IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 20, 2019, accepted May 6, 2019, date of publication May 9, 2019, date of current version May 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915988

Classification of ADHD Individuals and Neurotypicals Using Reliable RELIEF: A Resting-State Study

B. MIAO[1], L. L. ZHANG[1], J. L. GUAN[2], Q. F. MENG[1], AND Y. L. ZHANG[1]
[1]University of Jinan, Jinan, Shandong 250000, China
[2]University of New South Wales, Sydney, NSW 2000, Australia

Corresponding author: Y. L. Zhang (ise_zhangyl@ujn.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61671220.

ABSTRACT Fractional amplitude of low-frequency fluctuation (fALFF) can reflect the intensity of spontaneous neuronal activity. Feature selection on functional magnetic resonance imaging (fMRI) data combined with fALFF can be well used to study the pathology of attention deficit hyperactivity disorder (ADHD) and assist in its diagnosis. However, the unsatisfactory effect of feature selection limits the study of ADHD. In this study, a novel method is proposed to classify ADHD individuals and neurotypicals. This work introduces multiple linear regressions for reduction of confounding effects. After that, fALFF combined with principal component analysis (PCA), Shannon entropy (ShEn), and sample entropy (SampEn) are used to construct features, and a reliable RELIEF (R-RELIEF) algorithm is proposed to find the most prominent features. Public ADHD-200 dataset is used in this study to evaluate our method. The results in this study suggest that fALFF is a reliable fMRI marker for investigation of ADHD, and R-RELIEF shows good performance compared with RELIEF algorithm. Moreover, fALFF of many brain regions are found to differ between ADHD and neurotypicals, thus coinciding with the results of previous studies.

INDEX TERMS Medical diagnosis, image classification, feature extraction, magnetic resonance imaging.

I. INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a common brain-related disease, and it is characterized by lack of attention, hyperactivity, and difficulty in controlling one's behavior [1]. According to a meta-analysis, global morbidity rate of ADHD in children and adolescents totals 3.4%, and many symptoms still remain after children with ADHD reach adulthood [2]. ADHD not only causes difficulty for patients to organize their daily life but also poses a considerable burden on their families. The study of ADHD can help doctors in diagnosing and curing ADHD. Therefore, finding the pathology of and correctly diagnosing ADHD bears importance in improving the lives of affected individuals.

Functional magnetic resonance imaging (fMRI) combined with fractional amplitude of low-frequency fluctuation (fALFF) are widely used in the research of ADHD [3]. fMRI is a reliable technique based on rapid imaging technique with high field intensity and blood-oxygenation-level-dependent (BOLD) principle. BOLD signal change is mainly based on the rise and fall of concentrations of localized deoxyhemoglobin in blood vessels; it closely tracks the changes in cerebral blood flow that related to neural activities. fMRI indirectly measures changes in neural activities of each voxel according to BOLD signal; thus, it can quantitatively evaluate brain activity. Given its safety, reliability, and dynamics, research hotspots center on the methods based on fMRI. Spontaneous neuronal activities constantly occur in the brain, and BOLD signals can detect spontaneous fluctuations at all times. As a biomarker of fMRI, fALFF can effectively reflect spontaneous neuronal activities of the brain in resting-state, and it has been proven to be a reliable tool for neuroimaging studies of the brain [4].

On the research of fMRI data, confounding effects of age and sex are considerable impediments for accurate analysis and classification. Previous studies have shown that women have higher cerebral blood flow than men, and effects of brain disorders on men and women are different [5]. Age is also

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

an important factor that found to be inversely proportional to cerebral blood flow [6]. Thus, differences of age and sex between two groups will impact their fALFF. Moreover, the ratio of boys to ADHD is much higher than that of girls. Therefore, it is meaningful to regress out the influence of age and sex as far as possible.

Based on fMRI technology, some researchers have variously investigated the brain and discovered abnormities in its structure and function between ADHD individuals and neurotypicals. Reduced functional connectivity between anterior cingulate cortex (ACC) and posterior cingulate cortex was detected in ADHD [7]. A decrease in the gray matter volume of lenticular nucleus was observed in ADHD using a voxel-based morphological analysis [8]. A meta-analysis was also conducted to investigate the differences between ADHD and neurotypical subjects; subjects with ADHD showed consistent inferior frontal, cerebellar, and parietal dysfunctions, and volume differences were mostly noted in the accumbens nucleus, basal ganglia, and hippocampus [9], [10]. Given the low signal-to-noise ratio of fMRI data, traditional analysis methods present insufficient abilities to detect abnormalities in the human brain. Some researchers have discovered changes in brain activation patterns in ADHD based on fMRI technology, but their conclusions reveal inconsistencies [11], [12]. To date, machine learning and deep learning have been applied in numerous fields [13], [14]. The use of fMRI data analysis methods based on classification and prediction has drawn a growing attention [15], [16]. Sato *et al.* achieved the detection of ADHD using L2-regularized logistic regression and functional fMRI features [17]. Liang *et al.* found significant differences in brain regions between ADHD individuals and neurotypicals, and achieved the recognition of ADHD using the functional connectivity features extracted from these brain regions and Linear Discriminant Analysis [18]. Sarraf and Tofighi obtained a high accuracy in the recognition of Alzheimers disease using the convolutional Neural Network with LeNet-5 [19]. Suk *et al.* proposed a novel methodological architecture that combines deep learning and state-space modeling in order to recognize Mild Cognitive Impairment [20]. Firstly, they use a deep auto-encoder to hierarchical functional relations among brain regions. Then, a hidden Markov model is being used to diagnose Mild Cognitive Impairment. By using machine learning and deep learning methods, the multivariate nature of fMRI data can be fully excavated. Multi-voxel pattern analysis (MVPA) in machine learning and deep learning has been used to analyze fMRI data, enabling the decoding of cognitive state from neural signals [21]. MVPA is also used to determine the pathology of ADHD and assist in its diagnosis.

In summary, the study of ADHD can assist in diagnosis and treatment of such condition. MVPA based on fMRI data is a promising approach in the study of ADHD. Therefore, confounding effects of age and sex in fMRI data should be reduced, and a more effective feature selection method should be developed. However, a dearth of studies currently investigate effective feature selection methods in MVPA. This paper attempts to fill this gap. Multiple linear regression is a good method for controlling covariates such as age and sex and it is used in statistical analysis of neuroimaging, thus it can be used as a data processing step in fMRI image classification. RELIEF is an effective filter feature selection algorithm and it is used in classification of remote sensing images [22], [23], thus it may be able to be used in classification of fMRI data. However, the performance of RELIEF algorithm remains unstable given the problem of small sample size, this issue can be ameliorated by improving the traditional RELIEF algorithm to promote the effects of classification.

This study primarily aims to improve accuracy of ADHD diagnosis and to identify the brain regions that differ the most between ADHD and neurotypical individuals. To achieve these goals, this research proposes a novel method for classification of ADHD individuals and neurotypicals. In this method, fALFF in each voxel is calculated. Confounding effects of age and sex are then regressed out by using a multiple linear regression model. After that, principal component analysis (PCA), shannon entropy (ShEn), and sample entropy (SampEn) are introduced to construct features, and a reliable RELIEF (R-RELIEF) is proposed to find the most prominent features. Moreover, the performance of R-RELIEF is evaluated by comparing the classification results with RELIEF algorithm and minimum Redundancy Maximum Relevance (mRMR) algorithm [24]. Finally, classification is achieved using several classifiers, and classification results are compared with other state-of-art methods. Furthermore, brain regions that differ the most between ADHD and neurotypical individuals are found based on fALFF in each voxel and R-RELIEF algorithm. The major contributions and novelties of this work are as follows.

(1) A novel R-RELIEF algorithm is proposed for feature selection of ADHD fMRI data.

(2) A novel framework that applying the fALFF related features combined with R-RELIEF algorithm and classifiers is used for the classification of ADHD individuals and neurotypicals, and it achieved a good performance.

(3) R-RELIEF combined fALFF are used to find the brain regions that differ the most between ADHD and neurotypical individuals.

The remainder of this paper is organized as follows: Section 2 describes the details of the data used in this work and the proposed method. The analysis of experimental results are presented in section 3. Finally, the conclusion is presented in Section 4.

II. MATERIALS AND METHODS

In this study, we propose a framework for the classification of ADHD individuals and neurotypicals. Fig. 1 illustrates the framework of this study. Firstly, all fMRI data are preprocessed and fALFF is calculated. After that, confounding effects are regressed out. PCA, ShEn and SampEn are then used to construct features, and these combined features are

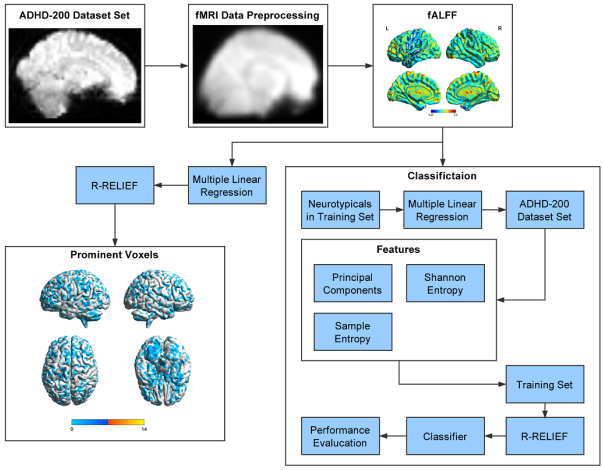

FIGURE 1. Framework for the classification of ADHD individuals and neurotypicals. In addition to classification, R-RELIEF is also used to find brain regions that differ the most between ADHD and neurotypical individuals at the voxel level.

TABLE 1. Participant characteristics.

	Training Set		Test Set	
	ADHD	Neurotypical	ADHD	Neurotypical
KKI	22	61	3	8
NI	23	25	11	14
NYU	118	98	29	12
Peking	78	116	24	27

being selected using R-RELIEF algorithm. Finally, classification is achieved using several classifiers. Furthermore, R-RELIEF is used to find brain regions that differ the most between ADHD and neurotypical individuals at the voxel level.

A. DATA SET

Neuroimaging data were obtained from the ADHD-200 data set, which is publicly available at http://fcon_1000.projects.nitrc.org/indi/adhd200. The data sets used in this work are obtained from Kennedy Krieger Institute (KKI), NeuroImage (NI), New York University Medical Center (NYU) and Peking University (Peking). Each data set uses different scan parameters, and contains different numbers of samples, which makes the data sets complex and diverse. Table 1 summarizes specific information of individuals in each data set. The subjects were prohibited from taking psychostimulant drugs within 24 h of data collection.

B. PREPROCESSING OF FMRI DATA

All data were preprocessed using SPM12 and DPARSF [25]. The measures of each voxel are 3.0 * 3.0 * 4.0 mm^3. The data were slice-timing corrected and realigned. The corrected data were then spatially normalized into a standard space (MNI305), and resampled to 3 mm isotropic voxels. An isotropic Gaussian kernel of 8 mm full width at half maximum was used to spatially smooth the normalized

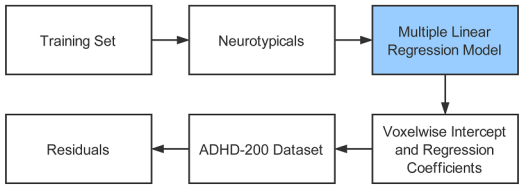

FIGURE 2. Illustration of age and sex adjustment. Age and sex information of neurotypical children in training set were used to construct a multiple linear regression model.

data, and linear trends were removed [26]. Finally, six head motion parameters, averaged signals from white matter, cerebrospinal fluid, and global signals were regressed out using voxelwise nuisance regression models.

C. FRACTIONAL AMPLITUDE OF LOW-FREQUENCY FLUCTUATION

fALFF was calculated by using DPARSF. Time series of each voxel was extracted and transformed into frequency domain by fast Fourier transform, and power spectrum was obtained. fALFF is calculated as the sum of the power spectrum's square root at each frequency across the interested frequency band (IFB) (0.01-0.08 Hz) divided across by the whole frequency band. Then the data were standardized by converting them into Z-score as follows:

$$zf_n = \frac{f_n - \bar{f}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - \bar{f})^2}} \tag{1}$$

where $\bar{f}$ represents the mean fALFF value across the brain, and N represents the number of voxels.

D. MULTIPLE LINEAR REGRESSION AND BINARIZATION

In this study, a standard multiple linear regression with age and sex as independent variables was used to reduce the effect of independent variables on fALFF. Least squares method was used to fit the linear regression equations of neurotypical group. Voxelwise intercept (β_0), regression coefficients of age and sex (β_1 and β_2, respectively), and the residuals (δ) were used for confounding factor correction. Fig. 2 illustrates the process flow; age data and sex data of neurotypical individuals in the training set were used to calculate voxelwise effects using the multiple linear regression model. Then, all the data (D) were used to subtract their voxelwise effects.

E. PRINCIPLE COMPONENTS AND ENTROPY-BASED FEATURES

In this work, principle components and entropy-based features were used for the classification of ADHD individuals and neurotypicals.

PCA is a feature extraction method that has strict mathematical foundation and has been widely adopted. PCA converts the observed values of potentially relevant variables into a set of linearly independent variables through orthogonal

transformation. Considering that each sample in fALFF data has tens of thousands of voxels, thus it is necessary to use PCA to extract the principal components as features. In this work, PCA is used to extract principal components from fALFF data, and the standard of dimension reduction is to save 99% of the original information. ShEn and SampEn can also be used to measure the uncertainty of data, and they have achieved good discriminating effect in the study of fMRI. To compute SampEn, embedding dimension m is set to 2, and similar tolerance r is set to 0.3 based on a previous study [27].

F. RELIEF FEATURE SELECTION ALGORITHM

RELIEF algorithm is efficient for binary classification [22]. RELIEF algorithm assigns each feature with a weight according to the Euclidean distance between chosen samples and their two nearest samples. Thus, all features can be sorted by their weight. Weight of a feature is directly proportional to its contribution to classification. Calculation of Euclidean distance between two samples can be obtained as follows:

$$d_{x,y} = \sqrt{\sum_{j=1}^{N}(x_j - y_j)^2} \tag{2}$$

where x and y represent two samples, and N represents the number of features.

In RELIEF algorithm, iterations m need to be initially set. At the beginning of each iteration, a sample X is being randomly selected as the starting point. Then, the nearest same-class sample and different-class sample are chosen based on Euclidean distance. Feature weights are updated as follows:

$$W_i = W_i + \frac{(X_i - H_i)^2 + (X_i - M_i)^2}{m} \tag{3}$$

where H and M are nearest same-class sample and different-class sample, respectively. W_i is accumulated m times as the final weight of ith feature.

G. R-RELIEF FEATURE SELECTION ALGORITHM

Due to the fact that feature weights are being updated only based on the difference between chosen samples and their two nearest neighbors in RELIEF algorithm, and randomness of chosen samples will result in poor reliability of selected features. Therefore, the performance of RELIEF algorithm are not always reliable. To overcome these shortcomings, we propose an R-RELIEF algorithm for the feature selection of fMRI data, which measures the overall difference by considering the difference between chosen sample's $2r$ nearest neighbors (r plus one is the minimum number of samples in two groups). In this method, each sample is treated as the starting point, and mean absolute error (MAE) is used to measure the difference between every sample pair in the $2r$ neighbors.

In R-RELIEF algorithm, the number of training samples is n. In the first step, R-RELIEF traverses each sample as the starting point. Then, the algorithm searches r nearest same-class samples and r nearest different-class samples from the starting point and forms a subset with $2r$ samples, with r plus one equal to the minimum number of samples in two groups. In the second step, the weight of each feature is being updated by computing MAE between samples in the subset. Weights are updated as follows (4), as shown at the bottom of the next page:

$$b_i = \frac{\sum_{j=1}^{r}\sum_{g=r+1}^{2r}|H_{j,i} - M_{g,i}|}{r^2} \tag{5}$$

$$W_i = W_i + \frac{b_i - a_i}{n} \tag{6}$$

where H and M represent the nearest same-class samples and different-class samples, respectively; a is the MAE between same-class samples, and b is the MAE between different-class samples. W_i is accumulated n times as the final weight of ith feature. At the beginning of R-RELIEF algorithm, data are standardized by converting them into Z-scores. Thus, we considered that all features are at the same level. Finally, T top ranked features composed the selected feature set. R-RELIEF is summarized in Algorithm I.

Input: The sample set $N = \{N_k, k = 1 : n\}$, the feature number m, the sample number n, and the threshold T.
Output: The selected feature set.
r plus one equal to the smallest sample size of two classes;
for $k = 1{:}n$ **do**
  Select N_k as starting point;
  Select r nearest same-class samples $\{H_1,..., H_r\}$;
  Select r nearest different-class samples $\{M_1,..., M_r\}$;
  $L = \{H_1,..., H_r, M_1,..., M_r\}$;
  for $i = 1{:}m$ **do**
    Calculate the MAE between same-class samples in L:
    $a_i = \frac{\sum_{j=1}^{r-1}\sum_{g=j+1}^{r}|H_{j,i}-H_{g,i}|+\sum_{p=r+1}^{2r-1}\sum_{q=p+1}^{2r}|M_{p,i}-M_{q,i}|}{r(r-1)}$;
    Calculate the MAE between different-class samples in L:
    $b_i = \frac{\sum_{j=1}^{r}\sum_{g=r+1}^{2r}|H_{j,i}-M_{g,i}|}{r^2}$;
    Update the weight of ith feature:
    $W_i = W_i + \frac{b_i-a_i}{n}$;
  end
end
Return T top ranked features;

Algorithm 1 Algorithm of R-RELIEF

H. PERFORMANCE EVALUATION METHODS

Many indicators can evaluate the performance of classifiers. In this experiment, we used accuracy, sensibility and specificity to evaluate classifier performance. TP is true positive, FP is false positive, TN is true negative, and

TABLE 2. Classification performance of our method.

	Data	Logistic Regression (%)			Linear-SVM (%)			RBF-SVM (%)			Random Forest (%)			Decision Tree (%)		
		Acc	Sen	Spec	Acc	Sen	Spec	Acc	Sen	Spec	Acc	Sen	Spec	Acc	Sen	Spec
mRMR	Peking	56.86	16.67	92.60	62.74	41.67	81.48	58.82	29.17	85.19	58.82	16.67	96.30	56.87	45.83	66.67
RELIEF	Peking	58.82	37.50	77.78	56.86	37.50	74.07	58.82	12.50	100.0	60.78	20.83	96.30	68.63	62.50	74.07
R-RELIEF	Peking	76.47	58.33	92.59	76.47	58.33	92.59	72.55	50.00	92.59	64.71	41.67	85.19	68.63	62.50	74.07
mRMR	NI	56.00	54.55	57.14	60.00	45.45	71.43	64.00	36.36	85.71	60.00	54.54	64.29	56.00	45.45	64.29
RELIEF	NI	68.00	63.64	71.43	72.00	63.64	78.57	72.00	45.45	92.86	56.00	72.73	42.86	68.00	81.82	57.14
R-RELIEF	NI	76.00	54.55	92.86	76.00	54.55	92.86	76.00	72.73	78.57	64.00	63.64	64.29	76.00	72.73	78.57
mRMR	NYU	60.98	62.07	58.33	65.85	62.07	75.00	58.54	65.52	41.67	70.73	86.21	33.33	58.54	68.97	33.33
RELIEF	NYU	68.29	65.52	75.00	70.73	96.55	8.33	68.29	79.31	41.67	68.29	79.31	41.67	68.29	65.52	75.00
R-RELIEF	NYU	68.29	68.97	66.67	68.29	65.52	75.00	68.29	89.66	16.67	68.29	79.31	41.67	70.73	65.52	83.33
mRMR	KKI	63.63	0.00	87.50	72.73	0.00	100.0	72.73	0.00	100.0	72.73	33.34	87.50	63.63	0.00	87.50
RELIEF	KKI	72.73	0.00	100.0	72.73	0.00	100.0	72.73	0.00	100.0	81.82	33.34	100.0	72.73	0.00	100.0
R-RELIEF	KKI	72.73	0.00	100.0	81.82	33.34	100.0	72.73	0.00	100.0	72.73	0.00	100.0	81.82	66.67	87.50

FN is false negative. These criteria can be achieved as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$SEN = \frac{TP}{TP + FN} \tag{8}$$

$$SPE = \frac{TN}{TN + FP} \tag{9}$$

III. RESULTS AND DISCUSSION

A. BINARY CLASSIFICATION RESULTS

Using appropriate features is one of the most important factors in determining the performance of classification. In this paper, fALFF in each voxel is calculated. Prominent components and entropy-based features are then being used to construct features. Finally, diagnosis of ADHD is achieved using several classifiers. In this study, default parameters are used. Moreover, the performance of R-RELIEF is compared with RELIEF algorithm and mRMR algorithm. Unlike R-RELIEF algorithm and RELIEF algorithm, which can select multiple features simultaneously through parallel computation, mRMR algorithm uses incremental search to select features and rates them according to mutual information. Table 2 shows the classification performance using the three feature selection algorithms and different classifiers. In order to further analysis the performance, a statistic analysis is used to compared the average performance of each feature selection algorithm. Fig. 3 shows the average performance of the three algorithms using different classifiers. Results of our study obviously show that our proposed R-RELIEF feature selection algorithm has a better performance than other two feature selection algorithms, and classification on KKI data set obtains the best performance. Moreover, decision tree has the better classification performance than logistic regression, linear-SVM, RBF-SVM and random forest. Results of this

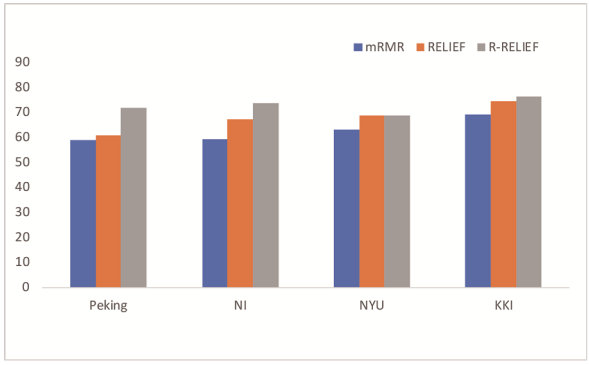

FIGURE 3. Average performance of the three algorithms using different classifiers.

study indicate that R-RELIEF performs better than RELIEF and mRMR. To further evaluate the classification performance obtained by our method, our results are compared with other previous methods. Table 3 shows a summary of studies that other current methods to the classification of ADHD and neurotypical individuals using the same data used in this work. Kuang *et al.* achieved ADHD diagnosis using fast Fourier transform algorithm and deep belief network [28]. In their study, the Gaussian-Bernoulli restricted Boltzmann machines are used to represent each layer of deep belief network. Zou *et al.* achieved ADHD diagnosis using a 3-D convolutional neural network since 3D convolutional layer is effective in learning local patterns and exploring spatial information [29]. The structure of their neural network includes four convolutional layers and two fully connected layers. Riaz *et al.* achieved ADHD diagnosis using elastic net based feature selection and support vector machine [15]. It can be clearly seen from Table 3 that it is the most difficult to identify ADHD using NI data set and NYU data set. Moreover, results show that our method is comparable.

$$a_i = \frac{\sum_{j=1}^{r-1} \sum_{g=j+1}^{r} |H_{j,i} - H_{g,i}| + \sum_{p=r+1}^{2r-1} \sum_{q=p+1}^{2r} |M_{p,i} - M_{q,i}|}{r\,(r-1)} \tag{4}$$

TABLE 3. Comparison of our results with average results of competition teams and other state-of-art methods.

Site	ADHD-200 Competition	Kuang 2014	Zou 2017	Riaz 2018	Proposed method
KKI	43.1%	71.82%	72.82%	81.8%	81.82%
NI	56.9%	-	-	44.0%	76.0%
NYU	32.3%	37.41%	70.5%	60.9%	70.73%
Peking	51.0%	54.0%	62.95%	64.7%	68.63%

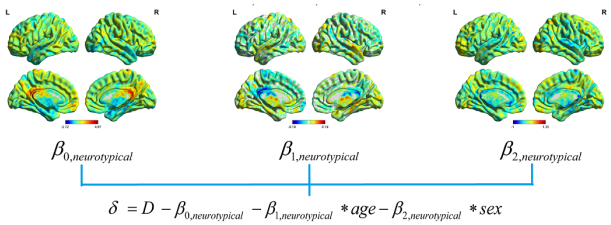

FIGURE 4. Voxelwise intercepts and regression coefficients of age and sex for neurotypicals.

B. REGRESSION PARAMETERS OF AGE AND SEX

In order to observe how age and sex affect the fALFF of brain. Voxelwise intercept and regression coefficients of age and sex for neurotypicals are obtained using multiple linear regression. Fig. 4 shows the regression parameters of each voxel for neurotypicals. In the neurotypicals, there is a significant decrease in fALFF with age in posterior cingulate. At the same time, the fALFF of women in posterior cingulate, caudate nucleus, and superior temporal gyrus (STG) are significantly higher than that in the men.

C. LOCATION OF INFORMATIVE VOXELS

In this study, R-RELIEF is also used for finding the brain regions that differ the most between ADHD and neurotypical individuals in addition to the classification of ADHD and neurotypical individuals. By using R-RELIEF algorithm, weights are calculated in voxel level, and voxels with positive weights are reserved. Fig. 5 shows the weight of each voxel calculated by R-RELIEF algorithm. Furthermore, reserved voxels are visualized in Fig. 6, these reserved voxels are distributed in many brain regions. The cluster with maximum weights and large volume is located in the cerebellum. Table 4 summarizes voxel weight distribution, coordinates in the table correspond to the MNI template. It can be observed that many regions differed between the ADHD group and the neurotypical group; these regions include the left inferior frontal gyrus (IFG), left middle frontal gyrus (MFG), right superior frontal gyrus (SFG), bilateral STG, right ACC, left putamen, right caudate nucleus, right insula, left precuneus, bilateral brainstem, and bilateral cerebellum.

Attention control and cognitive control are important components of executive function. According to impulsive behavior and attention deficit behavior, individuals with ADHD are assumed to exhibit executive function deficits. As the main brain region of executive function, prefrontal cortex (PFC) participates in voluntary movement and cognitive control.

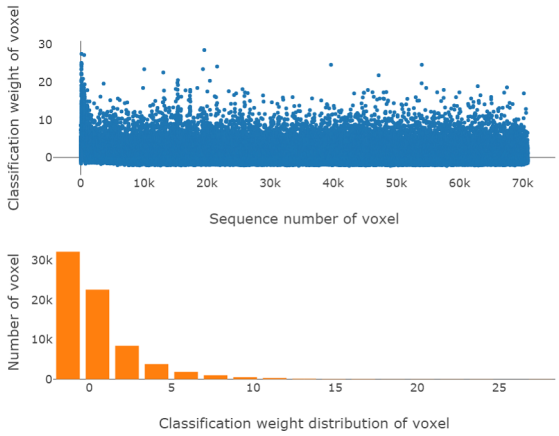

FIGURE 5. Weight of each voxel that being calculated by R-RELIEF algorithm.

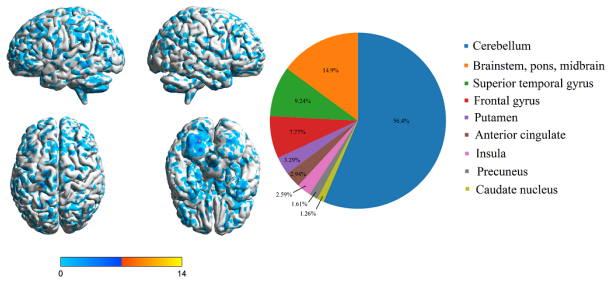

FIGURE 6. Weight distribution of voxels with positive weight. The non-white part of the brain's surface represents voxels with positive weight.

A study has shown that people with frontal lobe lesions are more willing to lie for their own benefit [30]. This result may be due to their cognitive control dysfunction; thus, they cannot restrain themselves and make decisions against their morality.

Intact executive function is not only related to PFC but also closely related to basal nucleus and cerebellum. We observed that fALFF of various brain regions differed between the ADHD group and the neurotypical group; these regions include left IFG, left MFG, right SFG, bilateral STG, right ACC, left putamen, right caudate nucleus, right insula, left precuneus, bilateral brainstem, and bilateral cerebellum. This pattern is consistent with a previous resting-state study [31], which showed that abnormalities between ADHD and neurotypicals involved the SFG, fronto-cerebellar, and fronto-striatal circuits. Another study based on structural MRI using machine learning methods also showed that children with ADHD present abnormalities in the prefrontal-striatal-cerebellar circuit [32].

In the fronto-cerebellar circuit, which is related to executive functions, PFC is connected to the cerebellum via pons. Brainstem is not only related to maintenance of heartbeat and breathing but also plays an important role in transmission of signals. Although considerable noise occurs in the skull base, some neuroimaging techniques are still used to study

TABLE 4. Voxel weight distribution in brain regions.

	Volume (mm^3)	Side	Peak Weight	Coordinates		
				x	y	z
Inferior frontal gyrus	486	L	8.45	-57	9	12
Middle frontal gyrus	1917	L	9.76	-39	48	-12
Superior frontal gyrus	594	R	13.24	15	57	-15
Superior temporal gyrus	1566	L	14.67	-39	15	-27
Superior temporal gyrus	1998	R	11.82	45	24	-24
Anterior cingulate	1134	R	11.37	9	21	24
Putamen	1269	L	8.78	-24	12	-6
Caudate nucleus	486	R	8.21	12	21	3
Insula	999	R	7.21	36	9	-12
Precuneus	621	L	6.80	-12	-72	36
Brainstem, pons, midbrain	5751	Bilateral	12.03	14	-30	-36
Cerebellum	13203	L	16.50	-18	-69	-48
Cerebellum	8532	R	12.97	27	-64	-58

the brainstem. We observed a large number of voxels in right midbrain and bilateral pons differed between two groups, indicating that fALFF of ADHD children changes significantly in the brainstem. This result is partially consistent with a previous resting-state study [33]. By investigating voluntary movement in mice, the midbrain was shown to be involved in regulating voluntary movements [34]. However, the cluster with maximum weight and large volume was located in the bilateral cerebellum. Neurons mainly function in integration and transmission of information, and cerebellum contains nearly half of all neurons in the human brain and thus plays a very important role. More than half of the cerebellar cortex was found to be associated with communication regions in cerebrum [35]. Cerebellum performs various important functions; it not only plays an important role in maintaining body balance and coordinating exercise but also being considered to participate in emotion regulation and cognitive functions, such as attention and language.

Our study discovered a considerable number of voxels in the left putamen and right caudate nucleus differed between two groups. Putamen and caudate nucleus are important components of striatum, which is the primary input channel of basal ganglia. Current studies have shown that manifestation of ADHD symptoms is closely related to dopamine and norepinephrine [36]. Lesions in fronto-striatal circuit will cause imbalance of dopamine and norepinephrine and then directly affect executive and motor functions.

For STG, we noted that many voxels in bilateral STG possess large weights. Temporal gyrus processes sounds in different frequencies and can extract useful information from the surrounding noise; it is also involved in social cognitive processes. Abnormality of STG may worsen auditory discrimination ability of children with ADHD.

We also discovered that numerous voxels in right ACC and right insula exhibit large weights in identifying ADHD individuals. Both right ACC and right insula act as hubs in the salience network. The salience network primarily functions in integrating various sensory information and completing attention orientation by coordinating and switching to associated processing systems [37]. When salience network is damaged, it may produce cognitive distortion, thus affecting individual behavior. Insula is the origin of social feelings; it shows relationship with different brain functions. Insula can accurately predict the body's internal situation and transfer demand information to decision-making-related regions, such as ACC and PFC. Lesions in insula will cause ADHD individuals to exhibit abnormal emotional states, such as failing to produce conscience, thus affecting their performance, or cause conduct disorder. The specific function of ACC is assumed to be related to decision making, impulse control, attention, and emotional regulation [38]. As a high-level regulatory region, ACC provides control signals from conflict monitoring for corresponding regions. Lesions in ACC will affect their ability to weigh the pros and cons and cause difficulty in ADHD individuals to control their impulses. These dysfunctions are all related to self-control.

In our current study, we identified ADHD individuals and obtained good results. Results in our study indicate that fALFF is an effective fMRI marker for investigation of ADHD. Moreover, brain regions that differ the most between ADHD and neurotypical individuals are found based on fALFF in each voxel and R-RELIEF algorithm. These results show that feature selection algorithm possesses huge potential in decoding the cognitive state from neural signals and exploring ADHD pathology. In future studies, more fMRI data can be used to establish more effective models to decode the neural activity.

IV. CONCLUSION

Finding prominent features and correctly detecting ADHD is crucial in diagnosing and curing ADHD. In our present study, a novel method is proposed to classify ADHD individuals and neurotypicals. In this work, multiple linear regression is introduced to reduce confounding effects of age and sex. Prominent components and entropy-based features combined fALFF are used to construct features. R-RELIEF algorithm is then proposed for the classification of ADHD and neurotypicals. We further compare our findings with other studies, comparison results show that our proposed method is comparable. In this work, R-RELIEF is also used for finding the brain regions that differ the most between ADHD and neurotypical individuals in addition to the classification of ADHD and neurotypical individuals. fALFF of many brain regions are found to differ between ADHD and neurotypicals based on R-RELIEF algorithm. Results of our study suggest that our method has potential for recognizing ADHD individuals, and R-RELIEF algorithm is a promising feature selection method in the research of the mechanism of ADHD.

REFERENCES

[1] I. Singh, "Beyond polemics: Science and ethics of ADHD," *Nature Rev. Neurosci.*, vol. 9, no. 12, pp. 957–964, Dec. 2008.

[2] G. V. Polanczyk, G. A. Salum, L. S. Sugaya, A. Caye, and L. A. Rohde, "Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents," *J. Child Psychol. Psychiatry*, vol. 56, no. 3, pp. 345–365, Feb. 2015.

[3] M. O. Sokunbi, W. Fung, V. Sawlani, S. Choppin, D. E. J. Linden, and J. Thome, "Resting state fMRI entropy probes complexity of brain activity in adults with ADHD," *Psychiatry Res., Neuroimag.*, vol. 214, no. 3, pp. 341–348, Oct. 2013.

[4] Q.-H. Zou *et al.*, "An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF," *J. Neurosci. Methods*, vol. 172, no. 1, pp. 137–141, Apr. 2008.
[5] D. G. Amen *et al.*, "Gender-based cerebral perfusion differences in 46,034 functional neuroimaging scans," *J. Alzheimer's Disease*, vol. 60, no. 2, pp. 605–614, Sep. 2017.
[6] M. P. Pase, N. A. Grima, C. Stough, A. Scholey, and A. Pipingas, "Association of pulsatile and mean cerebral blood flow velocity with age and neuropsychological performance," *Physiol. Behav.*, vol. 130, no. 6, pp. 23–27, Mar. 2014.
[7] F. X. Castellanos *et al.*, "Cingulate-precuneus interactions: A new locus of dysfunction in adult attention-deficit/hyperactivity disorder," *Biol. Psychiatry*, vol. 63, no. 3, pp. 332–337, Sep. 2008.
[8] T. Nakao, J. Radua, K. Rubia, and D. Mataix-Cols, "Gray matter volume abnormalities in ADHD: Voxel-based meta-analysis exploring the effects of age and stimulant medication," *Amer. J. Psychiatry*, vol. 168, no. 11, pp. 1154–1163, Nov. 2011.
[9] H. Hart, J. Radua, D. Mataix-Cols, and K. Rubia, "Meta-analysis of fMRI studies of timing in attention-deficit hyperactivity disorder (ADHD)," *Neurosci. Biobehav. Rev.*, vol. 36, no. 10, pp. 2248–2256, Aug. 2012.
[10] M. Hoogman *et al.*, "Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: A cross-sectional mega-analysis," *Lancet Psychiatry*, vol. 4, no. 4, pp. 310–319, Feb. 2017.
[11] A. Mueller, D. S. Hong, S. Shepard, and T. Moore, "Linking ADHD to the neural circuitry of attention," *Trends Cogn. Sci.*, vol. 21, no. 6, pp. 474–488, May 2017.
[12] L. Sørensen, T. Eichele, H. van Wageningen, K. J. Plessen, and M. C. Stevens, "Amplitude variability over trials in hemodynamic responses in adolescents with ADHD: The role of the anterior default mode network and the non-specific role of the striatum," *NeuroImage, Clin.*, vol. 12, pp. 397–404, Feb. 2016.
[13] A. M. Syed, M. U. Akram, T. Akram, M. Muzammal, S. Khalid, and M. A. Khan, "Fundus images-based detection and grading of macular edema using robust macula localization," *IEEE Access*, vol. 6, pp. 58784–58793, 2018.
[14] H. Ghanadian, M. Ghodratigohar, and H. A. Osman, "A machine learning method to improve non-contact heart rate monitoring using an RGB camera," *IEEE Access*, vol. 60, pp. 57085–57094, 2018.
[15] A. Riaz, M. Asad, E. Alonso, and G. Slabaugh, "Fusion of fMRI and non-imaging data for ADHD classification," *Computerized Med. Imag. Graph.*, vol. 65, pp. 115–128, Aug. 2018.
[16] R. Zafar *et al.*, "Prediction of human brain activity using likelihood ratio based score fusion," *IEEE Access*, vol. 5, pp. 13010–13019, 2017.
[17] 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.
[18] S.-F. Liang *et al.*, "Differentiation between resting-state fMRI data from ADHD and normal subjects: Based on functional connectivity and machine learning," in *Proc. Int. Conf. Fuzzy Theory Appl.*, Nov. 2012, pp. 294–298.
[19] S. Sarraf and G. Tofighi, "Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data," in *Proc. Future Technol. Conf.*, Dec. 2016, pp. 816–820.
[20] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fMRI," *NeuroImage*, vol. 129, pp. 292–307, Apr. 2016.
[21] F. Wang, Y. Li, and Z. Gu, "An MVPA method based on sparse representation for pattern localization in fMRI data analysis," *Neurocomputing*, vol. 269, pp. 206–211, Dec. 2017.
[22] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Proc. 10th Nat. Conf. Artif. Intell.*, 2017, pp. 129–134.
[23] J. Jia, N. Yang, C. Zhang, A. Yue, J. Yang, and D. Zhu, "Object-oriented feature selection of high spatial resolution images using an improved Relief algorithm," *Math. Comput. Model.*, vol. 58, nos. 3–4, pp. 619–626, Aug. 2013.
[24] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[25] Y. Chao-Gan and Z. Yu-Feng, "DPARSF: A MATLAB toolbox for 'pipeline' data analysis of resting-state fMRI," *Frontiers Syst. Neurosci.*, vol. 4, no. 13, p. 13, May 2010.
[26] M. Mikl *et al.*, "Effects of spatial smoothing on fMRI group inferences," *Magn. Reson. Imag.*, vol. 26, no. 4, pp. 490–503, 2008.
[27] M. O. Sokunbi *et al.*, "Sample entropy reveals high discriminative power between young and elderly adults in short fMRI data sets," *Front Neuroinform.*, vol. 8, p. 69, Jun. 2014.
[28] D. Kuang, X. Guo, X. An, Y. Zhao, and L. He, "Discrimination of ADHD based on fMRI data with deep belief network," in *Proc. Int. Conf. Intell. Comput.*, Oct. 2014, pp. 225–232.
[29] 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.
[30] L. Zhu *et al.*, "Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest," *Nature Neurosci.*, vol. 17, no. 10, pp. 1319–1321, Aug. 2017.
[31] F. Li *et al.*, "Intrinsic brain abnormalities in attention deficit hyperactivity disorder: A resting-state functional MR imaging study," *Radiology*, vol. 272, no. 2, pp. 514–523, Apr. 2014.
[32] L. Lim *et al.*, "Disorder-specific predictive classification of adolescents with attention deficit hyperactivity disorder (ADHD) relative to autism using structural magnetic resonance imaging," *PLoS ONE*, vol. 8, no. 5, May 2013, Art. no. e63660.
[33] Y. F. Zang *et al.*, "Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI," *Brain Develop.*, vol. 29, no. 2, pp. 83–91, Mar. 2007.
[34] E. M. Kolb *et al.*, "Mice selectively bred for high voluntary wheel running have larger midbrains: Support for the mosaic model of brain evolution," *J. Exp. Biol*, vol. 216, no. 4, pp. 515–523, Feb. 2013.
[35] C. J. Stoodley, "Distinct regions of the cerebellum show gray matter decreases in autism, ADHD, and developmental dyslexia," *Frontiers Syst. Neurosci.*, vol. 8, p. 92, May 2014.
[36] D. J. Chandler, B. D. Waterhouse, and W.-J. Gao, "New perspectives on catecholaminergic regulation of executive circuits: Evidence for independent modulation of prefrontal functions by midbrain dopaminergic and noradrenergic neurons," *Frontiers Neural Circuits*, vol. 8, p. 53, May 2014.
[37] L. Q. Uddin, "Salience processing and insular cortical function and dysfunction," *Nature Rev. Neurosci.*, vol. 16, no. 1, pp. 55–61, Jan. 2015.
[38] G. Bush, P. Luu, and M. I. Posner, "Cognitive and emotional influences in anterior cingulate cortex," *Trends Cognit. Sci.*, vol. 4, no. 6, pp. 215–222, Jun. 2000.

B. MIAO received the bachelor's and master's degrees from the School of Information Science and Engineering, University of Jinan, in 2015 and 2018, respectively. He is currently an Algorithm Engineer with Sina Corporation. His current research interests include the application of machine learning to neurosciences, including feature selection and classification.

L. L. ZHANG received the bachelor's and master's degrees from the School of Information Science and Engineering, University of Jinan, in 2015 and 2018, respectively, where he is currently a Research Assistant. His current research interests include image processing, machine learning, hybrid computational intelligence, and mathematical modeling.

J. L. GUAN is currently pursuing the bachelor's degree in mathematics and statistics with the University of New South Wales (UNSW). Her current research interest includes statistical methods for neurosciences.

Q. F. MENG received the master's and Ph.D. degrees from the School of Information Science and Engineering, Shandong University, in 2005 and 2008, respectively. Since 2008, she has been an Associate Professor with the University of Jinan. She has published more than 40 research papers, where 11 are indexed by SCI and more than 30 are indexed by EI. Her research interests include nonlinear time series analysis, biomedical signal processing, and computational intelligence.

Y. L. ZHANG received the M.Sc. and Ph.D. degrees from Chongqing University, Chongqing, China, in 1996 and 2002, respectively. He is currently a Full Professor with the School of Information Science and Engineering, University of Jinan (UJN). His current research interests include the applications of signal processing to medical image, machine learning, and deep learning.