classification to cope with the insufficient data problem of ADHD database. Meanwhile, to overcome noise disturbance, an $l_{2,1}$-norm LDA model is adopted in feature extraction to robustly learn the ADHD features, where the corresponding optimization solution is given. Moreover, an energy-based decision strategy is proposed for the learned features in different hypotheses, which efficiently predicts the label of test data to identify ADHD individuals. On the platform of ADHD-200 database, the experiments show that our classification method outperforms other state-of-the-art methods with the higher average accuracy of 97.6%. The corresponding result analysis with ADHD symptom score and the explanation of discriminative FCs of ADHD are also given to verify the validity of our classification method.

II. MATERIALS AND METHODS

A. IMAGE DATASET AND PREPROCESSING

In our work, all resting-state fMRI data is from the ADHD-200 consortium (http://fcon_1000.projects.nitrc.org /indi/adhd200). We use four ADHD databases to investigate the binary classification between ADHD and healthy control subjects. These databases are from NeuroImage (NI), New York University medical center (NYU), Kennedy Krieger Institute (KKI) and Peking University (PU), with their information shown in Table 1. For ADHD data preprocessing, we obtain the time course values of BOLD signals from the connectome website (http://www.preprocessed-connectomes-project.org/adhd200). The preprocessing procedure of these time courses includes removing of first four time points, slice time correction, motion correction (first image taken as the reference), registration on $4 \times 4 \times 4$ voxel resolution using the Montreal Neurological Institute (MNI)

TABLE 1. Summary of several ADHD-200 data.

Site	Age	Female	Male	Control	ADHD	Total
NYU	7-18	77	145	99	123	222
KKI	8-13	37	46	61	22	83
NI	11-22	17	31	23	25	48
PU	8-17	52	142	116	78	194
PU_1*	8-17	36	48	62	24	86

*PU includes three subsets, and PU_1 is the first subset of PU.

space, filtration (band pass filter $0.009Hz < f < 0.08Hz$) and smoothing using 6mm FWHM Gaussian filter. Brains are parceled in accordance with Automated Anatomical Labeling-116 (AAL-116) atlas, where 90 Regions of Interest (ROI) are used. The FCs are generalized by Pearson correlation between regional BOLD signals. Finally, a Fisher's r-to-z transform is utilized on FCs to transform their sampling distribution of correlation coefficients for normality.

B. ADHD CLASSIFICATION FRAMEWORK WITH BINARY HYPOTHESIS

Different from the training and testing scheme, the binary hypothesis approach provides an alternative way to detect ADHD individuals [18], [19]. Its basic idea is to let the FCs of test data (without seeing its label) affect the selected FCs of training data. In detail, we cast the FCs of labeled training data and test data labeled under an ADHD or control hypothesis into the feature selection procedure. The discriminative selected FCs of training data are only generated under the true hypothesis of test data, while the obscure ones are under the false hypothesis. Such influence is delivered to feature extraction. For the true hypothesis, the effective subspace is learned by the $l_{2,1}$-norm LDA model. The strong projected components of selected FCs are then obtained in the learned subspace with their large energies. Conversely, the projected components keep small energies with the subspace learning of $l_{2,1}$-norm LDA model under the false hypothesis. Thus, the projected-energy difference between binary hypotheses is generated, which forms a significant measure for the label prediction. By comparing the projected energies of FCs of training data under different hypotheses, we can predict the label of test data with the hypothesis of large projected energy. Note that, since the binary hypothesis approach is not necessary to meet the aforementioned space coverage requirement for training data, it becomes effective to overcome the insufficient data problem.

We give the proposed ADHD classification framework in Figure 1. In term of the ML-based classification, our scheme includes three phases, i.e., feature selection, feature extraction and ADHD decision. We employ a binary hypothesis model to initialize the label of test data. The test data is in advance labeled as a healthy control ($\mathcal{H}_0$) or ADHD ($\mathcal{H}_1$) subject. In the feature selection, both FCs of training and test data are utilized to compute the reliability value on each FC by the SVM-RFE approach. By sorting these values descendingly, two feature ranking orders, $\boldsymbol{R}^{\mathcal{H}_0}$ and $\boldsymbol{R}^{\mathcal{H}_1}$, are
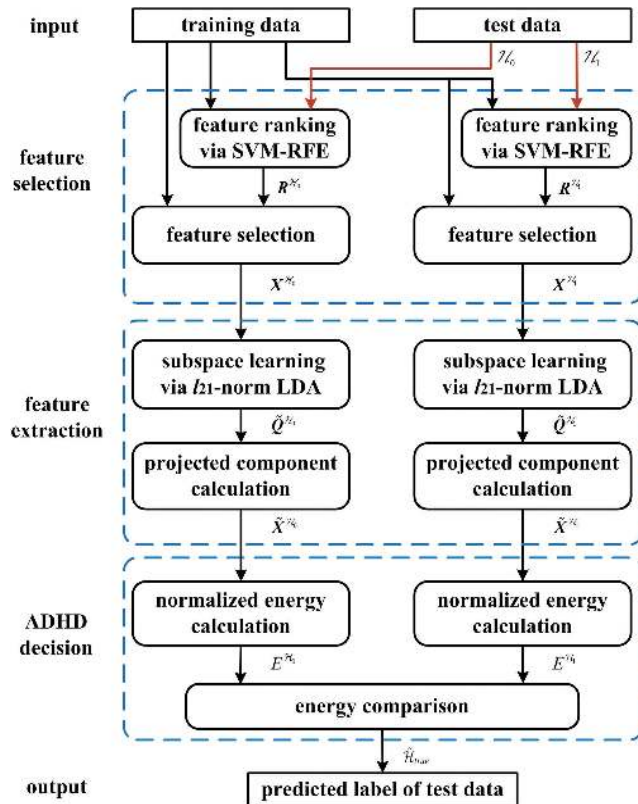
FIGURE 1. Framework of proposed ADHD classification.

generated in binary hypotheses. Hereafter, we only focus on the FCs of training data. The selected feature (selected FC) sets, $X^{\mathcal{H}_0}$ and $X^{\mathcal{H}_1}$, with their feature number N are obtained from training data by the ranking orders. In the feature extraction, subspaces are learned from the corresponding selected feature sets by the $l_{2,1}$-norm LDA model. In these subspaces, the projected components of selected features are respectively calculated as $\tilde{X}^{\mathcal{H}_0}$ and $\tilde{X}^{\mathcal{H}_1}$. ADHD decision is finally performed by the comparison between the normalized $l_{2,1}$-norm energies, $E^{\mathcal{H}_0}$ and $E^{\mathcal{H}_1}$. The true hypothesis is estimated as $\tilde{\mathcal{H}}_{true}$ with the large energy, and thus the predicted label is given to the test data.

It is worth mentioning that, compared with [18], [19], there exist some differences in our framework. In [18], [19], an energy balance operation is employed in the feature selection phase. It aims to avoid the violent energy fluctuation of selection features between binary hypotheses and generate the comparable projected energies. However, by consideration of noise, some selected FCs may be contaminated and viewed as outliers with their large offset values. In this case, the energy balance is improper to some extent. Therefore, we remove this operation and use the $l_{2,1}$-norm LDA model to deal with these outliers. But we still need the projected energies normalized in the ADHD decision to achieve the fair energy comparison under different hypotheses.

C. FEATURE EXTRACTION VIA $l_{2,1}$-NORM LDA MODEL

Some notations are firstly introduced as follows. We define x_a and x_c as the selected features of ADHD and healthy control subjects, which derive from the individual FC data. The selected feature sets of ADHD and control groups are represented as $X_a = [x_a^1 \ x_a^2 \ \dots \ x_a^{n_a}]$ and $X_c = [x_c^1 \ x_c^2 \ \dots \ x_c^{n_c}]$ with their subject numbers n_a and n_c. Thus, the LDA model is used on these feature sets to extract most discriminative projected features for a better subspace clustering. The classical LDA is described as

$$\tilde{Q} = \arg \max_{Q^T Q = I} \frac{tr\left(Q^T S_b Q\right)}{tr\left(Q^T S_w Q\right)}, \tag{1}$$

where the transformation matrix Q is regarded as a subspace with its K orthogonal bases. The between-class scatter matrix S_b and within-class scatter matrix S_w are defined as

$$S_b = n_a \left(\bar{x}_a - \bar{x}\right)\left(\bar{x}_a - \bar{x}\right)^T + n_c \left(\bar{x}_c - \bar{x}\right)\left(\bar{x}_c - \bar{x}\right)^T, \tag{2}$$

$$S_w = \sum_{i=1}^{n_a} \left(x_a^i - \bar{x}_a\right)\left(x_a^i - \bar{x}_a\right)^T$$
$$+ \sum_{i=1}^{n_c} \left(x_c^i - \bar{x}_c\right)\left(x_c^i - \bar{x}_c\right)^T, \tag{3}$$

where $\bar{x}_a$ and $\bar{x}_c$ are respectively the means of selected features for ADHD and healthy control subjects, $\bar{x}$ is the mean value of all selected features. Problem (1) can be turned with its l_2-norm form as

$$\tilde{Q} = \arg \max_{Q^T Q = I} \frac{\|Q^T G_b\|_2^2}{\|Q^T G_w\|_2^2}, \tag{4}$$

where the between-class and within-class kernels are given as $G_b = [\sqrt{n_a}\left(\bar{x}_a - \bar{x}\right) \ \sqrt{n_c}\left(\bar{x}_c - \bar{x}\right)]$ and $G_w = [X_a - \bar{x}_a \mathbf{1}_{n_a}^T \ X_c - \bar{x}_c \mathbf{1}_{n_c}^T]$, $\mathbf{1}_{n_a}$ and $\mathbf{1}_{n_c}$ are the vectors of ones with their corresponding lengths of n_a and n_c.

Unfortunately, the above l_2-norm LDA model is sensitive to outliers. It always attempts to force the outliers close to their centers with heavy weights, wherein these weights are computed from the distance metric of squared l_2-norm. But as a general rule, the outliers usually contain serious noise. The clustering operation on outliers becomes meaningless and conversely deteriorates the clustering performance of other selected features. Therefore, more measures, i.e., nuclear-norm [33], l_1-norm [34] and $l_{2,p}$-norm [35], [36], are proposed to improve the robustness of LDA. Among them, the $l_{2,1}$-norm LDA model has shown its strong ability to cope with outliers, due to $l_{2,1}$-norm well exploits the geometric structure of data rather than l_2-norm [37]. Its model is written as

$$\tilde{Q} = \arg \max_{Q^T Q = I} \frac{\|Q^T G_b\|_{2,1}}{\|Q^T G_w\|_{2,1}}, \tag{5}$$

where the between-class kernel now is turned to $G_b = [n_a\left(\bar{x}_a - \bar{x}\right) \ n_c\left(\bar{x}_c - \bar{x}\right)]$. The $l_{2,1}$-norm for a matrix A is

defined as

$$\|A\|_{2,1} = \sum_j \|A(:, j)\|_2, \quad (6)$$

where $A(:, j)$ is the j-th column vector of A.

As we know, problem (5) can be treated as a ratio trace optimization problem. It has no closed-form solution and is very often approximated by the solution of determinant ratio problem. Though the approximated solution works well on low-dimension data, its solution error sharply increases for high-dimension data. To address this issue, a non-greedy method is also referred in [37] to give an exquisite solution for the $l_{2,1}$-norm LDA model. More specifically, it firstly transforms the numerator of (5) as

$$\|Q^T G_b\|_{2,1} = tr\left(Q^T G_b D_b G_b^T Q\right), \quad (7)$$

where $D_b = diag\left(\frac{1}{\|Q^T G_b(:,1)\|_2}, \frac{1}{\|Q^T G_b(:,2)\|_2}\right)$ is the diagonal matrix. The similar operation is done on the denominator of (5) as

$$\|Q^T G_w\|_{2,1} = tr\left(Q^T G_w D_w G_w^T Q\right), \quad (8)$$

where

$$D_w = diag\left(\frac{1}{\|Q^T G_w(:,1)\|_2}, \frac{1}{\|Q^T G_w(:,2)\|_2}, \cdots, \frac{1}{\|Q^T G_w(:,n)\|_2}\right)$$

is the diagonal matrix with the total subject number $n = n_a + n_c$.

Turn problem (5) into a simple convex problem as

$$\tilde{Q} = \arg \max_{Q^T Q = I} tr\left(Q^T G_b D_b G_b^T Q\right) - tr\left(Q^T G_w D_w G_w^T Q\right). \quad (9)$$

Now, any gradient descent approach can be performed on (9) to obtain its optimal subspace $\tilde{Q}$. However, since the matrices D_b and D_w are the functions of Q, it makes the gradient calculation more complicated. Therefore, an iteration algorithm is adopted to solve (9) with two-phase operation. In each iteration, given D_b and D_w, a new subspace Q is firstly learned by the gradient descent method. Then the matrices D_b and D_w are updated with this learned subspace for the next iteration. After several iterations, the subspace can converge to the optimal one $\tilde{Q}$ (see more details in [37]).

As for our binary hypothesis framework, we input the selected feature sets, $X^{\mathcal{H}_0}$ and $X^{\mathcal{H}_1}$, into the $l_{2,1}$-norm LDA model to obtain their corresponding optimal subspaces, $\tilde{Q}^{\mathcal{H}_0}$ and $\tilde{Q}^{\mathcal{H}_1}$. The projected components of selected features are sequentially achieved by

$$\begin{cases} \tilde{X}^{\mathcal{H}_0} = (\tilde{Q}^{\mathcal{H}_0})^T X^{\mathcal{H}_0} \\ \tilde{X}^{\mathcal{H}_1} = (\tilde{Q}^{\mathcal{H}_1})^T X^{\mathcal{H}_1}. \end{cases} \quad (10)$$

D. ADHD DECISION

We employ a comparison operation in the ADHD decision to replace the traditional classifier. To remove the energy fluctuation among selected features, the normalized energies, $E^{\mathcal{H}_0}$ and $E^{\mathcal{H}_1}$, are calculated for $\tilde{X}^{\mathcal{H}_0}$ and $\tilde{X}^{\mathcal{H}_1}$ by

$$E^{\mathcal{H}_i} = \frac{\|\tilde{X}^{\mathcal{H}_i}\|_{2,1}}{\|X^{\mathcal{H}_i}\|_{2,1}}, \quad i \in \{0, 1\}. \quad (11)$$

Thus, the true hypothesis is estimated with

$$\tilde{\mathcal{H}}_{true} = \begin{cases} \mathcal{H}_1, & \triangle E > T_{thr} \\ \mathcal{H}_0, & otherwise, \end{cases} \quad (12)$$

where the energy difference is set as $\triangle E = E^{\mathcal{H}_1} - E^{\mathcal{H}_0}$ with the decision threshold $T_{thr} = 0$. In (12), the true hypothesis is identified with the large normalized energy. It is because, under the false hypothesis, the test data with wrong predicted label is treated as a noise to disturb the selected features of training data. In this situation, the subspace cannot be well learned from the $l_{2,1}$-norm LDA such that more energy of selected features leaks out of the subspace.

III. RESULTS AND DISCUSSION

We give a set of performance evaluations on the ADHD databases of Table 1. The classification accuracy is obtained with the Leave-One-Out Cross Validation (LOOCV), where in each test iteration one subject is chosen from the given database as test data and the rest subjects are all used as training data. The parameter setting of our method, i.e., selected feature number N and subspace dimension K, is given in Table 2. Several state-of-the-art methods are employed for the comparison, including machine learning methods as R-Relief [11], L$_1$BioSVM [27] and Fusion fMRI [22], deep learning methods as FCNet [29], 3D-CNN [31] and Deep fMRI [30], and two binary hypothesis methods with subspace learning and dual subspace learning respectively named as SP-BH [18] and Dual-SP-BH [19].

TABLE 2. Parameter setting of ADHD databases.

	NYU	PU	PU_1	KKI	NI
N	110	110	110	110	110
K	70	75	63	60	24

A. ADHD CLASSIFICATION COMPARISON

Various measures of ADHD classification, i.e., specificity, sensitivity and accuracy, are given in Figure 2. It shows our method achieves the significant classification performance with the high average accuracy of 97.6%. We also find the unbalance problem on PU_1 and KKI is well solved. Existing reports show the sensitivity of these two databases usually keeps in a low level [11], [27], since the ADHD subjects are seriously less than the control ones as shown in Table 1. In this case, unreliable features are extracted from their limited ADHD subjects with less statistical information.
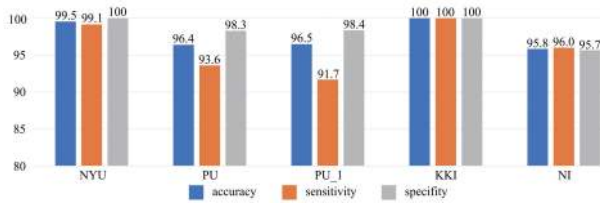
FIGURE 2. Classification comparison on ADHD databases (%).

TABLE 3. Accuracy comparison with various state-of-the-art methods (%).

	NYU	PU	PU_1	KKI	NI
Machine learning					
Fusion fMRI (2017)	52.7	–	85.8	86.7	72.9
L_1 BioSVM (2018)	–	81.1	86.7	81.3	–
R-Relief (2019)	70.7	68.6	–	81.8	76.0
Deep learning					
FCNet (2017)	58.5	–	62.7	–	60.0
3D-CNN (2017)	70.5	63.0	–	72.8	–
Deep fMRI (2018)	73.1	–	62.7	–	67.9
Binary hypothesis					
SP-BH (2019)	96.2	95.8	91.7	86.7	91.6
Dual-SP-BH (2020)	92.4	92.3	89.4	85.5	81.2
Our method	**99.5**	**96.3**	**96.4**	**100**	**95.8**

These features are impacted by more uncertain factors and tend to be outliers. But our method gives a remarkable solution to the unbalance problem. It is thanks to the $l_{2,1}$-norm LDA model for the robust feature learning. As a result, the sensitivity on these databases is impressively improved in Figure 2, where the 100% sensitivity is even obtained on KKI.

We further compare the accuracy of our method with those of state-of-the-art ones in Table 3. The first three methods, i.e., Fusion fMRI, R-Relief and L_1 BioSVM, are all ML-based methods. Though different approaches are incorporated to improve the accuracy, they suffer the general problem of ML-based methods, that is, the features of training data are difficult to cover the features of test data. Therefore, the accuracy on NI is much lower than those on PU_1 and KKI, due to the small size of NI. We also note that the accuracies on NYU and PU are unsatisfied. The selected features are now diffused, since more features are disturbed with noise and then turn into the outliers. Without considering the robust feature learning, it makes the classification ineffective. As for the DL-based methods, 3D-CNN and Deep fMRI achieve a better performance on NYU. They benefit from the multilayered and nonlinear learning strategy. Unfortunately, their accuracies on PU, PU_1, KKI and NI are rather bad, and even worse than those of the ML-based methods. An over-fitting problem appears on these databases, which is also caused by the insufficient data. Both binary hypothesis methods achieve remarkable performances, where

the binary hypothesis framework is adopted to overcome the lack of data. However, our method further develops the binary hypothesis methods. By taking advantage of the robust feature learning with $l_{2,1}$-norm LDA, its accuracy is the best among all listed methods. Furthermore, it shows our method well fits for various databases with less effect on the database size.

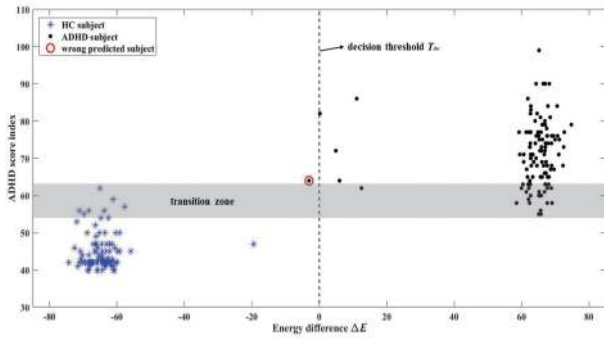### B. DECISION RESULTS COMBINED WITH ADHD SYMPTOM SCORE

To evaluate the validity of our ADHD decision strategy, in Figure 3, we plot the corresponding energy difference $\triangle E$ for each subject that is used as test data in its test turn of LOOCV. The ADHD score indexes are further given for these data as symptom score. In Figure 3, the decision threshold T_{thr} well divides the ADHD and healthy control subjects on NYU and PU, where subjects stay in their right positions and are hard to cross the decision threshold. It proves that the binary hypothesis approach is useful for classification. More specifically, for training data, the stronger projected components of selected FCs are obtained with the right predicted label of subject than those components with wrong predicted label. As for the wrong predicted subject, we find most of them are located in or nearby the transition zone between ADHD and non-ADHD. Now, they are also difficult to recognize by the symptomatological diagnosis. Moreover, we observe there are a number of subjects in the transition zone from the aspect of ADHD score index. It means these subjects may be misdiagnosed. However, our method only has a few of wrong predicted subjects, which shows the diagnosis advantage.

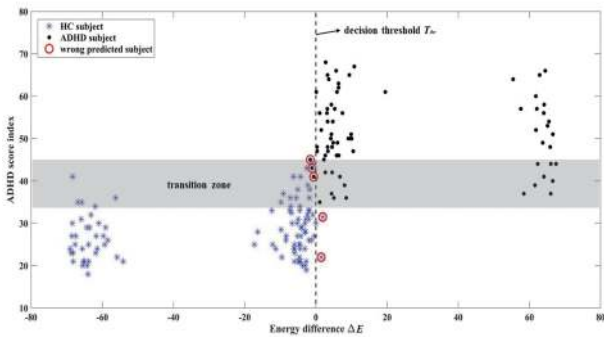### C. ANALYSIS ON DISCRIMINATIVE FCs

Some post-processing is performed on the classification results to discover the discriminative FCs between ADHD and control groups. We calculate the appearance probability for each FC to measure its importance. In detail, for each subject used as test data in its test turn of LOOCV, we record the selected FCs corresponding to its predicted label. Then the appearance possibility of FC is obtained with

$$F(i) = \frac{\sum_j F_p(i,j)}{\sum_j N_s(j)} \times 100\%, \tag{13}$$

where $F(i)$ is the appearance probability of i-th FC, $F_p(i,j)$ is the total appearance number of i-th FC in the j-th database with the right label prediction, and $N_s(j)$ is the subject number of j-th database. Therefore, the appearance probability $F(i)$ provides the statistical information on all used databases of Table 1. Several discriminative FCs with the highest appearance probabilities are consequently achieved. The result of FC probability is shown in Figure 4, where 29 highest-probability FCs are chosen with the appearance probability threshold $T_F = 55\%$. Here, the 90 ROIs are categorized into six lobes, i.e., medial temporal, subcortical, occipital, frontal, temporal, and parietal (pre)motor lobes [38]. In Figure 4, it shows a large number of discriminative FCs exist in parietal

(a) NYU

(b) PU

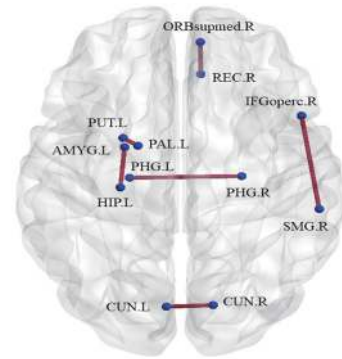**FIGURE 3.** Evaluation on ADHD decision with ADHD score index.

FIGURE 4. Discriminative FCs with highest appearance probabilities.

(pre)motor lobe which is directly associated with movement intention and motor awareness as mentioned in [39].

We further list several high-probability discriminative FCs in Table 4 with their P-values obtained by two-sample t-test. The locations of these FCs are shown in Figure 5. Now we focus on verifying whether these discriminative FCs are in

FIGURE 5. Several discriminative FCs with their locations.

TABLE 4. Discriminative FCs combined with their P-values.

FC	$F(i)$ (%)	P-value
left hippocampus — left amygdala	78.54	<0.001
left cuneus — right cuneus	77.66	0.004
left putamen — left globus pallidus	77.66	<0.001
right inferior frontal gyrus, pars opercularis — right supramarginal gyrus	59.18	0.022
left parahippocampal gyrus — right parahippocampal gyrus	58.52	0.034
right medial orbitofrontal cortex — right gyrus rectus	58.52	0.026

line with the existing discovery of ADHD disease, which provides another way to evaluate the validity of our method. However, it is difficult to directly check these FCs, due to FC is a mathematical concept derived from the correlation operation between brain regions. Thus we try to find more physical evidences to support these discriminative FCs. In detail, the discriminative FC is verified when the corresponding brain regions are abnormal under some physical measures, e.g., cortical thickness and gray matter probability. In other words, the discriminative FCs here can be viewed as abnormal FCs to identify ADHD subjects. Because ADHD is characterized by the executive function deficits of movement and emotion controls, we discuss the discriminative FCs of Table 4 mainly from three aspects, i.e., basic movement control, senior movement control and emotion control.

1) For the basic movement control, it reports inferior frontal gyrus, pars opercularis and supramarginal gyrus greatly affect on body movement, especially hand movement [40]. The regional gray matter abnormalities are appeared in pars opercularis and supramarginal gyrus with their reduced cortical thickness [41], [42], which leads the abnormal FC between right inferior frontal gyrus, pars opercularis and right supramarginal gyrus.

2) For the senior movement control, cuneus and hippocampal gyrus play an important role. Cuneus receives visual information and is known for its
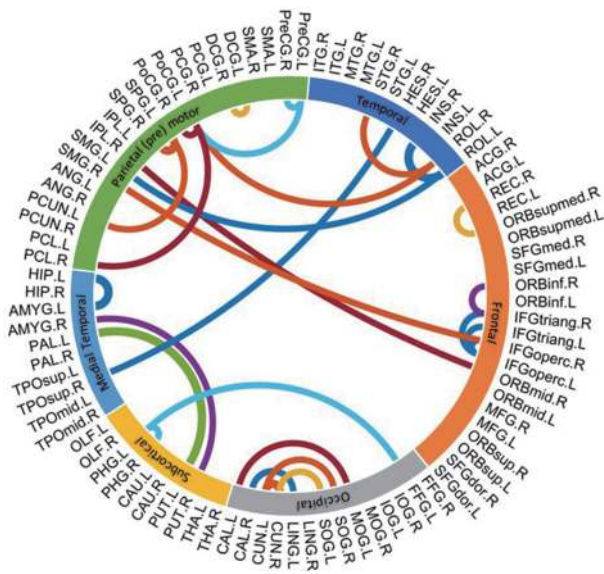
involvement in basic visual processing. Besides, some mid-level visual processing, just as attention, working memory and reward expectation, occurs in cuneus. It proves there are significant changes of cortical thickness on bilateral cuneus, which causes the high negative correlation of inattentive score [43], [44]. Besides, more abnormal biosignals, i.e., ReHo and ALFF, are detected on cuneus [45], [46]. On the other hand, hippocampal gyrus is in charge of memory encoding and retrieval with the recognition of environmental scenes. The hippocampal gyrus of ADHD children is often with smaller volumes [47] and hard to be activated during the go/no-go task [48]. The above evidences well support the abnormal FCs exist within bilateral cuneus and hippocampal gyrus. As for the FC between putamen and globus pallidus, putamen and globus pallidus are the key parts of striatum which coordinates fine and complex movements. ADHD is also known as a dysfunction of the striatum [49]. Moreover, the putamen and globus pallidus of boys with ADHD suffer from volumetric reduction and inward deformation [50], [51].

3) For the emotion control, the amygdala and hippocampus are respectively related to emotion and memory. Many researches disclose an altered effective connectivity between amygdala and hippocampus in ADHD children, where the negative emotion is often enlarged among these children [47], [52]. It also reports the gray matter volumes of amygdala and hippocampus are shrunk [53], [54]. It shows the FC between amygdala and hippocampus is changed for ADHD subjects.

Besides, the FC between right medial orbitofrontal cortex and right gyrus rectus is found in Table 4. The altered gray matter analysis of ADHD on gyrus rectus is given in [55], [56]. Interestingly, there has a significant effect of gender on rectus, that is, rectus can be more easily activated by male than female [57]. It well explains the phenomenon that the ADHD risk is 2.3 times higher for boys than that for girls [58].

IV. CONCLUSION

We propose a high-accuracy ADHD classification method. In this method, we use a binary hypothesis testing framework to effectively solve the insufficient data problem of ADHD databases. Meanwhile, an $l_{2,1}$-norm LDA model is employed for the robust feature learning to alleviate noise disturbance. The experiments show our method significantly outperforms the existing classification methods. The average accuracy achieves 97.6% with the LOOCV. Moreover, the corresponding analysis with ADHD symptom score and the explanation of discriminative FCs are also given, which well verifies the validity of our classification method.

REFERENCES

[1] G. Polanczyk and P. Jensen, "Epidemiologic considerations in attention deficit hyperactivity disorder: A review and update," *Child Adolescent Psychiatric Clinics North Amer.*, vol. 17, no. 2, pp. 245–260, Apr. 2008.

[2] S. A. Safren, M. W. Otto, S. Sprich, C. L. Winett, T. E. Wilens, and J. Biederman, "Cognitive-behavioral therapy for ADHD in medication-treated adults with continued symptoms," *Behaviour Res. Therapy*, vol. 43, no. 7, pp. 831–842, Jul. 2005.

[3] J. DellaBadia Jr, W. L. Bell, J. W. Keyes, Jr., V. P. Mathews, and S. S. Glazier, "Assessment and cost comparison of sleep-deprived EEG, MRI and PET in the prediction of surgical treatment for epilepsy," *Seizure*, vol. 11, no. 5, pp. 303–309, Jul. 2002.

[4] J.-P. Lachaux, P. Fonlupt, P. Kahane, L. Minotti, D. Hoffmann, O. Bertrand, and M. Baciu, "Relationship between task-related gamma oscillations and BOLD signal: New insights from combined fMRI and intracranial EEG," *Hum. Brain Mapping*, vol. 28, no. 12, pp. 1368–1375, Dec. 2007.

[5] D. Dai, J. Wang, J. Hua, and H. He, "Classification of ADHD children through multimodal magnetic resonance imaging," *Frontiers Syst. Neurosci.*, vol. 6, p. 63, Sep. 2012.

[6] E. Hoekzema, S. Carmona, J. A. Ramos-Quiroga, V. R. Fernández, R. Bosch, J. C. Soliva, M. Rovira, A. Bulbena, A. Tobeña, M. Casas, and O. Vilarroya, "An independent components and functional connectivity analysis of resting state fMRI data points to neural network dysregulation in adult ADHD," *Hum. Brain Mapping*, vol. 35, no. 4, pp. 1261–1272, Apr. 2014.

[7] K. Konrad and S. B. Eickhoff, "Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder," *Hum. Brain Mapping*, vol. 31, no. 6, pp. 904–916, Jun. 2010.

[8] J. B. Colby, J. D. Rudie, J. A. Brown, P. K. Douglas, M. S. Cohen, and Z. Shehzad, "Insights into multimodal imaging classification of ADHD," *Frontiers Syst. Neurosci.*, vol. 6, no. 6, p. 59, 2012.

[9] Y. Zhao, H. Chen, and R. T. Ogden, "Wavelet-based weighted LASSO and screening approaches in functional linear regression," *J. Comput. Graph. Statist.*, vol. 24, no. 3, pp. 655–675, Jul. 2015.

[10] M. Nuñez-Garcia, S. Simpraga, M. A. Jurado, M. Garolera, R. Pueyo, and L. Igual, "FADR: Functional-anatomical discriminative regions for rest fMRI characterization," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2015, pp. 61–68.

[11] B. Miao, L. L. Zhang, J. L. Guan, Q. F. Meng, and Y. L. Zhang, "Classification of ADHD individuals and neurotypicals using reliable RELIEF: A resting-state study," *IEEE Access*, vol. 7, pp. 62163–62171, 2019.

[12] M. Wang, B. Jie, W. Bian, X. Ding, W. Zhou, Z. Wang, and M. Liu, "Graph-kernel based structured feature selection for brain disease classification using functional connectivity networks," *IEEE Access*, vol. 7, pp. 35001–35011, 2019.

[13] D. Yao, H. Sun, X. Guo, V. D. Calhoun, L. Sun, and J. Sui, "ADHD classification within and cross cohort using an ensembled feature selection framework," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1265–1269.

[14] S. Dey, A. R. Rao, and M. Shah, "Exploiting the brain's network structure in identifying ADHD subjects," *Frontiers Syst. Neurosci.*, vol. 6, p. 75, Nov. 2012.

[15] A. Tabas, E. Balaguer-Ballester, and L. Igual, "Spatial discriminant ICA for RS-fMRI characterisation," in *Proc. Int. Workshop Pattern Recognit. Neuroimag.*, Jun. 2014, pp. 1–4.

[16] F. M. Grisales-Franco, J. M. Medina-Salcedo, D. M. Ovalle-Martinez, J. D. Martinez-Vargas, D. G. Garcia-Murillo, and G. Castellanos-Dominguez, "EEG source imaging based on dynamic sparse coding as ADHD biomarker," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Comput.*, 2017, pp. 416–425.

[17] Y. Zhang, Y. Tang, Y. Chen, L. Zhou, and C. Wang, "ADHD classification by feature space separation with sparse representation," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.

[18] Y. Tang, C. Wang, Y. Chen, N. Sun, A. Jiang, and Z. Wang, "Identifying ADHD individuals from resting-state functional connectivity using subspace clustering and binary hypothesis testing," *J. Attention Disorders*, early access, Apr. 2, 2019, doi: 10.1177/1087054719837749.

[19] Y. Chen, Y. Tang, C. Wang, X. Liu, L. Zhao, and Z. Wang, "ADHD classification by dual subspace learning using resting-state functional connectivity," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101786.

[20] S. Dey, A. R. Rao, and M. Shah, "Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects," *Frontiers Neural Circuits*, vol. 8, p. 64, Jun. 2014.

[21] W. M. Ewing, S. M. Hays, R. Hatfield, W. E. Longo, and J. R. Millette, "Abnormal functional resting-state networks in ADHD: Graph theory and pattern recognition analysis of fMRI data," *Biomed. Res. Int.*, vol. 2014, no. 4, 2014, Art. no. 380531.

[22] A. Riaz, M. Asad, E. Alonso, and G. Slabaugh, "Fusion of fMRI and non-imaging data for ADHD classification," *Comput. Med. Imag. Graph.*, vol. 65, pp. 115–128, Apr. 2018.

[23] J. Du, L. Wang, B. Jie, and D. Zhang, "Network-based classification of ADHD patients using discriminative subnetwork selection and graph kernel PCA," *Comput. Med. Imag. Graph.*, vol. 52, pp. 82–88, Sep. 2016.

[24] J. R. Sato, M. Q. Hoexter, A. Fujita, and L. A. Rohde, "Evaluation of pattern recognition and feature extraction methods in ADHD prediction," *Frontiers Syst. Neurosci.*, vol. 6, p. 68, Sep. 2012.

[25] D. A. Fair *et al.*, "Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data," *Frontiers Syst. Neurosci.*, vol. 6, p. 80, Feb. 2013.

[26] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme learning machine-based classification of ADHD using brain structural MRI data," *PLoS ONE*, vol. 8, no. 11, 2013, Art. no. e79476.

[27] L. Shao, Y. Xu, and D. Fu, "Classification of ADHD with bi-objective optimization," *J. Biomed. Informat.*, vol. 84, pp. 164–170, Aug. 2018.

[28] V. Sachnev, S. Suresh, N. Sundararajan, B. S. Mahanand, M. W. Azeem, and S. Saraswathi, "Multi-region risk-sensitive cognitive ensembler for accurate detection of attention-Deficit/Hyperactivity disorder," *Cognit. Comput.*, vol. 11, no. 4, pp. 545–559, Aug. 2019.

[29] A. Riaz, M. Asad, S. M. M. R. Al-Arif, E. Alonso, D. Dima, P. Corr, and G. Slabaugh, "FCNet: A convolutional neural network for calculating functional connectivity from functional MRI," in *Proc. Int. Workshop Connectomics Neuroimag.*, 2017, pp. 70–78.

[30] A. Riaz, M. Asad, S. M. M. R. A. Arif, E. Alonso, D. Dima, P. Corr, and G. Slabaugh, "Deep fMRI: AN end-to-end deep network for classification of fMRI data," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1419–1422.

[31] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI," *IEEE Access*, vol. 5, pp. 23626–23636, 2017.

[32] Z. Mao, Y. Su, G. Xu, X. Wang, Y. Huang, W. Yue, L. Sun, and N. Xiong, "Spatio-temporal deep learning method for ADHD fMRI classification," *Inf. Sci.*, vol. 499, pp. 1–11, Oct. 2019.

[33] F. Zhang, J. Yang, J. Qian, and Y. Xu, "Nuclear norm-based 2-DPCA for extracting features from images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2247–2260, Oct. 2015.

[34] N. Kwak, "Principal component analysis based on L_1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.

[35] Y. Luo, Y. Yang, F. Shen, Z. Huang, P. Zhou, and H. T. Shen, "Robust discrete code modeling for supervised hashing," *Pattern Recognit.*, vol. 75, pp. 128–135, Mar. 2018.

[36] Y. Yang, Z. Ma, Y. Yang, F. Nie, and H. Tao Shen, "Multitask spectral clustering by exploring intertask correlation," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1083–1094, May 2015.

[37] S. Liao, Q. Gao, Z. Yang, F. Chen, F. Nie, and J. Han, "Discriminant analysis via joint euler transform and $\ell_{2,1}$-norm," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5668–5682, Nov. 2018.

[38] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, Sep. 2005.

[39] M. Desmurget and A. Sirigu, "A parietal-premotor network for movement intention and motor awareness," *Trends Cognit. Sci.*, vol. 13, no. 10, pp. 411–419, Oct. 2009.

[40] J. H. Hong and S. H. Jang, "Neural network related to hand movement: A combined study of diffusion tensor tractography and functional MRI," *J. Phys. Therapy Sci.*, vol. 23, no. 1, pp. 97–101, 2011.

[41] K. A. McLaughlin, M. A. Sheridan, W. Winter, N. A. Fox, C. H. Zeanah, and C. A. Nelson, "Widespread reductions in cortical thickness following severe early-life deprivation: A neurodevelopmental pathway to attention-Deficit/Hyperactivity disorder," *Biol. Psychiatry*, vol. 76, no. 8, pp. 629–638, Oct. 2014.

[42] M. J. Batty, E. B. Liddle, A. Pitiot, R. Toro, M. J. Groom, G. Scerif, M. Liotti, P. F. Liddle, T. Paus, and C. Hollis, "Cortical gray matter in attention-Deficit/Hyperactivity disorder: A structural magnetic resonance imaging study," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 49, no. 3, pp. 229–238, Mar. 2010.

[43] H. Sun, Y. Chen, Q. Huang, S. Lui, X. Huang, Y. Shi, X. Xu, J. A. Sweeney, and Q. Gong, "Psychoradiologic utility of MR imaging for diagnosis of attention deficit hyperactivity disorder: A radiomics analysis," *Radiology*, vol. 287, no. 2, pp. 620–630, May 2018.

[44] X. Wang, Y. Jiao, T. Tang, H. Wang, and Z. Lu, "Altered regional homogeneity patterns in adults with attention-deficit hyperactivity disorder," *Eur. J. Radiol.*, vol. 82, no. 9, pp. 1552–1557, Sep. 2013.

[45] C.-Y. Shang, H.-Y. Lin, and S. S.-F. Gau, "The norepinephrine transporter gene modulates intrinsic brain activity, visual memory, and visual attention in children with attention-deficit/hyperactivity disorder," *Mol. Psychiatry*, Oct. 8, 2019, doi: 10.1038/s41380-019-0545-7.

[46] B. de Celis Alonso, S. Hidalgo Tobón, P. Dies Suarez, J. García Flores, B. de Celis Carrillo, and E. Barragán Pérez, "A multi-methodological MR resting state network analysis to assess the changes in brain physiology of children with ADHD," *PLoS ONE*, vol. 9, no. 6, 2014, Art. no. e99119.

[47] J. Van Dessel, E. Sonuga-Barke, M. Moerkerke, S. Van der Oord, J. Lemiere, S. Morsink, and M. Danckaerts, "The amygdala in adolescents with attention-deficit/hyperactivity disorder: Structural and functional correlates of delay aversion," *World J. Biol. Psychiatry*, to be published.

[48] S. Spinelli, S. Joel, T. E. Nelson, R. A. Vasa, J. J. Pekar, and S. H. Mostofsky, "Different neural patterns are associated with trials preceding inhibitory errors in children with and without attention-Deficit/Hyperactivity disorder," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 50, no. 7, pp. 705–715, Jul. 2011.

[49] H. Lou, "Etiology and pathogenesis of attention-deficit hyperactivity disorder (ADHD): Significance of prematurity and perinatal hypoxic-haemodynamic encephalopathy," *Acta Paediatrica*, vol. 85, no. 11, pp. 1266–1271, Nov. 1996.

[50] X. Tang, K. E. Seymour, D. Crocetti, M. I. Miller, S. H. Mostofsky, and K. S. Rosch, "Response control correlates of anomalous basal ganglia morphology in boys, but not girls, with attention-deficit/hyperactivity disorder," *Behav. Brain Res.*, vol. 367, pp. 117–127, Jul. 2019.

[51] K. S. Rosch, D. Crocetti, K. Hirabayashi, M. B. Denckla, S. H. Mostofsky, and E. M. Mahone, "Reduced subcortical volumes among preschool-age girls and boys with ADHD," *Psychiatry Res., Neuroimag.*, vol. 271, pp. 67–74, Jun. 2018.

[52] T. Villemonteix, D. Purper-Ouakil, and L. Romo, "Is emotional dysregulation a component of attention-deficit/hyperactivity disorder (ADHD)?" *Encephale-Revue de Psychiatrie Clinique Biologique et Therapeutique*, vol. 41, no. 2, pp. 108–114, 2015.

[53] B. Bonath, J. Tegelbeckers, M. Wilke, H. H. Flechtner, and K. Krauel, "Regional gray matter volume differences between adolescents with ADHD and typically developing controls: Further evidence for anterior cingulate involvement," *J. Attention Disorders*, vol. 22, no. 7, pp. 627–638, 2018.

[54] Y. Wang, Q. Xu, S. Li, G. Li, C. Zuo, S. Liao, Y. Long, S. Li, and R. M. Joshi, "Gender differences in anomalous subcortical morphology for children with ADHD," *Neurosci. Lett.*, vol. 665, pp. 176–181, Feb. 2018.

[55] K. R. Griffiths, S. M. Grieve, M. R. Kohn, S. Clarke, L. M. Williams, and M. S. Korgaonkar, "Altered gray matter organization in children and adolescents with ADHD: A structural covariance connectome study," *Transl. Psychiatry*, vol. 6, no. 11, p. e947, Nov. 2016.

[56] M. Stevens and E. Haney-Caron, "Comparison of brain volume abnormalities between ADHD and conduct disorder in adolescence," *J. Psychiatry Neurosci.*, vol. 37, no. 6, pp. 389–398, 2012.

[57] M. E. M. Benwell, D. J. K. Balfour, and J. M. Anderson, "Evidence that tobacco smoking increases the density of (–)-[³H]nicotine binding sites in human brain," *J. Neurochem.*, vol. 50, no. 4, pp. 1243–1247, 1988.

[58] J. J. Bauermeister, P. E. Shrout, L. Chávez, M. Rubio-Stipec, R. Ramírez, L. Padilla, A. Anderson, P. García, and G. Canino, "ADHD and gender: Are risks and sequela of ADHD the same for boys and girls?" *J. Child Psychol. Psychiatry*, vol. 48, no. 8, pp. 831–839, Aug. 2007.
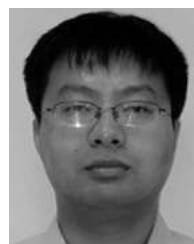
YIBIN TANG received the B.S. and M.S. degrees in information and communication engineering from the Nanjing University of Posts and Telecommunications, China, in 2004 and 2007, respectively, and the Ph.D. degree in information and signal processing from Southeast University, China, in 2010. He is currently an Associate Professor with the College of Internet of Things Engineering, Hohai University, Changzhou, China. His research interests include imaging and speech processing, intelligent signal processing, and machine learning.

XUFEI LI received the B.S. degree in information and communication engineering from Hohai University, China, in 2017, where she is currently pursuing the M.S. degree in information and communication engineering. Her research interests include biomedical signal processing and deep learning.

YING CHEN received the B.S. and M.S. degrees in optical engineering from the Nanjing University of Science and Technology, in 2010 and 2013, China, respectively. She is currently pursuing the Ph.D. degree in information and communication engineering with Southeast University, China. Her research interests include imaging processing and graph signal processing.

YUAN ZHONG received the Ph.D. degree in biomedical engineering from the Nanjing University of Aeronautics and Astronautics, China, in 2010. He is currently an Associate Professor with the School of Psychology, Nanjing Normal University, China. His research interests include the action mechanisms of mental disorders in brain.

AIMIN JIANG received the Ph.D. degree in electrical engineering from the University of Windsor, Canada, in 2010. He is currently a Professor with the College of Internet of Things Engineering, Hohai University, China. His research interests include mathematical optimization and its applications to digital signal processing, and graph networks.

CHUN WANG received the M.B. degree in clinical medicine from the School of Medicine, Southeast University, China, in 2002, the M.S. degree in applied psychology from Nanjing Normal University, China, in 2005, and the Ph.D. degree in psychiatry and mental health from the Xiangya School of Medicine, Central South University, China, in 2009. She is currently the Deputy Director of the Department of Mood Disorders, Nanjing Brain Hospital, Nanjing Medical University, China. Her research interest includes the understanding of mind-body mechanisms with mental disorders and the corresponding therapeutic strategies.

● ● ●