Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques

Yuyang Luo[a], Tara L. Alvarez[a], Jeffrey M. Halperin[b], Xiaobo Li[a,c,*]

[a] Department of Biomedical Engineering, New Jersey Institute of Technology, NJ, USA
[b] Department of Psychology, Queens College, City University of New York, NY, USA
[c] Department of Electrical and Computer Engineering, New Jersey Institute of Technology, NJ, USA

A B S T R A C T

Attention-deficit/hyperactivity disorder (ADHD) is a highly prevalent and heterogeneous neurodevelopmental disorder, which is diagnosed using subjective symptom reports. Machine learning classifiers have been utilized to assist in the development of neuroimaging-based biomarkers for objective diagnosis of ADHD. However, existing basic model-based studies in ADHD report suboptimal classification performances and inconclusive results, mainly due to the limited flexibility for each type of basic classifier to appropriately handle multi-dimensional source features with varying properties. This study applied ensemble learning techniques (ELTs), a meta-algorithm that combine several basic machine learning models into one predictive model in order to decrease variance, bias, or improve predictions, in multimodal neuroimaging data collected from 72 young adults, including 36 probands (18 remitters and 18 persisters of childhood ADHD) and 36 group-matched controls. All currently available optimization strategies for ELTs (i.e., voting, bagging, boosting and stacking techniques) were tested in a pool of semifinal classification results generated by seven basic classifiers. The high-dimensional neuroimaging features for classification included regional cortical gray matter (GM) thickness and surface area, GM volume of subcortical structures, volume and fractional anisotropy of major white matter fiber tracts, pairwise regional connectivity and global/nodal topological properties of the functional brain network for cue-evoked attention process. As a result, the bagging-based ELT with the base model of support vector machine achieved the best results, with significant improvement of the area under the receiver of operating characteristic curve (0.89 for ADHD vs. controls and 0.9 for ADHD persisters vs. remitters). Features of nodal efficiency in right inferior frontal gyrus, right middle frontal (MFG)-inferior parietal (IPL) functional connectivity, and right amygdala volume significantly contributed to accurate discrimination between ADHD probands and controls; higher nodal efficiency of right MFG greatly contributed to inattentive and hyperactive/impulsive symptom remission, while higher right MFG-IPL functional connectivity strongly linked to symptom persistence in adults with childhood ADHD. Considering their improved robustness than the commonly implemented basic classifiers, findings suggest that ELTs may have the potential to identify more reliable neurobiological markers for neurodevelopmental disorders.

1. Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a highly prevalent heterogeneous neurodevelopmental disorder. Diagnostic standards for ADHD are clinical symptom-based, and rely primarily on subjective reports collected from multiple sources, which often cause biases and inconsistencies of the diagnoses.

Multimodal neuroimaging techniques have been widely implemented to investigate the neural substrates of ADHD. A number of structural MRI and Diffusion Tensor Imaging (DTI) studies have suggested that gray- and/or white-matter (GM/WM) structural underdevelopment in frontal lobe, thalamus, and striatum significantly contribute to the emergence of ADHD during childhood (Ellison-Wright et al., 2008; Xia et al., 2012). Furthermore, functional aberrations in the fronto-thalamo/fronto-striatal circuitries have also been frequently reported to link with symptom onset in children with ADHD. For instance, altered task-driven or spontaneous neural activity in prefrontal cortex, thalamus, and striatum, and their functional

connectivity, have been found to be significantly associated with increased inattentive and/or hyperactive-impulsive symptoms in children with ADHD (Bush et al., 2005; Cubillo et al., 2012; Durston, 2003; Li et al., 2012; Rubia et al., 1999; Yang et al., 2011). Increasingly, neuroimaging studies have found that more optimal structural/functional development in fronto-subcortical pathways may contribute to symptom reduction and remission of ADHD in adulthood. For instance, a longitudinal study found that persistently decreased GM thickness in dorsolateral prefrontal, middle frontal, and inferior parietal regions, and reduced WM fractional anisotropy (FA) in left uncinate and inferior frontal-occipital fasciculi were associated with a greater number of ADHD symptoms persisting into adulthood (Shaw et al., 2013, 2015). Proal et al. (2011) reported that adults with persistent ADHD had thinner cortical thickness relative to the remitted ADHD in prefrontal region. In addition, greater prefronto–thalamo functional connectivity during a cue-evoked attention task (Clerkin et al., 2013), and greater within-frontal functional connectivity during resting-state (Francx et al., 2015), have been observed in adult ADHD remitters relative to the persisters. However, neuroimaging findings are widely inconsistent, partially due to the sample biases, differences of the implemented imaging and analytic techniques, and the limitations of the traditional parametrical models for group comparisons. Indeed, traditional statistical methods (e.g. t-tests, analysis of variance (ANOVA), correlation, etc.) estimate group differences only within a voxel or region of interest (ROI) at a time without having the capacity to explore how ROIs interact in linear and/or non-linear ways, as they quickly become overburdened when attempting to combine predictors and their interactions from high dimensional imaging data sets (Sun et al., 2009).

Compared to traditional parametrical models, multivariate machine learning techniques are able to leverage high dimensional information simultaneously to understand how variables jointly distinguish between groups (Greenstein et al., 2012). Support vector machine (SVM) is the most frequently applied machine learning classifier in neuroimaging data from children with ADHD, which has been aided by recursive feature elimination (RFE), temporal averaging, principle component analysis (PCA), fast Fourier transform (FFT), independent component analysis (ICA), 10-fold cross-validation (CV), hold-out, and leave-one-out cross-validation (LOOCV) techniques, to distinguish children with ADHD from normal controls (Brown et al., 2012; Chang et al., 2012; Cheng et al., 2012; Colby et al., 2012; Du et al., 2016; Fair et al., 2012; Iannaccone et al., 2015; Johnston et al., 2014; Sen et al., 2018; Yasumura et al., 2017). The commonly reported most important features (according to importance score) that contribute to successful group discrimination included functional connectivity of bilateral thalamus, functional connectivity, surface area, cortical curvature and/or voxel intensity in frontal lobe, cingulate gyrus, temporal lobe, etc. (Brown et al., 2012; Colby et al., 2012; Iannaccone et al., 2015). SVM has also been applied to structural MRI and DTI data collected from adults with ADHD and controls, which reported between-group differences in widespread GM and WM regions in cortices, thalamus, and cerebellum (Chaim-Avancini et al., 2017). Neural network-based techniques, including deep belief network, fully connected cascade artificial neural network, convolutional neural network, extreme learning machine, and hierarchical extreme learning machine, have also been applied to structural MRI and resting-state functional MRI (fMRI) data in children with ADHD and controls (Deshpande et al., 2015; Kuang and He, 2014; Peng et al., 2013; Qureshi et al., 2016; 2017; Zou et al., 2017). The most important group discrimination predictors identified by these neural network studies included functional connectivity within cerebellum, surface area, cortical thickness and/or folding indices of frontal lobe, temporal lobe, occipital lobe and insula (Deshpande et al., 2015; Peng et al., 2013; Qureshi et al., 2017). In addition, principle component-based Fisher discriminative analysis (PC-FDA) (Zhu et al., 2008), Gaussian process classifiers (GPC) (Hart et al., 2014; Lim et al., 2013), and multiple kernel learning (Dai et al., 2012; Ghiassian et al., 2016) have also been used in functional and structural MRI data to discriminate children with ADHD from controls. Supplementary Table 1 provides more details of existing machine learning studies in ADHD. These existing studies have either utilized features representing regional/voxel brain properties collected from only single imaging modality, or the combination of two modalities (mostly structural MRI and resting-state fMRI) (Brown et al., 2012; Fair et al., 2012; Hart et al., 2014; Iannaccone et al., 2015; Johnston et al., 2014), or reported poor accuracy (Dai et al., 2012; Sen et al., 2018; Zou et al., 2017). Some studies did not conduct the necessary step of estimating the most important features that contribute to accurate classifications (Chang et al., 2012; Dai et al., 2012; Kuang and He, 2014; Qureshi et al., 2016; Sen et al., 2018; Tenev et al., 2014; Zou et al., 2017). Systems-level functional and structural features, such as global and regional topological properties from functional brain networks during cognitive processes and WM tract properties have not been considered. In addition, relations between the suggested predictors from imaging features and clinical/behavioral symptoms in samples of ADHD patients, which can provide important clinical context, have not been studied.

Ensemble learning techniques (ELTs), which integrate results from multiple basic classifiers by using voting (Lam and Suen, 1997; Ruta and Gabrys, 2005), bagging (Breiman, 1996), stacking (Wolpert, 1992), or boosting (Johnston et al., 2014; Schapire, 1990a; Yoav Freund and Schapire, 1997) strategies, have been recently developed in the big data science field, to deal with complicated feature variations, biases, and optimized prediction performances (Deng and Platt, 2014; Wang et al., 2011). ELTs have been applied in three recent studies to discriminate patients with ADHD from controls (Eloyan et al., 2012; Tenev et al., 2014; Zhang-James et al., 2019). Eloyan et al. (2012) applied a voting-based ELT, along with hold-out technique for CV, in structural MRI and resting-state fMRI data from children with ADHD and controls, and reported an important group discrimination predictor of dorsomedial-dorsolateral functional connectivity in the motor network. Voting-based ELT has also been applied in electro-encephalogram (EEG) data collected from adults with ADHD and controls, without reporting the most important discrimination predicators (Tenev et al., 2014). Very recently, Zhang-James et al. (2019) applied ELTs in structural MRI data from patients with ADHD (both adults and children) and controls, and suggested that GM volume of bilateral caudate and thalamus and orbitofrontal surface area significantly contribute to successful group discrimination. However, clarifications about optimization strategies was lacking and low accuracy of discriminations (< 0.65) was reported.

The current study applied ELTs to structural MRI, DTI, and task-based fMRI data collected from a sample of adults with childhood ADHD who were clinically followed from ages 7–11 years and never-ADHD controls who have been followed since adolescence. All currently available optimization strategies (i.e., voting, bagging, boosting and stacking techniques) were tested in a pool of semifinal classification results generated by seven basic classifiers (including K-Nearest Neighbors (KNN), SVM, logistic regression (LR), Naïve Bayes (NB), linear discriminant analysis (LDA), random forest (RF), and multilayer perceptron (MLP)). A nested CV including an inner LOOCV and an outer 5-fold CV were applied with grid search to tune the hyperparameters and minimize the overfitting. The high-dimensional neuroimaging features for classification included regional cortical GM thickness and surface area, GM volume of subcortical structures estimated from structural MRI data, volume and FA of major WM fiber tracts derived from DTI data, the pair-wise regional connectivity and global/nodal topological properties (i.e., global-, local-, and nodal-efficiency, etc.) of the cue-evoked attention processing network computed from task-based fMRI data. Based on findings from existing studies (Clerkin et al., 2013; Francx et al., 2015; Luo et al., 2018; Proal et al., 2011; Shaw et al., 2013, 2015), we hypothesized that structural and functional alterations in frontal, parietal, and subcortical areas and their interactions would significantly contribute to accurate discrimination of ADHD probands (adults diagnosed with ADHD in childhood) from controls; while

abnormal fronto-parietal hyper-communications in right hemisphere would play an important role in inattentive and hyperactive/impulsive symptom persistence in adults with childhood ADHD. Finally, we hypothesized that classification performance parameters (accuracy, area under the curve (AUC) of the receiver operating characteristics (ROC), etc.) derived from ELT-based procedures would be superior to those of basic model-based procedures.

2. Materials and methods

2.1. Participants

Seventy-two young adults [mean (SD) age 24.4 (2.1) years] who provided good quality data from multimodal neuroimaging and clinical assessments, participated in this study. There were 36 ADHD probands diagnosed with ADHD combined-type (ADHD-C) in childhood and 36 group-matched comparison subjects with no history of ADHD. Among the 36 ADHD probands, 18 were classified as ADHD remitters, who were endorsed no more than 3 inattentive or 3 hyperactive/impulsive symptoms in adulthood and had no more than 5 symptoms in total. The other 18 probands were classified as ADHD persisters, endorsing at least five inattentive and/or hyperactive/impulsive symptoms in their adulthood and at least 3 symptoms in each domain.

Those with ADHD were recruited when they were 7–11 years-old and subsequently clinically followed. Childhood diagnoses were based on teacher ratings using the IOWA Conners' Teachers Rating Scale (Loney and Milich, 1982) and parent interview using the Diagnostic Interview Schedule for Children version 2 (DISK-2) (Shaffer et al., 1989). Exclusion criteria in childhood were chronic medical illness; neurological disorder; diagnosis of schizophrenia, autism spectrum disorder, or chronic tic disorder; Full Scale IQ < 70; and not speaking English. The never-ADHD comparison group was recruited in adolescence, as part of an adolescent follow-up of the ADHD sample, and history of ADHD was ruled out using the ADHD module of the DISK-2, the IOWA Conners, and the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS) (Kaufman et al., 1997), which was administered to both the parent and adolescent. Adult psychiatric status was assessed using the Structured Clinical Interview for DSM-IV Axis I Disorders (First et al., 2002), supplemented by a semi-structured interview for ADHD that was adapted from the K-SADS and the Conners' Adult ADHD Diagnostic Interview for DSM-IV (Epstein et al., 2006). Raw scores of inattentive and hyperactive/impulsive symptoms from the Conner's Adult Self-Rating Scale were normalized into T scores based on DSM-IV standard, and were used as dimensional measures for inattentive and hyperactive/impulsive behaviors. Exclusion criteria in adulthood were psychotropic medication that could not be discontinued and conditions that would preclude MRI (e.g., metal in body, pregnancy, too obese to fit in scanner). Clinical and demographic information are listed in Table 1.

The study received Institutional Review Board approval at the participating institutions. Participants provided signed informed consent and were reimbursed for their time and travel expenses.

2.2. Multimodal neuroimaging data acquisition protocol

Multimodal neuroimaging data of each participant were collected using the same 3.0T Siemens Allegra (Siemens, Erlangen, Germany) whole body MRI scanner. High resolution 3-dimensional T1-weighted structural MRI data was acquired using magnetization prepared rapid gradient echo pulse sequence with TR = 2050 ms, TE = 4.38 ms, inversion time (TI) = 1.1 s, flip angle = 8°, field of view (FOV) = 256 mm × 256 mm × 256 mm, voxel size = 0.94 mm × 0.94 mm × 1 mm. DTI data were collected using an echo planar imaging (EPI) pulse sequence with a b-value = 1250 s/mm^2 along 12 independent non-collinear orientations, as well as one reference volume without diffusion-weighting $b = 0$ s/mm^2

(TR = 5200 ms, TE = 80 ms, flip angle = 90°, FOV = 128 mm × 128 mm, voxel size = 1.875 mm × 1.875 mm × 4 mm, matrix = 128 × 96, number of slices = 63). FMRI data were acquired using a gradient-echo EPI sequence with TR = 2500 ms, TE = 27 ms, flip angle = 82°, matrix = 64 × 64, slice thickness = 4 mm, 40 slices, in-plane resolution = 3.75 mm^2. Images were acquired with slices positioned parallel to the anterior commissure-posterior commissure line. The total duration of fMRI data acquisition was 20 min, which contained 4 runs of a cued attention task (CAT) with stimuli counter-balanced.

2.3. The cued attention task (CAT) for fMRI

The CAT was developed and described in detail in (Clerkin et al., 2013, 2009; Luo et al., 2018). Briefly, it began and ended with a 30-second fixation period. The task contained a series of 120 letters, including 24 targets ("X"), 12 cues ("A"), and 84 non-cue letters ("B" through "H"). For the 24 targets, half of them were preceded by a cue, and the other half were preceded by a non-cue letter. Participants were told that a cue ("A") was always followed by the target letter ("X"), but not all targets were preceded by a cue letter. The letters were presented individually for 200 ms with a pseudorandom inter-stimulus interval which ranged from 1550 to 2050 ms (mean = 1800 ms/run). Participants were instructed to respond to each target as rapidly as possible using their right index finger. Before entering the scanner, detailed instructions and practice trials of the task were provided to each participant to ensure satisfactory performance.

2.4. Multimodal imaging data processing for feature extractions

For each subject, the T1-weighted data was reconstructed into a 3-dimensional cortical model for thickness and area estimations using FreeSurfer v.5.3.0 (https://surfer.nmr.mgh.harvard.edu). Each volume was first registered to the Talairach atlas. Intensity variations caused by magnetic field in homogeneities were corrected and non-brain tissue was removed. A cutting plane was used to separate the left and right hemispheres and to remove the cerebellum and brainstem. Two mess surfaces (mess of grids created using surface tessellation technique) were generated between GM and WM (WM surface), as well as between GM and cerebrospinal fluid (pial surface). The distance between the two closest vertices of the WM and pial surfaces represented the cortical thickness at that specific location. Cortical subregions were parcellated based on the Desikan atlas. A total of 202 structural MRI features, including regional cortical GM thickness, surface area, and GM volume of subcortical structures were extracted from each subject.

The DTI data was corrected for eddy current-induced distortions due to the changing gradient field directions. Head motion was corrected with non-diffusion-weighted reference image (b0 image) using an affine, 12 degrees of freedom registration. Then the FA value and principle diffusion direction at each brain voxel were calculated. WM probabilistic tractography between each pair of 18 ROIs (bilateral thalami, putamen and caudate nuclei from striatum, hippocampus, and frontal, parietal, occipital, temporal, and insular cortices based on the Harvard-Oxford Cortical Atlases and Julich Histological Atlas) were constructed using the FSL/BEDPOSTX tool. The multi-fiber probabilistic connectivity-based method was applied to determine the number of pathways between the seed and each of the target clusters, with the default setting of parameters for the Markov Chain Monte Carlo estimation of the probabilistic tractography. At the end, a total of 120 DTI-based features, including the volume and FA of cortico-cortical and subcortico-cortical WM fiber tracts were extracted for each subject.

The fMRI data from each participant was preprocessed using Statistical Parametric Mapping version 8 (SPM8, Wellcome Trust center for Neuroimaging, London, United Kingdom; http://www.fil.ion.ucl.ac.uk/spm/) implemented on a MATLAB platform. The preprocessing procedures included slice timing correction, realignment, co-registration, segmentation, normalization, and spatial smoothing. The first-

Table 1
Demographic and clinical characteristics in groups of controls and ADHD probands (and further in the sub-groups of remitters and persisters of the ADHD probands).

	Controls ($N = 36$) Mean (SD)	ADHD ($N = 36$) Mean (SD)	p	Remitted ($N = 18$) Mean (SD)	Persistent ($N = 18$) Mean (SD)	p
Age	24.3 (2.3)	24.66 (2.0)	0.48	24.79 (2.2)	24.52 (2.0)	0.7
Full-scale IQ	103.83(15.4)	97.96 (14.1)	0.1	99.22 (14.9)	96.71 (13.6)	0.6
Conners' Adult ADHD Rating Scale (T score)						
Inattentive	45.75 (8.8)	56.5 (13.2)	<0.001	49.83 (10.9)	63.17 (12.0)	0.001
Hyperactive/impulsive	42.97 (6.2)	53.64 (12.9)	<0.001	46.17 (9.0)	61.11 (12.0)	<0.001
ADHD Total	43.89 (8.2)	56.5 (14.7)	<0.001	42.61 (7.5)	54.33 (8.8)	<0.001
ADHD semistructured interview (number of symptoms)	0.79 (1.6)	6.17 (5.2)	<0.001	2.64 (2.0)	10.24 (3.6)	<0.01
	N (%)	N (%)	p	N (%)	N (%)	p
Male	31 (86.1)	30 (83.3)	0.74	16 (88.9)	14 (77.8)	0.37
Right-handed	32 (88.9)	32 (88.9)	1	15 (83.3)	16 (88.9)	0.63
Race			0.17			0.59
Caucasian	15 (41.7)	21 (58.3)		9 (50.0)	12 (66.7)	
African American	13 (36.1)	7 (19.4)		4 (22.2)	3 (16.7)	
More than one race	6 (16.7)	8 (22.2)		5 (27.8)	3 (16.7)	
Asian	2 (5.6)	0 (0.0)		0 (0.0)	0 (0.0)	
Ethnicity			0.09			0.74
Hispanic/Latino	10 (27.8)	17 (47.2)		8 (44.4)	9 (50.0)	
Task performance measures	Mean (SD)	Mean (SD)	p	Mean (SD)	Mean (SD)	p
Reaction time average	395.8 (53.1)	422.8 (74.3)	0.08	431.1 (67.0)	439.1 (107.8)	0.79
Reaction time std	129.6 (24.8)	137.2 (29.9)	0.25	136.2 (27.6)	138.2 (32.8)	0.84
Anticipation error	1.86 (2.1)	1.74 (1.6)	0.78	1.69 (1.6)	1.78 (1.7)	0.88
Commission error	0.33 (0.8)	0.85 (1.4)	0.07	0.75 (1.6)	0.94 (1.3)	0.7
Omission error	4.97 (5.8)	8 (10.8)	0.15	4.38 (4.0)	11.22 (13.8)	0.06

level analyses were conducted using general linear model (GLM) to generate the activation map responding to the cues. The group average activation maps for ADHD probands and controls were generated, respectively. A total of 52 cortical and subcortical seed regions, which was parceled according to the structural and functional connectivity-based Brainnetome atlas, were determined based on the results of the combination of the functional activation maps of the groups of ADHD probands and controls (Fan et al., 2016). To construct the cue-evoked attention processing network, the single-trial beta value series from the 48 cue-related events in the four runs were extracted. Among all the voxels in each of the 52 node ROIs, the average beta value series was calculated and used to create a 52 × 52 pair-wise Pearson correlation matrix. Then the graph theoretic techniques (GTTs) were carried out. More details of the fMRI data processing can be found in (Luo et al., 2018). A total of 200 fMRI features, including the global- and local-efficiency of the entire network, the nodal efficiency, degree, and be-tweenness-centrality measures of the 52 nodes, as well as their pair-wise functional connectivities, were generated for each subject.

2.5. Modeling of ensemble learning architecture (ELT)

Modeling of the ELTs for classifications between ADHD probands and controls, as well as between ADHD persisters and remitters respectively, is described in Fig. 1. Specifically, Part A of Fig. 1 presents feature selection and preparation flow. In order to decrease the risk of overfitting, two-sample t-tests were applied and a total of 20 neuroimaging features that showed the largest between group differences were first selected from the 522 multimodal neuroimaging features derived from structural MRI, DTI, and fMRI data. Then each value of the 20 selected features was normalized by using a z-score transformation in the feature-specific space. The normalized 20 top-ranked neuroimaging features were then entered to the training and validation procedures (Part B of Fig. 1), which consisted of a nested CV (there were two CV loops, including an outer 5-fold CV loop to split the data into training set and validation set, and an inner loop to tune the hyperparameters for 7 basic models and 4 ELTs-based models using grid search in combination with LOOCV). More specifically, the 20 neuroimaging features were split into a total of 5 stratified folds such that each fold consisted of balanced 20% of the entire data. The five-fold CV was performed by using these 5 stratified folds, where each trial

dedicated four folds for training data and the remaining one for validation. Then for each iteration in 5-fold CV, the corresponding training set was sent into the LOOCV processing. In each iteration, one subject was extracted from the training set to act as a validation data, and the remaining subjects were trained to construct the models. According to the classification performance of the validation data, the hyperparameters for each model were tuned and the optimal hyperparameters setup were selected using grid search. More details of the hyperparameters are described in Table 2. We utilized the LOOCV to tune the hyperparameters of 7 basic models, including KNN, SVM, LR, NB, LDA, RF, and MLP. Based on the hyperparameters of basic models, we applied LOOCV to tune the hyperparameters of 4 ELTs-based models, including max Voting, Bagging, AdaBoost, Stacking. As shown in Part C of Fig. 1, during iterations of 5-fold CV outer loop, the performance of each basic and ELTs-based models with the optimal hyperparameters derived from LOOCV inner loop iterations was evaluated. The group average of classification performance of each classifier derived from each iteration of 5-fold CV was generated. The 7 basic and 4 ELTs-based models according to the group average value of AUC of the ROC from iterations of 5-fold CV outer loop. The basic and ELTs-based models with the highest average AUC were selected as optimal classifiers. Based on the types of ELT-based models we evaluated and selected, the importance score corresponding each feature was then calculated using the ELT-based model and the corresponding basic models.

We also applied unsupervised learning (i.e. the hierarchical clustering) in our dataset. The hyperparameters, including the metric used to compute the linkage (affinity), the linkage criterion used to determine which distance between sets of observation (linkage) were also tuned by using grid search. Then the model with best classification performance (i.e. accuracy) was selected. All the procedures were conducted by in-house codes developed in Python 3.7.

2.6. Regression models

Following the classification procedures, we constructed the regression models to identify the relations between the neuroimaging features and the clinical inattentive and hyperactive/impulsive symptom T-scores. Based on the ELT-based classification results, the top three neuroimaging features were selected based on the weight of each feature in the optimal discriminators between ADHD and normal controls,
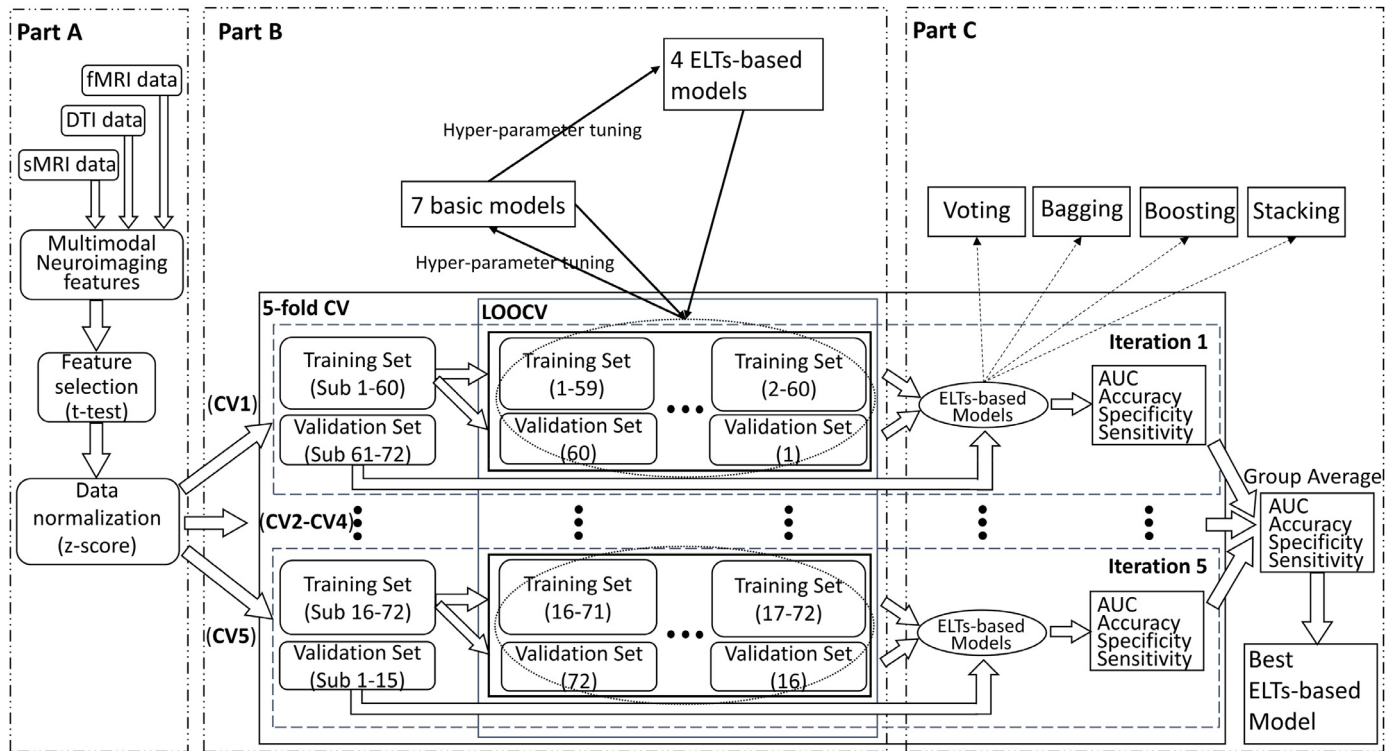
Fig. 1. The ensemble learning flowchart. (sMRI: structural MRI; DTI: diffusion tensor imaging; fMRI: functional MRI; CV: cross-validation; LOOCV: leave-one-out cross-validation; AUC: the area under the receiver operating characteristic curve; ELTs: ensemble learning techniques).

as well as between ADHD persisters and remitters. Then, we applied Ordinary Least Squares (OLS) (Hutcheson, 1999), Ridge regression (Hoerl and Kennard, 1970), least absolute shrinkage and selection operator (LASSO) regression (Santosa and Symes, 1986; Tibshirani, 1996), Elastic Net regression (Zou and Hastie, 2005) to construct the prediction models for inattentive and hyperactive/impulsive T-scores, respectively. The same nested CV utilized in previous steps were also conducted in regression model construction. The hyperparameters included the regularization strength (alpha), solver to use in the computational routines (solver) for Ridge regression, the constant that multiplies the L1 term (alpha) for LASSO regression, the constant that multiplies the penalty terms (alpha), the Elastic Net mixing parameter (l1_ratio) for Elastic Net regression. During the iteration of 5-fold CV outer loop, the performance of each regression model with the optimal hyperparameters derived from LOOCV inner loop iterations was evaluated. The group average of regression performance, including the Pearson correlation coefficient and mean squared error (MSE) between predicted and observed values, of each regression model derived from each iteration of 5-fold CV were calculated. Again, all the regression

analyses were conducted by in-house codes developed in Python 3.7.

2.7. Evaluation measures

The performance of each classification procedure classifier was measured in terms of classification accuracy, sensitivity, and specificity. The accuracy of a machine learning classification algorithm is to measure how often the algorithm classifies a data point correctly. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Sensitivity describes the proportion of actual positive cases that are correctly identified as positive. It implies that there will be another proportion of actual positive cases, which would get predicted incorrectly as negative. The sensitivity is defined as:

Table 2
The hyperparameters of 7 basic models and 4 ELTs-based models. (ELTs: ensemble learning techniques; KNN: k-nearest neighbors; SVM: support vector machine; LR: logistic regression; RF: random forest; LDA: linear discriminant analysis; MLP: multilayer perceptron).

Classifiers	Hyperparameters
KNN	n_neighbors: [1, 3, 5, 7, 9]; algorithm: ['auto', 'ball_tree', 'kd_tree', 'brute']; p: [1, 2, 3]
SVM	C: [0.001, 0.01, 0.1, 1, 10, 100, 1000]; gamma: ['auto', 'scale']; kernel: ['linear', 'rbf', 'poly', sigmoid]
LR	solver: ['newton-cg', 'lbfgs', 'sag', 'saga']; multi_class: ['ovr', 'multinomial', 'auto']
RF	n_estimators: list(range(3, 60, 5)); criterion: ['gini', 'entropy']; min_samples_leaf: [3, 5, 10]; max_depth: [3, 4, 5, 6]; min_samples_split: [3, 5, 10]; bootstrap: [True, False]
LDA	solver: ['svd', 'lsqr', 'eigen']
MLP	activation: ['identity', 'logistic', 'tanh', 'relu']; solver: ['lbfgs', 'sgd', 'adam']; hidden_layer_sizes: np.arange(1, 72, 10); max_iter: [4000]
ELT-Voting	estimators; voting: ['hard', 'soft']
ELT-Bagging	base_estimator; n_estimators: list(range(10, 150, 10)); max_samples = [0.2, 0.3, 0.4, 0.5]; max_features = [0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
ELT-Boosting	base_estimator; n_estimators: list(range(10, 150, 10)); learning_rate: list(range(0.01, 1, 0.01))
ELT-Stacking	classifiers; meta_classifiers

$$Sensitivity \ (Recall) = \frac{TP}{TP + FN}$$

Specificity is a measure of the proportion of actual negatives, which got predicted as the negative. It implies that there will be another proportion of actual negative, which got predicted as positive and could be termed as false positives. It is defined as:

$$Specificity = \frac{TN}{TN + FP}$$

In addition, a ROC curve was plotted to illustrate the diagnostic ability of a binary classifier system as its discrimination threshold is varied. In the classification case, we calculated the confusion matrix for each iteration cycle of the classifier and calculated the AUC of ROC. AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example. The AUC of ROC is defined as:

$$AUC = 2 \ \times \frac{Precision \times Recall}{Precision + Recall}$$

Among the equation of AUC, Precision and Recall are defined, respectively, as:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

For the regression model, the Pearson correlation coefficient and MSE between predicted values and actual values were calculated. The Pearson correlation coefficient is a measure of the linear correlation between two variables. It is defined as:

$$\rho_{X,Y} = \frac{cov(X, \ Y)}{\sigma_X \sigma_Y}$$

where cov is the covariance, σ_X is the standard deviation of X, σ_Y is the standard deviation of Y.

The MSE of an estimator measures the average squared difference between the estimated values and the actual value, which is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y_i})^2$$

Where Y_i and $\hat{Y_i}$ represent the actual and predicted value.

3. Results

3.1. Demographic, clinical and behavioral measures

The demographic, clinical information and behavioral performance of all groups are summarized in Table 1. There were no significant demographic between-group differences. Moreover, all participants achieved $a > 85\%$ rate for response accuracy when performing the fMRI task. Task performance measures, including reaction time, response accuracy rate, omission error rate, commission error rate did not show between-group differences ($p > 0.05$).

3.2. Classification model performance

The Table 3 (Part I) summarizes the ADHD probands vs. controls classification performances of the basic models and ELTs. Additional details of ADHD probands vs. controls classification performance of ELTs are shown in Supplementary Table 2 (Part I). The classifier of SVM performed the best among the seven basic models regarding the AUC, accuracy, and specificity (AUC = 0.87, accuracy = 0.816, specificity = 0.942). Furthermore, the bagging-based ELT with SVM as the basic model performed the best among all ensemble models

Table 3

The results of 7 basic and 4 ELTs-based classifications between the groups of ADHD and normal controls (Part I) as well as between the groups of ADHD persisters and ADHD remitters (Part II). (ELT: ensemble learning technique; KNN: k-nearest neighbors; SVM: support vector machine; LR: logistic regression; RF: random forest; LDA: linear discriminant analysis; MLP: multilayer perceptron; AUC: the area under the receiver operating characteristic curve; ADHD: attention deficit/hyperactivity disorder; NC: normal controls; ADHD-R: ADHD remitters; ADHD-P: ADHD persisters).

Classifiers	Specificity	Sensitivity	Accuracy	AUC
Part I: ADHD vs. NC				
KNN	0.72	0.66	0.689	0.69
SVM	0.942	0.69	0.816	0.87
LR	0.756	0.742	0.75	0.85
NB	0.778	0.718	0.748	0.86
RF	0.866	0.75	0.705	0.82
LDA	0.734	0.774	0.754	0.78
MLP	0.782	0.746	0.764	0.84
ELT-Voting	0.808	0.718	0.763	0.87
ELT-Bagging	0.734	0.798	0.766	0.89
ELT-Boosting	0.67	0.77	0.721	0.88
ELT-Stacking	0.756	0.742	0.75	0.82
Part II: ADHD-P vs. ADHD-R				
KNN	0.4	0.934	0.667	0.72
SVM	0.65	0.75	0.7	0.85
LR	0.6	0.682	0.642	0.85
NB	0.734	0.65	0.692	0.77
RF	0.734	0.6	0.667	0.76
LDA	0.568	0.518	0.542	0.63
MLP	0.634	0.75	0.692	0.84
ELT-Voting	0.8	0.65	0.725	0.82
ELT-Bagging	0.75	0.582	0.67	0.90
ELT-Boosting	0.75	0.682	0.717	0.86
ELT-Stacking	0.884	0.684	0.783	0.82

(AUC = 0.89). Table 3 (Part II) summarizes ADHD persisters vs. remitters classification performances of the basic models and ELTs, and again demonstrated that SVM performed the best among all the basic models regarding the AUC and accuracy (AUC = 0.85, accuracy = 0.7), while the bagging-based ELT with SVM as the basic model performed the best among all ensemble models (AUC = 0.9). Supplementary Table 2 (Part II) provided more details of ADHD persisters vs. remitters classification performance of ELTs.

3.3. ROC curves of classification models

The ROC curve for each classification procedure, including the unsupervised hierarchical clustering is plotted in Fig. 2. Results showed that classification performance parameters of the ELTs-based procedures were greatly improved compared to those of the basic model-based procedures. In addition, relative to the performance improvement between ensemble learning and basic models of the classification between ADHD and normal controls, the performance improvement of classification between ADHD persisters and remitter is greater.

4. Importance score of the classification model

The importance score of top three features for the classifications between ADHD probands and normal controls, and between ADHD persisters and remitters are shown in Table 4. More specifically, the nodal efficiency of right inferior frontal gyrus (IFG), the functional connectivity between right middle frontal gyrus (MFG) and right inferior parietal lobule (IPL), the volume of right amygdala served as the top three important features in the classification model between ADHD and normal controls. The nodal efficiency of right MFG, functional connectivity between right MFG and right IPL, and betweenness-centrality of left putamen played the three most important characteristics in the classification between ADHD persisters and remitters.
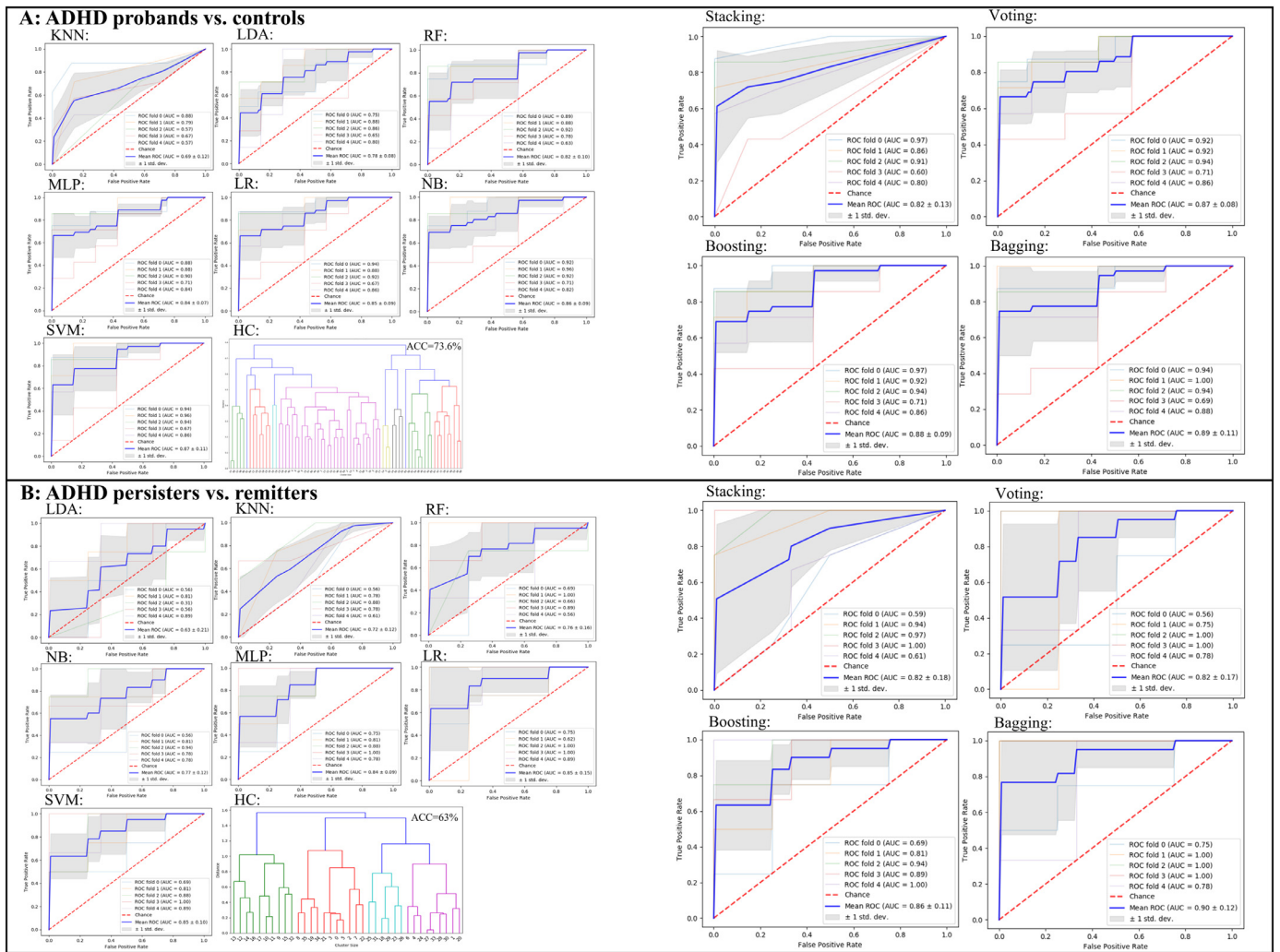
Fig. 2. The AUC of each classification procedure for discrimination between ADHD probands and normal controls (A), and between ADHD persisters and remitters (B). (KNN: k-nearest neighbors; SVM: support vector machine; LR: logistic regression; RF: random forest; LDA: linear discriminant analysis; MLP: multilayer perceptron; HC: hierarchical clustering; ROC: the receiver operating characteristic curve; AUC: the area under the ROC curve).

Table 4

The importance scores of top three features of classifications between ADHD probands and normal controls, as well as between ADHD persisters and ADHD remitters. (FC: functional connectivity; NC: normal controls; ADHD: attention deficit/hyperactivity disorder; ADHD-P: ADHD persisters; ADHD-R: ADHD remitters).

Feature	Importance Score
ADHD vs. NC	
Nodal efficiency of right Inferior Frontal gyrus	0.134
FC between right Middle Frontal gyrus and right Inferior Parietal lobule	0.111
Volume of right amygdala	0.1
ADHD-P vs. ADHD-R	
Nodal efficiency of right Middle Frontal gyrus	1.028
FC between right Middle Frontal gyrus and right Inferior Parietal lobule	0.852
Betweenness-centrality of left putamen	0.677

Table 5

Pearson correlation coefficient and mean squared error performance of regression models.

Regression	Pearson Correlation Coefficient	MSE
T-Inattentive		
OLS	$r = 0.4603; p < 0.001$	126.3
LASSO	$r = 0.4592; p < 0.001$	124.6
Ridge	$r = 0.4605; p < 0.001$	126.1
Elastic Net	$r = 0.4689; p < 0.001$	121.1
T-Hyperactive/Impulsive		
OLS	$r = 0.3329; p = 0.0043$	126.5
LASSO	$r = 0.3395; p = 0.0035$	123.3
Ridge	$r = 0.3334; p = 0.0042$	126.3
Elastic Net	$r = 0.3488; p = 0.0027$	119.8

4.1. Regression model and importance score

The regression results (Table 5) indicated that Elastic Net regression performed the best for the prediction of both inattentive and hyperactive/impulsive T-scores. Table 6 shows the importance scores of the top three features of Elastic Net regression for inattentive and hyperactive/impulsive symptom T-scores. Specifically, the top three features for the prediction of inattentive T-score were the nodal efficiency of right IFG, the functional connectivity between MFG and IPL in right hemisphere, the volume of right amygdala. The top three features for the prediction of hyperactive/impulsive T-score included the nodal efficiency in right IFG, the functional connectivity between right MFG and right IPL, the nodal efficiency of right MFG.

(OLS: ordinary least square; LASSO: least absolute shrinkage and selection operator; T-Inattentive: inattentive T-score; T-Hyperactive/

Table 6

The importance scores of top three features of Elastic Net regression for inattentive and hyperactive/impulsive symptom T-scores. (FC: functional connectivity).

Feature	r	p	Importance Score
Inattentive			
Nodal efficiency of right Inferior Frontal gyrus	− 0.399	0.001	3.471
FC between right Middle Frontal gyrus and right Inferior Parietal lobule	0.405	< 0.001	2.126
Volume of right Amygdala	− 0.011	0.928	1.819
Hyperactive/Impulsive			
Nodal efficiency of right Inferior Frontal gyrus	− 0.345	0.003	2.289
FC between right Middle Frontal gyrus and right Inferior Parietal lobule	0.361	0.002	2.134
Nodal efficiency of right Middle Frontal gyrus	− 0.333	0.004	1.997

Impulsive: Hyperactive/Impulsive T-score; MSE: mean squared error)

5. Discussion

To the best of our knowledge, this is the first study to apply ELT to multimodal neuroimaging features generated from structural MRI, DTI, and task-based fMRI data collected from a sample of adults with childhood ADHD and controls, who have been clinically followed up since childhood. We found that the nodal efficiency in right IFG, functional connectivity between MFG and IPL in right hemisphere, and right amygdala volume were the most important features for discrimination between the ADHD probands and controls, while the nodal efficiency of right MFG, functional connectivity between right MFG and right IPL, and betweenness-centrality of left putamen played the most important roles for discrimination between the ADHD persisters and remitters. Moreover, the classification performance parameters of ELT-based procedures were superior to those of the basic classifiers.

5.1. Neurobiological markers for discriminations

5.1.1. Classification between ADHD probands and controls

Our current study observed the important roles of nodal efficiency in right IFG and functional connectivity between right MFG and right IPL for discrimination between ADHD probands and normal controls. The abnormalities of these regions have been supported by a variety of existing neuroimaging and machine learning studies. Specifically, both task-based and resting-state fMRI studies have consistently reported the decreased functional activation in right IFG (Cao et al., 2006; Konrad et al., 2006; Rubia et al., 2019, 1999; Silk et al., 2005; Smith et al., 2006) and reduced functional connectivity between right MFG and right IPL (Lin et al., 2015; Vance et al., 2007) in children with ADHD as compared with normal controls. In addition, multivariate machine learning and ELT-based studies have commonly reported that functional activation and connectivity in frontal and parietal areas are associated with improved classification between children with ADHD and normal controls (Brown et al., 2012; Colby et al., 2012; Deshpande et al., 2015; dos Santos Siqueira et al., 2014; Fair et al., 2012; Iannaccone et al., 2015; Qureshi et al., 2017). They have supported the hypothesis that functional abnormalities in frontal and parietal areas, which are critical components of the attention network in human brain, especially stimulus-driven top-down control, are associated with the symptom emergence of childhood ADHD (Posner and Rothbart, 2009). Additionally, we found that the volume of right amygdala played a vital role in discrimination of ADHD probands and controls. The findings of amygdala anatomical abnormities in children with ADHD were supported by many previous studies. Amygdala plays as a critically important role in emotion regulation (Banks et al., 2007; Davidson et al., 2000; Domes et al., 2010) and thus structural anomalies associated amygdala have been widely observed in children (Van Dessel et al., 2019) and adults with ADHD (Tajima-Pozo et al., 2018), which suggests that the aberrant structure of amygdala may be associated with less control of impulsivity and delay aversion (Van Dessel et al., 2018).

5.1.2. Classification between ADHD persisters and remitters

Additionally, our findings point to the important roles of nodal efficiency in right MFG, functional connectivity between right MFG and right IPL for discrimination between ADHD persisters and remitters, and findings were supported by a variety of existing neuroimaging studies. More specifically, reduced activation and functional connectivity in IFG, MFG, and fronto-parietal regions were observed in ADHD persisters when compared to ADHD remitters (Clerkin et al., 2013; Mattfeld et al., 2014; Schulz et al., 2017). The functional activation and connectivity in frontal and parietal regions during cognitive control were associated with the diverse adult outcomes of ADHD diagnosed in childhood, with symptom persistence linked to reduced activation or symptom recovery associated with higher connectivity within frontal areas (Clerkin et al., 2013; Francx et al., 2015; Mattfeld et al., 2014; Schulz et al., 2017). Although several existing multivariate machine learning and ELT-based studies have commonly reported that the anatomical features in frontal and parietal areas are associated with the classification performance between adults with ADHD and group-matched normal controls (Chaim-Avancini et al., 2017; Zhang-James et al., 2019), no machine learning study has been conducted to identify the classification pattern for discrimination between ADHD persisters and remitters. We further observed that the features of nodal efficiency in right IFG, functional connectivity between right MFG and right IPL, and right amygdala volume were associated with inattentive symptom severity T-score, while the nodal efficiencies of right IFG and MFG and functional connectivity between MFG and IPL in right hemisphere were associated with hyperactive/impulsive symptom severity T-score. These findings suggest the significant involvement of frontal and parietal lobes in right hemisphere for both inattentive and hyperactive symptom persistence of childhood ADHD (Francx et al., 2015).

5.2. Performance of classification and regression models

5.2.1. Ensemble learning (ELT)

Moreover, we found that the classification performance parameters of ELT-based procedures were improved compared to those of basic models. The ELTs have been developed in the big data science field to adaptively combine multiple basic classifiers in order to strategically deal with feature variance and bias, and optimize prediction performances (Balakrishnan et al., 2012; Dror et al., 2011; Hansen and Salomon, 1990; Schapire, 1990b). According to our classification results, bagging, sampling with replacement, would help to reduce the chance overfitting complex models. In our study, bagging with the basic model of SVC was applied to train our model and proved to be the best classifier for both discriminations. In addition, we used AUC statistic for model evaluation, instead on commonly used accuracy, which can be influenced by case-control imbalance in data sets (Fawcett, 2006; Hanley and McNeil, 1982). Our study showed a satisfactory performance of AUC with 0.89 and 0.9 for the discrimination between groups of ADHD and normal controls, and between the groups of ADHD persisters and remitters, respectively. Although we had a relatively small sample size, our findings suggest that ELT-based models

performed better than basic models.

5.2.2. Elastic Net regression

In addition, the Elastic Net-based regression model demonstrated the best performance parameters when investigating the relations between the neuroimaging features and clinical symptom measures in the ADHD probands. The reason Elastic Net regression had the best performance was that it compromised the LASSO penalty (L1) and the ridge penalty (L2) (Zou and Hastie, 2005). The LASSO (L1) penalty function performs variable selection and dimension reduction by shrinking coefficients (Tibshirani, 1996), while the ridge (L2) penalty function shrinks the coefficients of correlated variables toward their average (Hoerl and Kennard, 1970). Therefore, as for the study with relatively small sample size, the combined method obviously performed better than isolated ones, e.g. LASSO regression and ridge regression.

5.3. Limitations

Although the ELTs improved the model classification performance, especially for the cases when the base models had weak classification results, the current study has some limitations. First, our cohort consisted of both male and female subjects, but many more males. It is still unclear whether the discrimination models of ADHD differ between males and females. The future work may focus on constructing the classification models for both males and females. Second, the sample size of this study is relatively small. It is noted that among the existing studies in adults with ADHD, the variability of clinical characteristics inside the group/subgroups of the current study is relatively small (Engelhardt et al., 2019; Groom et al., 2015; Harrison et al., 2007; Shaw et al., 2013; Solanto et al., 2010). Nevertheless, although our study provided a considerable robust algorithm to reduce potential overfitting issues that can happen in studies with small sample sizes, future work in a much larger cohort is still expected to further test the performance of the ELTs.

6. Conclusions

In summary, together with existing findings, results of this study suggest that structural and functional alterations in frontal, parietal, and amygdala areas and their functional interactions significantly contribute to accurate discrimination of ADHD probands from controls; while abnormal fronto-parietal functional communications in the right hemisphere plays an important role in symptom persistence in adults with childhood ADHD. Furthermore, the classification performance parameters (accuracy, AUC of the ROC, etc.) of the ELT-based procedures were improved than those of basic model-based procedures, which suggests that ELTs may have the potential to identify more reliable neurobiological markers for neurodevelopmental disorders.

CRediT authorship contribution statement

Yuyang Luo: Methodology, Software, Validation, Writing - original draft, Writing - review & editing. **Tara L. Alvarez:** Writing - review & editing. **Jeffrey M. Halperin:** Writing - review & editing. **Xiaobo Li:** Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

Acknowledgements

Supplementary materials

References

Balakrishnan, S., Wang, R., Scheidegger, C., MacLellan, A., Hu, Y., Archer, A., Krishnan, S., Applegate, D., Ma, G.Q., Au, S.T., 2012. Combining predictors for recommending music: the false positives' approach to KDD cup track 2. In: Gideon, D., Yehuda, K., Markus, W. (Eds.), Proceedings of KDD Cup 2011. PMLR, Proceedings of Machine Learning Research, pp. 199–213.

Banks, S.J., Eddy, K.T., Angstadt, M., Nathan, P.J., Phan, K.L., 2007. Amygdala-frontal connectivity during emotion regulation. Soc. Cogn. Affect Neurosci. 2, 303–312.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 18.

Brown, M.R., Sidhu, G.S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P.H., Greenshaw, A.J., Dursun, S.M., 2012. ADHD-200 global competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. Front. Syst. Neurosci. 6, 69.

Bush, G., Valera, E.M., Seidman, L.J., 2005. Functional neuroimaging of attention-deficit/ hyperactivity disorder: a review and suggested future directions. Biol. Psychiatry 57, 1273–1284.

Cao, Q., Zang, Y., Sun, L., Sui, M., Long, X., Zou, Q., Wang, Y., 2006. Abnormal neural activity in children with attention deficit hyperactivity disorder: a resting-state functional magnetic resonance imaging study. NeuroreportNeuroreport 17, 1033–1036.

Chaim-Avancini, T.M., Doshi, J., Zanetti, M.V., Erus, G., Silva, M.A., Duran, F.L.S., Cavallet, M., Serpa, M.H., Caetano, S.C., Louza, M.R., Davatzikos, C., Busatto, G.F., 2017. Neurobiological support to the diagnosis of ADHD in stimulant-naive adults: pattern recognition analyses of MRI data. Acta Psychiatr. Scand. 136, 623–636.

Chang, C.W., Ho, C.C., Chen, J.H., 2012. ADHD classification by a texture analysis of anatomical brain MRI data. Front. Syst. Neurosci. 6, 66.

Cheng, W., Ji, X., Zhang, J., Feng, J., 2012. Individual classification of ADHD patients by integrating multiscale neuroimaging markers and advanced pattern recognition techniques. Front. Syst. Neurosci. 6, 58.

Clerkin, S.M., Schulz, K.P., Berwid, O.G., Fan, J., Newcorn, J.H., Tang, C.Y., Halperin, J.M., 2013. Thalamo-cortical activation and connectivity during response preparation in adults with persistent and remitted ADHD. Am. J. Psychiatry 170, 1011–1019.

Clerkin, S.M., Schulz, K.P., Halperin, J.M., Newcorn, J.H., Ivanov, I., Tang, C.Y., Fan, J., 2009. Guanfacine potentiates the activation of prefrontal cortex evoked by warning signals. Biol. Psychiatry 66, 307–312.

Colby, J.B., Rudie, J.D., Brown, J.A., Douglas, P.K., Cohen, M.S., Shehzad, Z., 2012. Insights into multimodal imaging classification of ADHD. Front. Syst. Neurosci. 6, 59.

Cubillo, A., Halari, R., Smith, A., Taylor, E., Rubia, K., 2012. A review of fronto-striatal and fronto-cortical brain abnormalities in children and adults with attention deficit hyperactivity disorder (ADHD) and new evidence for dysfunction in adults with ADHD during motivation and attention. Cortex 48, 194–215.

Dai, D., Wang, J., Hua, J., He, H., 2012. Classification of ADHD children through multimodal magnetic resonance imaging. Front. Syst. Neurosci. 6, 63.

Davidson, R.J., Putnam, K.M., Larson, C.L., 2000. Dysfunction in the neural circuitry of emotion regulation–a possible prelude to violence. ScienceScience 289, 591–594.

Deng, L., Platt, J., 2014. Ensemble deep learning for speech recognition. In: Proc. Interspeech..

Deshpande, G., Wang, P., Rangaprakash, D., Wilamowski, B., 2015. Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. IEEE Trans. Cybern. 45, 2668–2679.

Domes, G., Schulze, L., Bottger, M., Grossmann, A., Hauenstein, K., Wirtz, P.H., Heinrichs, M., Herpertz, S.C., 2010. The neural correlates of sex differences in emotional reactivity and emotion regulation. Hum. Brain Mapp. 31, 758–769.

dos Santos Siqueira, A., Biazoli Junior, C.E., Comfort, W.E., Rohde, L.A., Sato, J.R., 2014. Abnormal functional resting-state networks in ADHD: graph theory and pattern recognition analysis of fMRI data. Biomed. Res. Int. 2014, 380531.

Dror, G., Koenigstein, N., Koren, Y., Weimer, M., 2011. The Yahoo! music dataset and KDD-Cup'11. In: Proceedings of the 2011 International Conference on KDD Cup 2011. 18. pp. 3–18 JMLR.org.

Du, J., Wang, L., Jie, B., Zhang, D., 2016. Network-based classification of ADHD patients using discriminative subnetwork selection and graph kernel PCA. Comput. Med. Imaging Graph. 52, 82–88.

Durston, S., 2003. A review of the biological bases of ADHD: what have we learned from imaging studies? Ment. Retard. Dev. Disabil. Res. Rev. 9, 184–195.

Ellison-Wright, I., Ellison-Wright, Z., Bullmore, E., 2008. Structural brain change in attention deficit hyperactivity disorder identified by meta-analysis. BMC Psychiatry 8, 51.

Eloyan, A., Muschelli, J., Nebel, M.B., Liu, H., Han, F., Zhao, T., Barber, A.D., Joel, S., Pekar, J.J., Mostofsky, S.H., Caffo, B., 2012. Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. Front. Syst. Neurosci. 6, 61.

Engelhardt, P.E., Nobes, G., Pischedda, S., 2019. The relationship between adult symptoms of attention-deficit/hyperactivity disorder and criminogenic cognitions. Brain Sci. 9 (6), pii:E128.

Epstein, J.N., Johnson, D., Conners, C.K., 2006. Conners' adult ADHD diagnostic interview for DSM-IV.

Fair, D.A., Nigg, J.T., Iyer, S., Bathula, D., Mills, K.L., Dosenbach, N.U., Schlaggar, B.L., Mennes, M., Gutman, D., Bangaru, S., Buitelaar, J.K., Dickstein, D.P., Di Martino, A., Kennedy, D.N., Kelly, C., Luna, B., Schweitzer, J.B., Velanova, K., Wang, Y.F., Mostofsky, S., Castellanos, F.X., Milham, M.P., 2012. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. Front. Syst. Neurosci. 6, 80.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T., 2016. The human braintnetome atlas: a new brain atlas based on connectional architecture. Cereb. Cortex 26, 3508–3526.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit Lett 27, 861–874.

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 2002. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version.

Francx, W., Oldehinkel, M., Oosterlaan, J., Heslenfeld, D., Hartman, C.A., Hoekstra, P.J., Franke, B., Beckmann, C.F., Buitelaar, J.K., Mennes, M., 2015. The executive control network and symptomatic improvement in attention-deficit/hyperactivity disorder. Cortex 73, 62–72.

Ghiassian, S., Greiner, R., Jin, P., Brown, M.R., 2016. Using functional or structural magnetic resonance images and personal characteristic data to identify adhd and autism. PLoS ONE 11, e0166934.

Greenstein, D., Malley, J.D., Weisinger, B., Clasen, L., Gogtay, N., 2012. Using multi-variate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. Front. Psychiatry 3, 53.

Groom, M.J., van Loon, E., Daley, D., Chapman, P., Hollis, C., 2015. Driving behaviour in adults with attention deficit/hyperactivity disorder. BMC Psychiatry 15, 175.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Hansen, L.K., Salamon, P., 1990. Neural network ensembles. IEEE Trans. Pattern Anal. Mach. Intell 12, 993–1001.

Harrison, A.G., Edwards, M.J., Parker, K.C., 2007. Identifying students faking ADHD: preliminary findings and strategies for detection. Arch. Clin. Neuropsychol. 22, 577–588.

Hart, H., Chantiluke, K., Cubillo, A.I., Smith, A.B., Simmons, A., Brammer, M.J., Marquand, A.F., Rubia, K., 2014. Pattern classification of response inhibition in ADHD: toward the development of neurobiological markers for ADHD. Hum. Brain Mapp. 35, 3083–3094.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

Hutcheson, G.D., 1999. The multivariate social scientist.

Iannaccone, R., Hauser, T.U., Ball, J., Brandeis, D., Walitza, S., Brem, S., 2015. Classifying adolescent attention-deficit/hyperactivity disorder (ADHD) based on functional and structural imaging. Eur. Child Adolesc. Psychiatry 24, 1279–1289.

Johnston, B.A., Mwangi, B., Matthews, K., Coghill, D., Konrad, K., Steele, J.D., 2014. Brainstem abnormalities in attention deficit hyperactivity disorder support high accuracy individual diagnostic classification. Hum. Brain Mapp. 35, 5179–5189.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., Ryan, N., 1997. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J. Am. Acad. Child Adolesc. Psychiatry 36, 980–988.

Konrad, K., Neufang, S., Hanisch, C., Fink, G.R., Herpertz-Dahlmann, B., 2006. Dysfunctional attentional networks in children with attention deficit/hyperactivity disorder: evidence from an event-related functional magnetic resonance imaging study. Biol. Psychiatry 59, 643–651.

Kuang, D., He, L., 2014. Classification on ADHD with deep learning. 2014 international conference on cloud computing and. Big Data 6.

Lam, L., Suen, C.Y., 1997. Application of majority voting to pattern recognition: an analysis of its behavior and performance. IEEE Trans. Syst. Man Cybern. 27, 16.

Li, X., Sroubek, A., Kelly, M.S., Lesser, I., Sussman, E., He, Y., Branch, C., Foxe, J.J., 2012. Atypical pulvinar-cortical pathways during sustained attention performance in children with attention-deficit/hyperactivity disorder. J. Am. Acad. Child Adolesc. Psychiatry 51, e1194 1197-1207.

Lim, L., Marquand, A., Cubillo, A.A., Smith, A.B., Chantiluke, K., Simmons, A., Mehta, M., Rubia, K., 2013. Disorder-specific predictive classification of adolescents with attention deficit hyperactivity disorder (ADHD) relative to autism using structural magnetic resonance imaging. PLoS ONE 8, e63660.

Lin, H.Y., Tseng, W.Y., Lai, M.C., Matsuo, K., Gau, S.S., 2015. Altered resting-state frontoparietal control network in children with attention-deficit/hyperactivity disorder. J. Int. Neuropsychol. Soc. 21, 271–284.

Loney, J., Milich, R., 1982. Hyperactivity, inattention, and aggression in clinical practice. In: Wolraich, M.R.D (Ed.), Advances in Developmental and Behavioral Pediatrics, pp. 113–147.

Luo, Y., Schulz, K.P., Alvarez, T.L., Halperin, J.M., Li, X., 2018. Distinct topological properties of cue-evoked attention processing network in persisters and remitters of childhood ADHD. Cortex 109, 234–244.

Mattfeld, A.T., Gabrieli, J.D., Biederman, J., Spencer, T., Brown, A., Kotte, A., Kagan, E., Whitfield-Gabrieli, S., 2014. Brain differences between persistent and remitted attention deficit hyperactivity disorder. Brain 137, 2423–2428.

Peng, X., Lin, P., Zhang, T., Wang, J., 2013. Extreme learning machine-based classification of ADHD using brain structural MRI data. PLoS ONE 8, e79476.

Posner, M.I., Rothbart, M.K., 2009. Toward a physical basis of attention and self regulation. Phys. Life Rev. 6, 103–120.

Proal, E., Reiss, P.T., Klein, R.G., Mannuzza, S., Gotimer, K., Ramos-Olazagasti, M.A., Lerch, J.P., He, Y., Zijdenbos, A., Kelly, C., Milham, M.P., Castellanos, F.X., 2011. Brain gray matter deficits at 33-year follow-up in adults with attention-deficit/hyperactivity disorder established in childhood. Arch. Gen. Psychiatry 68, 1122–1134.

Qureshi, M.N., Min, B., Jo, H.J., Lee, B., 2016. Multiclass classification for the differential diagnosis on the adhd subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI study. PLoS ONE 11, e0160697.

Qureshi, M.N.I., Oh, J., Min, B., Jo, H.J., Lee, B., 2017. Multi-modal, multi-measure, and multi-class discrimination of ADHD with hierarchical feature extraction and extreme learning machine using structural and functional brain MRI. Front. Hum. Neurosci. 11, 157.

Rubia, K., Criaud, M., Wulff, M., Alegria, A., Brinson, H., Barker, G., Stahl, D., Giampietro, V., 2019. Functional connectivity changes associated with fMRI neurofeedback of right inferior frontal cortex in adolescents with ADHD. NeuroimageNeuroimage 188, 43–58.

Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S.C., Simmons, A., Bullmore, E.T., 1999. Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: a study with functional MRI. Am. J. Psychiatry 156, 891–896.

Ruta, D., Gabrys, B., 2005. Classifier selection for majority voting. Inf. Fusion 6, 19.

Santosa, F., Symes, W.W., 1986. Linear inversion of band-limited reflection seismograms. SIAM J. Sci. Stat. Comput. 7, 1307–1330.

Schapire, R.E., 1990a. The strength of weak learnability. Mach. Learn. 5, 31.

Schapire, R.E., 1990b. The strength of weak learnability. Mach. Learn. 5, 197–227.

Schulz, K.P., Li, X., Clerkin, S.M., Fan, J., Berwid, O.G., Newcorn, J.H., Halperin, J.M., 2017. Prefrontal and parietal correlates of cognitive control related to the adult outcome of attention-deficit/hyperactivity disorder diagnosed in childhood. Cortex 90, 1–11.

Sen, B., Borle, N.C., Greiner, R., Brown, M.R.G., 2018. A general prediction model for the detection of ADHD and AUTISM using structural and functional MRI. PLoS ONE 13, e0194856.

Shaffer, D., Fisher, P., Piacentini, J., Schwab-Stone, M., Wicks, J., 1989. Diagnostic interview schedule for children-parent version (DISC-2.1P).

Shaw, P., Malek, M., Watson, B., Greenstein, D., de Rossi, P., Sharp, W., 2013. Trajectories of cerebral cortical development in childhood and adolescence and adult attention-deficit/hyperactivity disorder. Biol. Psychiatry 174, 599–606.

Shaw, P., Sudre, G., Wharton, A., Weingart, D., Sharp, W., Sarlls, J., 2015. White matter microstructure and the variable adult outcome of childhood attention deficit hyperactivity disorder. Neuropsychopharmacology 40, 746–754.

Silk, T., Vance, A., Rinehart, N., Egan, G., O'Boyle, M., Bradshaw, J.L., Cunnington, R., 2005. Fronto-parietal activation in attention-deficit hyperactivity disorder, combined type: functional magnetic resonance imaging study. Br. J. Psychiatry 187, 282–283.

Smith, A.B., Taylor, E., Brammer, M., Toone, B., Rubia, K., 2006. Task-specific hypoactivation in prefrontal and temporoparietal brain regions during motor inhibition and task switching in medication-naive children and adolescents with attention deficit hyperactivity disorder. Am. J. Psychiatry 163, 1044–1051.

Solanto, M.V., Marks, D.J., Wasserstein, J., Mitchell, K., Abikoff, H., Alvir, J.M., Kofman, M.D., 2010. Efficacy of meta-cognitive therapy for adult ADHD. Am. J. Psychiatry 167, 958–968.

Sun, D., van Erp, T.G., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., Hardt, M.E., Nuechterlein, K.H., Toga, A.W., Cannon, T.D., 2009. Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. Biol. Psychiatry 66, 1055–1060.

Tajima-Pozo, K., Yus, M., Ruiz-Manrique, G., Lewczuk, A., Arrazola, J., Montanes-Rada, F., 2018. Amygdala abnormalities in adults with ADHD. J. Atten. Disord. 22, 671–678.

Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanov, J., Muller, A., Candrian, G., 2014. Machine learning approach for classification of ADHD adults. Int. J. Psychophysiol. 93, 162–166.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc.. Series B (Methodological) 58, 267–288.

Van Dessel, J., Sonuga-Barke, E., Mies, G., Lemiere, J., Van der Oord, S., Morsink, S., Danckaerts, M., 2018. Delay aversion in attention deficit/hyperactivity disorder is mediated by amygdala and prefrontal cortex hyper-activation. J. Child Psychol. Psychiatry 59, 888–899.

Van Dessel, J., Sonuga-Barke, E., Moerkerke, M., Van der Oord, S., Lemiere, J., Morsink, S., Danckaerts, M., 2019. The amygdala in adolescents with attention-deficit/hyperactivity disorder: structural and functional correlates of delay aversion. World J. Biol. Psychiatry 1–12.

Vance, A., Silk, T.J., Casey, M., Rinehart, N.J., Bradshaw, J.L., Bellgrove, M.A., Cunnington, R., 2007. Right parietal dysfunction in children with attention deficit hyperactivity disorder, combined type: a functional MRI study. Mol. Psychiatry 12, 826–832 793.

Wang, G., Hao, J., Ma, J., Jiang, H., 2011. A comparative assessment of ensemble learning for credit scoring. Expert Syst Appl 38, 8.

Wolpert, D.H., 1992. Stacked generalization. Neural Netw. 5, 19.

Xia, S., Li, X., Kimball, A.E., Kelly, M.S., Lesser, I., Branch, C., 2012. Thalamic shape and connectivity abnormalities in children with attention-deficit/hyperactivity disorder. Psychiatry Res. 204, 161–167.

Yang, H., Wu, Q.Z., Guo, L.T., Li, Q.Q., Long, X.Y., Huang, X.Q., Chan, R.C., Gong, Q.Y., 2011. Abnormal spontaneous brain activity in medication-naive ADHD children: a resting state fMRI study. Neurosci. Lett. 502, 89–93.

Yasumura, A., Omori, M., Fukuda, A., Takahashi, J., Yasumura, Y., Nakagawa, E., Koike, T., Yamashita, Y., Miyajima, T., Koeda, T., Aihara, M., Tachimori, H., Inagaki, M., 2017. Applied machine learning method to predict children with adhd using prefrontal cortex activity: a multicenter study in Japan. J. Atten. Disord 1087054717740632.

Freund, Yoav, Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55, 21.

Zhang-James, Y., Helminen, E.C., Liu, J., Franke, B., Hoogman, M., Faraone, S.V., 2019. Machine learning classification of attention-deficit/hyperactivity disorder using structural MRI data. bioRxiv, 546671.

Zhu, C.Z., Zang, Y.F., Cao, Q.J., Yan, C.G., He, Y., Jiang, T.Z., Sui, M.Q., Wang, Y.F., 2008. Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder. Neuroimage 40, 110–120.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Series B (Stat. Methodol.) 67, 301–320.

Zou, L., Zheng, J., Miao, C., Mckeown, M.J., Wang, Z.J., 2017. 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. IEEE Access 5.