(MIJ) 2023, Vol. No. 9, Jan-Dec **e-ISSN: 2454-924X; p-ISSN: 2454-8103**

Developing an Integrated Model for Classifying Attention Deficit Hyperactivity Disorder (ADHD) by Leveraging an Integrated Feature Selection Framework Both Within and Across a Cohort

Bahisht Samar

Amity University, Noida, India

[1]*Received: 19 September 2023; Accepted: 18 October 2023; Published: 02 November 2023*

ABSTRACT

Attention-deficit/hyperactivity disorder (ADHD) is a childhood-onset neurodevelopmental disorder that often persists into adulthood. However, as lacking objective measures, several studies have questioned the stability in diagnosing of ADHD from childhood to adulthood. In this study, we propose a novel feature selection framework based on functional connectivity (FCs) pattern, the so-called 'FS_RIWEL,' which could classify ADHD from age matched healthy controls (HCs) with ~80% accuracy (both for children and adults). More importantly, the feature space learned from child ADHD dataset can discriminate adult ADHD from HCs at ~70% accuracy. To the best of our knowledge, this is the first attempt to perform a cross-cohort prediction between the adult and child ADHD using FC features. In addition, the most frequently selected FCs indicate that ADHD exhibit widely-impaired FC patterns in frontoparietal, basal ganglia, cerebellum network and so on suggesting that FCs may serve as potential biomarkers for ADHD diagnosis.

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorder defined as a combination of age-inappropriate levels of inattention, hyperactivity and impulsive behaviour [1]. According to reports [2, 3], ADHD affects 5% children and persists into adults' life two-thirds of the case, either as the full condition or persistence of some symptoms associated with impairments in psychosocial functioning. Decades of research has firmly established modern conceptualization of ADHD as a chronic disorder. However, the persistence of ADHD into adulthood was questioned reflecting instability in diagnosing [4, 5]. Furthermore, the etiological bases and neural substrates of ADHD are far from being fully understood. Therefore, a more accurate discriminative method to identify objective imaging biomarkers is crucial for diagnosis and prognosis of ADHD, which may facilitate better intervention and more effective treatment. Driven by ADHD-200 Consortium, researches on children ADHD have made substantial achievements, whereas only a few studies paid attention to adults ADHD [6-8], let alone using a data-driven method to investigate children ADHD and adults ADHD together. On the other hand, developing novel machine learning algorithm to identify MRI features that may sever as potential biomarkers for brain disorders' diagnosis and treatment has been a hot topic. As reported in [6-11], functional connectivity (FCs) derived from resting-state functional MRI (fMRI) data are effective tools to better depict brain dysfunctions and to discriminate brain disorders. However, currently high dimensional FC features are often derived from limited fMRI samples, which could include great redundancy and degrade the classification performance [12]. To deal with the challenges of high dimensional features associated with small samples, we proposed a new feature selection framework, i.e., FS_RIWEL, which consist of 3 steps: 1) feature ranking based on relative importance; 2) feature selection based on the forward-backward learning; 3) classification using a weighted ensemble classifier. The relative importance, calculated from different decision trees, reveals the degree of one feature contributes to the target label [17]. After ranking these features from different ensemble algorithm, a forward-backward selection algorithm is employed to increase the diversity of new feature space while it still can maintain the low dimensionality of FCs feature subspace. Finally, weighted ensemble classifiers will be tuned on the refined feature subspace. Furthermore, the refined feature subspace

[1] *How to cite the article:* Samar B. (2023) Developing an Integrated Model for Classifying Attention Deficit Hyperactivity Disorder (ADHD) by Leveraging an Integrated Feature Selection Framework Both Within and Across a Cohort; *Multidisciplinary International Journal,* Vol 9, 203-209

from children ADHD will be used to predict adults ADHD dataset to verify whether the stable biomarkers with FCs for the whole ADHD disorder exist.

MATERIALS AND METHODS

Participants

The adults ADHD dataset includes 112 ADHD patients and 77 age-matched HCs while 106 ADHD and 73 age-matched HCs constitute the children ADHD dataset. All participants were recruited from the Sixth Hospital of Peking University

(PKU6), who were scanned at either PKU or Beijing Normal University (BNU). Subjects were fully informed about the research and signed the informed consent. Demographic data provided in Table 1.

TABLE 1: DEMOGRAPHIC CHARACTERISTICS

	ADULTS DATASET		CHILDREN DATASET	
	HC	ADHD	HC	ADHD
SITE PKU6	43	73	45	72
SITE BNU	34	39	28	34
NUMBER	77	112	73	106
GENDER (F/M)	34F/43M	37F/75M	18F/55M	10F/96M
AGE (MEAN±SD)	26.0±3.9	25.9±4.9	10.1±1.1	10.0±1.9

Resting-State Functional Connectivity Analysis

Both children and adults ADHD datasets were pre-processed with the Data Processing Assistant for Resting-State fMRI (DPARSFA, http://rfmri.org/DPARSF [13]). Subsequent data preprocessing included first ten volumes discard, slice timing correction, head motion correction, normalizing to the MNI template, resampling to 3 × 3 × 3 mm3 , smoothing with a 4 mm Gaussian kernel, temporal band-pass filtering (0.01 Hz to 0.1 Hz), and regressing out nuisance signals of head motion parameters, white matter, CSF, and global signals. The registered functional MRI volumes with the MNI template were divided into 273 regions according to the Brainnetome Atlas [14] incorporating 210 cortical, 36 subcortical and 27 cerebellar regions.

Regional mean time series were acquired for each by averaging the functional MRI time series over all voxels in each of the 273 regions. Pearson correlation coefficients between pairs of node time courses were calculated and normalized to z score using Fisher transformation. Then each subject generated a 273×273 symmetric connectivity matrix. We used the upper triangle elements of the functional connectivity matrix, ignoring all diagonal elements, as prediction features, i.e., the feature space for prediction was spanned by (273×272)/2 =37128-dimensional feature vectors.

FS_RIWEL

In this study, a novel feature selection method based on relative importance and weighted ensemble learning (FS_RIWEL) is employed to address the critical issue “curse of dimensionality” [15]. The benefit of methods like Random Forest or XGBoost is that, after constructed those decision trees, it is relatively straight-forward to retrieve importance scores for each feature. In general, importance score is calculated using formula (1) that indicates how informative or valuable each feature was averaged across all of the decision trees in the constructions of the classifier. Relative importance from random forest or Extra-trees may be biased, yet different model counts not the same score on these features. Therefore, ensemble learning with other boosting models is employed to balance and establish a better-refined feature space [16].

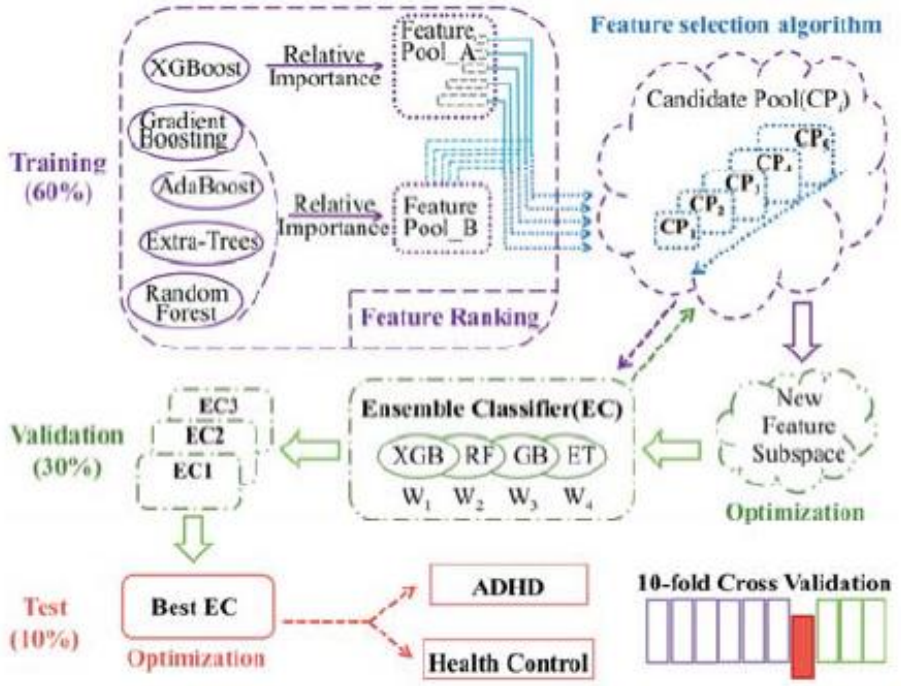

Figure 1 Flowchart of our proposed FS_RIWEL algorithm

The flowchart of FS_RIWEL algorithm is shown in Figure 1. Five different algorithms are employed to calculate relative importance on each feature. We can see from Figure 1 that Feature Pool_A which calculated from XGBoost [17] is composed of five different dimensional feature spaces. Rectangles of different length in Pool_A represent different numbers of features we employed (from 0.05% to 0.2% with the size of original features). Four different algorithms including randomized decision trees (a.k.a. ExtraTrees), Random Forest, AdaBoost, and Gradient Boosting were employed to generate Feature Pool_B, and top 2% importance features calculated from different models were assembled without any repetition. Each set in Feature Pool_ A with Feature Pool_B generates one Candidate Pool (CPi). All pools will feed into our forward backward feature selection algorithm which fused with different wrappers into model learning to maintain diversity and prevent overfitting on the new feature subspace.

Refined subspace would reduce redundant features, besides several wrappers guarantee more informative features would take into consideration. The purple dash line from feature selection algorithm to ensemble classifier present the contents that wrappers used in feature selection process

constitute next ensemble classifier. Details on forward backward feature selection algorithm are demonstrated in Figure 2.

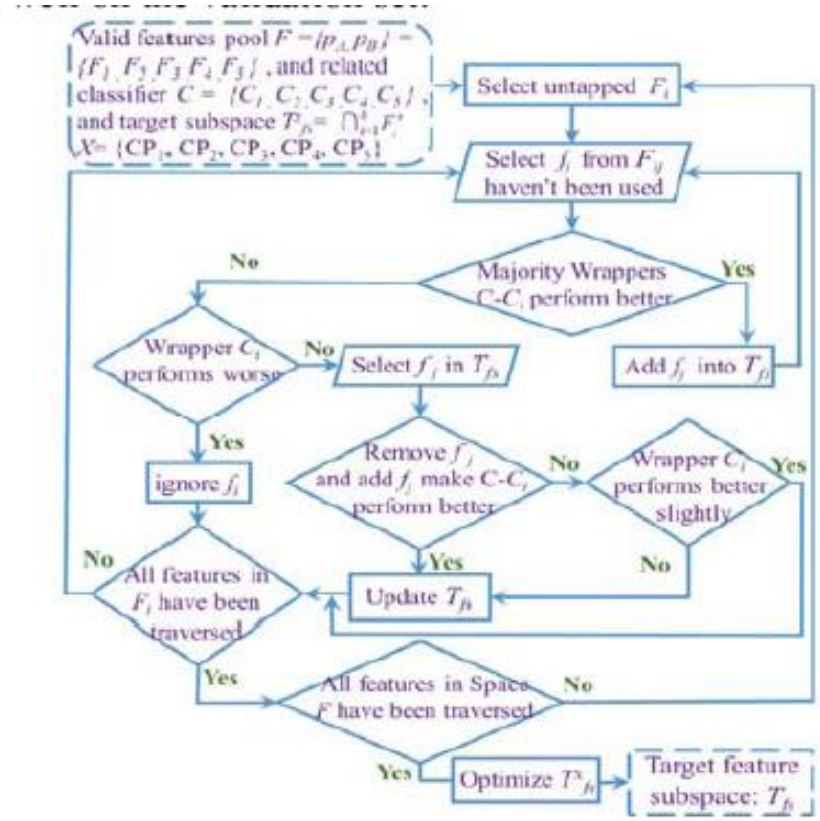

Figure 2 Details on forward-backward feature selection algorithm

A refined feature subspace is established after this valid feature selection strategy. Based on this feature subspace, ensemble classifier tunes the weight of each wrapper on validation data. Nested 3-fold cross-validation on the whole training and validation set generates three ensemble classifiers, and the classifier which performed best will be employed as the final model to discriminate ADHD from age-matched HCs on test data. The optimization criterion based on the number of final features and counts that majority wrappers work. The blue dash line from ensemble classifier to feature selection algorithm depicts a retrain process on base classifier when ensemble classifier performs not well on the validation set.

MRI Data Acquisition

Data scanned from PKU6 data were obtained from a GE Signa 3T Horizon HDx system, while data scanned from BNU were acquired on a Siemens Trio 3 T scanner.

Resting-state fMRI images were acquired using an echoplanar imaging sequence with the following parameters on the Siemens scanner: repetition time (TR) = 2,000 ms, echo time (TE) = 30 ms, flip angle= 90°, slice thickness/gap=3.5/4.2 mm, matrix = 64 × 64, field of view (FOV) = 200 mm × 200 mm, 33 axial slices, and 240 volumes. And the same parameters on GE scanner were: TR = 2,000 ms, TE = 30 ms, flip angle = 90°, slice thickness/gap = 3.2/0 mm, matrix = 64 × 64, FOV = 200 mm × 200 mm, 43 axial slices and 240 volumes.

RESULTS

In this study, we first use a nested cross-validation strategy to estimate the classification performance (ADHD vs. HC) for adults and children separately, then based on the learned features and model to implement the cross-cohort prediction.

Classification accuracy within each cohort

For intra-classification, we evaluate our model with 10-fold stratified cross validation strategy. As shown in Fig 1, 10% subjects were extracted as testing data; the remained data were further split into a training set (2/3, 60%) and validation set (1/3, 30%). Each loop generates a refined new feature subspace to predict subjects on the test data. As

shown in Fig 3, the proposed FS_RIWEL algorithm was compared with four popular feature selection methods including Lasso, ElasticNet, Fisher_score, and Trace_Ratio, of which the parameters were tuned using grid search strategy. The values of alpha in Lasso are chosen from {0.005, 0.006, …, 0.1}. The values of alpha in ElasticNet are varied from 0.1 to 2 with step 0.1 as l1_ratio is changing from 0.02 to 0.7 also with 0.01 step. After ranking all features under Trace_Ratio or Fisher_score algorithm, the number of remaining features ranged from 10 to 500. The candidate classifiers include linear/RBF SVM and XGBoost.

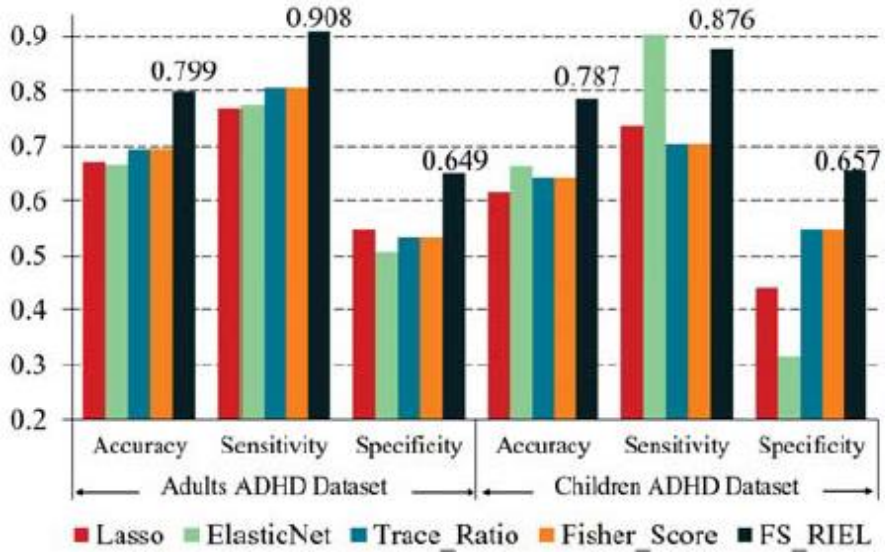

Figure 3 Three metrics for performance evaluation on different feature selection methods with both adults and children ADHD.

Figure 3 illustrates the average of accuracy, sensitivity, and specificity repeated 10 times for 10-fold cross validation on both adults and children dataset under optimized models mentioned above. It is remarkable that FS_RIEL achieved 10% higher on accuracy than all other four methods except for sensitivity with the ElasticNet algorithm on children ADHD. The most frequently selected FCs based on our method are shown in Figure 4. Lines on more width denotes more frequency were used in new subspace, and pink lines are the top two frequency FCs mined by FS_RIWEL. These FCs were mainly involved in the frontoparietal network, default network, salience network, basal ganglia network and cerebellum network, consistent with previous findings that large-scale brain networks were impaired in ADHD. Besides, two pictures exhibit some similar connectivity patterns enlighten us to give Inter-prediction from child to adult ADHD dataset a shot.

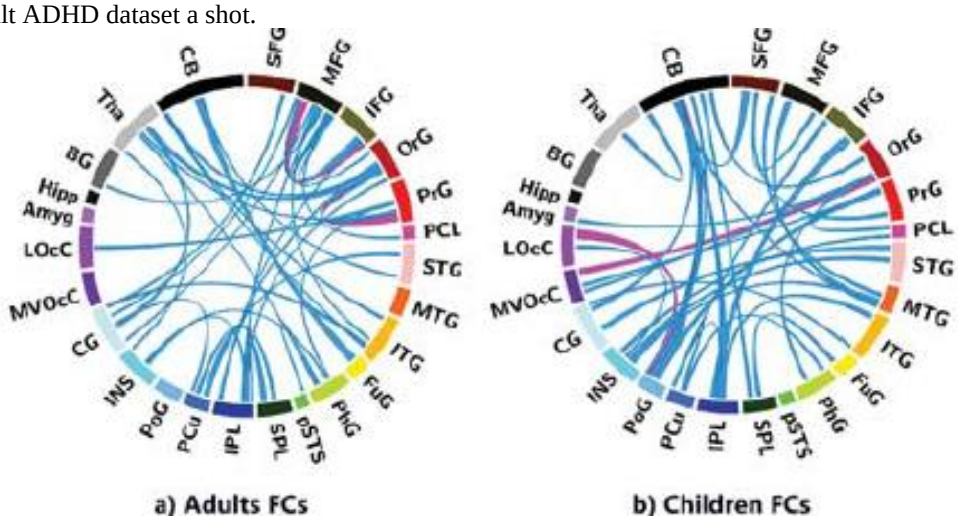

Figure 4. The number of FCs appeared greater than ten times that mined by our FS_RIWEL algorithm on two datasets

Child->Adult prediction on ADHD

For cross-cohort prediction, we collect FCs that occurring with 100% frequency during each fold with FS_RIWEL on children dataset. Then corresponding FCs on adults' subjects are obtained as input feature space. The specific classifiers (include Gradient Boost, Random Forest, Extra Tress, and XGBoost) are used to verify whether this feature space can discriminate on adults ADHD dataset efficiently or not. The reason why these classifiers are employed is that FS_RIWEL algorithm constructs and refined feature subspace on children dataset with them. Furthermore, an ensemble classifier comprises these mentioned models to classify adults ADHD dataset.

We also build one 1D-convolutional neural network firstly trained on children dataset (achieved 0.700 accuracy) and then transfer the learned model into adults ADHD dataset for comparison. The results based on four feature selection methods with adults ADHD dataset were also added into Figure 5.

As shown in Figure 5, The accuracy with one specific classifier shows lower performance than a model with simple discriminate criterion after feature selection process. Though the input feature spaces are different, results still illustrate the importance of feature reduction when facing the high dimensional and small samples issue. Ensemble classifier achieved 0.701 on accuracy value after combine these single classifier [GradientBoost, RandomForest, ExtraTrees and XGBoost] with weights [0.15, 0.30 ,0.10 and 0.45].

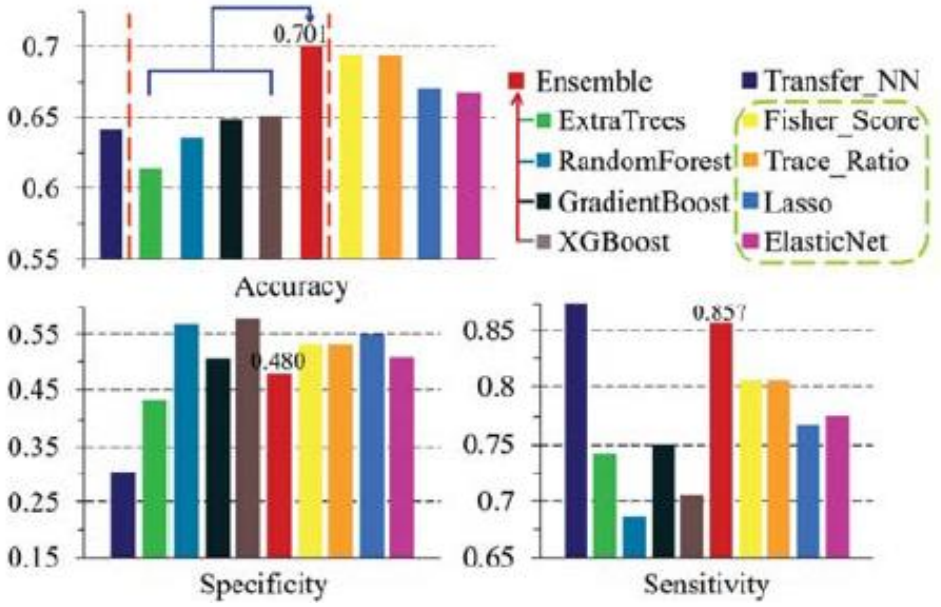

Figure 5. Different classifiers work on feature space derived from children ADHD dataset to predict adults ADHD dataset

The high sensitivity value indicates FCs from children ADHD may still work on adults ADHD while low value on specificity reflects the difference between adults HCs and children HCs on FCs features.

CONCLUSION

In this work, we proposed a new feature selection framework (FS_RIWEL) which performs better than the cutting-edge alternative methods when dealing with high dimensional features with limited samples size. Besides, the most frequently selected FCs not only implicate alternations in FCs between multiple brain regions in ADHD but also keep pace with previous findings that both children and adults ADHD patients suffer from large-scale brain networks dysfunction. To the best of our knowledge, this is the first attempt that employed machine learning methods to identify common impaired FCs between children ADHD and adults ADHD, compared with age-matched HCs. Moreover, 70% accuracy was achieved when using the FC features derived from children ADHD classification to guide prediction of adults ADHD, suggesting that FCs from rsfMRI may serve as a potential useful biomarker for ADHD diagnosis and prognosis.

REFERENCES

1. Biederman J, Faraone SV. Attention-deficit hyperactivity disorder [J]. CNS Drugs, 20(2):107-123, 2005

2. Kessler R C, Adler L, Barkley R, et al. The prevalence and correlates of adult ADHD in the United States: results from the National Comorbidity Survey Replication [J]. American Journal of Psychiatry, 163(4):716, 2006.

3. Kuntsi J, Mcloughlin G, Asherson P. Attention deficit hyperactivity disorder [J]. Practitioner, 8(4):461, 2006.

4. Zang Y F, He Y, Zhu C Z, et al. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI[J]. Brain & Development, 29(2):83-91, 2007.

5. Sibley M H, Mitchell J T, Becker S P. Method of adult diagnosis influences estimated persistence of childhood ADHD: a systematic review of longitudinal studies [J]. Lancet Psychiatry, 3(12):1157, 2016.

6. Deshpande G, Wang P, Rangaprakash D, et al. Fully Connected Cascade Artificial Neural Network Architecture for Attention Deficit Hyperactivity Disorder Classification From Functional Magnetic Resonance Imaging Data [J]. IEEE Transactions on Cybernetics, 45(12):2668-2679, 2015.

7. Du J, Wang L, Jie B, et al. Network-based classification of ADHD patients using discriminative subnetwork selection and graph kernel PCA[J]. Computerized Medical Imaging & Graphics, 52:82-88, 2016.

8. Sun H, Chen Y, Huang Q, et al. Psychoradiologic Utility of MR Imaging for Diagnosis of Attention Deficit Hyperactivity Disorder: A Radiomics Analysis. [J]. Radiology, 287 (Suppl 11): 170226, 2017.

9. Castellanos F X, Aoki Y. Intrinsic Functional Connectivity in Attention-Deficit/Hyperactivity Disorder: A Science in Development [J]. Biological Psychiatry Cognitive Neuroscience & Neuroimaging, 1(3):253-261, 2016.

10. Norman L J, Carlisi C, Lukito S, et al. Structural and functional brain abnormalities in attention-deficit/hyperactivity disorder and obsessive-compulsive disorder: a comparative metaanalysis[J]. JAMA Psychiatry, 73(8): 815-825, 2016.

11. Meng X, Jiang R, Lin D, Bustillo J, Jones T, Chen J, Yu Q, Du Y, Zhang Y, Jiang T, Sui J. Predicting individualized clinical measures by a generalized prediction framework and multimodal fusion of MRI data. Neuroimage. 145, 218-29, 2017.

12. Jie NF, Zhu MH, Ma XY, Osuch EA, Wammes M, Theberge J, Li HD, Zhang Y, Jiang TZ, Sui J, Calhoun VD. Discriminating Bipolar Disorder From Major Depression Based on SVM-FoBa: Efficient Feature Selection With Multimodal Brain Imaging Data. IEEE transactions on autonomous mental development 7, 320-331 (2015).

13. Chao-Gan Y, Yu-Feng Z. DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI [J]. Frontiers in Systems Neuroscience, 4(13):13, 2010

14. Fan L, Li H, Zhuo J, et al. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture [J]. Cerebral Cortex, 26(8):3508-3526, 2016.

15. Li J, Cheng K, Wang S, et al. Feature Selection: A Data Perspective [J]. Acm Computing Surveys, 50(6), 2016.

16. Yao, D., Calhoun, V. D., Fu, Z., Du, Y., & Sui, J. (2018). An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment. Journal of neuroscience methods, 302, 75-81.

17. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016: 785-94.