2017

A Structural and Functional Analysis of Human Brain MRI with Attention Deficit Hyperactivity Disorder

Arjun A. Watane
University of Central Florida

A STRUCTURAL AND FUNCTIONAL ANALYSIS OF HUMAN BRAIN MRI WITH ATTENTION DEFICIT HYPERACTIVITY DISORDER

by

ARJUN A. WATANE

A thesis submitted in partial fulfilment of the requirements
for the Honors in the Major Program in Computer Science
in the College of Engineering and Computer Science
and in the Burnett Honors College
at the University of Central Florida
Orlando, Florida

Spring Term 2017

Thesis Chair: Ulas Bagci

ABSTRACT

Attention Deficit Hyperactivity Disorder (ADHD) affects 5-10% of children worldwide. Its effects are mainly behavioral, manifesting in symptoms such as inattention, hyperactivity, and impulsivity. If not monitored and treated, ADHD may adversely affect a child's health, education, and social life. Furthermore, the neurological disorder is currently diagnosed through interviews and opinions of teachers, parents, and physicians. Because this is a subjective method of identifying ADHD, it is easily prone to error and misdiagnosis. Therefore, there is a clear need to develop an objective diagnostic method for ADHD.

The focus of this study is to explore the use of machine language classifiers on information from the brain MRI and fMRI of both ADHD and non-ADHD subjects. The imaging data are preprocessed to remove any intra-subject and inter-subject variation. For both MRI and fMRI, similar preprocessing stages are performed, including normalization, skull stripping, realignment, smoothing, and co-registration. The next step is to extract features from the data. For MRI, anatomical features such as cortical thickness, surface area, volume, and intensity are obtained. For fMRI, region of interest (ROI) correlation coefficients between 116 cortical structures are determined.

A large number of image features are collected, yet many of them may include redundant and useless information. Therefore, the features used for training and testing the classifiers are selected in two separate ways, feature ranking and stability selection, and their results are compared. Once the best features from MRI and fMRI are determined, the following classifiers are trained and tested through leave-one-out cross validation, experimenting with varying feature numbers, for each imaging modality and feature selection method: support vector machine, support vector regression, random forest, and elastic net.

Thus, there are four experiments (MRI-rank, MRI-stability, fMRI-rank, fMRI-stability) with four

classifiers in each for a total of 16 classifiers trained per each feature count attempted. The results of each classifier are the decisions of each subject, ADHD or non-ADHD. Finally, a classifier decision ensemble is created through the combination of the outputs of the best classifiers in a majority voting method that includes results of both the MRI and fMRI classifiers and keeps both feature selection results independent.

The results suggest that ADHD is more easily identified through fMRI because the classification accuracies are a lot higher using fMRI data rather than MRI data. Furthermore, significant activity correlation differences exist between the brain's frontal lobe and cerebellum and also the left and right hemispheres among ADHD and non-ADHD subjects. When including MRI decisions with fMRI in the classifier ensemble, performance is boosted to a high ADHD detection accuracy of 96.2%, suggesting that MRI information assists in validating fMRI classification decisions.

This study is an important step towards the development of an automatic and objective method for ADHD diagnosis. While more work is needed to externally validate and improve the classification accuracy, new applications of current methods with promising results are introduced here.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is one of the most common neurological brain disorders in children, affecting approximately 5-10% of their population worldwide. Because diagnosed children may suffer from learning difficulties, behavioral abnormalities, and disobedience or aggression towards authority, its effects may be detrimental to their health, education, and social skills [1].

Recently, there has been a lot of effort to discover the root cause of this problem, but at present there is no well known biological measure that exists to diagnose ADHD. Instead, physicians and psychologists rely on behavioral symptoms reported by parents and teachers to aid in identifying the disorder. They ask for subjective behavioral observations of inattention, impulsiveness, and hyperactivity. When asked to identify these symptoms however, a person may be subject to confirmation bias, which is the tendency to interpret any evidence as a confirmation of one's belief. As a result, many times the diagnosis may be inaccurate, especially if a parent or teacher believes the child has ADHD.

Nearly one in seven children within the United States and approximately one in five of male children are affected by ADHD, as reporeted by the Center for Disease Control and Prevention. Many physicians and scientists believe this ratio is an obvious mark of over-diagnosis, and therefore this motivates us to develop an objective ADHD diagnostic method. By relying on the brain's cortical structure and functional activity, we aim to standardize the detection process and reduce the dependency of subjective analysis. Furthermore, Dr. Thomas Insel, Director of the National Institute of Mental Health, agrees with this position. Speaking about the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), he states "We need to begin collecting the genetic, imaging, physiologic, and cognitive data to see how all the data - not just the symptoms - cluster and how these

clusters relate to treatment response."

Furthermore, the medication required to treat the inattention symptom of ADHD is prevalent amongst college students. [2] reported after studying 1,811 undergraduates at a large, public university that 34% anonymously admitted to the illegal use of ADHD stimulants. Especially during highly stressful academic periods such as final exams season, the extra "boost" that the medication provides attracts a large market for the underground passing of the drug. How this underground market is created is largely unknown, but the subjectivity of the ADHD diagnostic test may allow those who do not have ADHD to act as if they do in order to gain the medication.

These issues at hand motivate us to focus on this major topic of interest: the identification of specific structural or functional differences in brains with and without ADHD. Through this study and future investigations, multiple biological markers can be reasoned that, when assessed together, may point towards a definite and objective diagnosis of ADHD.

Objectives

In this study, I conduct a comparative analysis of the MRI and fMRI of brains with and without ADHD. By utilizing state-of-the-art software and statistical techniques on clinical images for a highly applicable purpose, I explore the possible correlates of ADHD within the brain.

My main objectives of this study include the diagnosis of ADHD via machine learning methods on **a)** a full structural analysis of MRI and **b)** a full functional connectivity analysis of fMRI, as well as the **c)** identification of significant functional correlations and **d)** identification of significant cortical structures, of the brain.

For a brief overview - first relevant features (such as cortical thickness, intensity, surface area, vol-

ume, etc.) are extracted from structural MRI in brains with ADHD matched by age and gender to normal controls. Next significant qualities separating the ADHD vs normal subjects are identified. Secondly, a full connectivity analysis of functional MRI in brains with ADHD matched by age and gender to normal controls is performed. Significant correlations in the networks of the brains of ADHD vs normal subjects are assessed. Finally, the features are ranked and ensembles of various combinations of classifiers are experimented with to achieve a state-of-the-art 96.2% prediction accuracy in discriminating an ADHD vs non-ADHD brain.

Functional Imaging Techniques

Most of the brain's cognitive activities are performed through communication between neurons via their synapses. This neural signaling is performed through the release and reception of specific neurotransmitter molecules. The transmission process of neural signals through the many branches (axons) of neurons is called conduction. Because these electrical signals often force molecules against their concentration gradients, the process requires an input of energy which is derived from the reactions of adenosine tri-phosphate (ATP). Glucose and oxygen is required for the production of ATP in the mitochondria of cells. Therefore, whenever a region of the brain is activated by a cognitive task, the increase in neural signaling amplifies the local energy requirement. In turn, this requires an uptake of chemicals such as oxygen and glucose in the region in order to fuel the signal's conduction.

The energy required for neural signaling is generated via the oxidation of glucose through a metabolic process called glycolysis. The glucose and oxygen is supplied via blood vessels throughout the brain. It has been observed that the activity in a region of the brain and its local blood flow, oxygen consumption, and glucose uptake are highly correlated. In other words, the increase of brain activity in a region coincides with the increase of their chemical and energy needs in that

specific area. Thus, the metabolism process of the brain is highly informative about its cognitive activities. With this information in mind, brain functional imaging techniques take advantage of these relationships to map activity levels of the brain by measuring its local blood flow and chemical consumptions through radiological tracers. Following are some examples of various functional imaging techniques.

Positron Emission Tomography

In Positron Emission Tomography (PET), the subject is injected with a radioactive isotope which is introduced via a biologically active molecule. After a short duration, the active molecules are concentrated in the desired tissue. The subject is then placed under a scanner which records the radioactive emission of the tracer. Through the process of radioactive decay, the tracer molecule releases beta particles (positrons) which the scanner detects. A remote computer calculates the location of the tracer molecule based on the collected data. While PET has a high spatial resolution (approximately 1-10mm), it is at the cost of a low temporal resolution. Still though in this way, PET can detect blood flow or glucose uptake rate, which as discussed previously, are indirect measures of cognitive activity.

Multichannel Electroencephalography

As described earlier, the neurons communicate with each other via electrical signals through the exchange of ionized particles through the synapses. This communication process causes electrical currents in the brain. Through Multichannel Electroencephalography (EEG), the brains electrical current is recorded for a short period of time. EEG can record the neuronal activity in a very high temporal frequency (in the range of milliseconds), however the spatial resolution is compromised.

Magnetoencephalography

The flow of ionized particles through neurons produces weak magnetic fields within the brain. Magnetoencephalography (MEG) is a functional neuroimaging technique which can record the magnetic fields produced by electrical currents due to neuronal activity. The brain activity level is then mapped with the information from the recorded magnetic fields. Since the magnetic fields are very weak, extremely sensitive magnetometers which use arrays of superconducting quantum interference devices are used. Similar to EEG, it has a very high temporal resolution and low spatial resolution.

Near Infrared Spectroscopic Imaging

Near Infrared Spectroscopic Imaging (NIRSI) is a non-invasive optical imaging technique to measure brain activity. As another functional brain imaging method, NIRSI uses near-infrared (800 nm to 2500 nm) light to measure blood oxygen saturation changes in blood vessels of the brain. It does so by measuring the absorption and attenuation of the near-infrared signals emitted by the source and received by the photodiodes. An advantage of NIRSI is that it is inexpensive and portable, allowing subjects to be measured while they are moving or performing tasks. NIRSI and functional magnetic resonance imaging (fMRI) produces similar data as previous studies have shown close spatial and temporal correlation.

Functional Magnetic Resonance Imaging

Functional Magnetic Resonance Imaging (fMRI) measures signal changes in the brain that are due to changing neural activity. Compared to a regular structural MRI, in fMRI the brain is scanned at a lower spatial resolution but through a higher temporal resolution. During increases of neural activ-

ity, there is an increased demand for oxygen in the localized neurological area. The vascular system compensates for this by increasing the amount of oxygenated blood in the area. This mechanism, referred to as blood-oxygen-level dependent (BOLD), changes the ratio of oxygenated hemoglobin versus deoxygenated hemoglobin. As a result, the varying ratio of the types of hemoglobin affects the MR signal, which is recognized and processed for visual representation.

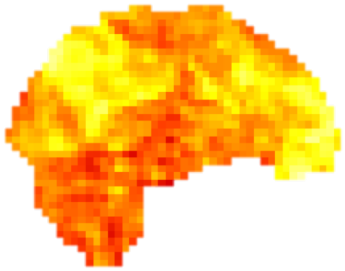

Figure 1.1: Brain fMRI

Structural Imaging Techniques

While functional imaging focuses on activity levels of the brain, structural imaging focuses directly and only on the anatomy. These normally give little to zero information about energy or chemical consumptions, however they do give detailed data about the brain's physical structure. Through state-of-the-art machine learning processes as well, the brain can be statistically analyzed to receive data on its cortical thickness, curvature, density, thickness, volume, intensity, and more. Especially in the past few decades, brain imaging techniques have significantly improved.

Computer Tomography

Computerized tomography (CT) refers to an x-ray imaging procedure in which an x-ray source rotates around a patients body. The x-ray beams produce signals that are processed by the machines

computer to generate cross-sectional images. After a number of successive slices are collected by the machines computer, they are digitally stacked to form a three-dimensional image. CT scans are useful in detecting tumors or lesions within the abdomen, heart disease, clots, and other condition. In particular, CT scans are advantageous when identifying bone fractures and other harder materials; however, it is difficult to distinguish the differences in soft tissue among organs.

Magnetic Resonance Imaging

Because we use Magnetic Resonance Imaging (MRI) for our ADHD analysis studies, we provide the basic principles behind the data capturing method without going into the mathematics. MRI scanners utilize strong magnetic fields, radio waves, and field gradients to generate images of inside the body. The core concept of MRI is based upon the idea of nuclear magnetic resonance (NMR). When placed within an external magnetic field, certain atomic nuclei can absorb and emit radio frequency energy.

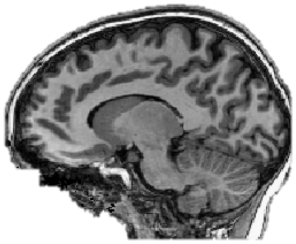

Figure 1.2: Brain MRI

In clinical MRI, hydrogen atoms are most-often used to generate a detectable radio-frequency signal that is recorded by receivers in close distance to the anatomy being examined. Since hydrogen atoms exist naturally in humans, especially in water and fat, most MRI scans essentially map the location of water and fat in the body. Pulses of radio waves excite the nuclear spin energy tran-

sitions, and magnetic field gradients localize the signal in space. By varying the parameters of the pulse sequence, different contrasts can be generated between tissues based on the relaxation properties of the hydrogen atoms. Because MRI does not use any ionizing radiation, it is generally favored in preference to CT or PET. Additionally, whereas CT is weak in deciphering soft tissue, MRIs advantages lie in its relatively strong resolution of soft tissues.

CHAPTER 2: Related Work

While literature on ADHD is vast, imaging studies on ADHD have increased almost exponentially over the past decade. In this chapter, I discuss literature on general computer vision methods on medical imaging, structural studies of ADHD, functional investigations, and finally machine learning and computer vision based medical imaging studies.

Computer Vision in Medical Imaging

Medical image segmentation is an important application of computer science. The simplest method is a threshold based process where a thresholding value is used to identify a region of interest. These methods are simple yet have its disadvantes - they rely mainly on pixel or voxel intensities, but are prone to intensity leakage and poor structural isolation. Because of noise, segmentation through structural means is usually used with assistance from other methods so organs and structures can clearly be identified.

Machine learning methods are very useful because they depend on firmly ground statistical analysis where control data are statistically compared with data positive for disease or abnormality. After creating a model by estimating these statistical variations, new data can be tested to determine the accuracy of the learned machine. Essentially, a high number of imaging features are iteratively sampled to identify image characteristics which may include minute local details or larger global patterns that show a relationship within the data and its label.

With many potential diagnostic and insightful applications, computer vision methods have been used in studies for organ detection, cancer diagnosis, fat quantification, and neural analysis. In [3], an atlas-based rib-bone detection is performed on X-rays. This allows lung function abnormalities

to be more easily detected, especially in countries where health resources are limited through the use of portable X-rays. Furthermore, in [4], the automatic segmentation and quantification of adipose tissue is investigated in PET and CT scans. This study helps the effort in identifying risk factors, prognoses, and long-term health outcomes due to various diseases. This is just the tip of the iceberg in applications of computer science in the medical field; there are a vast ocean of new technologies being developed to assist physicians.

Structural Studies

It is important to realize the structural differences in brains with and without ADHD. [5] showed a quadratic growth model of brain development and defined a growth trajectory of cortical points. In subjects with ADHD, he showed that there was a five-year delay in brains attaining their peak thickness in the cerebrum. The cerebrums development is important because it controls the cognitive functions needed to suppress inappropriate responses. Furthermore, [6] found that ADHD symptoms improve with age. According to their studies, compared to other diseases, delayed neurological maturation seems specific to ADHD. [7] used a global voxelwise approach to examine the hypothesis that ADHD severity is associated with WM microstructure deformalities within the subgenual cingulum. [8] identifies the diminished amount of dopamine receptors in the amygdala, a clinical symptom of ADHD that overlaps with symptoms of bipolar disorder.

ADHD may manifest itself in a structural way, and therefore it is important to understand any structural differences identified in brains of ADHD and non-ADHD subjects. In a study by [5], the growth development of the brain was tracked through a trajectory of specific cortical points and mapped to a quadratic growth model. The results showed that there was a three to five year delay in brains achieving their optimal thickness within the cerebrum. The cerebrum is an important structure pertaining to ADHD because it is a cognitive controller of appropriate and inappropriate

responses. Furthermore, the delayed neurological maturation was shown to be specific to ADHD in [6], and it was also found that its symptoms decrease as the children age. Another study produced by [7] determined using a global voxelwise method that the degree of ADHD is related to white matter structural deformality within the subgenual cingulum. Finally, the concentration of dopamine receptors was identified to be diminished in teh amygdala, which is a similar clinical symptom that overlaps with bipolar disorder [8].

Functional Studies

Functional activity in the brain, shown by oxygen uptake or labeled chemicals may also give some insight into the neurological differences of ADHD affected brains. In a series of studies, differences between ADHD and non-ADHD brains were identified. Altered relationships between larger scaled networks within default networks and task networks of the brain were found in [9]. The results show that there may be a weakened regulatory control of the default network of the brain within ADHD subjects. In another study, [10] reported that ADHD subjects showed a weakened amplitude of low-frequency fluctuations in their study on brain default network functions. While this study did not actually conclude a specific area which was abnormal, it also was only limited to 13 ADHD subjects. Furthermore, [11] quantified maturational effects of the brain on ADHD throughout 400,000 connections between structures in the brain. Ultimately, they found that there was a lag in the development of connections in the default brain network, and that subjects exhibiting more severe inattention symptoms displayed a longer lag in maturation of connections of the default mode network.

Other studies included tasks to test functionality changes. [12] asked subjects to perform the Counting Stroop and determined that ADHD subjects displayed a more significantly diminished activity in the anterior cingulate cortex. In another study, [13] found that when performing the

go/no-go task, ADHD subjects show a weaker activity in the frontostriatal region of the brain. Finally, structural MRI were found to be important in a study by [14] where it was determined that a higher T2 relaxation time in the putamen was found in boys with ADHD, a possible direct correlate to a child's impulsive response.

Furthermore, a review of functional connectivity was put together in [15]. They found ADHD subjects exhibited significantly different resting state functional connectivities bilaterally in the thalamus, cerebellum, insula, and pons. Another study demonstrated a decreased amplitude of low frequency fluctuations (ALFF) in the right inferior frontal cortex, left sensorimotor cortex, and bilateral brain stem. Other analyses revealed that ADHD subjects showed a decrease in the functional connectivity between the anterior cingulate and posterior cingulate, as well as significant alterations in the prefrontal temporal, and occipital cortex regions.

As it can be seen, there seems to be many different locations that play a role in ADHD's emergence. This makes it increasingly difficult to use only single markers as signifiers for ADHD. Thus, a plethora of information incorporated through learning-based classifiers intuitively should provide a better answer.

Machine Learning Studies

In terms of machine learning based classification, although group level statistics were successfully derived to identify various regions of abnormal function and structure of brains in ADHD subjects, an automatic method of diagnosis was lacking. There are relatively few studies exploring the individual level of classification of ADHD subjects.

For a short review of classification studies, a brief review is provided. In [16], the regional homogeneity of fMRI data is used as a feature to classify ADHD subjects. [17] achieved an 85.29%

accuracy on structural MRI data. Peng et. Al in [18] achieved a 90.18% accuracy by using extreme learning machine-based classification of ADHD using structural MRI data. While the ELM method looks promising, the dataset comprised only 55 ADHD subjects and 55 healthy controls. Furthermore, using the same data from the ADHD-200 competition, [19] identified latent dimensions in MRI and compared various non-negative matrix factorization algorithms to achieve On the other hand, a study by [20] received a classification accuracy of 95% by performing a PCA-based feature optimization with a fully connected cascade artificial neural network as the classifier. Finally, a study by [21] performs classification using only single features and achieves a relatively good score of 65.87% accuracy.

CHAPTER 3: Methods

In this study, we analyze structural and functional MRI and apply several statistical measures and machine learning methods to discriminate specific features that differ amongst ADHD and non-ADHD subjects. An array of experiments are performed on MRI features and fMRI features. After features are extracted, several machine learning classifiers are trained and evaluated in their accuracy of diagnosing subjects.

Dataset

Recently, a global competition named ADHD-200 was organized for researchers to develop methods to aid in the automatic diagnosis of ADHD subjects [22]. The organizers released datasets containing fMRI, MRI, and phenotypic data of a large number of ADHD and control subjects. In total, eight different data collection centers contributed to the data set. Since subjects from dfiferent demographic and experimental protocols were used by different data centers, the dataset is comprehensive and fair. In total, the ADHD-200 dataset consists of 776 resting-state fMRI and MRI, in which 491 images were obtained from normal control individuals and 285 from children and adolescents with ADHD. In accordance with the Health Insurance Portability and Accountability Act (HIPAA) guidelines, all data from ADHD-200 is anonymous and no protected health information has been included. Each subject went through a series of tests to accurately determine their labelling as ADHD or control.

For this study, a subset of the data from ADHD-200 is used. I evaluate my methods on a 54 subject sub-subset of the Kennedy Krieger Institute's (KKI) subset of 94 subjects. This sub-subset was selected because their fMRI all had the same number of time points (124), whereas the other 40

subjects had a different number of time points. In addition, the time cost of MRI feature extraction limited the amount of MRI subjects able to be tested. It takes approximately eight hours for one subject's MRI to be pre-processed and features extracted. Thus, for efficiency and to provide a controlled environment, only subjects with an equal number of time points from the same data center were included in the study.

The developers of the KKI dataset performed their measurements with a Siemens Magnetom TrioTim syngo MR B17 scanner, in which every subject was asked to keep their eyes closed. Other fMRI parameters included the time of repetition (TR) set at 2500 ms and time of echo set at 30 ms. MRI parameters included 47 number of slices each with a thickness of 3.0 mm. Of the 54 subjects, 17 were labeled as ADHD and 34 as Controls.

MRI Pre-Processing and Feature Extraction

For MRI preprocessing, we utilized the powerful FreeSurfer software to assist in the controlled adjustment of the data. FreeSurfer is a software package developed for the analysis and visualization of structural and functional neuroimaging data from cross-sectional or longitudinal studies. It was developed by the Laboratory for Computational Neuroimaging at the Athinoula A. Martinos Center for Biomedical Imaging. For our purposes, FreeSurfer provides a full processing stream for structural MRI data, including: skull stripping, gray-white matter segmentation, region labeling on cortical surfaces, and statistical analysis of group morphometric differences [23, 24, 25].

The original T1 anatomical MRI were processed using Freesurfer's *recon-all* command for entire brain segmentation and parcellation. The end result of this pipeline generates segmentations of white matter, gray matter, subcortical volumes, and various statistics. Furthermore, a mesh model of the cortical surface is developed and subdivided into many cortical regions according to

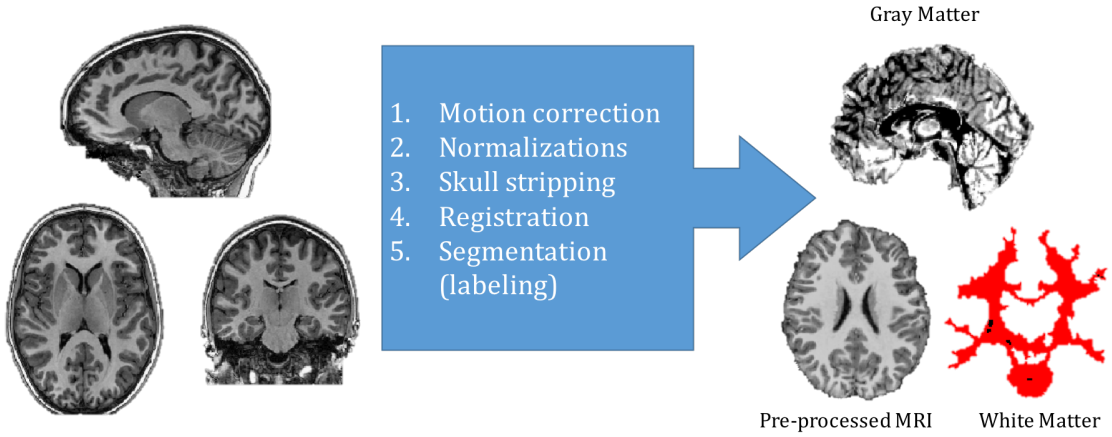

Figure 3.1: MRI Pre-Processing Stages

two atlases, the Desikan-Killiany and Destrieux. Within each region, the software calculates the surface area, gray matter volumes, cortical thicknesses, and cortical Gaussian curvatures. Also measured were non-cortical regions such as white-matter, ventricles, intensities, and CSF. All of these features were considered when selecting features to include in the MRI feature vector.

A number of preprocessing stages were performed including: 1) motion correction and conformation, 2) non-uniform intensity normalization, 3) Talairach transform computation, 4) first intensity normalization, 5) skull stripping, 6) linear volumetric registration, 7) CA intensity normalization, 8) CA non-linear volumetric registration, 9) neck removal, 10) LTA with skull, 11) volumetric labeling and statistics. A more detailed description along with the feature extraction steps can be found on the FreeSurfer website. The final extracted features included in the study are various structures' cortical thickness, intensity, surface area, and volume.

fMRI Pre-Processing and Feature Extraction

The recorded fMRI data must be preprocessed in order to account for variations and center the focus of the analysis on only the relevant structures. For all our fMRI experiments, we used the preprocessed fMRI data released by the ADHD-200 competition organizers. The pre-processing is done using the AFNI [26] and FSL [27] tools and computed on the Athena computer clusters.

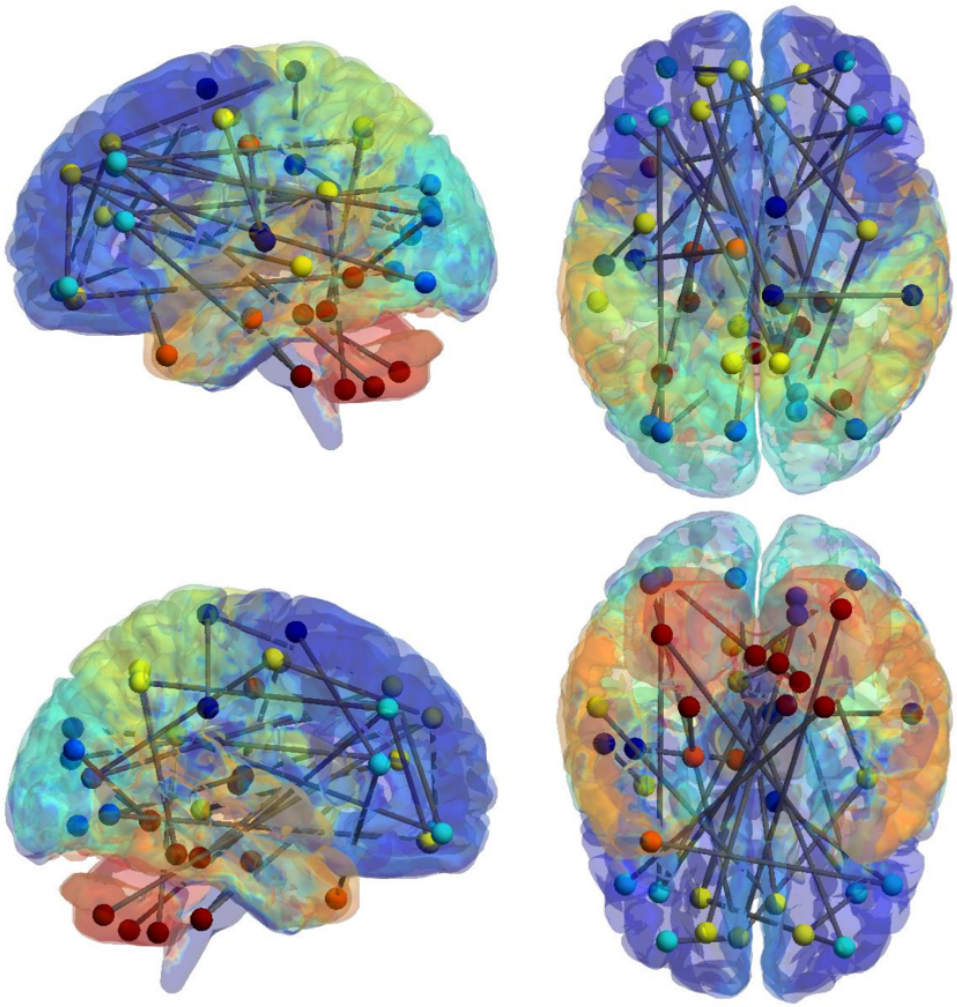

Figure 3.2: fMRI ROI Connection Correlations

The Athena functional data preprocessing pipeline includes 1) removal of the first four EPI vol-

umes, 2) slice timing correction, 3) deoblique of the dataset, 4) reorientation, 5) motion correction to the first image of the time series, 6) masking to exclude non-cortical structures (skull and neck stripping), 7) averaging of the volumes to create a mean image, 8) co-registering the fMRI to its corresponding T1 image, 9) writing fMRI data and mean image into a template space, 10) down-sampling the WM and CSF masks (from the anatomical preprocessing that occurs in parallel but not used for our experiments), 11) time-course extraction for the WM and CSF, 12) regressing out WM, CSF, and motion time courses, 13) band-pass filtering voxel timecourses to exclude frequencies not implicated in functional connectivity, and blurring of the filtered and unfiltered data using a 6-mm FWHM Gaussian filter.

For feature extraction, a toolbox provided by University College London known as Statistical Parametric Mapping (SPM) is used in the analysis of the brain fMRI data sequences. Since the data has already been preprocessed to control for any unwanted variability and unnecessary structures within the scans, SPM is used to extract features. The chosen feature extraction method is a region of interest correlation matrix consisting of the correlation coefficients of the activity between structures of the brain. The Automated Anatomical Labeling (AAL) atlas is used to parcellate the volumes into 116 substructures, and the correlation coefficients between each structures' level of activity is calculated. Thus, there are 116x116 or 6670 correlations which serve as the features used in the classification of ADHD vs non-ADHD subjects.

Feature Selection

A fundamental problem in machine learning and even more so a problem in the application of machine learning to medical imaging is that the dimensionality of the data (the number of pixels and voxels) far outnumber the amount of data that is helpful in identifying a specific class, be it a disorder or disease. Therefore, solutions have been optimized to provide feature selection

algorithms in order to improve classification performance by throwing out non-informative features and including only significant features as part of the training and testing data.

Furthermore, recent studies have shown that in a cluster-wise analysis of fMRI, there is almost a 70% false-positive rate. Thus, identifying valuable features are important in order to improve specificity and sensitivity, and ultimately, the accuracy performance. Especially when considering the complexities and commonalities between brains, many redundant and extraneous features are reported. As a result, they usually degrade the performance of a classifier. Thus, an important component of our method was to only utilize features that are significantly different between ADHD and Control subjects. In this exploratory study, two main feature selection methods are evaluated and compared for their effectiveness: the Wilcoxon Rank Sum t-test feature ranking, and the Stability Selection algorithm.

Feature Ranking

In order to rank the features, the Wilcoxon Rank Sum t-test is performed on each feature of the ADHD set versus the Control set. The Wilcoxon Rank Sum t-test is a non-parametric statistical method, which means that it does not assume that the populations are normally distributed. It is ideal in identifying if a feature is significantly different between the populations. Based on the z-score that is produced by the equation below, the p-value is calculated. Normally, any feature with p-value less than 0.05 is considered significant. In this study however, numerous tests are conducted by ranking the features by p-value from lowest to highest and then selecting a varied amount of features for each experiment. The number of features selected will be explained further on, however for the MRI features, we set a threshold of p-value <0.25 to be included in the ranked list. For the fMRI features, there is no threshold because of the 6670 features, more than 1000

features have a p-value less than 0.25.

$$z = \frac{T_1 - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} \tag{3.1}$$

Stability Selection of Features

Stability Selection is another form of feature selection that capitalizes on a high frequency of selected data based on a regression analysis. In the algorithm, subjects are randomly sampled and a regression is performed on this subsample. The most discriminative features based on their variance of this subsample are selected. This process is repeated 1000 times, and the features receive selection probabilities. Finally, these selection probabilities are ranked, based on their stability score, from highest to lowest.

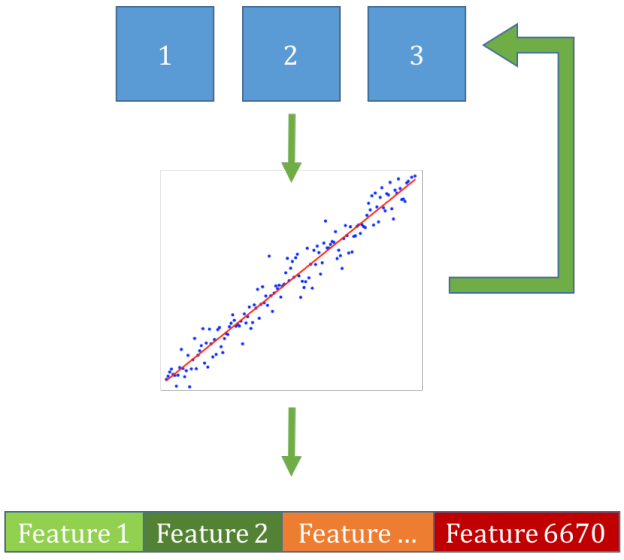

Figure 3.3: Stability Selection Algorithm

Machine Learning

The general method to the machine learning classification approach included first performing either feature ranking or stability selection on the fMRI and MRI features, then carrying out a leave-one-out cross-validation testing approach on four different classifiers while also varying the number of features selected. This provides four sets of data: MRI feature ranking results and MRI stability selection results, as well as fMRI feature ranking results and fMRI stability selection results. Finally, within feature selection method, another set of results were generated by combining the decisions of the classifiers in various combinations in a majority voting fashion, using the number of features for each classifier that generated the best results in the previous four experiments. Therefore, in total there are five sets of results reported.

Number of Features

The classifiers were tested with a varying number of features for both MRI and fMRI. There were less MRI features available than fMRI features, and as a result, a less number of MRI features were used for training and testing the classifiers. For MRI classification, experiments were performed on the top-ranked (both, feature ranked and stability selected) *n* features, with *n* being tested as: *1, 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 75, 90, 100*. For fMRI classification, experiments were performed on the top-ranked *n* features with *n* being tested as: *1, 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, 300, 400, 500, 1000*.

Classifiers

The goal of machine learning is to provide a computer with the ability to learn from data without explicitly programming it. This form of artificial intelligence allows the program to create pre-

dictive models based on past data. For our purposes, we experiment with four different types of machine learning classifiers - elastic net, random forest, support vector machine, and support vector regression - in order to diagnose a subject with ADHD. These classifiers are ”trained” on input data, and then given ”test” data to test the model's accuracy. Each method is founded in statistical analysis, and their general approaches are explained next.

Elastic Net

The elastic net algorithm was first proposed in 2004 by [28] as a regularization and variable selection method. It is quite similar to the lasso algorithm; it does automatic variable selection with a continuous shrinkage, and also can select groups of correlated variables. [28] states “It is like a stretchable fishing net that retains ‘all the big fish’.” In other words, the elastic net promotes the grouping effect where highly correlated features tend to be included or not included in the model together. Especially in the field of image processing and medical imaging, the elastic net is useful when the number of features is much larger than the number of observations (subjects in our case).

Random Forest

Random forests are an ensemble prediction method which is similar to a divide-and-conquer approach [29]. The main principle is that many individual learners are aggregated together to create a single output through a majority voting method. In random forest, multiple decision trees are created by randomly sub-sampling features from the data. Each tree has a different set of weights corresponding to different features. Ultimately, after creating the various decision trees from the training data, the same model is utilized when experimenting with the testing data. As each decision tree outputs a diagnosis of ADHD, the ensemble aggregates the decisions for a final diagnosis. The entire collection of trees is called a forest, and the random forest name comes from the random

feature sub-sampling.

Support Vector Machine

The support vector machine (SVM) algorithm was first described in the 1960s, developed in Russia, and is firmly grounded in statistical theory as it is continuously being developed [30]. Given input data that is labeled as one of two categories (in our case ADHD or Control), the SVM training algorithm essentially builds a model that plots the data onto a space. The space may be highly multi-dimensional, especially when the quantity of features are high [30]. The SVM then finds an optimal hyperplane that divides the two categories of data with the highest separation between the two classes. Once this model is developed, any new observations are similarly plotted onto this space. Depending on which side of the hyperplane the test data falls, the SVM will assign the observation a class (diagnosis of ADHD or Control). Thus, the SVM algorithm is a *non-probabilistic linear binary classifier*.

Support Vector Regression

Similar to the support vector machine, a support vector regressor (SVR) builds a model by plotting the past observations or training data onto a space. The higher quantity of features correlates to a higher dimensionality feature space. Unlike an SVM however, the SVR algorithm does not separate the data based on a linear hyperplane. Instead, it is extended via nonlinear functions to create a model that predicts a probability ratio for the classification of new data [30]. The threshold limit for deciding if an observation is of a certain class (ADHD or Control) can then be manipulated at the user's discretion.

The general testing pipeline is displayed in Figure. First, MRI and fMRI features are ranked or stabilized. Next, the top n features of either the (1) ranked MRI, (2) stabilized MRI, (3) ranked fMRI, or (4) stabilized MRI are selected for classification. A leave-one-out cross validation approach is performed on all 54 subjects. In other words, the classifier or classifier(s) are trained with 53 subjects, and tested on one subject. In the first iteration, the first subject is the test subject and the rest are training subjects. In the second iteration, the second subject is the test subject, and the rest are training subjects (including the first subject), and so on. After 54 iterations, each subject will have been tested once and trained classifiers 53 times. The classifier or classifier ensemble outputs one-by-one the predicted diagnosis of the test subject.

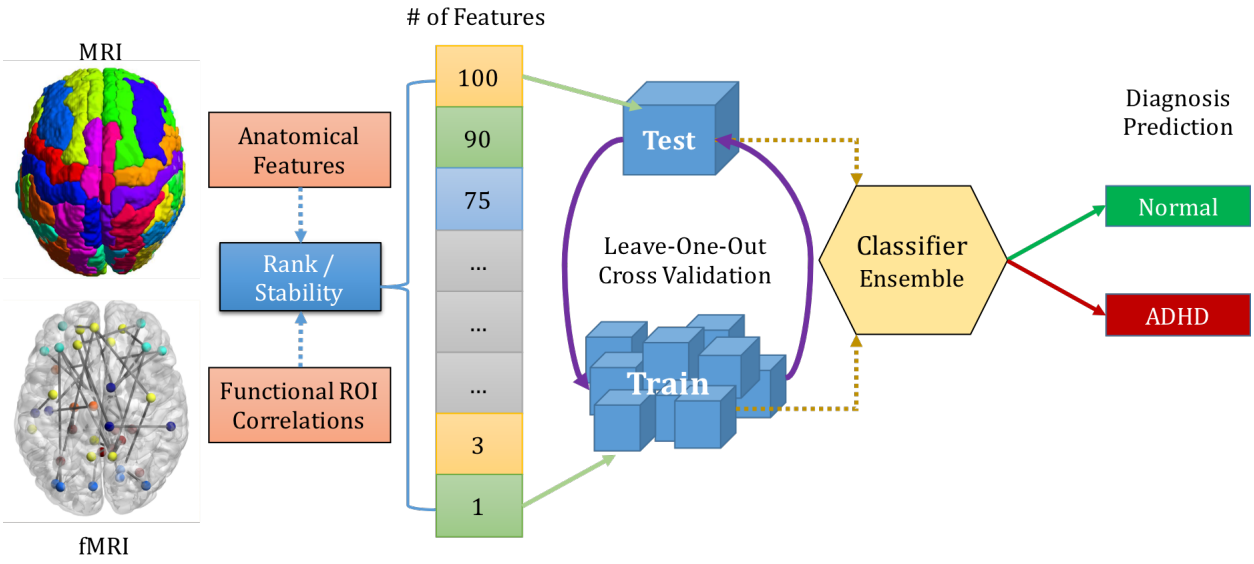

Figure 3.4: Pipeline from Features to Diagnosis

Ensemble Classifier

There are four classifiers being tested and compared with each other: the SVM, SVR, Elastic Net, and Random Forest algorithms. While each respectively learns valuable information in order to predict the test subject's diagnosis, it is not always optimal. Because each classifier analyzes the features in different ways, we believe that combining the decisions of each classifier and ultimately deciding the diagnosis of a subject through a voting process can create more confident predictions of ADHD or Controls.

After saving the predictions of each of the independent classifier experiments, the predictions of the best performing number of features for each classifier are used in the classifier ensemble. The classifier ensemble consists of some or all of the four classifiers for each imaging modality. Therefore for each feature selection method, at a maximum there are eight classifiers used to decide the final prediction, four from MRI and four from fMRI. Different feature selection methods are not included within the classifier ensemble. Various combinations of classifiers are investigated, with the goal of seeing if combining information from both imaging modalities improves performance.

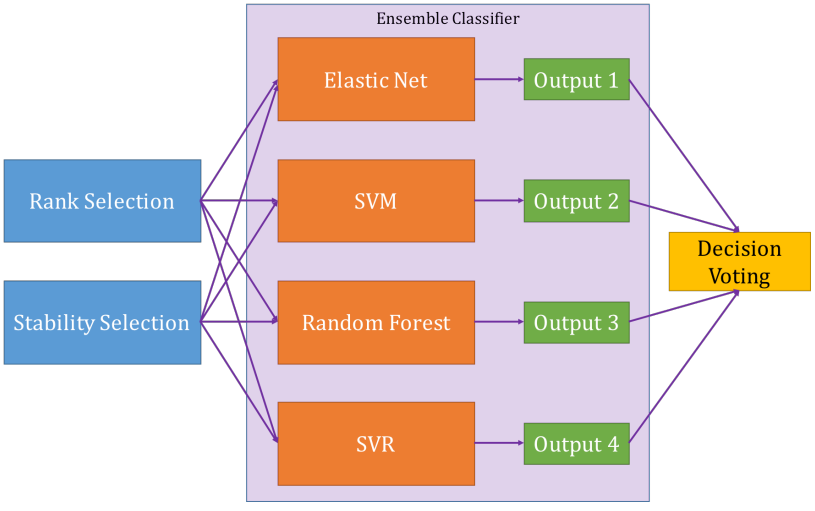

Figure 3.5: Classifier Ensemble Experiments

The classifier ensemble simply adds each subject's predictions from the various classifiers included in that particular experiment. Only if the sum of predictions is greater than two (in some cases, three), then the final prediction for that particular subject is ADHD. In other words, at least two classifiers within the ensemble have to had labeled the subjects as ADHD for the diagnosis to be positive.

CHAPTER 4: Results

An array of experiments were performed and large number of results were generated. For each method of feature selection, four classifiers were trained and tested with the data from MRI and then repeated again for fMRI. This allows us to compare the effects of feature ranking on MRI and fMRI data with stability selection on MRI and fMRI data. Next, the predictions of various combinations of classifiers (including cross-modality combinations) are summed together, and a final label is decided upon each subject based on a majority voting of the particular combination of classifiers. Additionally, the feature ranking algorithm through the use of Wilcoxon's Rank Sum t-test was used to identify the top significant features of the MRI and the top ROI correlations. These results are also reported.

Structural Results

The Wilcoxon Rank Sum t-test was used to compare each features' difference in distributions between the ADHD and Control subject sets. In terms of MRI features, there were very few features that reported significantly different (p-value ¡ 0.05) amongst the two sets. While there were a total of 109 features that achieved a p-value less than 0.25, the Table 4.1 reports the top 10 significant features on the MRI data. In Table 4.1, R = right, L = left, Ant. = Anterior, Sup. = Superior, Mid. = Middle, WM = White Matter.

The independent classifier results for both feature selection methods, the feature ranking method and stability selection method, are graphed below for each experiment on the top *n* features. For detailed results that include the accuracy, specificity, and sensitivity for each experiment, please refer to the appendix.

Functional Results

In fMRI, the Wilcoxon Rank Sum t-test was again used to compare and rank each connectivity correlations' difference in distributions between the ADHD and Control subject sets. There were a maximum of 6670 correlations available because the 116 structures from the AAL atlas were compared amongst each other. Relative to the MRI features, there were far more significantly different (p-value < 0.05) fMRI features amongst the ADHD and Control sets. While there were nearly 300 activity correlation coefficients between structures' that reported a p-value less than 0.05, Table 4.2 reports the top 10 significant features from the fMRI data. In Table 4.2, R = right, L = left, Ant. = Anterior, Sup. = Superior, Mid. = Middle, WM = White Matter.

The independent classifier results for both feature selection methods, the feature ranking method and stability selection method, are graphed below for each experiment on the top n features. For detailed results that include the accuracy, specificity, and sensitivity for each experiment, please refer to the appendix.

Classifier Ensemble

Between a total of eight classifiers (four for MRI and four for fMRI) while keeping the feature selection method results independent of each other, various combinations of majority voting predictions are evaluated. Since there a large number of possible combinations (8! to be exact), the combinations are incrementally and intuitively decided to understand if combining MRI and fMRI predictions improves performance. The results for feature ranking and stability selection are compared.The integers per each bar grouping correspond to the type of classifier included in the ensemble. 1-4 are from MRI, 5-8 are from fMRI. Classifiers 1 & 5 = Elastic Net; 2 & 6 = SVM; 3 & 7 = Random Forest; 4 & 8 = SVR.

Finally, Table 4.3 provides the best results from our experiments and also includes results from other studies.

MRI Features			
Hemisphere	Structure	Attribute	*p-value*
R.	Fusiform	Surface Area	0.00389
R.	Choroid Plexus	Volume	0.00692
R.	Transverse Temporal	Surface Area	0.00937
R.	Rostral Ant. Cingulate	Surface Area	0.01469
L.	Cuneus	Surface Area	0.01846
L.	WM Caudal Ant. Cingulate	Volume	0.02092
L.	Paracentral	Surface Area	0.02196
L.	Sup. Temporal	Surface Area	0.02197
R.	Lateral Ventricle	Intensity	0.02797
R.	Rostral Mid. Frontal	Surface Area	0.03374

Table 4.1: Top ten structural features.

MRI Features				
Structure 1	Hemi. 1	*p-value*	Hemi. 2	Structure 2
Precentral	R.	0.00009	R.	Calcarine
Frontal Mid. Orbital	L.	0.00024	L.	Frontal Sup. Med.
Frontal Mid.	L.	0.00086	R.	Precuneus
Frontal Sup.	L.	0.00086	R.	Frontal Mid. Orbital
Precuneus	L.	0.00092	R.	Cerebellum 4,5
Frontal Mid.	L.	0.00105	R.	Cerebellum 3
Cingulum Ant.	R.	0.00119	L.	Cuneus
Precentral	L.	0.00119	L.	Temporal Sup.
Frontal Sup. Med.	L.	0.00119	R.	Cerebellum Crus 2
Cingulum Post.	L.	0.00119	R.	Cerebellum 9

Table 4.2: Top ten functional ROI connection correlation features.

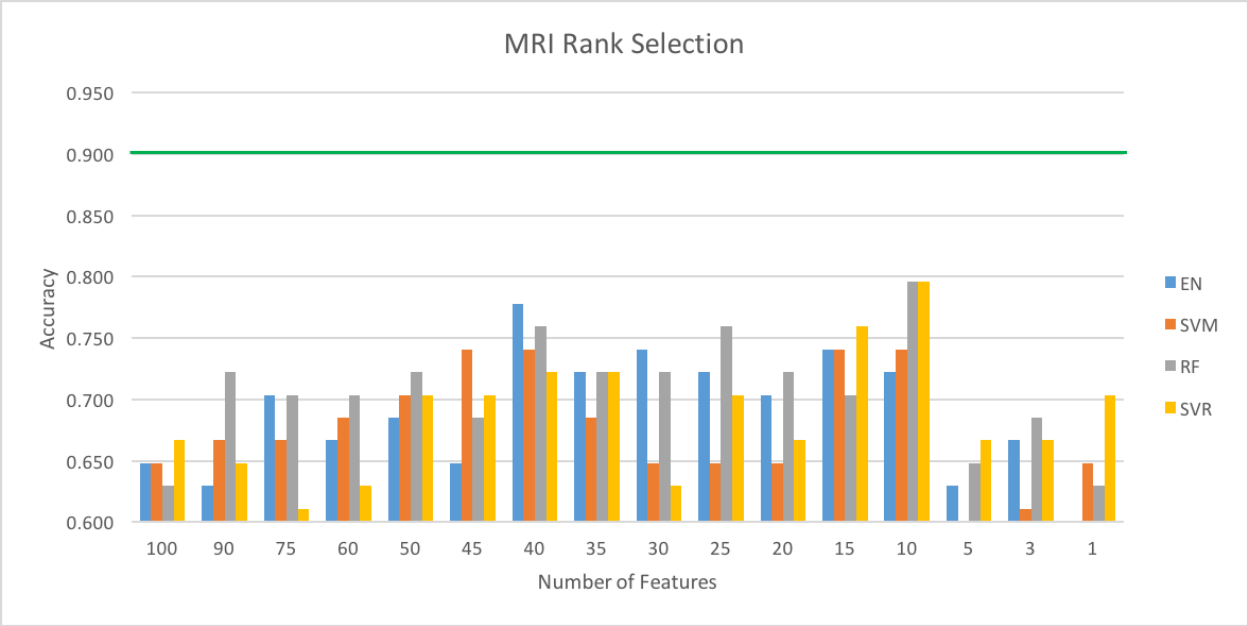

Figure 4.1: Independent classifiers on MRI features selected through the Wilcoxon Rank Sum t-test. Maximum accuracy is 79.6% from Random Forest and SVR classifiers.

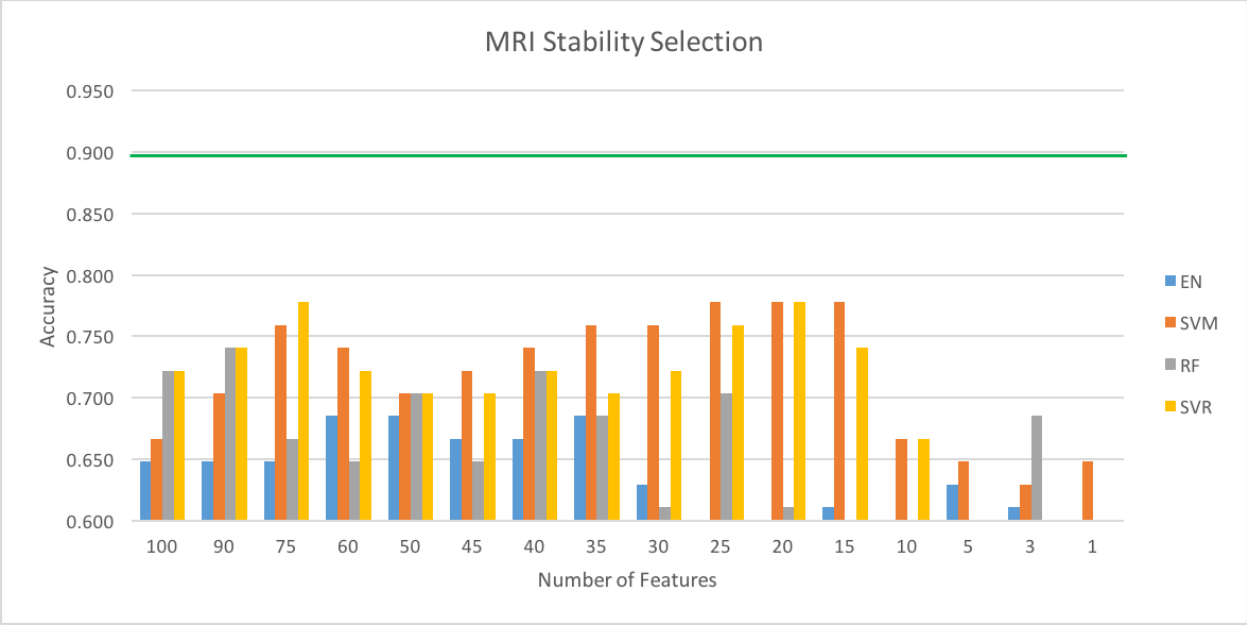

Figure 4.2: Independent classifiers on MRI features selected through the Stability Selection. Maximum accuracy is 77.8% from SVM and SVR classifiers.

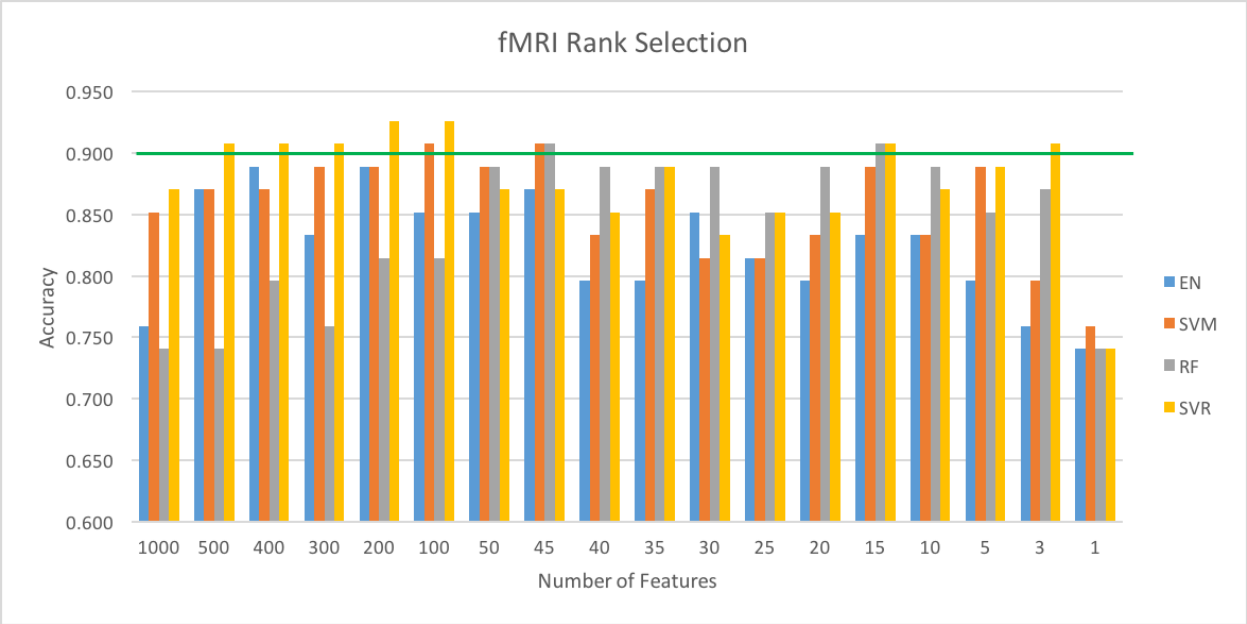

Figure 4.3: Independent classifiers on fMRI features selected through the Wilcoxon Rank Sum t-test. Maximum accuracy is 92.6% from the SVR classifier.

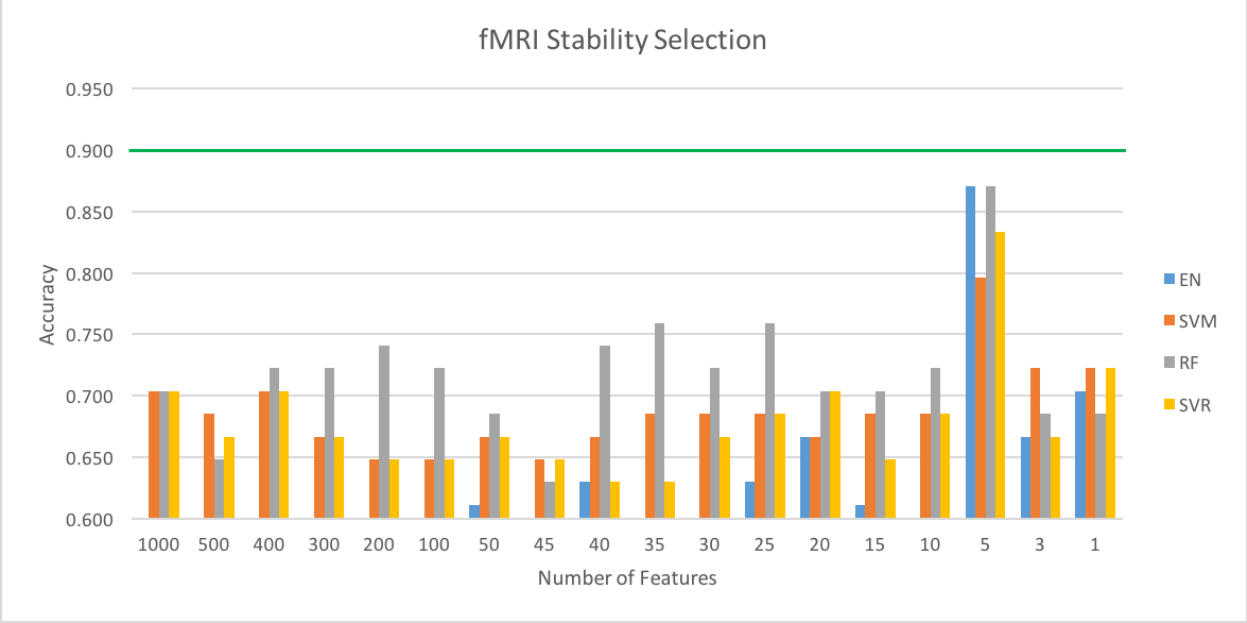

Figure 4.4: Independent classifiers on fMRI features selected through the Stability Selection. Maximum accuracy is 87.0% from the Elastic Net and Random Forest classifiers.

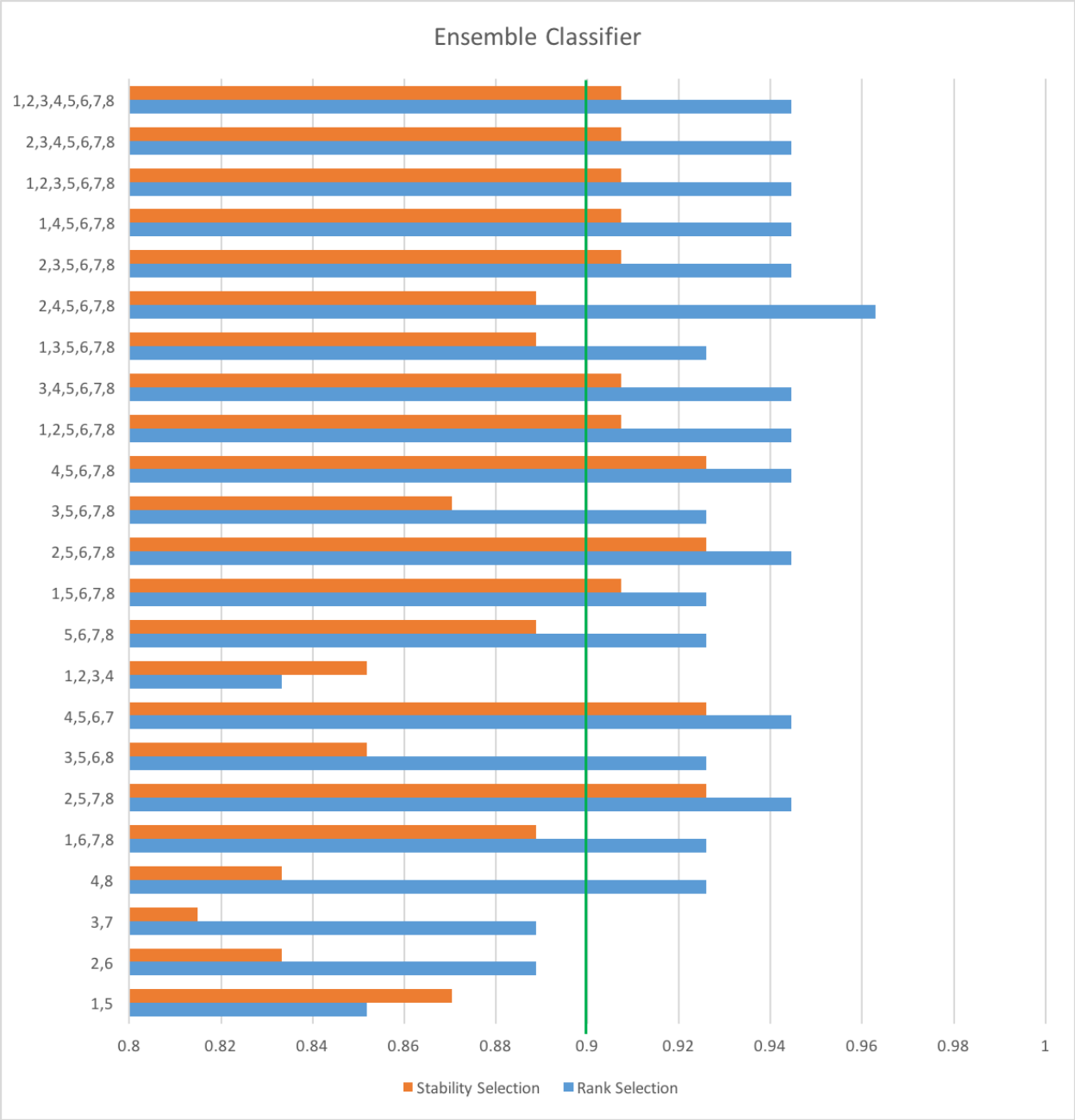

Figure 4.5: Ensemble of classifiers including both MRI and fMRI for both feature selection methods. Classifiers 1 & 5 = Elastic Net; 2 & 6 = SVM; 3 & 7 = Random Forest; 4 & 8 = SVR. Classifiers 1-4 are trained from MRI, 5-8 from fMRI.

Modality	Method	Accuracy	Sensitivity	Specificity
MRI	Stability Selection	77.8%	94.6%	41.2%
MRI	Rank Selection	79.6%	94.6%	47.1%
fMRI	Stability Selection	87.0%	94.6%	70.6%
fMRI	Rank Selection	92.6%	100%	76.5%
Both	Classifier Ensemble (Stability)	92.59%	91.89%	94.11%
Both	**Classifier Ensemble (Rank)**	**96.29%**	100%	88.24%
MRI	Recursive Feature Elimination [17]	85.29%	NP	NP
MRI	Extreme Learning Machine [18]	90.18%	NP	NP
fMRI	Centrality Measures [19]	73%	63%	83%
fMRI	Fully Connected Neural Network [20]	90%	NP	NP
Both	Multi-Kernel Learning [21]	67.79%	38.29%	84.08%

Table 4.3: Overall results and comparison to other studies. Max accuracy is 96.2%. NP = Not Provided. Note - different studies used different data sets and testing methods (i.e. K-fold or leave-one-out cross validation).

CHAPTER 5: Conclusion

Machine learning has proven to be useful in many medical applications, and medical imaging is a field with no exception. We leverage its statistical foundations to train objective models that can accurately and reliably distinguish between ADHD and non-ADHD patients.

In short, structural and functional information from brain MRI and fMRI, respectively, are sources for the features analyzed. The features are selected via two different methods, rank selection and stability selection. Their respective selected feature compositions are tested through the classification system through an array of feature counts. The classification system includes four different machine learning classifiers: support vector machine, support vector regression, random forest, and elastic net. The outputs of the classifiers are the diagnostic decisions, and they are taken either independently or through a majority voting procedure between various combinations of classifiers.

The methods explored in this project serve as a seed for future studies on ADHD classification. While very promising results were received, a few limitations exist in this study. The results would be further strengthened if validated on a larger data set, which is available but was not used because of possible effects from different scanning machines and parameters. With a larger dataset, better training and testing models can be created, allowing for pipeline that can easily be generalized to the larger population, instead of the leave-one-out cross validation approach used here. Additionally, the demographics of the patients matter. This study included both males and females aged 8 to 13. During this young age, males and females' bodies are undergoing many changes, which may affect their brain's structure or function. Separating the genders in building ADHD models will most likely improve results. Finally, methods for combining features between MRI and fMRI should also be explored, as they are currently tested separately. It will be interesting to see if combining MRI features with fMRI features improves results just as voting the results

between MRI classifiers and fMRI classifiers improves results.

A few general points can be understood as a result of the outcomes of this project, and they can be applied in the general fields of neurological disorder diagnostics and medical imaging. First, feature selection methods are important in determining the exact information that should be tested. This study finds that rank selection via the Wilcoxon Rank Sum *t*-test outperforms the stability selection algorithm when using the subset of features on the classifiers. A significant increase in the accuracy of diagnosis was found when using features chosen through rank selection instead of stability selection in both MRI and fMRI data. Additionally, there seems to be a significant activity correlation between and within the brain's frontal lobe and cerebellum amongst ADHD vs. non-ADHD subjects, as well as between the left and right hemispheres, as noted from the top ten fMRI features.

Furthermore, the results favor the statement that ADHD is mostly a functional problem within the brain, but it also has a few anatomical elements. This can be seen in the vast number of functional ROI correlation features with a p-value less than 0.05, as opposed to the approximately ten structural features under that threshold, found through the Wilcoxon Rank Sum *t*-test. This means that there are very few significant anatomical differences, and a lot of significant connection correlation differences between ADHD and non-ADHD subjects. When the features are subsequently tested within the classifiers, the fMRI features outperform the MRI features in diagnostic accuracies.

Because the fMRI features generally produce stronger results for ADHD classification, we tested to see if the predictions from MRI features added any value to the fMRI predictions if they were combined through a voting procedure. The results show that summing the predictions of of fMRI classifiers and MRI classifiers separately improves the results for the two imaging modalities. When summing predictions of different combinations of classifiers from both MRI and fMRI, the results improve even more to 96.2% in one instance. Therefore, multi-modal decision boosts

performance, especially when MRI decisions are added to fMRI decisions.

Objective ADHD diagnostic methods will more than likely be implemented in the near future as technology develops. However, it is clear that machine learning methods provide useful tools in its diagnosis. Whether it is affirming a physician's decision or used as the sole basis for classification, the future is bright for this neurological disorder. This study is a strong step in the right direction for early and accurate diagnosis of ADHD, which will allow preventative treatments before the symptoms affect a child's social, academic, and personal growth.

LIST OF REFERENCES

[1] J. Biederman, "Attention-deficit/hyperactivity disorder: a selective overview.," *Biological psychiatry*, vol. 57, pp. 1215–20, jun 2005.

[2] A. D. DeSantis, E. M. Webb, and S. M. Noar, "Illicit use of prescription ADHD medications on a college campus: a multimethodological approach.," *Journal of American college health : J of ACH*, vol. 57, no. 3, pp. 315–24, 2007.

[3] S. Candemir, S. Jaeger, S. Antani, U. Bagci, L. R. Folio, Z. Xu, and G. Thoma, "Atlas-based rib-bone detection in chest X-rays," *Computerized Medical Imaging and Graphics*, vol. 51, pp. 32–39, 2016.

[4] S. Hussein, A. Green, A. Watane, D. Reiter, X. Chen, G. Z. Papadakis, B. Wood, A. Cypess, M. Osman, and U. Bagci, "Automatic Segmentation and Quantification of White and Brown Adipose Tissues from PET/CT Scans," *IEEE Transactions on Medical Imaging*, vol. 0062, no. c, pp. 1–1, 2016.

[5] P. Shaw, K. Eckstrand, W. Sharp, J. Blumenthal, J. P. Lerch, D. Greenstein, L. Clasen, A. Evans, J. Giedd, and J. L. Rapoport, "Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 49, pp. 19649–54, 2007.

[6] K. Rubia, A. Smith, and E. Taylor, "Performance of Children with Attention Deficit Hyperactivity Disorder (ADHD) on a Test Battery of Impulsiveness," *Child Neuropsychology*, vol. 13, pp. 276–304, apr 2007.

[7] M. Cooper, A. Thapar, and D. K. Jones, "ADHD severity is associated with white matter microstructure in the subgenual cingulum," *NeuroImage: Clinical*, vol. 7, pp. 653–660, 2015.

[8] T. Dresler, B. Barth, T. Ethofer, K.-P. Lesch, A.-C. Ehlis, and A. J. Fallgatter, "Imaging genetics in adult attention-deficit/ hyperactivity disorder (ADHD): a way towards pathophysiological understanding?," *Borderline Personality Disorder and Emotion Dysregulation*, vol. 1, no. 6, pp. 1–9, 2014.

[9] C. S. Sripada, D. Kessler, and M. Angstadt, "Lag in maturation of the brain's intrinsic functional architecture in attention-deficit/hyperactivity disorder," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 39, pp. 14259–14264, 2014.

[10] Z. Yu-Feng, H. Yong, Z. Chao-Zhe, C. Qing-Jiu, S. Man-Qiu, L. Meng, T. Li-Xia, J. Tian-Zi, and W. Yu-Feng, "Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI," *Brain and Development*, vol. 29, pp. 83–91, mar 2007.

[11] C. Sripada, D. Kessler, Y. Fang, R. C. Welsh, K. Prem Kumar, and M. Angstadt, "Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder," *Human Brain Mapping*, vol. 35, pp. 4693–4705, sep 2014.

[12] G. Bush, J. A. Frazier, S. L. Rauch, L. J. Seidman, P. J. Whalen, M. A. Jenike, B. R. Rosen, and J. Biederman, "Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the Counting Stroop.," *Biological psychiatry*, vol. 45, no. 12, pp. 1542–1552, 1999.

[13] S. Durston, N. T. Tottenham, K. M. Thomas, M. C. Davidson, I. M. Eigsti, Y. Yang, A. M. Ulug, and B. J. Casey, "Differential patterns of striatal activation in young children with and without ADHD," *Biological Psychiatry*, vol. 53, no. 10, pp. 871–878, 2003.

[14] S. L. Andersen and M. H. Teicher, "Sex differences in dopamine receptors and their relevance to ADHD.," *Neuroscience and biobehavioral reviews*, vol. 24, pp. 137–41, jan 2000.

[15] K. Konrad and S. B. Eickhoff, "Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder," *Human Brain Mapping*, vol. 31, no. 6, pp. 904–916, 2010.

[16] C.-Z. Zhu, Y.-F. Zang, Q.-J. Cao, C.-G. Yan, Y. He, T.-Z. Jiang, M.-Q. Sui, and Y.-F. Wang, "Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder.," *NeuroImage*, vol. 40, pp. 110–20, mar 2008.

[17] M. N. I. Qureshi, B. Min, H. J. Jo, and B. Lee, "Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: Structural MRI study," *PLoS ONE*, vol. 11, no. 8, pp. 1–20, 2016.

[18] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme learning machine-based classification of ADHD using brain structural MRI data," *PLoS ONE*, vol. 8, no. 11, 2013.

[19] A. Dos Santos Siqueira, C. E. Biazoli Junior, W. E. Comfort, L. A. Rohde, and J. R. Sato, "Abnormal Functional Resting-State Networks in ADHD: Graph Theory and Pattern Recognition Analysis of fMRI Data," *BioMed Research International*, vol. 2014, 2014.

[20] G. Deshpande, P. Wang, D. Rangaprakash, and B. Wilamowski, "Fully Connected Cascade Artificial Neural Network Architecture for Attention Deficit Hyperactivity Disorder Classification From Functional Magnetic Resonance Imaging Data.," *IEEE transactions on cybernetics*, vol. 45, pp. 2668–79, dec 2015.

[21] D. Dai, J. Wang, J. Hua, and H. He, "Classification of ADHD children through multimodal magnetic resonance imaging," *Front Syst Neurosci*, vol. 6, no. September, p. 63, 2012.

[22] ADHD-200 Consortium, "The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience.," *Frontiers in systems neuroscience*, vol. 6, no. September, p. 62, 2012.

[23] F. Ségonne, A. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn, and B. Fischl, "A hybrid approach to the skull stripping problem in MRI.," *NeuroImage*, vol. 22, pp. 1060–75, jul 2004.

[24] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis. I. Segmentation and surface reconstruction.," *NeuroImage*, vol. 9, pp. 179–94, feb 1999.

[25] B. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen, and A. M. Dale, "Automatically Parcellating the Human Cerebral Cortex," *Cerebral Cortex*, vol. 14, no. 1, pp. 11–22, 2004.

[26] R. W. Cox, "AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages," *Comput. Biomed. Res.*, vol. 29, pp. 162–173, jun 1996.

[27] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL.," *NeuroImage*, vol. 62, pp. 782–90, aug 2012.

[28] H. Zou and T. Hastie, "Regularization and variable selection via the elastic-net," *Journal of the Royal Statistical Society*, vol. 67, no. 2, pp. 301–320, 2005.

[29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[30] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, pp. 199–222, 2003.