

RESEARCH ARTICLE

Multiclass Classification for the Differential Diagnosis on the ADHD Subtypes Using Recursive Feature Elimination and Hierarchical Extreme Learning Machine: Structural MRI Study

Muhammad Naveed Iqbal Qureshi¹, Beomjun Min¹, Hang Joon Jo², Boreom Lee^{1*}

1 Department of Biomedical Science and Engineering (BMSE), Institute of Integrated Technology (IIT), Gwangju Institute of Science and Technology (GIST), Gwangju, Republic of Korea, **2** Department of Neurologic Surgery, Mayo Clinic, Rochester, Minnesota, United States of America

* leebr@gist.ac.kr



OPEN ACCESS

Citation: Qureshi MNI, Min B, Jo HJ, 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(8): e0160697. doi:10.1371/journal.pone.0160697

Editor: Quan Zou, Tianjin University, CHINA

Received: March 2, 2016

Accepted: July 22, 2016

Published: August 8, 2016

Copyright: © 2016 Qureshi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All the MRI files are available from the ADHD-200 database (<http://fcon.1000.projects.nitrc.org/indi/adhd200>).

Funding: This work was supported by the GIST Research Institute (GRI) in 2016. It was also supported by the Pioneer Research Center Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (grant no. 2012-0009462). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The classification of neuroimaging data for the diagnosis of certain brain diseases is one of the main research goals of the neuroscience and clinical communities. In this study, we performed multiclass classification using a hierarchical extreme learning machine (H-ELM) classifier. We compared the performance of this classifier with that of a support vector machine (SVM) and basic extreme learning machine (ELM) for cortical MRI data from attention deficit/hyperactivity disorder (ADHD) patients. We used 159 structural MRI images of children from the publicly available ADHD-200 MRI dataset. The data consisted of three types, namely, typically developing (TDC), ADHD-inattentive (ADHD-I), and ADHD-combined (ADHD-C). We carried out feature selection by using standard SVM-based recursive feature elimination (RFE-SVM) that enabled us to achieve good classification accuracy (60.78%). In this study, we found the RFE-SVM feature selection approach in combination with H-ELM to effectively enable the acquisition of high multiclass classification accuracy rates for structural neuroimaging data. In addition, we found that the most important features for classification were the surface area of the superior frontal lobe, and the cortical thickness, volume, and mean surface area of the whole cortex.

Introduction

The neurodevelopmental disorder of attention deficit hyperactivity disorder (ADHD) has a prevalence of 3.4% among children and adolescents [1]. Although the statistic itself seems to have decreased from 5.3% as reported by [2], the burden ADHD places on patients, their families, and the societies to which they belong remains significant. Usually, the disorder begins to influence patients from an early age and affects their entire life, and without appropriate

Competing Interests: The authors have declared that no competing interests exist.

treatments, the illness leads to a poor prognosis. Thus, the diagnosis of ADHD is important. Children affected by the disorder have characteristic symptoms such as attention deficit, hyperactivity, and impulsiveness. Currently, it is believed that this characteristic manifestation of symptoms originates from the dysfunction of related cognitive processes [3]. In addition, the underlying mechanisms of ADHD seem to be associated with delayed cortical development [4]. On the other hand, according to the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5), there are three types of ADHD, based on the predominant symptoms: 1) predominantly inattentive presentation, 2) predominantly hyperactive-impulsive presentation, and 3) combined presentation [5]. In addition, the results of many studies that have investigated these subtypes of ADHD such as [6–8] have been published. However, the outcomes of a small number of studies that attempted to distinguish among the features of these subtypes with neuroimaging methods were inconclusive [9, 10].

In recent years, many studies based on classification using neuroimaging data for discriminating the clinical diagnosis have been published. In fact, the clinical experience of the psychiatrist, a detailed record of the history of the patient, and other information resulting from interviews are important. However, the past decade has seen an increase in interest in the use of machine-learning techniques, not only in the field of psychiatry, but also in other medical fields such as protein folds and bioinformatics data prediction and classification [11–13]. Previously, a number of studies using neuroimaging methods in the psychiatric research field have been reported [14–16]. Consequently, with the familiarity and understanding of neuroimaging methods, a series of candidate features such as cortical thickness, volumetric data, demographic information, and other fMRI data have been used for machine learning in studies related to this field. This kind of machine-learning-based approach serves as a very powerful tool for neuroimaging studies and plays an important role in terms of the integration of machine learning and neuroscience. The comparatively higher resolution of structural MRI data makes these data a more appropriate choice for use in classification experiments. The most abundantly used machine-learning tool in the neuroimaging community is SVM. Our work involves a comparative study by using a recently proposed H-ELM [17] classification framework in comparison to SVM and basic ELM, of which H-ELM achieved the highest accuracies both for binary and multiclass settings.

As previously reported in several studies, certain numbers of anatomical regions were identified as being affected by ADHD. A recent study by [18] reported the abnormal surface morphology of the central sulcus in children with ADHD; specifically, they found that the average depth and maximum depth of the left central sulcus, as well as the average cortical thickness value of the bilateral central sulcus in the ADHD group, were significantly larger than those in healthy children. ADHD is responsible for a reduction in the surface area and cortical volume up to 7% bilaterally and 7–8%, respectively [19]. In the anterior brain region, posterior brain regions including the left and right superior temporal and parietal lobes, temporal junction, and insula, evident anatomical abnormalities can be observed from the cortical thickness and folding analysis [20, 21]. The studies by [22] showed a total reduction of 3–5% in cerebral volume. Abnormalities in several specific regions including the callosum, striatum, cerebellum, lateral prefrontal cortex, and cingulate cortex were reported by [22]. Thus, these regions can be used as potential biomarkers for the automatic diagnosis of ADHD by using machine-learning approaches.

Our work focused on the differential diagnosis of the subtypes of ADHD based on machine learning. Currently, one of the reasons for the increasing tendency to implement machine learning for the diagnosis of psychiatric disease is that the diagnostic process mainly depends on what is determined by experienced psychiatrists. The present study utilized the implementation of an H-ELM classifier to distinguish ADHD-combined-type and ADHD-inattentive-

type patients from typically developing children. Additionally, to improve the classification accuracy, we utilized the standard RFE-SVM algorithm proposed by [23] and available at http://people.kyb.tuebingen.mpg.de/spider/help_rfe.html. This algorithm accepts five different types of cortical feature data as input. The cortical feature measures include mean cortical thickness, surface area, folding indices, mean curvature indices, and the volume of the pre-defined “Desikan Killiany Tourville” (DKT) atlas-based ROI [24] segments. In addition, we performed an ANOVA test based on group analysis to identify the significantly different anatomical regions among the three groups namely, ADHD-I, ADHD-C, and TDC.

The remainder of this paper is organized as follows. Section 2 provides details of the dataset, subject selection, preprocessing of the MRI data, an introduction of the classification algorithms, feature selection and optimization, permutation testing for validation of the results, and statistical analysis. Section 3 presents the comparative results of both the multiclass and binary classifiers by using H-ELM, ELM, and SVM. In addition, this section also includes the ANOVA test results. Section 4 includes the discussion and conclusion of the article.

Materials and Methods

MRI Dataset

We obtained neuroimaging data of ADHD patients from the ADHD-200 MRI dataset, which is publicly available at http://fcon_1000.projects.nitrc.org/indi/adhd200. These T-1-weighted MRI scans were acquired at six different institutes with participant ages ranging from 7 to 14 years. The sites were Brown University (BU), New York University Child Study Center (NYU), Beijing Normal University (BNU), the Kennedy Krieger Institute (KKI), Oregon Health and Science University (OHSU), and Washington University in St. Louis (WU). All participants were scanned using 3.0-Tesla scanners. The other technical details about the scanner parameters from each participating site are already available at the above-mentioned URL of the ADHD-200 global competition. In addition, all the sites contributing to ADHD-200 had approval from their local institutional review board (IRB) and complied with local IRB protocols.

Subjects

The ADHD-200 dataset included more than 1000 subjects; however, we followed a balanced design approach and only chose 53 subjects from each of the three groups: typically developing children (TDC), ADHD inattentive type (ADHD-I), and ADHD combined type (ADHD-C). The main focus of the current study was to elaborate the significance of a supervised multiclass classification between TDC, ADHD-C, and ADHD-I based on the H-ELM framework and all the subjects were chosen from the training part of the ADHD-200 dataset, which consists of 776 subjects. Each group contained 9 female and 44 male subjects. There may have existed some intrinsic bias in the data regarding batch effects as well as hardware bias due to multisite data collection [25]. However, we retained the bias as default for the criteria for selecting subjects for this study. The IQ information of a few subjects was not provided by some of the collection sites; therefore, we did not use this information in the present study. More detailed demographic information can be acquired from the same resources listed above. Information regarding our selected subjects is summarized in [Table 1](#).

Preprocessing of Structural MRI Data

The raw data on their own provide very little information; therefore, preprocessing is required to extract the patterns of useful knowledge [26]. All of the structural MRI images were

Table 1. Demographic variables of the participants.

Groups	TDC	ADHD-I	ADHD-C
Number of subjects	53	53	53
Age (<i>mean ± SD</i>)	12.75 ± 3.86	12.42 ± 2.23	11.83 ± 3.52
Full IQ (<i>mean ± SD</i>)	114.86 ± 13.86	102.47 ± 13.11	110.10 ± 13.88
Handedness	Right only	Right only	Right only

Abbreviation: TDC = typically developing children; ADHD-I = attention-deficit/hyperactivity disorder, inattentive type; ADHD-C = attention-deficit/hyperactivity disorder, combined type; SD = standard deviation

doi:10.1371/journal.pone.0160697.t001

preprocessed using the fully automated pipeline of FreeSurfer 5.3.0 [27–31] for volumetric segmentation and cortical reconstruction. The software proceeds with motion correction, T1-weighted image averaging, and registration to the Talairach space, followed by skull stripping with a deformable template model. The white and pial surface was generated for each hemisphere. A cortical surface-based atlas (DKT atlas) was mapped to a sphere aligning the cortical folding patterns, which provided accurate matching of the morphologically homologous cortical locations across subjects. For each of the DKT31 protocol-based segments, FreeSurfer calculated nine different measures, including, surface vertices, surface area, gray matter volume, average cortical thickness, cortical thickness standard deviation, cortical mean curvature, cortical Gaussian curvature, cortical folding index, and cortical curvature indices [25]. We used five of the above-mentioned measures in this study. The surface area was calculated by computing the area of every triangle in a standardized spherical surface tessellation. The average shortest distance between white and pial surfaces denotes the cortical thickness at each vertex of the cortex. The local curvature was computed using the registration surface based on the folding patterns. The folding index over the whole cortical surface was measured for each subject [19].

After preprocessing, the subcortical regions were masked to separate the significant cortical data by using AFNI. Finally, the cortical thickness data were converted into surface maps using the AFNI program MapIcosahedron. These maps were then subjected to heat-kernel-based smoothing using the AFNI program SurfSmooth with a 30-mm kernel.

ANOVA for Cortical Thickness

First, we performed an ANOVA to examine and visualize the differences of cortical thickness among the three groups. The AFNI program 3dANOVA2 [32] was used for the analysis. A multiple comparison test was performed using the AFNI program 3drefit with the FDR correction. In addition, the AFNI program SurfClust was used to find regions that have significant differences between groups with an area of more than 50mm^2 . The p -value threshold was set to be 0.001.

Classification

In this study, three machine-learning classification algorithms were used namely, H-ELM, ELM, and SVM. After comparing the results of all the classifiers, H-ELM proved to be the most efficient algorithm both in terms of computation time and accuracy. A brief description of each method is provided in the following paragraphs.

Hierarchical-Extreme Learning Machines. The extreme learning machine, which was originally proposed by [33], is a comparatively newer machine-learning algorithm. It was

adopted by many previous neuroimaging studies [19, 34, 35] as the main tool for data discrimination in the binary setting. However, to the best of our knowledge, the present study is the first in this domain that utilized the recently proposed H-ELM framework for the multiclass classification of neuroimaging data.

The H-ELM framework consists of two main parts, an ELM-based unsupervised sparse multilayer autoencoder for feature encoding and supervised ELM for classification. The ELM is an effective solution for single-layer feed-forward networks (SLFNs). Unlike SVM and other back-propagation methods, the parameters of the hidden layer need not be tuned and can be generated randomly before the training samples are acquired. Unlike conventional ELM, which is used for comparison purposes in this study, H-ELM is a multi-layer learning architecture and uses an ELM-based sparse autoencoder for feature learning via ℓ_1 -norm optimization to generate more sparse and meaningful hidden layers of features before classification by the original ELM.

Mathematically, in the first step, the output of each hidden layer can be represented as

$$\mathbf{H}_j = \lambda(\mathbf{H}_{j-1} \cdot \boldsymbol{\beta}), \tag{1}$$

where \mathbf{H}_j is the output of the j th layer ($j \in [1, K]$), \mathbf{H}_{j-1} is the output of the $(j-1)$ th layer, $\lambda(\cdot)$ denotes the activation function of the hidden layers and $\boldsymbol{\beta}$ represents the output weights.

In the second step, an original SLFN ELM with m hidden nodes is used for classification. It can be represented as

$$f_m(\boldsymbol{\rho}) = \sum_{j=1}^m \eta_j(\boldsymbol{\rho}, \boldsymbol{\theta}_j, \gamma_j) \cdot \boldsymbol{\beta}_j, \quad \boldsymbol{\theta}_j \in \mathbf{R}^d, \gamma_j, \boldsymbol{\beta}_j \in \mathbf{R}^1 \tag{2}$$

where $\eta_j(\cdot)$ denotes the j th hidden node activation function, $\boldsymbol{\theta}_j$ is the input weight vector connecting the input layer to the j th hidden layer, γ_j is the bias weight of the j th hidden layer, and $\boldsymbol{\beta}_j$ is the output weight. \mathbf{R}^d is the d -dimensional and \mathbf{R}^1 is the one-dimensional Euclidean space. For the additive nodes with activation function λ , η_j can be defined as

$$\eta_j(\boldsymbol{\rho}, \boldsymbol{\theta}_j, \gamma_j) = \lambda(\boldsymbol{\theta}_j \cdot \boldsymbol{\rho} + \gamma_j) \tag{3}$$

The dot in the Eq (3) denotes the inner product of $\boldsymbol{\theta}_j$ and $\boldsymbol{\rho}$. Eqs (2) and (3) can be compactly written as

$$\mathbf{B} = \mathbf{H}^\dagger \mathbf{T}, \tag{4}$$

where $\mathbf{B} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m]^T$ is the weight matrix for m hidden nodes, \mathbf{H} is the hidden layer output matrix of each of the n arbitrary distinct training data samples,

$$\mathbf{H} = \begin{bmatrix} \lambda(\boldsymbol{\theta}_1 \cdot \boldsymbol{\rho}_1 + \gamma_1) & \cdots & \lambda(\boldsymbol{\theta}_m \cdot \boldsymbol{\rho}_1 + \gamma_m) \\ \vdots & \ddots & \vdots \\ \lambda(\boldsymbol{\theta}_1 \cdot \boldsymbol{\rho}_n + \gamma_1) & \cdots & \lambda(\boldsymbol{\theta}_m \cdot \boldsymbol{\rho}_n + \gamma_m) \end{bmatrix} \tag{5}$$

$\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ is the Moore-Penrose inverse of \mathbf{H} , and $\mathbf{T} = [\mathbf{t}_1^T, \dots, \mathbf{t}_m^T]^T$ is the target matrix of training data where $\mathbf{t}_1^T = [t_{11}, \dots, t_{1n}]$ [17].

The H-ELM framework has a certain number of hyperparameters that need to be tuned to achieve the maximum performance of the classifier in terms of accuracy. In this study, we used the MATLAB implementation of the H-ELM as proposed by [17]. The hyperparameters were the number of nodes in all three layers N_1 , N_2 , and N_3 , C was the ℓ_2 penalty of the 3rd layer of

ELM, and s was the scaling (regularization) parameters. Since there is no proper guideline available for the adjustment of these parameters, we have used a greedy search method to tune these parameters for achieving the maximum validation accuracy before training the classifier. In this study, the search scale for selecting these parameters was set to $N1 = N2 = 48$, $N3 = [49, 50, \dots, 500]$, $C = [2^{-31}, 2^{-32}, \dots, 2^{-40}]$, and $s = [0.1, 0.2, \dots, 1.0]$. In addition, the data labels were encoded as $class1 = [1 \ -1 \ -1]$, $class2 = [-1 \ 1 \ -1]$, and $class3 = [-1 \ -1 \ 1]$. All the features were normalized and scaled to values between -1 and 1 to improve the performance of the H-ELM framework.

ELM Classifier. We used the basic ELM classifier [33] with a sigmoid activation function in this study. The number of nodes was selected by a greedy search method. The search scale was set to $N = [1, 2, \dots, 50]$.

SVM Classifier. SVM, which was originally proposed by [36], is useful for classification in the neuroimaging field and has been one of the most popular machine-learning tools in the neuroscience domain in the last decade. It is a supervised classification algorithm, and, intrinsically, it is a binary classification algorithm. We have used this algorithm in both binary and multiclass (more commonly known as one-versus-all, or OVA) settings. It maps features in higher dimensional space using linear and nonlinear functions known as kernels. In this study we used both linear and nonlinear radial basis function (RBF) kernels. The SVM classifier discriminates between the two classes in the feature space by constructing a hyperplane that discriminates the training data from the maximum margin between the two data clusters. In a multiclass setting, an SVM algorithm was used to classify each individual into one of the three diagnostic groups. The multiclass classification was conducted by building a K OVA classification technique. The OVA is an alternative method of applying the SVM in the case of $K > 2$ classes. It functions by comparing one of the K classes to the remaining $K-1$ classes. The OVA chooses the class that is selected by the largest number of classifiers as it amounts to a high level of confidence that the test class belongs to the K^{th} class rather than to any of the other classes.

Feature Selection and Optimization

In most studies involving neuroimaging analysis, the number of features per subject greatly outnumbers the number of subjects. This concept is commonly referred to as the *curse of dimensionality* [37]. Feature selection can generally be subdivided into two categories: the filter model and the wrapper model. Optimal feature selection provides a subset of features that leads towards optimal classification accuracy.

Usually, an *optimal* feature has relevance to the class concept and represents no overlap with any other feature in the feature set and/or subset. We used a very straightforward filter model for feature measure selection in this study [19, 38]. The FreeSurfer pipeline was used to automatically generate the cortical features of five feature measure types as indicated in Table 2.

Table 2. Feature Measures and Cortical Feature Index Information.

Feature Measure (<i>fM</i>)	Feature Measure Type	Indices of Cortical Features
<i>fM1</i>	Mean Cortical thickness	1–64
<i>fM2</i>	Surface Area	65–128
<i>fM3</i>	Folding Indices	129–192
<i>fM4</i>	Mean Curvature Indices	193–256
<i>fM5</i>	Volume	257–320

doi:10.1371/journal.pone.0160697.t002

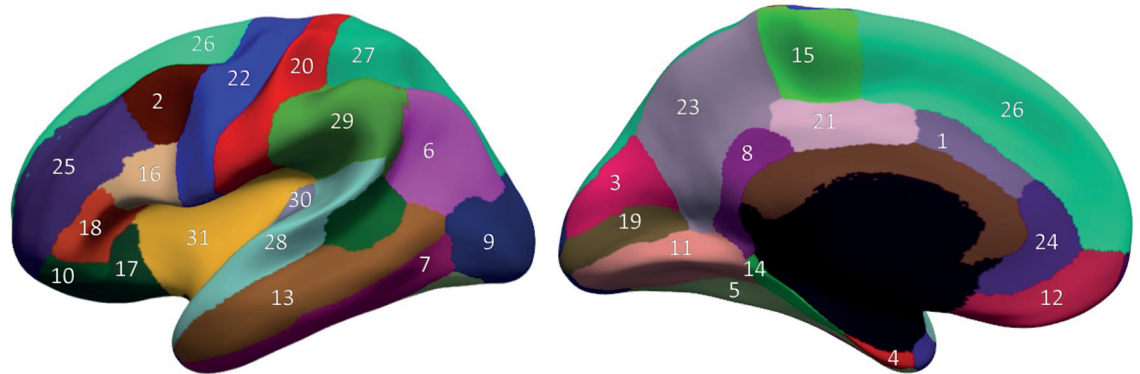


Fig 1. Cortical segmentation based on the DKT atlas. Each of the 31 anatomical ROIs are shown with a specific color and index (see Table 3 for indices). The figures on the left and right show the lateral and medial views, respectively.

doi:10.1371/journal.pone.0160697.g001

The “feature measures (fM)” were utilized to acquire 64 cortical features of each type. These 64 cortical features were obtained as a pair of 32 features from each hemisphere. The 32 cortical features consisted of the values of 31 ROIs of DKT atlas-based brain segments and the remaining cortical feature was generated by obtaining the average value of all 31 ROIs of the corresponding hemisphere. Mathematically, this is expressed as follows:

$$Mean\ Value\ (MV) = \frac{1}{31} \sum_{i=1}^{31} (fM\ value\ in\ ROI_i) \quad (6)$$

$$5fM \times (31\ ROI + MV) \times 2\ Hemispheres = 320\ CorticalFeatures \quad (7)$$

The segmentation of cortical ROIs was based on the comparatively new DKT atlas [24]. This atlas is provided as a built-in tool in the Freesurfer software package [28]. These cortical ROIs are shown in Fig 1 with the indices referring to those listed in Table 3.

The use of a wrapper to achieve model-based feature selection usually consists of two steps: 1) feature subset selection, using the accuracy of the base classifier and 2) using the best feature subset for learning and testing. The wrapper approach consists of using the prediction performance of a base classifier. It performs selection by taking into account the classifier as a black box and ranking the subset of features by their predictive power [26]. The preselected 320 cortical features categorized in Table 2 are subsequently fed into the wrapper model (RFE-SVM) to generate the feature-ranking list in the order of significance.

RFE-SVM. We improved the classification accuracy by applying another fold of feature selection by using RFE-SVM on the preselected cortical features provided in Table 2. This is a wrapper-based method and selects feature subsets by utilizing greedy backward selection. A linear SVM classifier serves as the base algorithm and removes the features with the smallest absolute value it returns in each cycle of the training iteration. RFE-SVM continues the iteration until it reaches the specified feature count in the base algorithm call. During this training process, RFE-SVM generates a feature-ranking list. RFE-SVM uses the accuracy of cross-validation to generate the feature subset, thereby possibly avoiding the over-fitting problem as well [22]. This method was successfully deployed in many previous research efforts related to ADHD-200 dataset classification, including [22], and a detailed motivation for choosing this

Table 3. DKT31 Protocol-based Cortical ROIs.

No.	Label	No.	Label
1	Caudal Anterior Cingulate	17	Pars orbitalis (IFG)
2	Caudal middle frontal gyrus	18	Pars triangularis (IFG)
3	Cuneus	19	Pericalcarine cortex
4	Entorhinal cortex	20	Postcentral gyrus
5	Fusiform gyrus	21	Posterior cingulate
6	Inferior parietal lobule	22	Precentral gyrus
7	Inferior temporal gyrus	23	Precuneus
8	Isthmus cingulate	24	Rostral anterior cingulate
9	Lateral occipital cortex	25	Rostral middle frontal
10	Lateral orbitofrontal gyrus	26	Superior frontal
11	Lingual gyrus	27	Superior parietal lobule
12	Medial orbitofrontal gyrus	28	Superior temporal gyrus
13	Middle temporal gyrus	29	Supramarginal gyrus
14	Parahippocampal gyrus	30	Transverse temporal gyrus
15	Paracentral gyrus	31	Insula
16	Pars opercularis (IFG)	32	Mean measure

Abbreviation: IFG = Inferior frontal gyrus

doi:10.1371/journal.pone.0160697.t003

method for feature selection can be found in [25]. Fig 2 depicts the overall framework of the current study.

Permutation Testing

Permutation testing is used to assess the statistical significance of the classifier performance [39]. Briefly, the assessment proceeds as follows. First, we choose the test statistic of the classifier, permute the class labels for the training dataset randomly and assign them to the classifier, and check for cross-validation. If the classifier can learn the relationship between the data and their labels with a p-value exceeding 0.05, the test result is considered as insignificant and the null hypothesis is considered to be true. Generally, a lower p-value of the permuted prediction rate against the prediction rate with the original data labels indicates a higher significance of the classifier performance. As there is no fixed rule for setting the number of permutations, in the current study, we permuted the data 10,000 times.

Validation, Training, and Testing Subjects

We used two methods of validation, i.e., conventional 70/30 partitioning, and k-fold ($k = 10$) cross-validation. In the case of conventional partitioning, we divided the data into training and testing datasets and used the first 36 randomly chosen subjects from each group for training and the remaining 17 subjects for testing the classifier. In the 10-fold cross-validation, all 159 subjects were randomly divided into 10 equally sized subsets, i.e., testing (10%) and training (90%) subjects for each of the 10 folds of the cross-validation. In addition, for nested validation, we repeated the classification experiment 10 times in the case of 10-fold cross-validation and 100 times in the case of conventional validation to ensure the robustness of the classification results. The mean accuracy of all the repetitions was reported as the final result.

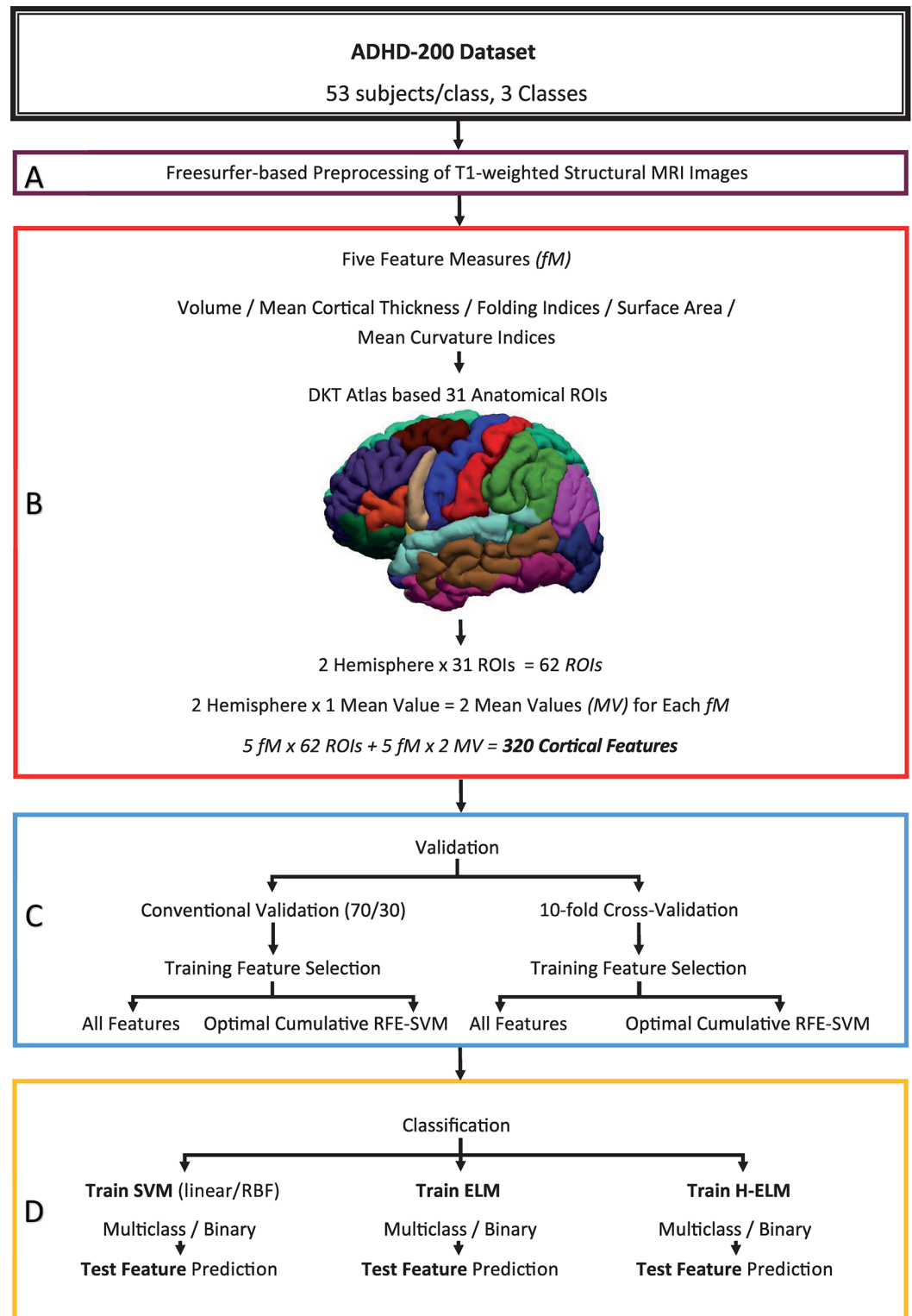


Fig 2. Overall framework of the study. The level-A block represents subject selection from the ADHD-200 dataset and briefly describes the preprocessing step, the level-B block provides information about the 31 DKT atlas-based ROIs, selection of the five feature measures and the total count of the cortical features. The level-C block elaborates the data validation methodologies and feature selection for the classifiers. Level-D represents the classifier choice for both the multiclass and binary settings.

doi:10.1371/journal.pone.0160697.g002

Performance Evaluation Methods

The performance of the classifier was measured in terms of classification accuracy and the Kappa score for multiclass classification. In addition, we calculated the confusion matrix for each iteration cycle of the classifier. In the binary case, besides accuracy, there exist many measures, including the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, specificity, F-score, and J-Statistic. **Classification accuracy** can be measured as

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100, \quad (8)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. These values were obtained from the confusion matrix. **Kappa score** is considered as a robust statistical measure for the assessment of multiclass classifiers. It is defined by [40] as

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}, \quad (9)$$

where $Pr(a)$ is the relative observer agreement between raters and $Pr(e)$ is the hypothetical probability of chance agreement. The value of k ranges between +1 and -1, where $k = 1$ represents perfect classification, $k = 0$ represents chance level, and $k = -1$ represents completely erroneous classification [41].

Results

Multiclass Classification

The OVA class setting was used to perform the multiclass classification. We performed the classification on both the complete and reduced cortical features for the multiclass (3 classes) classification and applied an RFE-SVM feature selector on the cortical data to achieve a high classification score. As a result, after 10-fold cross-validation, the average overall accuracy was 57.81%, whereas the use of conventional validation enabled us to achieve 60.78% accuracy with H-ELM. Details of the classification scores (accuracy and Kappa) using all of H-ELM, ELM, and SVM (with RBF and linear kernels) are presented in Table 4. We also performed a similar classification experiment with 10-fold cross-validation and conventional validation on the complete dataset with all the features; overall, the average maximum accuracy for the latter part with 10-fold cross validation was 51.59%, whereas the use of conventional validation achieved 56.87% accuracy with H-ELM. Hence, in view of the results presented in Table 4 we can suggest that, by using H-ELM, the accuracy of the testing classification between multiple diagnosis groups can be increased.

Binary Classification

Although the main focus of the present study is to prove the significance of multiclass classification results, we have also shown binary classification results to justify the multiclass classification results. In general, the multiclass classification should preferably show at most equal and ideally lower overall accuracy than any of the binary combinations of data groups. Our binary classification results help us to demonstrate this fact.

Three binary classifications were performed on the preprocessed cortical dataset by using all of the H-ELM, ELM, and SVM (with RBF and linear kernel) classifiers. We applied the RFE-SVM-based significant feature selector on the cortical data to acquire the optimal features. We achieved maximum accuracy of 85.29%, 79.40%, and 73.58% in the case of H-ELM, ELM,

Table 4. Mean Multiclass (3 classes) classification results.

RFE-SVM feature selection				
Classifier	H-ELM	ELM	SVM linear	SVM RBF
10-fold CV (%)	57.81 ($p < 0.0001$)	54.02	49.06	49.69
Kappa score	0.3675	0.3118	0.2358	0.2453
Conventional Validation (%)	60.78 ($p < 0.0001$)	54.90	43.14	45.10
Kappa score	0.4118	0.3235	0.1471	0.1769
All features				
Classifier	H-ELM	ELM	SVM linear	SVM RBF
10-fold CV (%)	51.59 ($p < 0.0001$)	40.75	43.40	42.77
Kappa score	0.2747	0.1101	0.1509	0.1415
Conventional Validation (%)	56.87 ($p < 0.0001$)	39.22	31.37	39.22
Kappa score	0.3529	0.0882	-0.0294	0.0882

Abbreviations: H-ELM = hierarchical extreme learning machine; ELM = extreme learning machine; SVM = support vector machine; RBF = radial basis function; RFE = recursive feature elimination; CV = cross-validation

doi:10.1371/journal.pone.0160697.t004

and SVM, respectively. Details of the classification score (accuracy) for each disease group are presented in [Table 5](#).

Classification Performance on Selected Features

The RFE-SVM feature selection method was used to arrange the 320 cortical features of the data in descending order according to the feature rank indicated in each cycle of the iteration. We combined each feature with all preceding feature rows as a ranked-classification dataset (RCD) for each iteration cycle. Even though the feature ranking changes in each cycle of the iteration, all the features selected in the previous iteration cycle stay in the selected subgroup but with a different rank for each cycle according to their significance level. Next, conventional 70/30 and 10-fold cross-validation was applied to compare the performance of all of the

Table 5. Binary classification results.

RFE-SVM feature selection, 10-fold Cross-Validation			
Classifier	ADHD-C-ADHD-I	ADHD-C-TDC	ADHD-I-TDC
H-ELM (%)	80.30 ($p < 0.0001$)	77.64 ($p < 0.0001$)	80.30 ($p < 0.0001$)
ELM (%)	74.24	74.73	74.58
SVMlinear (%)	73.58	71.70	68.86
SVMRBF (%)	66.98	64.15	57.54
RFE-SVM feature selection, Conventional Validation			
Classifier	ADHD-C-ADHD-I	ADHD-C-TDC	ADHD-I-TDC
H-ELM (%)	79.41 ($p < 0.0001$)	79.40 ($p < 0.0001$)	85.29 ($p < 0.0001$)
ELM (%)	76.47	79.40	79.40
SVMlinear (%)	73.52	64.71	70.59
SVMRBF (%)	58.82	64.71	61.76

Abbreviation: H-ELM = hierarchical extreme learning machine; ELM = extreme learning machine; TDC = typically developing children; ADHD-I = attention-deficit/hyperactivity disorder-inattentive type; ADHD-C = attention-deficit/hyperactivity disorder combined type; SVM = support vector machine; RBF = radial basis function; RFE = recursive feature elimination

doi:10.1371/journal.pone.0160697.t005

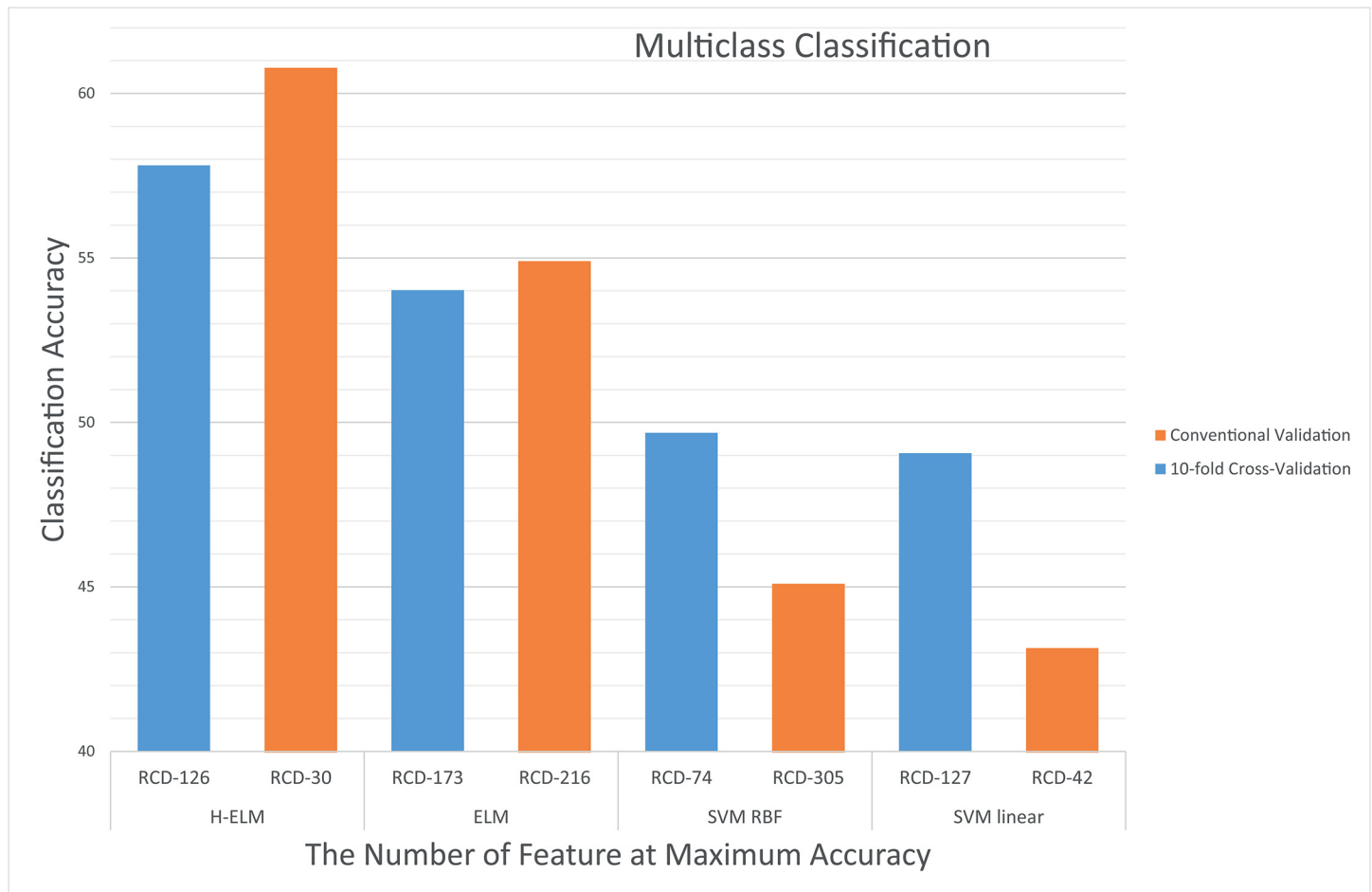


Fig 3. Comparison of the testing accuracy of H-ELM, ELM, SVM linear, and SVM RBF in a multiclass setting using conventional validation and 10-fold cross-validation with RFE-SVM. *RCD* represents the ranked classification dataset size acquired by cumulative RFE-SVM to achieve the highest accuracy.

doi:10.1371/journal.pone.0160697.g003

H-ELM, ELM, and SVM classifiers in ADHD classification. The first 46 *RCD*s include all the peak classification rates in the binary settings. In the case of a multiclass setting, *RCD*₁₂₆ has the highest accuracy with 10-fold cross-validation.

[Fig 3](#) shows a comparison of the multiclass classification rates of all four of the H-ELM, ELM, SVM linear, and SVM-RBF classifiers used in this study. The classification rates are shown for each of the experimental datasets until we achieve the highest accuracy at the corresponding *RCD*.

The permutation testing was performed with 10,000 iterations using the top *RCD* features. The maximum accuracy rate with the corresponding type of validation was used as the test statistic in both binary and multiclass classifications. The test reflected that the H-ELM classifier learned the relationship between the data and the labels with high significance of $p < 0.0001$ in both binary and multiclass classifications.

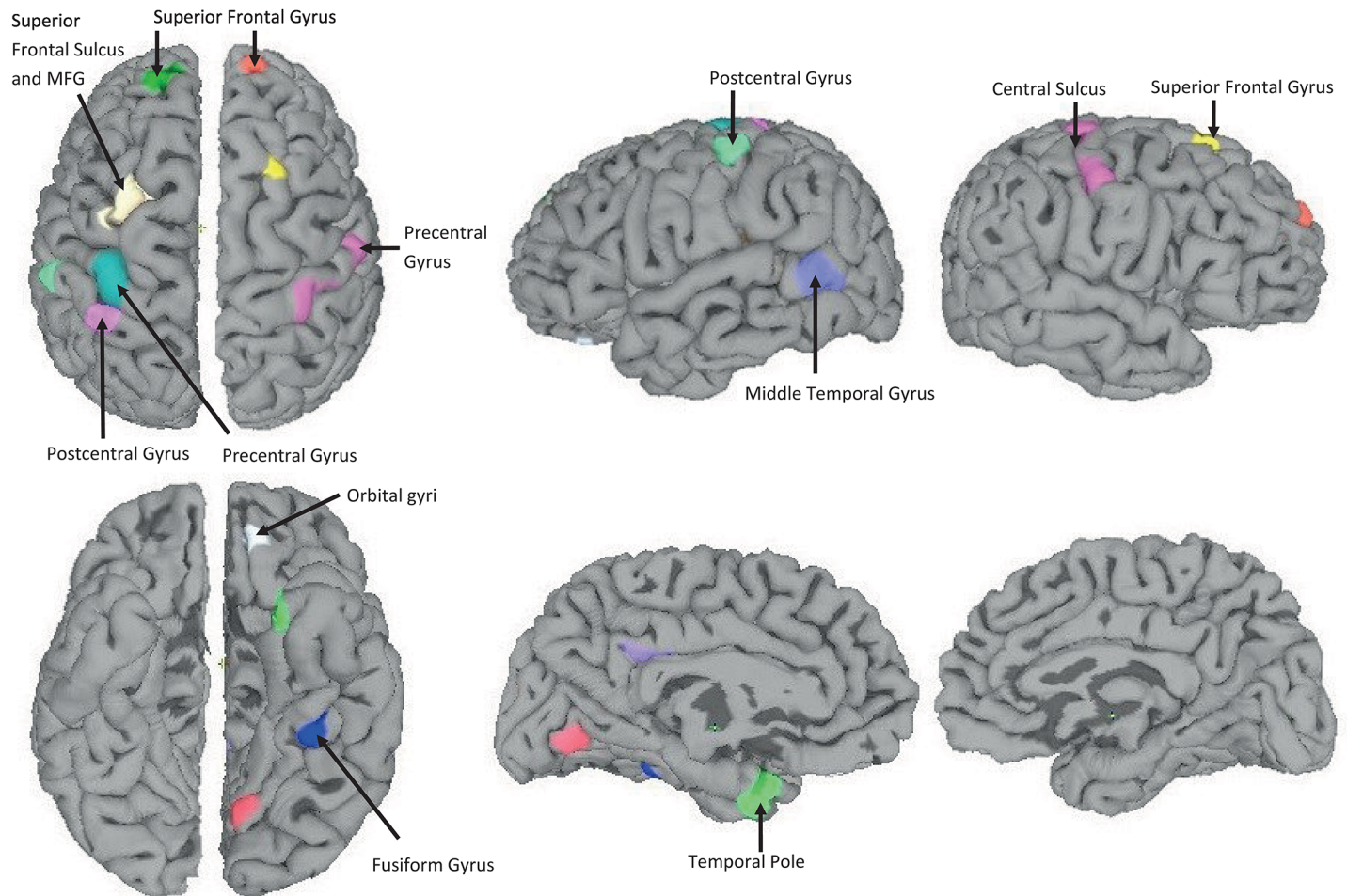


Fig 4. Twelve regions with significant differences as determined by ANOVA. The left column shows the transverse view. The middle and right figures in the first row show the lateral view. The middle and right figures of the second row show the medial view. The most significant results are located in the superior frontal gyrus.

doi:10.1371/journal.pone.0160697.g004

ANOVA Analysis for Cortical Thickness

ANOVA tests showed 10 areas of significant thickness change in the left hemisphere and 2 areas in the right hemisphere ($p < 0.001$, cluster size $50mm^2$) (see Fig 4). In the post-hoc test, we found that the TDC group had 8 areas (left: middle frontal gyrus, precentral gyrus, dorsal and middle postcentral gyrus, middle temporal gyrus, temporal pole, intraparietal sulcus; right: precentral gyrus and central sulcus) that had the highest thickness value as compared to other groups. In contrast, the ADHD-I group had 3 areas (left: orbital gyri and fusiform gyrus; right: superior frontal gyrus) with the highest thickness value. However, none of the highest values was found in any region within the ADHD-C group.

Our study also confirms the previous findings of anatomical abnormalities as mentioned in the introduction of this paper. These regions can be used as potential biomarkers for the automatic diagnosis of ADHD by machine-learning approaches. We identified the surface area of the superior frontal lobe (ROI-26) and mean surface area of the whole cortex as being the

major contributors to differential diagnosis of ADHD along with the cortical thickness and volume of the whole cortex by using multiclass classification.

Discussion

The main result of this study was to achieve high multiclass accuracy by using the H-ELM framework on RFE-SVM-based selected features. To the best of our knowledge, this is the first study in which the H-ELM framework was used for multiclass classification on structural MRI data from the ADHD-200 competition dataset. Our results suggest that we can achieve high classification accuracy between the patient and control groups. In every test setting, the H-ELM framework outperformed basic ELM and two SVM classifiers.

In our study, we utilized the SVM-based recursive feature elimination algorithm. This algorithm selects features that represent higher degrees of significance based on the internal linear SVM-based classification scores, thus making classification significantly more accurate. The RFE feature selector exhibited 85.29% accuracy in binary classification settings. However, in the multiclass classification setting, we obtained 60.78% accuracy.

We utilized the H-ELM framework as the main classification algorithm. Moreover, our noteworthy results and approaches for multiclass classification could assist with the establishment of a clinical diagnosis.

Comparison with Previous Studies

Many studies have been conducted to classify ADHD subjects. Studies based on the use of a binary setting reported accuracies of 77% [42], 65% [43], and 65% [44]. Specifically, a study by [19], for which an accuracy of 90.18% was reported, was conducted only on subjects from the Peking University contribution to the ADHD-200 dataset for binary classification. In our study we used H-ELM, which is the most advanced classification framework for both binary and multiclass classification and has not been used by any other neuroimaging study to date. This framework has the benefit of an embedded ELM-based sparse feature autoencoder. Although the performance of the basic ELM classifier was found to be superior to that of the SVMs, it exhibits lower performance than H-ELM in every setting for both binary and multiclass classification. On the other hand, in a recent study by [45], a classification score of 95% by PCA-based feature optimization and a fully connected cascade artificial neural network classifier for binary settings was claimed. Furthermore, [46] reported a classification score of 82.7% in the binary case and 69.2% in multiclass settings with SVM by using the resting state functional connectivity maps of fMRI data. In addition, one study performed multiclass classification on ADHD, autistic, and normal children with an accuracy rate of 68.2% [47]. Another study by [22] reported 56.87% accuracy in multiclass settings by using single features and 57.71% by using multimodal features and it was ranked 6th in the ADHD-200 competition. However, our results showed higher performance than any study on ADHD in multiclass classification settings by using only structural data, to the best of our knowledge. We used a balanced design approach to select the number of subjects (53 from each diagnosis group) from the ADHD-200 dataset. Among the aforementioned studies, there is no other similar experimental setting for multiclass classification of MRI cortical data; therefore, we cannot directly compare the accuracy of our results to those obtained in any previous research.

Clinical Significance of ANOVA Results

The significant regions from our ANOVA result play an important role to obtain highly accurate differential diagnosis when used in combination with the other significant features that

did not form part of the outcome of the ANOVA analysis. The following description of the regions obtained from our ANOVA analysis may have clinical significance:

In most neuroimaging studies regarding ADHD, the frontal-striato-parietal circuitry and default mode network has been suggested to be involved in ADHD [48]. Especially, the lateral fronto-striato-parietal circuit relates to attention/cognition [49], and the prefrontal cortex relates to working memory [50], executive function [51], motor control [52], and inhibitory control [53]. In addition, reward/motivation were shown to be related to the ventral striatum and orbitofrontal cortex [54]. Moreover, the central sulcus was also reported to be associated with ADHD [18]. These regions showed significant differences between the control and ADHD groups. In our ANOVA result, multiple areas of the frontal cortex, including the MFG, precentral gyrus, and postcentral gyrus in the parietal cortex showed significantly higher cortical thickness in the TDC group. Thus, our findings are consistent with these results, and this might contribute to the high classification accuracy of our model. On the other hand, although not many reports were published, in the temporal lobe, it is evident from [55] that the cortical thickness differs significantly in ADHD patients as compared to the normal control group. The middle temporal gyrus is associated with language abilities, visual perception, multimodal sensory integration, and semantic memory processing. Our findings are consistent with these results.

Significant Cortical Features for ADHD Multiclass Classification

Many previous studies mention that, even if we may find some specific regions with a high order of significance through group differences, still, the features extracted from those regions may not be able to yield high classification accuracies [56]. Group analysis-based features selection does not take into account those regions that have an area smaller than the group analysis output cluster size at a certain threshold value of the significance level. There exists a strong possibility that the regions discarded during group analysis may contain significant information that can be used to discriminate the diagnosis groups. However, the significant features found in our study for classification overlap with those regions that are found to be significant in terms of group differences through the ANOVA test.

In our ANOVA and post-hoc tests, we found that the TDC group showed a higher cortical thickness than the other two groups overall. Moreover, most regions we identified were included in the frontal lobe. Therefore, the most important regions for classification purposes were included in the frontal lobe. A previous study [4] reported that ADHD children showed delayed cortical maturation at an average of 3 years later than that of normal children [48]. Furthermore, most ADHD studies consistently demonstrated the finding that prefrontal and precentral areas were reduced in ADHD patients [57]. Our findings agree with these previous results. Table 6 presents all the significant features for multiclass classification. Indices of the cortical features are based on the DKT31 labeling protocol [24], as provided in Table 3.

Table 6. Most Significant Features for Classification.

Feature Measure (<i>fM</i>)	Left Hemisphere (index)	Right Hemisphere (index)
Mean Cortical thickness (1 - 64)	All (1 - 32)	All (1 - 32)
Surface Area (65 - 128)	26 and 32	26 and 32
Folding Indices (129 - 192)	None	None
Mean Curvature Indices (193 - 256)	None	None
Volume (257 - 320)	All except 4, 30, and 32	All except 4, 30, and 32

doi:10.1371/journal.pone.0160697.t006

Second, during adolescence, normal children experience a cortical thinning phase [58]. Consistent with these findings, the top cortical features in our results can be used as promising biomarkers for automated ADHD classification. We found the surface area of the superior frontal lobe and mean surface area of the whole cortex to be the major contributors for the differential diagnosis of ADHD along with the cortical thickness and volume of the whole cortex by using multiclass classification. The folding indices and the curvature indices were found to be the most insignificant features.

In addition to the multiclass features reported in this paper, one important finding of the current study is that the significant cortical features are dataset dependent therefore, we cannot label any cortical feature as a global biomarker for ADHD classification. Our binary classification results indicate that the significant features are subject to change when we change the sub-group of ADHD for feature optimization. Therefore, it is presently very difficult to claim that any of the cortical features would be appropriate global biomarkers for ADHD discrimination.

Limitations

The major limitation of this study was the imbalanced number of subjects from each diagnosis group, shared from each site. Site bias was intrinsically present in the dataset and it was treated as default. IQ data were missing for some of the subjects; therefore, we did not use IQ information in our study. We only selected the 53 subjects that were statistically matched from each diagnosis sub-group to enable us to conduct a balanced design study. To date, there is no clear guideline available as to the significant anatomical regions that can be used for the differentiation of ADHD subgroups by machine-learning methods; thus, we had to use grid search methodology to identify the most significant anatomical regions for the differential classification of ADHD.

Conclusion

In conclusion, RFE feature selection was applied to ELM, SVM, and H-ELM classifiers, of which H-ELM achieved higher accuracy for both binary and multiclass classification with ADHD neuroimaging data. In addition, our noteworthy multiclass result could assist with clinical diagnosis of ADHD for subtype separation. Moreover, the top cortical features were consistent with previous neuroimaging studies involving ADHD subjects. This kind of multiclass classification approach is challenging; however, we hope that our study will inspire further studies with other clinical data. In a future study, we will perform a multimodal classification with only ANOVA-based features acquired from structural and functional brain MRI data to assess whether they can serve as a good predictor of the test data and increase the classification performance.

Acknowledgments

Authors thankfully acknowledge the ADHD-200 consortium for the released dataset.

Author Contributions

Conceptualization: MQ BL.

Methodology: MQ HJ.

Software: MQ HJ.

Validation: MQ HJ.

Formal analysis: MQ BM HJ.

Investigation: MQ BM BL.

Resources: MQ.

Writing original draft: MQ BM.

Writing review & editing: BM BL.

Visualization: MQ HJ.

Supervision: BL.

Project administration: BL.

Funding acquisition: BL.

References

1. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*. 2015; 56(3):345–365. doi: [10.1111/jcpp.12381](https://doi.org/10.1111/jcpp.12381) PMID: [25649325](https://pubmed.ncbi.nlm.nih.gov/25649325/)
2. Guilherme Polanczyk BLHJB Mauricio Silva de Lima, Rohde LA. The Worldwide Prevalence of ADHD: A Systematic Review and Metaregression Analysis. *American Journal of Psychiatry*. 2007; 164(6):942–948. doi: [10.1176/ajp.2007.164.6.942](https://doi.org/10.1176/ajp.2007.164.6.942) PMID: [17541055](https://pubmed.ncbi.nlm.nih.gov/17541055/)
3. Diamond A. Attention-deficit disorder (attention-deficit/ hyperactivity disorder without hyperactivity): A neurobiologically and behaviorally distinct disorder from attention-deficit/hyperactivity disorder (with hyperactivity). *Development and Psychopathology*. 2005;p. 807–825. doi: [10.1017/S0954579405050388](https://doi.org/10.1017/S0954579405050388) PMID: [16262993](https://pubmed.ncbi.nlm.nih.gov/16262993/)
4. Shaw P, Eckstrand K, Sharp W, Blumenthal J, Lerch JP, Greenstein D, et al. Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proceedings of the National Academy of Sciences*. 2007; 104(49):19649–19654. doi: [10.1073/pnas.0707741104](https://doi.org/10.1073/pnas.0707741104)
5. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5*. 5th ed.; 2013.
6. Castellanos TR F Xavier. Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Reviews Neuroscience*. 2002; 3:617–628. doi: [10.1038/nrn896](https://doi.org/10.1038/nrn896) PMID: [12154363](https://pubmed.ncbi.nlm.nih.gov/12154363/)
7. Solanto MV. The Predominantly Inattentive Subtype of Attention-Deficit/Hyperactivity Disorder. *CNS Spectr*. 2000; 5:45–51. PMID: [18268458](https://pubmed.ncbi.nlm.nih.gov/18268458/)
8. Willcutt EG, Nigg JT, Pennington BF, Solanto MV, Rohde LA, Tannock R, et al. Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes. *Journal of abnormal psychology*. 2012; 121(4):991–1010. doi: [10.1037/a0027347](https://doi.org/10.1037/a0027347) PMID: [22612200](https://pubmed.ncbi.nlm.nih.gov/22612200/)
9. Pineda D, Restrepo M, Sarmiento R, Gutierrez J, Vargas S, Quiroz Y, et al. Statistical analyses of structural magnetic resonance imaging of the head of the caudate nucleus in Colombian children with attention-deficit hyperactivity disorder. *Journal of Child Neurology*. 2002; 17(2):97–105. doi: [10.1177/088307380201700202](https://doi.org/10.1177/088307380201700202) PMID: [11952084](https://pubmed.ncbi.nlm.nih.gov/11952084/)
10. Miller SR, Miller CJ, Bloom JS, Hynd GW, Craggs JG. Right Hemisphere Brain Morphology, Attention-Deficit Hyperactivity Disorder (ADHD) Subtype, and Social Comprehension. *Journal of Child Neurology*. 2006; 21(2):139–144. doi: [10.1177/08830738060210021901](https://doi.org/10.1177/08830738060210021901) PMID: [16566878](https://pubmed.ncbi.nlm.nih.gov/16566878/)
11. Zou Q, Zeng J, Cao L, Ji R. A novel features ranking metric with application to scalable visual and bioinformatics data classification. *Neurocomputing*. 2016; 173:346–354. doi: [10.1016/j.neucom.2014.12.123](https://doi.org/10.1016/j.neucom.2014.12.123)
12. Li D, Ju Y, Zou Q. Protein Folds Prediction with Hierarchical Structured SVM. *Current Proteomics*. 2016; 13(2):79–85. doi: [10.2174/157016461302160514000940](https://doi.org/10.2174/157016461302160514000940)
13. Lin C, Zou Y, Qin J, Liu X, Jiang Y, Ke C, et al. Hierarchical classification of protein folds using a novel ensemble classifier. *PLoS one*. 2013; 8(2):e56499. doi: [10.1371/journal.pone.0056499](https://doi.org/10.1371/journal.pone.0056499) PMID: [23437146](https://pubmed.ncbi.nlm.nih.gov/23437146/)
14. Klouppel S, Abdulkadir A, Jr CRJ, Koutsouleris N, Mourão-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. *NeuroImage*. 2012; 61(2):457–463. doi: [10.1016/j.neuroimage.2011.11.002](https://doi.org/10.1016/j.neuroimage.2011.11.002)

15. Castro E, Gómez-Verdejo V, Martínez-Ramón M, Kiehl KA, Calhoun VD. A multiple kernel learning approach to perform classification of groups from complex-valued fMRI data analysis: Application to schizophrenia. *NeuroImage*. 2014; 87:1–17. doi: [10.1016/j.neuroimage.2013.10.065](https://doi.org/10.1016/j.neuroimage.2013.10.065) PMID: [24225489](https://pubmed.ncbi.nlm.nih.gov/24225489/)
16. Schnack HG, Nieuwenhuis M, van Haren NEM, Abramovic L, Scheewe TW, Brouwer RM, et al. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *NeuroImage*. 2014; 84:299–306. doi: [10.1016/j.neuroimage.2013.08.053](https://doi.org/10.1016/j.neuroimage.2013.08.053) PMID: [24004694](https://pubmed.ncbi.nlm.nih.gov/24004694/)
17. Tang J, Deng C, Huang GB. Extreme Learning Machine for Multilayer Perceptron. *Neural Networks and Learning Systems, IEEE Transactions on*. 2015;PP(99):1–1.
18. Li S, Wang S, Li X, Li Q, Li X. Abnormal surface morphology of the central sulcus in children with attention-deficit/hyperactivity disorder. *Frontiers in Neuroanatomy*. 2015; 9(114).
19. Peng X, Lin P, Zhang T, Wang J. Extreme Learning Machine-Based Classification of ADHD Using Brain Structural MRI Data. *PLoS ONE*. 2013; 8(11):e79476. doi: [10.1371/journal.pone.0079476](https://doi.org/10.1371/journal.pone.0079476) PMID: [24260229](https://pubmed.ncbi.nlm.nih.gov/24260229/)
20. Grant JA, Duerden EG, Courtemanche J, Cherkasova M, Duncan GH, Rainville P. Cortical thickness, mental absorption and meditative practice: Possible implications for disorders of attention. *Biological Psychology*. 2013; 92(2):275–281. doi: [10.1016/j.biopsycho.2012.09.007](https://doi.org/10.1016/j.biopsycho.2012.09.007) PMID: [23046904](https://pubmed.ncbi.nlm.nih.gov/23046904/)
21. Hyatt CJ, Haney-Caron E, Stevens MC. Cortical Thickness and Folding Deficits in Conduct-Disordered Adolescents. *Biological Psychiatry*. 2012; 72(3):207–214. Oxytocin and Social Bonds in Development doi: [10.1016/j.biopsych.2011.11.017](https://doi.org/10.1016/j.biopsych.2011.11.017) PMID: [22209639](https://pubmed.ncbi.nlm.nih.gov/22209639/)
22. Dai D, Wang J, Hua J, He H. Classification of ADHD children through multimodal Magnetic Resonance Imaging. *Frontiers in Systems Neuroscience*. 2012; 6(63). doi: [10.3389/fnsys.2012.00063](https://doi.org/10.3389/fnsys.2012.00063) PMID: [22969710](https://pubmed.ncbi.nlm.nih.gov/22969710/)
23. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*. 2002; 46(1):389–422. doi: [10.1023/A:1012487302797](https://doi.org/10.1023/A:1012487302797)
24. Klein A, Tourville J. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*. 2012; 6(171). doi: [10.3389/fnins.2012.00171](https://doi.org/10.3389/fnins.2012.00171) PMID: [23227001](https://pubmed.ncbi.nlm.nih.gov/23227001/)
25. Colby JB, Rudie JD, Brown JA, Douglas PK, Cohen MS, Shehzad Z. Insights into multimodal imaging classification of ADHD. *Frontiers in Systems Neuroscience*. 2012; 6(59). doi: [10.3389/fnsys.2012.00059](https://doi.org/10.3389/fnsys.2012.00059) PMID: [22912605](https://pubmed.ncbi.nlm.nih.gov/22912605/)
26. Yusta SC. Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters*. 2009; 30(5):525–534. doi: [10.1016/j.patrec.2008.11.012](https://doi.org/10.1016/j.patrec.2008.11.012)
27. Dale A, Fischl B, Sereno MI. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*. 1999; 9(2):179–194. doi: [10.1006/nimg.1998.0395](https://doi.org/10.1006/nimg.1998.0395) PMID: [9931268](https://pubmed.ncbi.nlm.nih.gov/9931268/)
28. Fischl B, Sereno MI, Dale A. Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage*. 1999; 9(2):195–207. doi: [10.1006/nimg.1998.0396](https://doi.org/10.1006/nimg.1998.0396) PMID: [9931269](https://pubmed.ncbi.nlm.nih.gov/9931269/)
29. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*. 2000; 97(20):11050–11055. doi: [10.1073/pnas.200033797](https://doi.org/10.1073/pnas.200033797)
30. Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Medical Imaging*. 2001; 20(1):70–80. doi: [10.1109/42.906426](https://doi.org/10.1109/42.906426) PMID: [11293693](https://pubmed.ncbi.nlm.nih.gov/11293693/)
31. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33:341–355. doi: [10.1016/S0896-6273\(02\)00569-X](https://doi.org/10.1016/S0896-6273(02)00569-X) PMID: [11832223](https://pubmed.ncbi.nlm.nih.gov/11832223/)
32. Cox RW. AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Comput Biomed Res*. 1996; 29(3):162–173. doi: [10.1006/cbmr.1996.0014](https://doi.org/10.1006/cbmr.1996.0014) PMID: [8812068](https://pubmed.ncbi.nlm.nih.gov/8812068/)
33. Huang GB, Zhu QY, Siew CK. Extreme learning machine: Theory and applications. *Neurocomputing*. 2006;70(1–3):489–501. *Neural Networks Selected Papers from the 7th Brazilian Symposium on Neural Networks (SBRN'04) 7th Brazilian Symposium on Neural Networks*.
34. Termenon M, Graña M, Barrós-Loscertales A, Ávila C. Extreme Learning Machines for Feature Selection and Classification of Cocaine Dependent Patients on Structural MRI Data. *Neural Processing Letters*. 2013; 38(3):375–387. doi: [10.1007/s11063-013-9277-x](https://doi.org/10.1007/s11063-013-9277-x)
35. Zhang W, Shen H, Ji Z, Meng G, Wang B. Identification of Mild Cognitive Impairment Using Extreme Learning Machines Model. In: Huang DS, Jo KH, Hussain A, editors. *Intelligent Computing Theories and Methodologies: 11th International Conference, ICIC 2015, Fuzhou, China, August 20–23, 2015, Proceedings, Part II*. Cham: Springer International Publishing; 2015. p. 589–600.

36. Cortes C, Vapnik V. Support-Vector Networks. *Mach Learn.* 1995; 20(3):273–297. doi: [10.1023/A:1022627411411](https://doi.org/10.1023/A:1022627411411)
37. Bellman RE. *Adaptive control processes—A guided tour.* Princeton, New Jersey, U.S.A.: Princeton University Press; 1961.
38. Gori I, Giuliano A, Muratori F, Saviozzi I, Oliva P, Tancredi R, et al. Gray matter alterations in young children with autism spectrum disorders: comparing morphometry at the voxel and regional level. *Journal of Neuroimaging.* 2015; 25(6):866–874. doi: [10.1111/jon.12280](https://doi.org/10.1111/jon.12280) PMID: [26214066](https://pubmed.ncbi.nlm.nih.gov/26214066/)
39. Golland P, Fischl B. Permutation Tests for Classification: Towards Statistical Significance in Image-Based Studies. In: Taylor C, Noble JA, editors. *Information Processing in Medical Imaging: 18th International Conference, IPMI 2003, Ambleside, UK, July 20-25, 2003. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 330–341.
40. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement.* 1960; 20(1):37. doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)
41. Wang D, Miao D, Blohm G. Multi-class motor imagery EEG decoding for brain-computer interfaces. *Frontiers in Neuroscience.* 2012; 6(151).
42. Hart H, Chantiluke K, Cubillo AI, Smith AB, Simmons A, Brammer MJ, et al. Pattern classification of response inhibition in ADHD: Toward the development of neurobiological markers for ADHD. *Human Brain Mapping.* 2014; 35(7):3083–3094. doi: [10.1002/hbm.22386](https://doi.org/10.1002/hbm.22386) PMID: [24123508](https://pubmed.ncbi.nlm.nih.gov/24123508/)
43. Anderson dos Santos Siqueira WECLAR Claudinei Eduardo Biazoli Junior, Sato JR. Abnormal Functional Resting-State Networks in ADHD: Graph Theory and Pattern Recognition Analysis of fMRI Data. *BioMed Research International.* 2014;.
44. Solmaz B, Dey S, Rao AR, Shah M. ADHD classification using bag of words approach on network features; 2012.
45. Deshpande G, Wang P, Rangaprakash D, Wilamowski B. Fully Connected Cascade Artificial Neural Network Architecture for Attention Deficit Hyperactivity Disorder Classification From Functional Magnetic Resonance Imaging Data. *Cybernetics, IEEE Transactions on.* 2015;(99):1–1.
46. Fair D, Nigg JT, Iyer S, Bathula D, Mills KL, Dosenbach NU, et al. Distinct Neural Signatures Detected for ADHD Subtypes After Controlling for Micro-Movements in Resting State Functional Connectivity MRI Data. *Frontiers in Systems Neuroscience.* 2013; 6(80). doi: [10.3389/fnsys.2012.00080](https://doi.org/10.3389/fnsys.2012.00080) PMID: [23382713](https://pubmed.ncbi.nlm.nih.gov/23382713/)
47. Lim L, Marquand A, Cubillo AA, Smith AB, Chantiluke K, Simmons A, et al. Disorder-Specific Predictive Classification of Adolescents with Attention Deficit Hyperactivity Disorder (ADHD) Relative to Autism Using Structural Magnetic Resonance Imaging. *PLoS ONE.* 2013; 8(5):e63660. doi: [10.1371/journal.pone.0063660](https://doi.org/10.1371/journal.pone.0063660) PMID: [23696841](https://pubmed.ncbi.nlm.nih.gov/23696841/)
48. Durston S. A review of the biological bases of ADHD: what have we learned from imaging studies? *Mental retardation and developmental disabilities research reviews.* 2003; 9(3):184–195. doi: [10.1002/mrdd.10079](https://doi.org/10.1002/mrdd.10079) PMID: [12953298](https://pubmed.ncbi.nlm.nih.gov/12953298/)
49. Cubillo A, Halari R, Smith A, Taylor E, Rubia K. 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.* 2012; 48(2):194–215. doi: [10.1016/j.cortex.2011.04.007](https://doi.org/10.1016/j.cortex.2011.04.007) PMID: [21575934](https://pubmed.ncbi.nlm.nih.gov/21575934/)
50. Ehlis AC, Bähne CG, Jacob CP, Herrmann MJ, Fallgatter AJ. Reduced lateral prefrontal activation in adult patients with attention-deficit/hyperactivity disorder (ADHD) during a working memory task: a functional near-infrared spectroscopy (fNIRS) study. *Journal of psychiatric research.* 2008; 42(13):1060–1067. doi: [10.1016/j.jpsychires.2007.11.011](https://doi.org/10.1016/j.jpsychires.2007.11.011) PMID: [18226818](https://pubmed.ncbi.nlm.nih.gov/18226818/)
51. Makris N, Biederman J, Valera EM, Bush G, Kaiser J, Kennedy DN, et al. Cortical thinning of the attention and executive function networks in adults with attention-deficit/hyperactivity disorder. *Cerebral Cortex.* 2007; 17(6):1364–1375. doi: [10.1093/cercor/bhl047](https://doi.org/10.1093/cercor/bhl047) PMID: [16920883](https://pubmed.ncbi.nlm.nih.gov/16920883/)
52. Rubia K, Overmeyer S, Taylor E, Brammer M, Williams SC, Simmons A, et al. Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: a study with functional MRI. *American Journal of Psychiatry.* 1999;. doi: [10.1176/ajp.156.6.891](https://doi.org/10.1176/ajp.156.6.891) PMID: [10360128](https://pubmed.ncbi.nlm.nih.gov/10360128/)
53. Rubia K, Smith AB, Brammer MJ, Toone B, Taylor E. Abnormal brain activation during inhibition and error detection in medication-naïve adolescents with ADHD. *American Journal of Psychiatry.* 2005; 162(6):1067–1075. doi: [10.1176/appi.ajp.162.6.1067](https://doi.org/10.1176/appi.ajp.162.6.1067) PMID: [15930054](https://pubmed.ncbi.nlm.nih.gov/15930054/)
54. Ströhle A, Stoy M, Wrase J, Schwarzer S, Schlagenhauf F, Huss M, et al. Reward anticipation and outcomes in adult males with attention-deficit/hyperactivity disorder. *Neuroimage.* 2008; 39(3):966–972. doi: [10.1016/j.neuroimage.2007.09.044](https://doi.org/10.1016/j.neuroimage.2007.09.044) PMID: [17996464](https://pubmed.ncbi.nlm.nih.gov/17996464/)
55. Fernández-Jaén A, López-Martín S, Albert J, Fernández-Mayoralas DM, Fernández-Perrone AL, Tapia DQ, et al. Cortical thinning of temporal pole and orbitofrontal cortex in medication-naïve children and

- adolescents with ADHD. *Psychiatry Research: Neuroimaging*. 2014; 224(1):8–13. doi: [10.1016/j.psychres.2014.07.004](https://doi.org/10.1016/j.psychres.2014.07.004) PMID: [25085707](https://pubmed.ncbi.nlm.nih.gov/25085707/)
56. Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. *Neuroinformatics*. 2014; 12(2):229–244. doi: [10.1007/s12021-013-9204-3](https://doi.org/10.1007/s12021-013-9204-3) PMID: [24013948](https://pubmed.ncbi.nlm.nih.gov/24013948/)
 57. Valera EM, Faraone SV, Murray KE, Seidman LJ. Meta-analysis of structural imaging findings in attention-deficit/hyperactivity disorder. *Biological psychiatry*. 2007; 61(12):1361–1369. doi: [10.1016/j.biopsych.2006.06.011](https://doi.org/10.1016/j.biopsych.2006.06.011) PMID: [16950217](https://pubmed.ncbi.nlm.nih.gov/16950217/)
 58. Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, Toga AW. Longitudinal Mapping of Cortical Thickness and Brain Growth in Normal Children. *The Journal of Neuroscience*. 2004; 24(38):8223–8231. doi: [10.1523/JNEUROSCI.1798-04.2004](https://doi.org/10.1523/JNEUROSCI.1798-04.2004) PMID: [15385605](https://pubmed.ncbi.nlm.nih.gov/15385605/)