A Feature Selection Method for Classification of ADHD

Bo Miao
Shandong Key Laboratory of Intelligent Computing
Technology of network environment
University of Jinan
Jinan, China
e-mail: bomiaobbb@foxmail.com

Yulin Zhang*
Shandong Key Laboratory of Intelligent Computing
Technology of network environment
University of Jinan
Jinan, China
e-mail: ise_zhangyl@ujn.edu.cn

***Abstract*—At present, the classification of brain diseases through neuroimaging data is a hot topic. Attention deficit hyperactivity disorder (ADHD) is usually diagnosed by the standard scale. However, the traditional diagnostic methods have high misdiagnosis rate and time consuming. In this paper, we discussed the classification of ADHD by using the feature subset obtained by preprocessing and feature selection of fractional amplitude of low-frequency fluctuation (fALFF) in resting-state functional magnetic resonance imaging (rs-fMRI) data. We proposed a feature selection algorithm based on Relief algorithm and verification accuracy (VA-Relief). The experimental results show that fALFF can be used to realize the high accuracy classification of ADHD by using our feature selection algorithm and preprocessing method. Therefore, it is possible to use rs-fMRI data and machine learning methods to assist the diagnosis of brain diseases.**

Keywords-brain disease, classification, feature selection

I. INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a common childhood psychiatric disorder, with a prevalence of approximately 3.4% [1], and symptoms still existed after many children became adults. Boys have a higher rate of ADHD than girls [2]. Children with ADHD are mainly manifested as a lack of attention, hyperactivity, and emotional impulse [3,4]. These obstacles make it difficult for children to learn normally in school and get along with others; these not only have an impact on the family, but also detrimental to the healthy growth of children [5].

Current diagnosis of ADHD is based on the doctor's subjective experience and clinical diagnostic criteria, a survey found that the sensitivity of this method was 70-90% [6]. High misdiagnosis rate not only made a lot of children with ADHD could not get very good treatment in the early period of disease, but also hindered the normal development of their lives. Therefore, to find a more reliable diagnostic method is now one of research hotspots. Many studies have shown that ADHD is associated with brain dysfunction, and the patient's brain structure and function are abnormal [7-10]. Therefore, it is feasible to classify ADHD based on neuroimaging data and machine learning methods [11].

Fractional amplitude of low-frequency fluctuation (fALFF) is the most direct way to reflect the local brain activity in resting state [12]. Some researchers have found that children with ADHD have an abnormal spontaneous brain activity compared with healthy children [13,14]. At present, most of researches on ADHD classification are based on a specific design experiment or brain network; as the brain grows with age [15,16], they need no significant difference in age between the training set and the test set when selecting samples. In order to overcome these shortcomings, we used fALFF of all age-corrected samples as features. Then, we used Relief algorithm to obtain a subset of candidate features, and finally achieved the classification with SVM. We not only overcame the problem of sample selection, but also got high classification accuracy.

In this article, remaining chapters are arranged as follows. Section II introduced related works. In Section III, experimental information and feature selection algorithm are introduced. In Section IV, experimental results are presented. Section V summarized the experiment.

II. RELATED WORKS

There are two methods of feature reduction: feature selection and feature extraction. Feature selection is one of the key technologies in target recognition, it is to select the most effective features from original features to improve the efficiency of machine learning. Feature selection has been applied to many areas [17-19]. As one of the best feature selection algorithms, Relief algorithm was first proposed by Kira et al. [20]. Ye et al. [21] then proposed the Multi-Relief algorithm for multi-classification. Sun [22] proposed a heuristic Relief algorithm to solve the convex optimization problem based on the boundary value objective function. In recent years, it has been one of the hot spots that classification of brain diseases by using neuroimaging data and machine learning. Researchers designed experiments or selected certain features for machine learning to achieve the classification of brain diseases. Colby et al. [23] used the multiple support vector machine recursive feature elimination (SVM-RFE) algorithm combined with a radial basis function to achieve the classification of ADHD, the classification accuracy was 55%. Tenev et al. [24] measured the spectrum with EEG and designed the experiment to classify ADHD subtypes. Ghiassian et al. [25] extracted histogram of oriented gradients features from functional magnetic resonance imaging data for ADHD classification, the classification accuracy was 69.6%. Lim et al. [26] did a VBM analysis on structure images, and then used Gauss process classifier to classify ADHD, with a classification accuracy of 79.3%. Deshpande et al. [27] classified ADHD with the full connected cascade artificial neural network architecture, with a classification accuracy of 90%. Peng et al. [28] based on brain structure data, classified ADHD with extreme learning machine, with a classification accuracy of 90.08%.

III. MATERIALS AND METHODS

A. Participants

We obtained the neuroimaging data from the ADHD-200 dataset. The specific information is shown in Table I, all subjects were right-handed, and no other chronic diseases.

TABLE I. THE INFORMATION OF SUBJECTS

Status	Number	Age
Patients	82	11.19 (2.63)
Normal control	72	12.41 (3.28)

B. Preprocessing of Rs-fMRI Data

We removed initial 10 time points to ensure the stability of the fMRI signal. Then the head motion parameter is estimated, sudden head movements exceeded 3 mm or 3^{o} were excluded. We then used spatial normalization to transform each voxel size into isotropic 3 mm, and removed the linear trend. Finally, we calculated the fractional amplitude of low-frequency fluctuation (fALFF) at frequency band of 0.01-0.08 Hz. We adjusted age effect of fALFF data so as to keep samples as uniform as possible.

C. Feature selection based on VA-Relief

Relief algorithm not only has high efficiency, but also brings more satisfactory results, so it is widely used in different areas. Relief algorithm assigns the feature weight according to the size of correlation of each feature and category.

In this paper, we divided the data into training set, validation set, and test set. In the n samples of training set, we first took the first sample R as the starting point; and then calculated the Euclidean distance between each sample and R; then, we selected k nearest same class samples and k furthest different class samples, and constituted a subset. The Euclidean distance $d_{x,y}$ between sample $x(x_1,x_2,...,x_m)$ and $y(y_1,y_2,...,y_m)$ is:

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

The number of voxels per sample is m. Then start from the first voxel g, a_g is error sum of square between each sample in the same class and b_g is error sum of square between each sample in the different class. We use following formulas to calculate weight W_1 of the voxel g in current subset:

$$a_g = \frac{\sum_{i=1}^{k}\sum_{h=1}^{k}(g_i - g_h)^2 + \sum_{p=k+1}^{2k}\sum_{q=k+1}^{2k}(g_p - g_q)^2}{k} \tag{2}$$

$$b_g = \frac{\sum_{i=k+1}^{2k}\sum_{h=1}^{k}(g_i - g_h)^2 + \sum_{p=1}^{k}\sum_{q=k+1}^{2k}(g_p - g_q)^2}{k} \tag{3}$$

$$W_1 = b_g - a_g \tag{4}$$

After calculating the weight of each voxel in current subset, we took the other n-1 samples as starting points, respectively, and calculated weights. The final weight of voxel g is $W=W_1+W_2+...+W_n$. After we got the weight of each voxel, we ranked them in descending order according to the weight, and got the candidate feature subset C by setting the threshold.

In machine learning, support vector machine (SVM) is a supervised learning model related to the relevant learning algorithm, which has a good effect in the recognition of small sample and high dimension data. The key of SVM is the selection of kernel function, because the number of features in our experiment is much larger than the number of samples, we choose the linear kernel function. For the candidate feature subset $C=\{C_1,C_2,...,C_n\}$, we used C_i (i=1,2,...,n) corresponding data from training set and validation set to complete training and classification, respectively. According to the classification accuracy provided by each feature, the features within C are sorted in descending order, and features with classification accuracies less than 50% were removed. Finally we got the selected features. Fig. 1 gives the framework of our method.

IV. RESULTS

An overview of the experimental data is shown in Table II, a whole brain fALFF image can be divided into 67541 voxels, which means that we initially have 67541 features. Original Relief algorithm needs to select the parameters of nearest neighbor number k and iteration frequency u. Original Relief algorithm randomly selects a sample as the starting point, repeat u times, and then take the average weight as the final weight. In our algorithm, we set nearest neighbor number k which is determined by the least number of samples of two classes. Our algorithm selects each sample as the starting point, and obtains the sum of weights as the weight of the feature. Compared to random sampling, our approach has a better robustness.

TABLE II. EXPERIMENTAL DATA INFORMATION

Voxel size	Number of voxels (Features)	Number of subjects	Number of categories
$1.0\times1.0\times1.0$ mm^3	67541	155	2

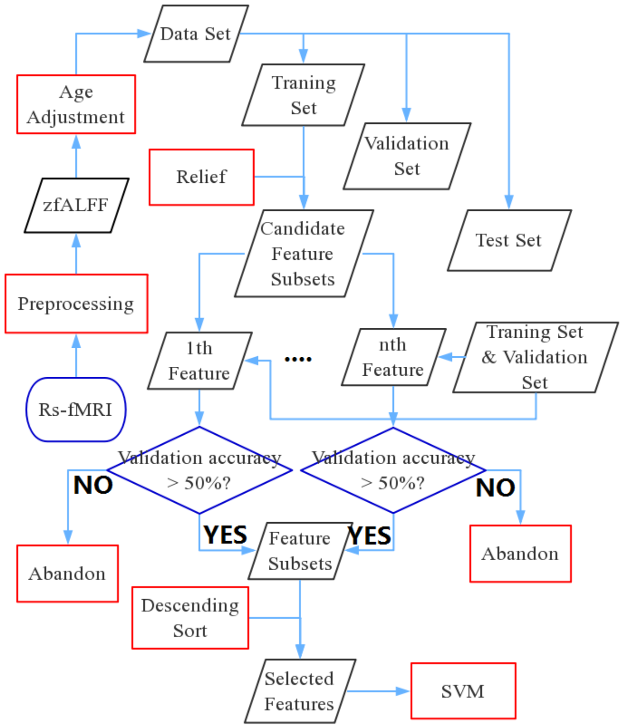

Figure 1. The framework of our proposed method. The process consists of three steps: data preprocessing, feature selection and classification.

Table III and Fig. 2 expressed the final experimental results, we used our algorithm to compare with the results of Relief algorithm and minimum redundancy maximum relevance (mRMR) algorithm, respectively. In Fig. 2 we can find that the classification accuracy is 92.16% based on the threshold set by us. The accuracy of our algorithm is better than that of Relief algorithm and mRMR algorithm in all feature dimensions.

TABLE III. THE CLASSIFICATION ACCURACY OF THREE ALGORITHMS

Feature Dimension	Relief	mRMR	VA-Releif
500	72.73%	67.53%	94.12%
1000	66.23%	64.94%	92.16%
1500	71.43%	71.43%	96.08%
2000	67.53%	72.73%	94.12%
2500	63.64%	64.04%	94.12%
2728	63.64%	59.74%	92.16%

Table IV shows the highest accuracy of the three algorithms, and the highest classification accuracy of our algorithm is up to 98%. So it can be shown that the classification accuracy of our algorithm for ADHD is higher than that of the other two algorithms.

Figure 2. Classification accuracy of three algorithms. The blue solid line indicates the classification accuracy of mRMR feature selection algorithm; the orange solid line represents the classification accuracy obtained by Relief feature selection algorithm; and the green solid line indicates the classification accuracy obtained by our feature selection algorithm.

TABLE IV. THE HIGHEST CLASSIFICATION ACCURACY OF THREE ALGORITHMS

	Relief	mRMR	VA-Releif
Highest accuracy	77.92%	80.52%	98.04%

V. CONCLUSION

In this paper, we proposed a VA-Relief feature selection algorithm for ADHD classification. Our algorithm first uses a slightly improved Relief algorithm to calculate the weight of each feature, and selects a candidate feature subset; then the final feature subset is selected according to the accuracy of verification. The experimental results show that our method can select a better feature subset, and achieve high accuracy classification of ADHD. Therefore, resting-state functional magnetic resonance imaging data and machine learning methods can be considered for clinical diagnosis.

There are other problems that need to be solved in further work. The data used in this article are from children and adolescents, and classification of adult patients is also worth studying. In addition, high accuracy classification of ADHD subtypes also needs to be realized.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant No. 61671220.

REFERENCES

[1] 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,” *Journal of Child Psychology and Psychiatry*, vol. 56, Dec. 2015, pp. 345-365.

[2] M. Rutter, A. Caspi, and T. E. Moffitt, “Using sex differences in psychopathology to study causal mechanisms: unifying issues and

research strategies," *Journal of Child Psychology and Psychiatry,* vol. 44, Dec. 2003, pp. 1092-1115.

[3] American Psychiatric Association, Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing, 2003.

[4] E. B. Gorman Bozorgpour, R. Klorman, and T. E. Gift, "Effects of subtype of attention-deficit/hyperactivity disorder in adults on lateralized readiness potentials during a go/no-go choice reaction time task," *Journal of Abnormal Psychology*, vol. 122, Dec. 2013, pp. 969-978.

[5] A. J. Feuer, A. Thai, R. T. Demmer, and M. Vogiatzi, "Association of Stimulant Medication Use With Bone Mass in Children and Adolescents With Attention-Deficit/Hyperactivity Disorder," *Jama Pediatrics,* Dec. 2016, pp. e162804.

[6] M. D. Weiler, D. K. Bellinger, E. K. Simmons, L. K. Rappaport, D. K. Urion, W. J. Mitchell, et al., "Reliability and validity of a DSM-IV based ADHD screener," *Child Neuropsychology A Journal on Normal & Abnormal Development in Childhood & Adolescence*, vol. 6, Dec. 2000, pp. 3-23.

[7] M. Hoogman, J. Bralten, D. P. Hibar, M. Mennes, M. P. Zwiers, L. S. J. Schweren, et al., "Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis," *The Lancet Psychiatry*, Dec. 2017.

[8] P. Shaw, K. Eckstrand, and W. Sharp, "Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation," *Proc Natl Acad Sci U S A*, vol. 104, Dec. 2007, pp. 19649-19654.d

[9] J. R. Andrews-Hanna, J. Smallwood, and R. N. Spreng. "The default network and self-generated thought: component processes, dynamic control, and clinical relevance," *Annals of the New York Academy of Sciences*, vol. 1316, Dec. 2014, pp. 29-52.

[10] 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 in Neural Circuits*, vol. 8, Dec. 2014, pp. 1983-1988.

[11] L. E. Collij, F. Heeman, J. P. Kuijer, R. Ossenkoppele, M. R. Benedictus, C. Möller, et al., "Application of Machine Learning to Arterial Spin Labeling in Mild Cognitive Impairment and Alzheimer Disease," *Radiology*, vol. 281, Dec. 2016, pp. 865-875.

[12] Q. H. Zou, C. Z. Zhu, Y. H. Yang, X. N. Zuo, X. Y. Long, Q. J. Cao, et al., "An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF," *Journal of Neuroscience Methods*, vol. 172, Dec. 2008, pp. 137-141.

[13] H. Yang, O. Q. Li, M. M. Zhang, and X. Y. Long, "WITHDRAWN: Spontaneous brain activity in medication-naive ADHD boys revealed by ALFF analysis," *Brain & Development*, Dec. 2010.

[14] J. R. Sato, M. Q. Hoexter, A. Fujita, and L. A. Rohde, "Evaluation of pattern recognition and feature extraction methods in ADHD prediction," *Frontiers in Systems Neuroscience*, vol. 6, Dec. 2012, pp. 68.

[15] P. Rzezak, P. Squarzoni, F. L. Duran, T. de Toledo Ferraz Alves, J. Tamashiro-Duran, C. M. Bottino, et al., "Relationship between Brain Age-Related Reduction in Gray Matter and Educational Attainment," *PLoS One*, vol. 10, Dec. 2015, pp. e0140945.

[16] E. Luders, N. Cherbuin, and C. Gaser, "Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners," *Neuroimage*, vol. 134, Dec. 2016, pp. 508-513.

[17] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, Dec. 2007, pp. 2507-2517.

[18] M. Pal, and G.M. Foody, "Feature Selection for Classification of Hyperspectral Data by SVM," *Geoscience & Remote Sensing IEEE Transactions on*, vol. 48, Dec. 2010, pp. 2297-2307.

[19] G. Chandrashekar, and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, Dec. 2014, pp. 16-28.

[20] K. Kira and L. A. Rendell, "The Feature Selection Problem: Traditional Methods and a New Algorithm," *AAAI-92 Proceedings*, Dec. 1992, pp. 129-134.

[21] K. Ye, K. A. Feenstra, J. Heringa, A. P. Ijzerman, and E. Marchiori, "Multi-RELIEF: a method to recognize specificity determining residues from multiple sequence alignments using a Machine-Learning approach for feature weighting," *Bioinformatics*, vol. 24, Dec. 2007, pp. 18-25.

[22] Y. Sun, "Iterative RELIEF for feature weighting: algorithms, theories, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, Dec. 2007, pp. 1035-1051.

[23] 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 in Systems Neuroscience*, vol. 6, Dec. 2012, pp. 59.

[24] A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Müller, and G. Candrian, "Machine learning approach for classification of ADHD adults," *International Journal of Psychophysiology Official Journal of the International Organization of Psychophysiology*, vol. 93, Dec. 2014, pp. 162-166.

[25] S. Ghiassian, R. Greiner, P. Jin, and M. R. G. Brown, "Using Functional or Structural Magnetic Resonance Images and Personal Characteristic Data to Identify ADHD and Autism," *PLoS ONE*, vol. 11, Dec. 2016, pp. e0166934.

[26] L. Lim, A. Marquand, A. A. Cubillo, A. B. Smith, K. Chantiluke, A. Simmons, 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, Dec. 2013, pp. e63660.

[27] G. Deshpande, P. Wang, D. Rangaprakash, and B. Wilamowski, "Fully Connected Cascade Artificial Neural Network Architecture for Attention Deficit Hyperactivity Disorder Classification From Functional Magnetic Resonance Imaging Data," *IEEE transactions on cybernetics*, vol. 45, Dec. 2015, pp. 2668-2679.

[28] 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, Dec. 2013, pp. e79476.

AUTHORS' BACKGROUND

Your Name	Title*	Research Field	Personal website
Bo Miao	master student	Signal and information processing	
Yulin Zhang	full professor	Signal and information processing	

*This form helps us to understand your paper better, the form itself will not be published.

*Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor