Diagnosing Autism Spectrum Disorder through Brain Functional Magnetic Resonance Imaging

THESIS

MARCH 2016

Kyle A. Palko, Second Lieutenant, USAF

AFIT-ENC-MS-16-M-123

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENC-MS-16-M-123

DIAGNOSING AUTISM SPECTRUM DISORDER THROUGH BRAIN FUNCTIONAL MAGNETIC RESONANCE IMAGING

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Kyle A. Palko, BS

Second Lieutenant, USAF

MARCH 2016

AFIT-ENC-MS-16-M-123

DIAGNOSING AUTISM SPECTRUM DISORDER THROUGH BRAIN FUNCTIONAL MAGNETIC RESONANCE IMAGING

Kyle A. Palko, BS
Second Lieutenant, USAF

Committee Membership:

Lt Col Ryan D. Kappedal, PhD
Chair

J. O. Miller, PhD
Member

AFIT-ENC-MS-16-M-123

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental condition that can be debilitating to social functioning. Current diagnosis rates for ASD are 1 in 68 children, increasing in the recent years. Diagnosing ASD requires long-term behavioral observation by a specialist often costing thousands of dollars. Previous functional Magnetic Resonance Imaging (fMRI) classification studies have included only small subject sample sizes ($n < 50$) and have seen high classification accuracy. The recent release of the Autism Brain Imaging Data Exchange (ABIDE) provides fMRI data for over 1,100 subjects. In our research, we develop a regularized logistic regression classifier that derives a subject's functional network connectivity (FNC) from their fMRI data to determine whether a subject has autism. We obtained up to 65% classification accuracy, similar to other studies using the ABIDE dataset, suggesting that generalizing a classifier over a large number of subjects is much more difficult than smaller studies. The connectivity among several brain regions of ASD subjects were highlighted in the model as abnormal compared to the control subjects which potentially warrants future investigations about how these regions affect ASD. Although the classification accuracy was lower than what could be considered as clinically applicable, this research contributes to beginning the development of an automated classifier for diagnosing autism.

"You have no responsibility to live up to what other people think you ought to accomplish. I have no responsibility to be like they expect me to be. It's their mistake, not my failing."

Richard Feynman

Acknowledgements

I would like to express my deepest appreciation to my advisor Lt Col Ryan Kappedal, for his valuable guidance and support throughout this effort. By encouraging me to step out of my comfort zone, I was able to learn about different machine learning algorithms as well as improve my Python skills. To my reader, Dr. J. O. Miller. I would like to thank you for your valuable corrections and valuable insights to improve the readability of this thesis for those without a technical background in fMRI or ASD. With my new knowledge on fMRI data and autism, I hope that one day I can apply what I've learned here to other neurological diseases such as Amyotrophic Lateral Sclerosis, or ALS. Perhaps this is my first step of contributing to curing ALS.

Kyle A. Palko

Table of Contents

Page

List of Figures

List of Tables

DIAGNOSING AUTISM SPECTRUM DISORDER THROUGH BRAIN FUNCTIONAL MAGNETIC RESONANCE IMAGING

I. Introduction

According to the National Institute of Mental Health, Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by a wide range of symptoms and levels of impairment or disability in children. These symptoms range from difficulties or deficiencies in social communication and social interaction, restricted, repetitive patterns of behavior, or significant impairment of social functioning (National Institute of Mental Health, 2015). The Diagnostic and Statistical Manual-Fifth Edition, or DSM-V spectrum, currently defines three severity levels: Level 3, "Requiring very substantial support," Level 2, "Requiring substantial support," and Level 1, "Requiring support"(American Psychiatric Association and others, n.d.). Asperger syndrome, often associated as a mild form of autism, was removed from the spectrum in 2012 .

While the exact causes of ASD are unknown, research suggests that both genes and environmental factors play a role. Most children with ASD do not come from a family with a history of autism, suggesting that the cause may be sporadic genetic mutations. While some have hypothesized a link between ASD and vaccines, no studies have ever presented evidence supporting the link (National Institute of Mental Health, 2015).

Living with ASD can cause a substantial burden for both the patients and their families. According to AutismSpeaks, ASD costs a family $60,000 a year on average. Fortunately, public schools provide help for children and teens with ASD and can support them as they grow into adulthood. Once a person with ASD reaches 22

however, public schools' responsibility ends. Families then have to make a choice about living arrangements for their child. Those with mild forms of ASD may be able to live on their own with some support but others may have to continue living at home or at a group home. Professional services are available for those with severe symptoms.

This research is motivated by the potential application of applying automated methods to diagnosing medical disorders through the use of computational methods. These revolutionary methods can provide medical experts the tools to make a faster and more accurate diagnosis, saving both patients and the medical community time and money. Results from these algorithms may also provide insight into the causes of autism and narrow focus for future efforts.

Current ASD diagnosis is a two step process. General screening during routine health checkups provide pediatricians insight into potential developmental problems. Children who demonstrate potential problems are referred to experts for additional evaluation. This second stage is an evaluation by a team of experts and medical pro–fessionals who may diagnose the child with ASD or another developmental disorder (National Institute of Mental Health, 2015). While autism diagnoses are sufficiently accomplished through testing and doctor expertise, a significant problem in the med–ical field is consistency. According to Zijdenbos *et al.* (2002), manual analysis of multiple sclerosis (MS) found variabilities of up to 23% between raters. They found that raters could not distinguish between the automated classifier and a manual rater, suggesting that the automated techniques are just as accurate but less variable than human raters. Fully automating, or at least combining automation and specialist expertise, can hopefully reduce the variance and allow for reproducible analysis while increasing the diagnosis accuracy.

Several studies on the use of functional Magnetic Resonance Imaging (fMRI) for

diagnosing neurological disorders have suggested potential success by identifying key differences between the brains of healthy and afflicted individuals. A combination of fMRI data and applicable classification algorithms could significantly reduce the amount of time required to test and reliably diagnose a patient.

Lord & Jones (2012) are hesitant to rely solely on neuroimaging because although strong, evidence has not yet conclusively proven the links as universal or unique. Behavioral diagnosis must be utilized for a complete diagnosis. However, they do note that several studies have indicated links between brain structures and symptom severity. The use of neuroimaging, they say, must not be prioritized over behavior diagnosis. There has not been a study that can reliably discriminate children with autism from other neurological disorders when they have similar brain patterns. Be–havioral diagnosis may be required to complement imagery classification regardless of the performance of our classifier.

Although research has applied classification techniques to several types of medical disorders, ASD is underrepresented. Fortunately, a recent groundbreaking effort to combine several ASD studies and fMRI images has culminated into the Autism Brain Imaging Data Exchange (ABIDE), a public database with 1,112 total subjects. This database provides an enormous amount of data to help researchers identify key links and apply new techniques to ASD. It is our hope that we can effectively use this database to establish a base for automated classification.

Our goal is to convert our fMRI images to the necessary data structure to perform a thorough analysis. Through experimentation, we expect to develop a robust model using machine learning techniques that will allow for reproducible analysis. While perfect classification is desired, neuroimagery is a difficult field and can have poor results if the data is not properly prepared. Detailed information gained from this research may prove useful for further studies on the topic.

The remainder of the thesis is organized as follows. Chapter 2 presents the literature review and provides details on past research on ASD and machine learning techniques from which we can shape our methodology. The chapter introduces functional network connectivity, our proposed data processing technique, and several classification techniques. Chapter 3 provides the methodology utilized in this thesis. Chapter 4 includes the results and analysis of our work. Finally, chapter 5 presents our conclusions and future research.

II. Literature Review

This section provides the source and motivation for collecting the data. We also discuss functional Magnetic Resonance Imaging (fMRI) to give the reader a back–ground on the techniques employed to obtain the data. Next, preprocessing and converting the images into a usable data structure is presented. Classification tech–niques and methods to determine the most significant features is next, followed by results of studies using the ABIDE database.

2.1 ABIDE Database

Today, about 1 in 68 children are diagnosed with some level of autism spectrum disorder. The prevalence of ASD has increased from roughly 1 in 2000 in studies from the 1960s through the 1980s (Kogan *et al.* , 2009). Specialists attribute this to increased research into the disease, as well as standardized evaluation and diagnosis methods. Genetic research into ASD has established large, open access databases giving researchers the capability to compare raw data. Neuroimaging, however, has failed to keep pace until the recently released Autism Brain Imaging Data Exchange (ABIDE) database.

The neuroimagery data used in this study was obtained through the ABIDE database, a grassroots effort "dedicated to aggregating and sharing previously col–lected resting-state fMRI (R-fMRI) datasets from individuals with ASD" (Di Martino *et al.* , 2014). The database is part of the International Data-sharing Initiative's 1000 Functional Connectomes Project, which aims at collecting and sharing high quality data of several different neurological disorders such as attention deficit hyperactivity disorder.

The database is an aggregate collection of 1,112 subjects from 20 different studies

at 16 international sites. The available data includes raw fMRI as well as processed MP RAGE[1] images. All data within the ABIDE database was compiled through studies on autism.

All fMRI and corresponding phenotype data included in the database has been scrubbed of protected personal identifying info in accordance with HIPPA guidelines. Every image acquisition was with informed consent according to the human subjects research boards at each study's respective institution. Details about each study's guidelines can be found at `http://fcon_1000.projects.nitrc.org/indi/abide/`.

2.2 Functional Magnetic Resonance Imaging

There are two main methods employed in mapping the human brain. The first, electroencephalography and magnetoencephalography provide exceptional temporal resolution (10-100 ms) of neural processes but are limited in spatial resolution (one to several centimeters). The second, and our chosen method, is functional Magnetic Resonance Imaging (fMRI). fMRI techniques can detect changes in blood perfusion, volume, or oxygenation that are thought to accompany neurological activity throughout the brain (Matthews & Jezzard, 2004).

fMRI data is acquired through sequential two-dimensional images, or slices, of the target. The entire brain is imaged through several repetitions of these slices as the machine moves throughout the brain region. Depending on the desired resolution, the brain can be imaged quite quickly, in a range of hundreds of milliseconds to a few seconds (Sladky *et al.* , 2011). MRI maps the distribution and density of water in the brain to create a structural image like in Figure 1. However, fMRI can develop more than just structural data. The ABIDE data utilizes blood oxygenation level dependent (BOLD) contrast to identify activity in the subjects. This BOLD contrast

[1](Mugler & Brookeman, 1990)

data provides spatial resolution of a few millimeters but with temporal resolution of a few seconds; researchers forfeit smaller sample times for a higher resolution voxel. It is suggested that increased blood flow to specific regions in the brain is caused by neurotransmitter action, identifying local signaling (Matthews & Jezzard, 2004).

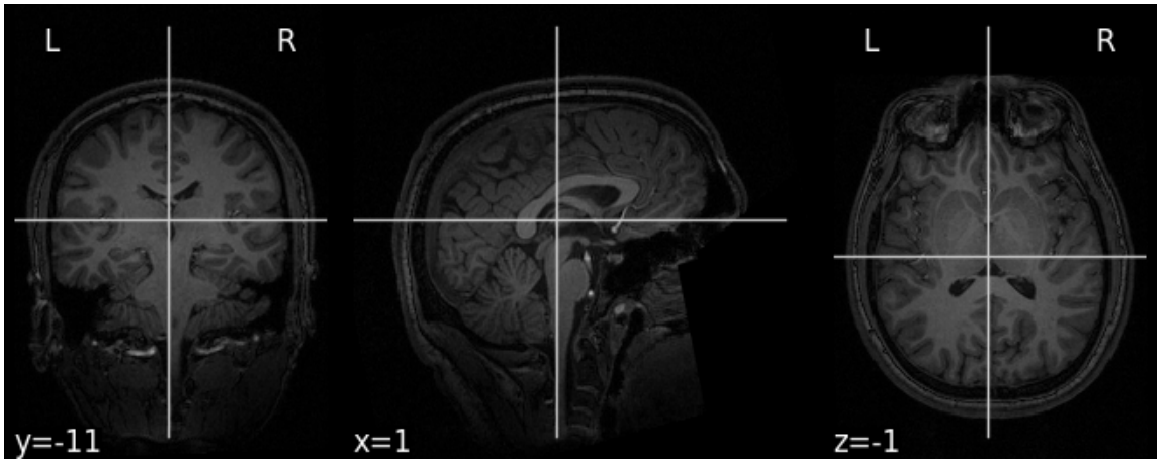


Figure 1. An example of an fMRI brain image after smoothening.

Similar to a flat image, fMRI data includes three dimensional pixel like structures called voxels. In Figure 1 above, each pixel in this two dimensional image corresponds to a single three dimensional voxel. These voxels vary in size depending on the desired resolution of the scan. Many fMRI studies use 3x3x3 mm voxels.

BOLD contrasts are identified from the oxygenation of blood within capillaries. Since an fMRI image is captured by mapping water within the brain, researchers initially sought to investigate changes in the water signal but discovered that these changes were minimal even with neuroactivity. However, when oxygen is captured from the bloodstream, the deoxygenated blood distorts the magnetic field in the surrounding tissue, creating interference of the water signal within the voxel (Ogawa *et al.* , 1990). Therefore, while oxygen extraction decreases as blood flow increases within a region of greater neuroactivity, the MRI signal intensity increases in comparison to the baseline (Matthews & Jezzard, 2004). The BOLD contrasts identify

these significant differences.

2.3 Preprocessing Data

Preprocessing fMRI data is a two step method. The first is cleaning the raw acquisition data. fMRI images are noisy and sometimes do not resemble brains. It is necessary to filter the noise and warp the brain volume to fix movement or other scan errors.

The second step is transforming fMRI data into the desired data type. While technically data, fMRI files can be gigabytes in size and most classification algorithms cannot use the file type. Therefore, the researchers must determine how to process the data into a structure that can be used.

There are several programs that are available for preprocessing data. A popular program is the Statistical Parametric Mapping[2] incorporated through MATLAB. This program is cited in several studies such as Ecker *et al.* (2010a) and Cetin *et al.* (2014). In the next paragraphs, we explain in detail the method of cleaning the fMRI images as accomplished by the Preprocessed Connectomes Project.

ABIDE Preprocessed.

In support of the Preprocessed Connectomes Project (PCP), five teams used four different common neuroimagery processing pipelines to preprocess the data from the ABIDE database. The PCP was developed to provide quality, open-source fMRI data to researchers that has been processed using a variety of pipeline methods. Due to discrepancies in literature, the PCP aims at providing several different preprocessing methods to researchers (Craddock & Bellec, 2015). The four different tools are located in Table 1. Our research uses the data preprocessed using the Configurable Pipeline

[2] `http://www.fil.ion.ucl.ac.uk/spm/`

for the Analysis of Connectomes (CPAC) tool.

Table 1. The four preprocessing pipeline tools used in the ABIDE Preprocessed Connectomes Project

Preprocessing Pipeline
Connectome Computation System
Configurable Pipeline for the Analysis of Connectomes
Data Processing Assistant for Resting-State fMRI
NeuroImaging Analysis Kit

While each software performs specific tasks differently, the overall method is roughly the same. Table 2 displays the first steps for each software. Dropping the first "N" volumes is sometimes used to remove what some consider the machine's warm up period. Slice timing correction is used to correct for the time it takes to acquire a full brain image, as a single pass of the whole volume can take several seconds. Sladky *et al.* (2011) demonstrates how slice timing correction compensates for the effects of these delays. Motion realignment attempts to keep the fMRI image consis–tent by compensating for small shifts in the image caused by either a subject's slight movement or some other force. Finally, intensity normalization is used to account for machine instability and global blood flow changes by normalizing the brightness of the fMRI data.

Table 2. First Preprocessing Steps for Each Pipeline Tool

Step	CCS	CPAC	DPARSF	NIAK
Drop first "N" volumes	4	0	4	0
Slice timing correction	Yes	Yes	Yes	No
Motion realignment	Yes	Yes	Yes	Yes
Intensity normalization	4D global mean	4D global mean	No	Non-uniformity correction using median volume

Each pipeline tool also uses some sort of nuisance variable regression to account

for variation due to physiological processes such as heart beat and respiration, head motion, and scanner drifts in the fMRI signal. Nielsen *et al.* (2014) found that autistic subjects moved significantly more in the scanner than control subjects. Accounting for these noise variables creates cleaner data.

To remove this noise, the pipeline tools correct for motion by identifying common parameters and extracting the signals from white matter (WM) and cerebrospinal fluid (CSF). Signal changes in these areas are generally regarded to be caused by external physiological processes, not brain signal activity (Dagli *et al.* , 1999). The pipelines use the signal from here to correct for motion. Finally, low frequency scanner drifts were corrected. Linear trends are likely caused by the scanner heating up during the process, while quadratic trends may be due to slow, subject movement (Craddock *et al.* , 2015). Table 3 displays the specifics of each tool for the nuisance signal removal.

Table 3. Nuisance Variable Extraction Techniques for each Pipeline Tool

Regressor	CCS	CPAC	DPARSF	NIAK
Motion	24-param	24-param	24-param	Scrubbing of 1st principal component of 6 motion parameters and their squares
Tissue signals	Mean WM and CSF signals	CompCor (5 PCs)	Mean WM and CSF signals	Mean WM and CSF signals
Motion realignment	Yes	Yes	Yes	Yes
Low-frequency drifts	Linear and quadratic trends	Linear and quadratic trends	Linear and quadratic trends	Discrete cosine basis with a 0.01 Hz high-pass cut-off

After nuisance variable regression, each of the pipelines performed four different filtering strategies. These strategies are in Table 4. Band-pass filtering is used to remove signal frequencies that are thought to include mostly noise to increase the signal to noise ratio (Della-Maggiore *et al.* , 2002). The band-pass filtering employed

a 0.01-0.1 Hz filter to remove the noise. Global signal regression is a debated prepro–cessing step in area of neuroimaging. Fox *et al.* (2009) found this step can improve the data quality, while other studies such as Murphy *et al.* (2009) suggest that global signal regression may introduce artificial features in the data. The PCP facilitates experimentation with these controversial strategies by applying the four combinations of the two methods to each fMRI.

Table 4. Band-pass filtering and global signal regression combinations for preprocessed fMRI data

Strategy	Band-Pass Filtering	Global Signal Regression
filt_global	Yes	Yes
filt_noglobal	Yes	No
nofilt_global	No	Yes
nofilt_noglobal	No	No

Finally, each brain was transformed from the original raw fMRI to a template brain (MNI152). Brain volumes differ from person to person in terms of shape and size. By transforming each volume to the same template, the data is standardized and can be compared across subjects. The images were then smoothed using a 6mm Gaussian kernel (Craddock & Bellec, 2015).

The researchers also performed manual quality analysis. Three independent raters screened the data for anomalies and gave their assessment for each subject. The first rater examined the general quality of the preprocessed functional data and derivatives. The other two raters focused on the quality of the raw images and functional data. The data that was identified by raters as potentially failing quality assessment was not included in our study.

Furthermore, there are seven brain region of interest (ROI) atlases available with accompanying time-series data. These seven atlases correspond to many different small regions in the brain that are combined instead of individual voxels. These

ROIs have been highlighted by several years of research as potential regions that play a role in brain connectivity. For example, the Talairach and Tournoux (TT) atlas by Talairach & Tournoux (1988) is one of the standard atlases in the medical field developed by labeling the large folds, or gyrii, found in the brain. Other atlases, such as the Craddock 200, were developed with a statistical mindset rather than labeling based on the biological activity in the brain.

Utilizing all seven atlases available allows for experimentation to determine the strengths and shortcomings in the atlas' applicability to autistic patients. Each atlas is different in resolution, number of ROIs, and location of ROIs, providing a range of different anatomical labeling.

Processing fMRI image files into a usable data format is the next step in classi–fication. Feature extraction transforms the original data into a set of features that can be used by the classification techniques. For neuroimaging, a four-dimensional image may be converted into a vector of features that correspond to the times series of the intensity of a single voxel. These vectors encode either gray or white mat–ter volume for structural data or brain activation for functional data (Orrù *et al.* , 2012). Feature selection may then be applied to reduce the dimension of the data by removing data considered redundant or unimportant while retaining important features. This can be done through several methods, including selecting regions of interest based on previous literature, or to run the data through advanced feature selection algorithms, such as principle component analysis (PCA), which select and combine similar features automatically.

Feature selection is an important step for classifying imagery because training on a reduced number of features increases the classifier accuracy through noise reduction. Furthermore, positive results from feature selection can be used to locate previously unknown areas of interest in the brain. Correctly identifying these regions potentially

allows for further scrutiny by other researchers while reducing computational effort. Finally, by removing irrelevant features, we reduce the dimensionality of the data, reducing computational load and increasing training efficiency (Orrù *et al.* , 2012).

Subtle and variable differences in the fMRI data between autistic and healthy subjects plagued early brain imaging studies. While several studies have concluded significant differences in brain scans, there has been difficulty reproducing the results (Lord *et al.* , 2000). Recent improvements of classification algorithms have contributed to increased success.

Literature suggests correctly processing the brain images increases successful classification. Ecker *et al.* (2010b) normalized high-resolution brain scans of 22 autistic and 22 control subjects using MATLAB-based SPM-5 into the Talairach and Tournoux standard space and partitioned into gray matter, white matter, and cerebrospinal fluid. They used a Support Vector Machine and classified ASD with 81.0% accuracy. They also reported the most important gray matter brain areas.

Ecker *et al.* (2010a) reports on the variety of studies with significant findings on brain structure abnormalities in persons with ASD. These include abnormalities in the frontal, parietal, and limbic regions. The differences between subjects with ASD and controls were found in volumetric and geometric features. Investigations into geometric brain features may identify abnormal intrinsic and extrinsic connectivity patterns (Van Essen, 1997). These patterns may be useful and provide insight into the biological cause of ASD.

Cetin *et al.* (2014) preprocessed images using SPM-5 and demonstrated that Group Independent Component Analysis through single subject PCA and Independent Component Analysis (ICA) in conjunction with feature identification increased schizophrenia identification. The preprocessing effectively reduced the data to 30 components. They primarily investigated cortical connectivity patterns, identified as

functional network connectivity (FNC), over several sensory tasks.

In a survey of autism spectrum disorder studies on brain connectivity, Vidaurre *et al.* (2013) suggested that the discrepancies that arose could perhaps be accounted for with separate models based on the subject's age. They separate the ages into three groups: late childhood (7-12 years), adolescence (12-18 years), and adult (18+ years). In the conflicting studies, it appears that there may be significant differences between the brains of a subject in one group with a subject in another. Several later studies such as Vigneshwaran *et al.* (2015) and Chen *et al.* (2016) include only subjects under the age of eighteen.

Furthermore, most studies report that autistic children under the age of twelve display hyperconnectivity versus the controls. The trend reverses in adulthood, with many reporting that autistic subjects showed a decrease in connectivity between regions (hypoconnectivity). Improperly accounting for age or developmental stage in studies may cause classification problems due to the switch between hypo and hyperconnectivity (Vidaurre *et al.* , 2013).

2.4 Brain Atlases

There are seven different atlases included for use with the preprocessed data by the ABIDE PCP. Each atlas provides a different interpretation of important brain regions based on previous research and the researcher's needs. The PCP researchers warped the Talairach and Tournoux (TT) atlas in Talairach & Tournoux (1988) to the standard template space and then performed nearest neighbor interpolation to fill the brain volume. The Talairach atlas was one of the first atlases developed for neuroimaging research. The TT atlas provided by the PCP researchers contains 97 regions of interest. Craddock *et al.* (2012) developed an atlas specifically for extract–ing neurological activity signals. This atlas, the Craddock 200 (CC200), provides 200

smaller ROIs than in the TT atlas. The seven atlases available for use are in Table 5.

Table 5. Brain atlases warped to MNI152 space for use with preprocessed fMRI images from ABIDE PCP

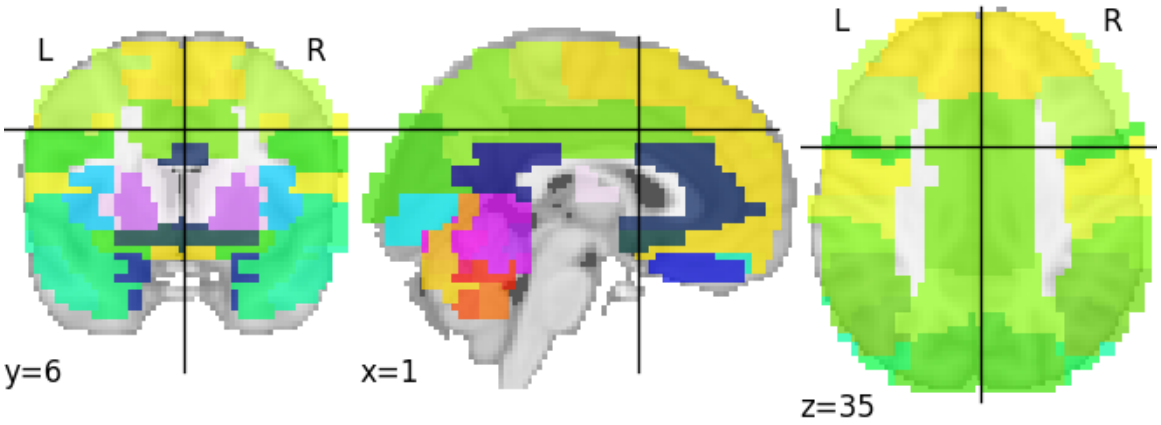
Atlas	Number of ROIs
Automated Anatomical Labeling	116
Eickhoff-Zilles	116
Harvard-Oxford	111
Talairach and Tournoux	97
Dosenbach 160	161
Craddock 200	200
Craddock 400	400

While each atlas segments the brain volume into different regions that are deemed important, many of the regions are similar in location. The following figure provides an example of the differences between the CC200 and TT atlases. Figure 2 illustrates the differences between the ROI segmentation of the two atlases. Each different color represents a distinct brain region.

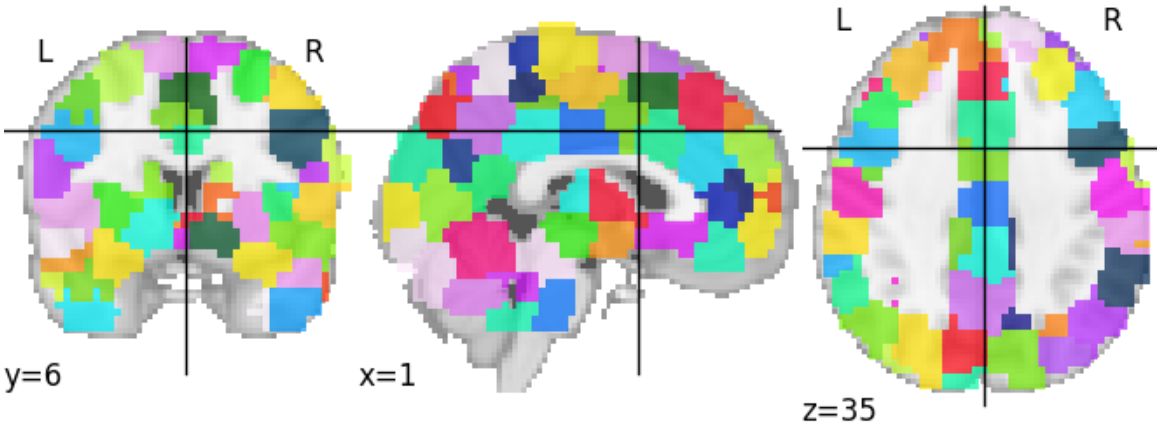
The use of these brain atlases is fundamental in extracting relevant data. Instead of treating each brain volume as thousands of voxels, we can use only a few hundred to present a representation of the entire volume. Since the TT atlas has the least number of ROIs of the available atlases, this serves as the baseline for our experiments. Larger atlases may be introduced over the course of time, limited by computational efficiency.

2.5 Functional Network Connectivity

A common goal among researchers using resting state fMRI data is to iden–tify biomarkers or phenotypes of neurological disorders. The discovery of accurate biomarkers could drastically increase accuracy and reduce time for diagnosis. Func–tional network connectivity is a common fMRI processing technique to extract po–



(a) Talairach and Tournoux



(b) Craddock 200

Figure 2. Comparison of the TT and CC200 brain atlas ROIs. The TT atlas was developed based on the biological structure of the brain whereas the CC200 was developed using statistics.

tential biomarkers.

Functional network connectivity (FNC) can be defined as the "observed temporal correlation between two electro/neurophysiological measurements from different parts of the brain" (Friston *et al.* , 1993). In simpler terms, FNC is how similar are the activity levels of two different regions in the brain. Figure 3 illustrates activity levels in two different voxels throughout the first fifty slices of the fMRI scan. This method can be quite useful as the large, four dimensional fMRI images reduce to a simple correlation structure. Correlation structure refers to the correlation of regions over

time in the same subject.

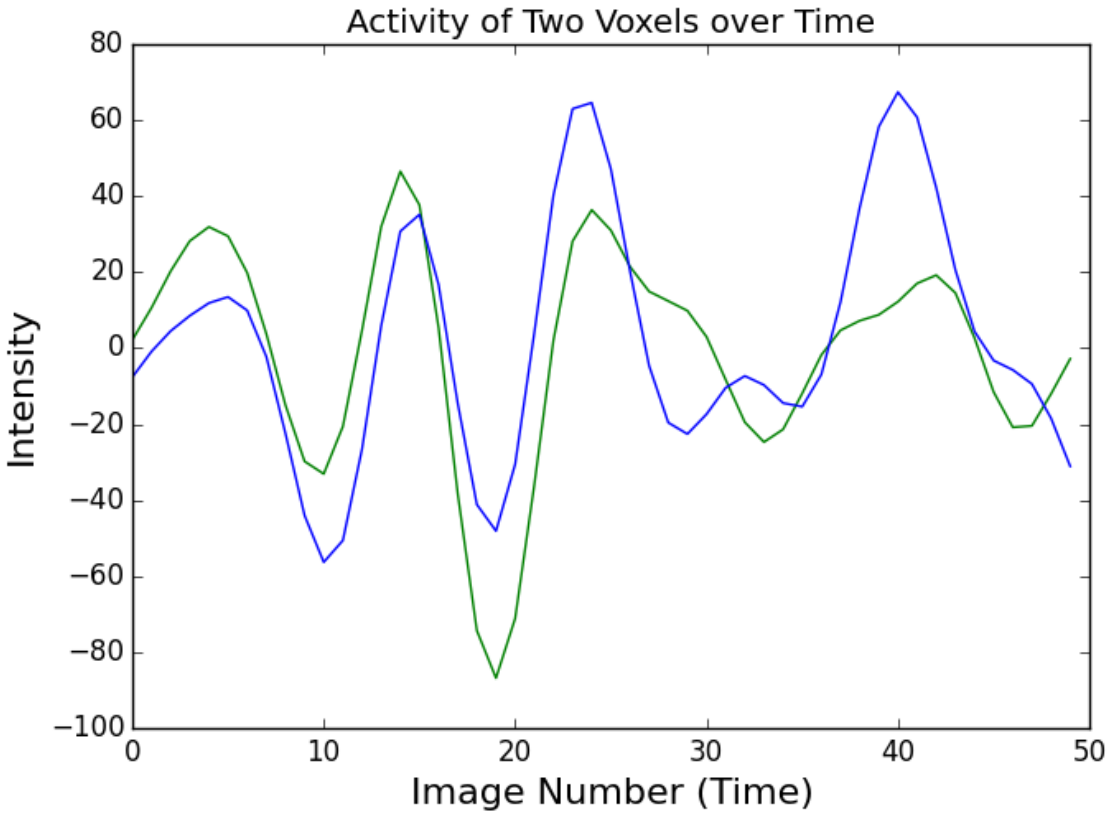


Figure 3. Comparison of BOLD contrast changes of two separate voxels over 50 fMRI slices.

Although not fully investigated, FNC may prove to be a powerful tool for classifying patients with ASD. Investigating how FNC differs between schizophrenic and control subjects while performing different tasks identified significant differences between dynamic FNC in several regions of the brain for schizophrenic patients (Cetin *et al.* , 2014). Research notes that the symptoms of autism are similar to the negative symptoms of schizophrenia, even suggesting that the two disorders share corresponding overlap in neural systems disruptions (Just *et al.* , 2007).

Advances in machine learning algorithms have provided data analysts the opportunity to apply these techniques to medical diagnoses through fMRI images. Recent efforts to classify schizophrenia using FNC data derived from fMRI brain images from

Cetin *et al.* (2014) yielded accuracies of over 92.8% in an online Kaggle competition. The winning model utilized Gaussian process classification where the parameters were tuned over only 83 labeled training data observations (Lebedev, 2014). After train–ing, the model predicted whether each 119,748 test observations were schizophrenic. The high accuracy of several teams using different classifier methods suggests that advanced machine learning techniques may be robust for properly processed fMRI data.

As explained before, an fMRI image is really a four dimensional data structure; multiple three dimensional images over time. Converting this image to the data structure required to compute the FNC values is completed by applying the chosen atlas to the fMRI data and extracting the mean intensity values for each ROI from each of the time slices. We now have a time series data structure for each of the ROIs from the atlas. The "shape," like in Figure 3, of the activity for each ROI is compared to one another to calculate correlation coefficients, our FNC values.

The correlation structure of the images depends on how the researchers approach the problem. For example, Nielsen *et al.* (2013) uses individual voxels for a total of 26.4 million correlation coefficients while Cetin *et al.* (2014) uses 34 regions of interest and 544 coefficients. The number of connections in the data is a function of the number of regions defined by the researcher. The structure flattens to an $n \times n$ matrix, where n is the number of regions of interest. Therefore, the total number of connections, N is defined as,

$$N = \frac{n^2 - n}{2} \tag{1}$$

where the matrix is symmetrical and all diagonals on the matrix correspond to the correlations between the same region (and will always be 1). We use the upper diagonal of the matrix without the diagonal values.

The maximum likelihood of the correlation coefficients is calculated by,

$$\rho_{ij} = \frac{\sum_{i=1}^{n}(p_i - \bar{p_i})(p_j - \bar{p_j})}{\sqrt{\sum_{i=1}^{n}(p_i - \bar{p_i})^2 \sum_{i=1}^{n}(p_j - \bar{p_j})^2}} \tag{2}$$

for $i \in \{1, ..., n\}, j \in \{i, ..., n\}$, or more simply, $\rho_{ij} = Cov(p_i, p_j)/\sigma_i\sigma_j$, where p is a node (ROI). A sample correlation matrix is below in Figure 4.

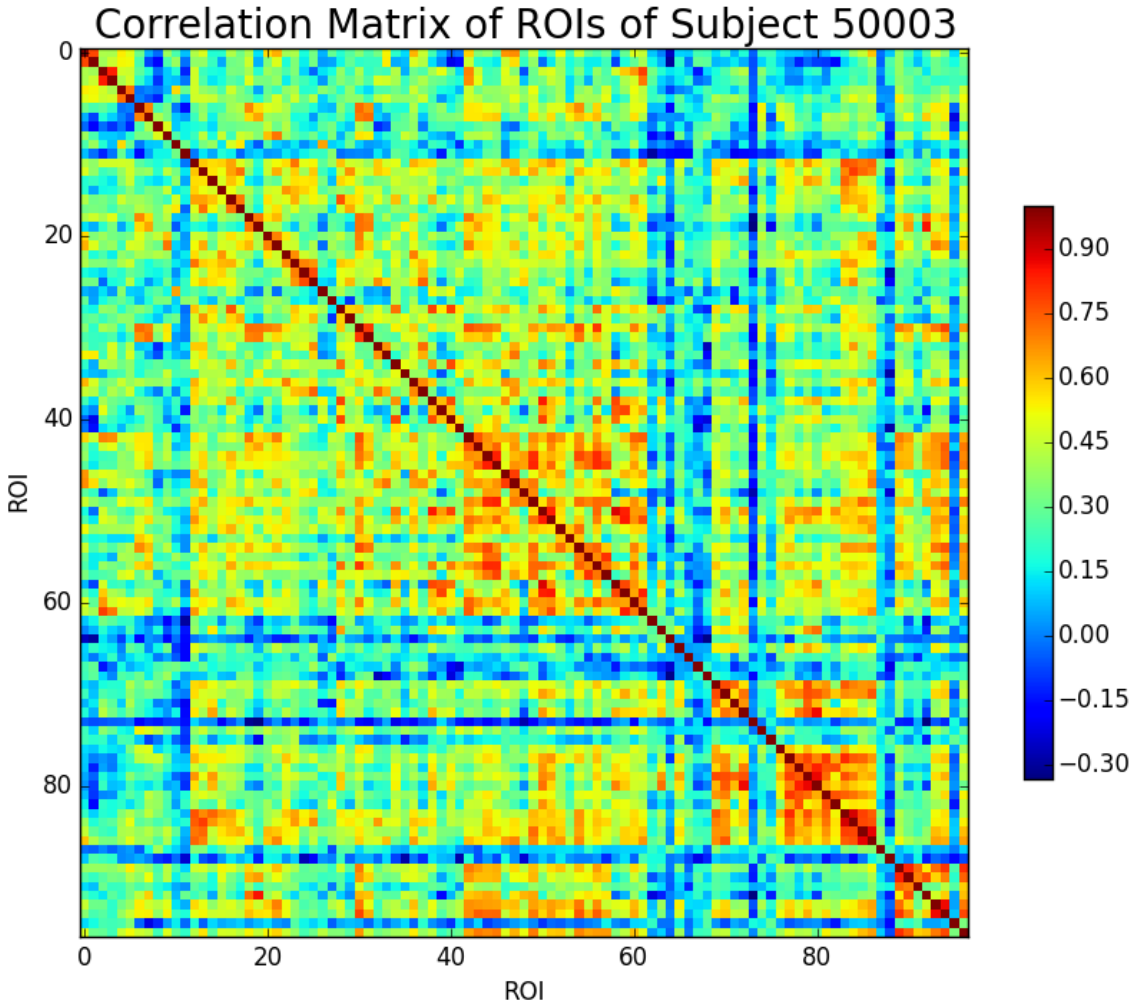


Figure 4. Subject 50003 correlation matrix using the Talairarch and Tournoux atlas represented as a heatmap.

Biswal *et al.* (1997) identifies an important shortfall of using FNC data. Correlation coefficients are calculated based on the similarity of the trends of two time

series vectors. The magnitude of the activity in the vectors is not accounted for by these coefficients, wherein we encounter a potential problem. Without accounting for the magnitude, two vectors with low but similar activity levels will have the same coefficient as two vectors that display high, similar activity. Without adjunct data, we cannot tell if two ROIs are both active or inactive.

Another challenge of functional connectivity data is the strength of the connections is only an estimate. Correlation magnitudes differ in variance between subjects and sometimes between a single subject at different points in the fMRI. As the fMRI data in this study are resting state, we cannot be certain that every subject remained fully motionless with a clear mind for the duration of the scan. While smoothening and motion realignment may be able to correct the major differences, there still remains variability within the data. Birn *et al.* (2013) found that increasing fMRI scan times from 5 to 13 minutes significantly increases the reliability of the connection estimates.

Identifying the coefficients that are most important across all of the subjects can potentially highlight the connections of most importance throughout the brain. Only a few of these connections are usually required to explain most of the variance in the data (Friston *et al.* , 1993).

2.6 Prior Autism Studies

There have been several smaller studies investigating autism using fMRI data. While many focus on building a model to classify autistic subjects from controls, several primarily focused on the differences between brain connections.

Anderson *et al.* (2011) investigated whether a "whole-brain distribution" of con–nectivity could indicate a subject with autism. The researchers used a whole-brain lattice with 7,266 regions of interest covering the entire gray matter. With 40 ASD subjects and 40 controls, their leave-one-out classifier had 79% accuracy, but con–

trolling for subjects under twenty years of age improved accuracy to 89%. Increasing the age of the subjects declined accuracy, indicating that early testing may provide more insight using developing brains.

Tyszka *et al.* (2014) investigated 19 high-functioning autistic and 20 control adults and found that the differences between FNC correlation structures were insignificant. They found that connection strength between different regions were not significantly abnormal. They did find some evidence that connections within brain regions were slightly lower in the autistic subjects versus the controls. This research challenges other such research suggesting a hypoconnectivity between regions in autistic sub–jects.

Calderoni *et al.* (2012) found $\text{AUC}_{max} = 0.80$ through the use of support vector machine on whole-brain analysis of female children. While not exactly classification accuracy, an AUC score of 1 is a perfect classification while a score of 0.5 represents a classifier that cannot distinguish between a binary classification. AUC scores relay general information about the ratio of true positives to false positives.

Ecker *et al.* (2010b) saw 86% classification accuracy with the use of SVM on structural MRI scans of patients. They found the method worked better with gray matter analysis compared to white matter. Further study concluded that SVM could also obtain success through a combination of volume and geometric brain features (Ecker *et al.* , 2010a).

2.7 Machine Learning Techniques

Although plagued by lack of success and generally unreplicable autism studies in the past, more powerful machine learning algorithms are constantly developed (Lord *et al.* , 2000). These improvements may provide an algorithm robust enough to successfully classify large, noisy data sets such as brain data. Not only do these

improvements boost classification accuracy, most also reduce the processing time required to train a model. This is also important as faster model training means more thorough parameter optimization, thereby increasing accuracy again.

The goal of classification in machine learning is to train a model to successfully map inputs, x, to outputs, y. With labeled data, we call this supervised learning. Classification can be summarized as training an estimated function, $y = \hat{f}(x)$, using our data and labels in order to approximate the true function. Our estimated function is then used to predict labels for novel data. Many of the techniques in this research provide the capability of returning a probabilistic prediction (Murphy, 2012). For ambiguous cases, probabilistic prediction can be more useful than outright classification. This may also be true for ASD, where a probabilistic prediction could be more useful for diagnosis and allow for expert judgment in difficult cases. Thresholding, defined as the probability value at which a subjects is classified as a zero or one (typically 0.5 classifies as a 1) could be used to diagnose these hard to classify cases. Weighing the tradeoff between false positives and false negatives could change the threshold value.

Support Vector Machine.

Support vector machine (SVM) is a classification technique that combines a kernel and a modified loss function to produce a solution that is sparse, requiring only a subset of training data, known as support vectors. The sparsity is enforced in the loss function where the machine seeks to use a minimum number of training data to create the maximum width of a boundary separating the two or more classes (i.e. autism vs control) (Murphy, 2012).

The strengths of SVMs are that they are effective in high dimensional spaces, even if the number of dimensions is greater than the number of samples. They are also

highly customizable as they are dependent on the kernel function. However, when the number of features is much greater than the number of observations, SVMs tend to produce poor results. They also do not provide probability estimates, although these can be calculated using cross-validation if necessary (Pedregosa *et al.* , 2011b).

A strength of SVM is the customizable kernel function. Perhaps the simplest is the linear kernel. The loss function relies on ℓ_2 distance to create a linear boundary separating the data. Unfortunately, data is rarely linearly separable, so other kernels such as radial basis functions (RBF) or polynomial functions may also be used.

The linear SVM uses a hinge loss function. This function tries to create the largest linear boundary between the classes of data. For the binary classification model, let $y_i \in \{0, 1\}$. Instead of a negative log likelihood, the hinge loss is defined as,

$$L_{\text{hinge}}(y, \eta) = \max(0, 1 - y\eta) = (1 - y\eta)_+ \tag{3}$$

where $\eta = f(\mathbf{x})$ is some function that represents the confidence of classifying an input as 1. The objective function is then,

$$\min_{\mathbf{w}, w_0} \frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{n} (1 - y_i f(\mathbf{x}_i))_+ \tag{4}$$

Due to the max term in Equation 4, slack variables ξ_i, transform it into the

quadratic program,

$$\begin{gathered} \min_{\mathbf{w},w_0,\boldsymbol{\xi}} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i \\ s.t. \\ \xi_i \geq 0 \\ y_i(\mathbf{x}_i^T\mathbf{w}+w_0) \geq 1-\xi_i \\ \text{for } i = 1,...,n \end{gathered} \tag{5}$$

where C is the hyperparameter that determines the tradeoff between sparsity and hinge loss.

The solution is in the form,

$$\hat{\mathbf{w}} = \sum_i \alpha_i x_i \tag{6}$$

where,

$$\alpha_i = \lambda_i y_i \tag{7}$$

and α is sparse. The x_i for which $\alpha_i > 0$ are the support vectors. These are the points that were incorrectly classified or are correct but inside the boundary of the model (Murphy, 2012; Cortes & Vapnik, 1995).

SVM is frequently cited in literature for brain anomaly detection. Orrù *et al.* (2012) surveyed the wide variety of neuroimaging problems utilizing SVM to identify biomarkers. SVMs have been successful for studies on Alzheimer's, schizophrenia, major depression, bipolar disorder, Huntington's disease, Parkinson's disease, and autism spectrum disorder.

While frequently cited in literature, advanced techniques exist that may be more useful. SVM can be an effective tool provided the parameters are properly tuned and the correct kernel function is used. But even when optimized, SVM performance

consistently falls short of other uncalibrated methods (Caruana & Niculescu-Mizil, 2006).

Boosting.

According to Schapire & Freund (2012), boosting is a "general method of converting rough rules of thumb into [a] highly accurate prediction rule." Given sufficient data and a weak learning algorithm with slightly better than random accuracy, boosting can *provably* construct a classifier with very high accuracy. The technique works by iteratively training the weak algorithm over the data and applying incremental weights to misclassified data (Murphy, 2012).

AdaBoosting is a very popular boosting technique that utilizes decision stumps as the weak classifier. After multiple iterations, the AdaBoost algorithm effectively creates a very complex barrier in the data separating the distinct classes. Furthermore, AdaBoost is slow to overfit the data (Murphy, 2012). Maintaining a generalized model is essential for positive performance for medical data, where data varies wildly between different patients.

A popular AdaBoost algorithm is the Stagewise Additive Modeling using a Multi-class Exponential loss function, or AdaBoost-SAMME, presented by Zhu *et al.* (2009). Let $T^{(m)}$ be any weak classifier that outputs a predicted class for $\mathbf{x}_i$. If the training data is misclassified, boost its weight and fit another classifier using the new weight. This is repeated M times or until the training set is perfectly fit. A score is assigned to each classifier and the final model is a combination of all prior classifiers. The algorithm for binary classification is shown in Algorithm 1.

Jiao *et al.* (2011) showed that decision trees, stumps, and boosted trees were equally or more accurate than SVM for classifying ASD. They also demonstrated that even with good accuracy, SVM had trouble separating the data. Trees and other

Algorithm 1 ADABoost-SAMME for binary classification

1: Initialize the observation weights $w_i = 1/n$.
2: **for** $m = 1$ to M **do**
3: Fit a classifier $T^{(m)}(\mathbf{x}_i)$ with weights w_i
4: Compute $err^{(m)} = \sum_{i=1}^{n} w_i \mathbb{I}\left(c_i \neq T^{(m)}(\mathbf{x}_i)\right) / \sum_{i=1}^{n} w_i$.
5: Compute $\alpha^{(m)} = \log \frac{1-err^{(m)}}{err^{(m)}}$.
6: Set $w_i \leftarrow w_i \cdot \exp\left(\alpha^{(m)} \cdot \mathbb{I}\left(c_i \neq T^{(m)}(\mathbf{x}_i)\right)\right)$ for $i = 1, 2, \ldots, n$.
7: Re-normalize w_i.
8: **end for**
9: **return** $C(\mathbf{x}) = \arg\max_k \sum_{m=1}^{M} \alpha^{(m)} \cdot \mathbb{I}(T^{(m)}(\mathbf{x}) = k)$

such techniques had less trouble separating the data with equal accuracy.

Uncalibrated boosting performs better than SVM over a variety of classical aca–demic machine learning datasets. The mean performance for uncalibrated boosted decision trees (similar to AdaBoost) was 0.828 versus 0.781 for uncalibrated SVM. However, parameter tuning increased performance to 0.896 versus 0.862 for calibrated SVM. Neural nets also demonstrate robustness over the problems for both calibrated and uncalibrated nets (Caruana & Niculescu-Mizil, 2006).

Logistic Regression.

For binary classification (i.e. a subject either has autism or does not), logistic regression is an obvious choice to try. Logistic regression uses the idea of Bernoulli random variables to produce a model that takes inputs to predict the likelihood of classifying to [0,1].

The logistic response function has the form,

$$E(y) = \frac{\exp(\mathbf{x}'\mathbf{w})}{1 + \exp(\mathbf{x}'\mathbf{w})} = \frac{1}{1 + \exp(\text{-}\mathbf{x}'\mathbf{w})} \tag{8}$$

where $\mathbf{x}'$ is an input vector and $\mathbf{w}$ is the vector of coefficients.

The coefficients can be estimated through the method of maximum likelihood.

The log-likelihood is expressed as,

$$\ln\ L(\mathbf{y}, \mathbf{w}) = \sum_{i=1}^{n} y_i \mathbf{x}_i{}' \mathbf{w} - \sum_{i=1}^{n} \ln[1 + \exp(\mathbf{x}_i{}' \mathbf{w})] \tag{9}$$

The algorithm to solve for $\mathbf{w}$ varies depending on software.

The model coefficients are odds ratios. Assume a trivial model with one coefficient,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{10}$$

then the odds ratio is,

$$\hat{O}_R = \frac{\text{odds}_{x_i+1}}{\text{odds}_{x_i}} = e^{\hat{\beta}_1} \tag{11}$$

The odds ratio is interpreted as the change in probability of classifying as a 1 per unit increase of the input variable (Montgomery *et al.* , 2012).

2.8 Feature Selection

With the abundance of available data, datasets are quickly becoming larger and larger. Datasets with terabytes of data are no longer unheard of. Classical machine learning techniques that may have been useful for smaller data may not scale well with such large datasets. With fMRI data, there is a problem of an over-determined system. There can be several thousand columns of data for only a few subjects. We have many more dimensions than we have subjects. Therefore, we want to find the smallest set of features that can accurately classify our response without overfitting a model.

Principal Component Analysis.

Factor analysis is a multivariate statistical technique which tries to identify an underlying structure within the data. It determines dimensions within the data and is used as a data reduction technique. Factor analysis aims to reduce the original set of variables to a reduced set while retaining as much information as possible (Dillon & Goldstein, 1984).

Principal component analysis (PCA) is a technique that transforms the original variables into a smaller set through linear combinations that account for most of the variance. PCA seeks to find the least number of orthogonal factors, or principal components, that explain the most amount of variance in the data. The principal components are extracted so that the first component explains the largest amount of variance. The m$^{\text{th}}$ principal component can be written as,

$$\text{PC}_m = c_{m,1}x_1 + c_{m,2}x_2 + \cdots + c_{m,n}x_n \tag{12}$$

where the weights, $c_{m,j}$, maximize the ratio of explained variation of the component to total remaining variation, subject to $\sum_{j=1}^{n} c_{m,j}^2 = 1$ (Dillon & Goldstein, 1984).

Murphy (2012) details how to extract the principal components. PCA would like to find the orthogonal set of L vectors, $\mathbf{w}_j \in \mathbb{R}^D$ and the corresponding component scores $\mathbf{z}_i \in \mathbb{R}^L$ such that we minimize the average reconstruction error,

$$J(\mathbf{W}, \mathbf{Z}) = \frac{1}{n}\sum_{i=1}^{n} ||\mathbf{x}_i - \hat{\mathbf{x}}_i||^2 \tag{13}$$

where $\hat{\mathbf{x}}_i = \mathbf{W}\mathbf{z}$ such that $\mathbf{W}$ is orthogonal. This can be written as

$$J(\mathbf{W}, \mathbf{Z}) = ||\mathbf{X} - \mathbf{W}\mathbf{Z}^T||_F^2 \tag{14}$$

where $\mathbf{Z}$ is an $n \times L$ matrix with $\mathbf{z}_i$ in the rows and where $||\mathbf{A}||_F$ is the Frobenius norm of matrix $\mathbf{A}$ defined as,

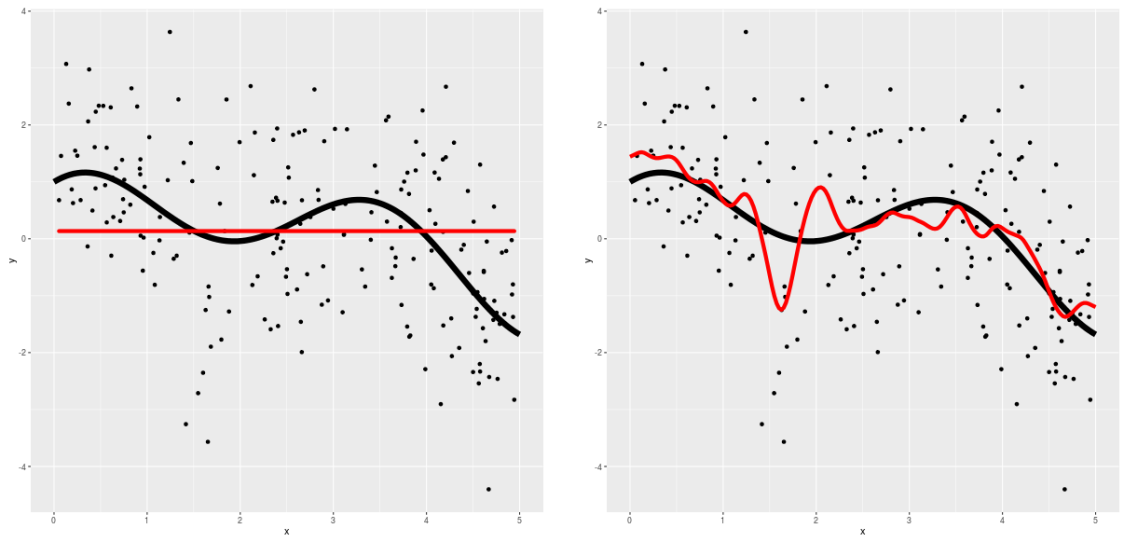
$$||A||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}^2} = \sqrt{\text{tr}(\mathbf{A}^T\mathbf{A})} = ||\mathbf{A}(:)||_2 \tag{15}$$

The optimal solution is obtained by setting $\hat{\mathbf{W}} = \mathbf{V}_L$, where $\mathbf{V}_L$ contains the L eigenvectors with the largest eigenvalues of the covariance matrix $\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$. The orthogonal projection of the data onto the column space spanned by the eigen–vectors is given by $\hat{\mathbf{z}}_i = \mathbf{W}^T\mathbf{x}_i$.

As the technique seeks to maximize the explained variance, data features with naturally high variance due to measurements will be weighted far heavier than features with lower variation. The data should be standardized before using PCA (Murphy, 2012).

Compressed Sensing.

Compressed sensing, or regularization, is the idea that true signals can be approx–imated by sparse samples. Sparse is defined as a small number of non-zero coefficients in the model. Take pictures for example. Instead of keeping every pixel's exact color structure, the JPEG extension clusters similarly colored pixels together to reduce the size of the file. Although the JPEG contains significantly less information, it provides an almost identical, new image to the true image. In signal processing environments, sparsity is the idea that the true signal is only a small portion of the true bandwidth. Lossy compression and compressed sensing takes advantage of this notion that sig–nals are generally redundant and not pure noise (Candès & Wakin, 2008). Figure 5 provides an example of how enforcing sparsity impacts the fit of a model when approximating the true signal.



Overregularized **Underregularized**

Figure 5. Regularization attempts to find the optimal tradeoff between model sparsity and accuracy. The true signal is shown with the bold, black line along with simulated data. The red lines indicate the model fit to the data. Too much regularization inhibits the model from fitting the signal, as seen in the overregularized model. Without enough regularization, the model attempts to fit all of the noise as illustrated by the underregularized model.

The theory of compressed sensing is that sparse vectors in high dimensions can be correctly recovered from incomplete information. Compressed sensing began in the field of signal processing, where the engineers sought to reconstruct the true signal from noisy samples. Obtaining complete information about a signal is a valid strategy but is often hindered by high costs, difficulties, or lack of time. Therefore, compressed sensing can be useful by obtaining an approximated signal from incomplete informa–tion (sampling) (Rauhut, 2010).

The original problem of recovering a sparse approximate solution to a linear system was proven NP-Hard by Natarajan (1995). However, Candes & Tao (2005) and Candes & Tao (2006) proved how exact recovery of the signal is possible given that the signal is sparse by using a convex optimization problem. The signal can be recovered in the condensed dataset by solving the bounded variance form of the compressed

sensing Linear Program,

$$\begin{aligned} &\min_{\beta \in \mathbb{R}^n} \|\beta\|_1 \\ &\text{s.t. } \|\mathbf{X}\beta - \mathbf{y}\|_\infty \leq c\sigma \end{aligned} \tag{16}$$

where β is our coefficient vector, c is some constant, and σ is the variance of the error of the signal. The constant c is a penalty parameter and restricts the size of the sparse set. If the true coefficient vector is sparse, this LP solves perfectly.

The regularized version of the bounded variance problem is,

$$\min_{\beta} \|\mathbf{X}\beta - \mathbf{y}\|_\infty + \alpha\|\beta\|_1 \tag{17}$$

where α is the Lagrange multiplier in the dual problem and $\alpha\|\beta\|_1$ is referred to as the regularizer. This equation can be modified in a variety of ways to fit the researcher's needs. For example, the regularization parameter can be included to restrict the size of the coefficient vector for linear models in which the objective of the model is to minimize the residual sum squares.

The restricted isometery property (RIP) is a sufficient, but not necessary, condi–tion for robust sparse recovery using ℓ_1 minimization. The property states that for a given sparsity, s, and matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, the following must hold,

$$(1 - \delta_s)||\boldsymbol{\beta}||_2^2 \leq ||\mathbf{X}\boldsymbol{\beta}||_2^2 \leq (1 + \delta_s)||\boldsymbol{\beta}||_2^2 \tag{18}$$

where $\delta_s < 1$ is the smallest value for which the inequality holds true for each vector, β with s nonzero entries. The property requires near orthogonality of the columns in each coefficient basis and generally uncorrelated features. If this holds, any sparse system with cardinality less than or equal to s will be recovered by the regularization (Candes & Tao, 2005).

2.9 ABIDE Studies

Previous studies on Autism Spectrum Disorder used small studies (generally 20-40 subjects). Until recently, data was unavailable for work on classifying a large scale study. With the release of the open-source ABIDE database, researchers could now work with over 1,000 subjects. Small studies may suffer from bias due to the small number of subjects. Large datasets provide researchers the opportunity to test for generalizability over a broader range of subjects.

Narayan & Allen (2015) selected subjects from UCLA and University of Michigan, for 98 and 140 subjects respectfully. They used R^2 and R^3 approaches to fit a novel two-level model. While they did not discuss classification, their results highlight different regions of importance versus the independent studies at UCLA or Michigan, suggesting that larger scale fMRI data may be more difficult to classify. Furthermore, their findings suggest that an abnormal decrease in connection between brain regions may be a cause of ASD.

Chen *et al.* (2015) selected 252 subjects (126 autistic and 126 controls) who were matched by age and motion. They scrubbed the data and removed excess motion for each scan before proceeding. They ran several machine learning algorithms over their FNC values. They found the particle swarm optimization support vector machine (SVM) achieved 58% accuracy on their holdout validation dataset. With a recursive feature selection SVM, they achieved 100% accuracy on their training dataset, but 66% on the holdout set. Finally, a random forest classified ASD with only 58% accuracy but when restricting the forest to only the top 100 most important features, they increased accuracy to 90.8%. It should be noted that a random forest is a bootstrapping method and uses the training data for its prediction accuracy. The classification accuracy of the random forest on a holdout validation set reduced to levels similar to the SVMs.

Chen *et al.* (2016) selected 112 autistic and 128 control adolescents (ages between 12 and 18) from six different study sites in the ABIDE database. The authors removed 18 of the 160 total ROIs in the Dosenbach brain atlas for 142 total ROIs. Furthermore, the authors divided the frequency range into five different bands and selected to use the FNC data within the range 0.01-0.073 Hz. They employed an F-score method to select the most important features in the FNC data. Using a leave-one-out linear SVM classifier, the researchers found an accuracy of up to 79.17%. This was the only study using the ABIDE dataset we found that restricted the data to a certain frequency range but it seemed to potentially increase classification accuracy. It would be interesting to redo the analysis using the same subjects but without restricting the frequencies.

Neural nets are another popular pattern recognition tool and can be quite pow–erful. A shortcoming of using neural nets is the complexity of the training algorithm does not always allow for explicit recovery of information from the data. They are often referred to as black boxes, where data is input and an answer is output with–out understanding why. Iidaka (2015) used a probabilistic neural network on the functional network connectivity of 312 ASD and 328 control subjects and claimed approximately 90% classification accuracy. The author also stated that there was no evidence of accuracy differences between study sites, sex, handedness (left or right hand dominant), or intellectual level (IQ).

Vigneshwaran *et al.* (2015) used 878 male subjects (443 autistic and 435 controls) from the ABIDE database to conduct their analysis. Using a Regional Homogeneity (ReHo) measure, the researchers were presented with 54,837 features. They used a Chi-square feature selection algorithm to reduce the size of the basis before classifying. They found that an SVM with a Gaussian kernel and the top 169 features classified all subjects with 63.03% accuracy while a Projection Based Learning Meta-cogni–

tive Radial Basis Function Network Classifier (a neural net utilizing an RBF kernel) classified the same set with 68.9% accuracy. The researchers also created models using 598 subjects only under eighteen years of age. The accuracy of the SVM model with 150 features significantly increased to 65.77% while the neural net remained mostly unchanged. Finally, a model trained only on 280 adult males ($\geq$ 18 years) saw accuracies of 75.95% and 79.4%, for the SVM and neural net, respectively.

Using the same methods as Anderson *et al.* (2011) but with a much larger sam–ple size, Nielsen *et al.* (2013) obtained roughly 60% classification accuracy using whole brain distribution over the 964 autistic and control subjects from the ABIDE database. They used a leave-one-out classifier and included several external categor–ical variables in their classification. The researchers suggested that fewer regions of interest may increase the classification accuracy.

Ghiassian *et al.* (2013) saw 61.88% accuracy over the 1,111 subjects they included in their study over the baseline accuracy of 51.57%. In the study, they converted each 4D fMRI image to a 3D image by averaging each voxel over the total time series. They then extracted features using a 3D HOG (histogram of oriented gradients), which computes spatial gradient information about each voxel to treat it as a feature. They used an SVM with 241 of the 116,480 features generated.

III. Methodology

Chapter two introduced how the data can be manipulated for use by machine learning classifiers. This chapter discusses the methodology behind the research. It also discusses the tools used throughout the research and is explicit enough that someone with a general knowledge of programming can replicate the results.

3.1 Programming Tools

The programming work was completed using Python 2.7. Several packages were instrumental throughout the process. NumPy was an integral part of the work and is a common module used for data science in Python (Van Der Walt *et al.* , 2011). Nilearn is an open-source Python module developed for applying machine learning to neuroimagery. Several Nilearn functions were integral to prepare the data and to produce the images in this research (Abraham *et al.* , 2014). The last essential mod–ule was scikit-learn, an open-source collection of several advanced machine learning algorithms and other data science tools (Pedregosa *et al.* , 2011a).

3.2 Functional Network Connectivity

Feature extraction began once the desired preprocessed fMRI images were down–loaded and stored. fMRI images are stored in a .nii file extension, meaning that typical image classifiers may not be able to import them. Specialized neuroimaging programs are designed to use the fMRI data and allow the user freedom to manipulate it. Nilearn provided the tools necessary to map the atlas to each brain volume.

The first step to extracting the FNC values is mapping the chosen atlas to the fMRI. Each atlas partitions the brain volume into separate groups called region of interests and information within the ROIs is captured. Each voxel is matched to the

ROI that covers that voxels location in the volume. The Talairach and Tournoux (TT) and Craddock 200 (CC200) atlases were both used to extract the FNC values to make two different datasets for comparison.

fMRI data is fundamentally time series and using atlases reduces the dimensions from four to two. As the ROIs encapsulate the voxels of the volume, the activity levels for each voxel within are registered and stored. The activity level for each voxel is used to calculate the mean activity within the ROI. This is repeated over each of the slices to build a time series for each ROI. The output of this process is data with a select number of columns and a row for each time slice. We can now calculate the functional network connectivity of the subject.

The correlation structure for each volume is simple to calculate. The correlation between each pair of ROI is calculated to form the matrix like in Figure 4 in chapter 2. Flattening the correlation matrix transforms the data into an input vector of features. With the Talairach atlas, the flattened vector had 4,656 correlation values (features).

Each subject in the ABIDE database was completed through these steps. The resulting data was exported as a comma separated value for storage and sharing capabilities. A great advantage of using FNC data is the reduced physical memory required to handle the data. Each fMRI file is about 425 MB while the TT atlas FNC text file is only 80 MB. If FNC values can capture all of the relevant information needed to classify autism, the size of the data can be reduced to only 0.023% of the physical memory required for all fMRI images.

ABIDE Preprocessed.

Extracting activity from ROIs was proven to work with the preprocessed fMRI images through the steps above. However, downloading the fMRI file for each subject was a slow process and took up a huge amount of memory. The ABIDE PCP has

prepared time-series data for each pipeline, nuisance variable scrubbing, and filtering method for the seven available atlases. The availability of this data allowed us to compare several different preprocessing strategies in a reasonable time.

Data that is marked `filt_noglobal` is our original FNC extraction data.

3.3 Subject Selection Criteria

Subjects were included in the analysis if they passed the manual inspection by the three ABIDE PCP researchers and were completely successful in feature engineering. There were 403 autistic and 468 subjects who passed initial fMRI quality inspection.

Feature engineering means converting the raw fMRI and BOLD contrasts into usable connectivity data. Several subjects recorded Not-a-Number (NaN) values for some connections, potentially indicating that an atlas ROI was outside of the nor–malized brain scan or asymptotically small variances between regions. If the variance of an ROI is effectively zero, ρ_{ij} in Equation 2 approaches infinity, rendering an NaN value. Some other research noted some ROIs with missing signals, possibly the cause of our NaN values (Chen *et al.* , 2015).

There was some debate about either removing the columns that were corrupted in multiple subjects or to remove the subjects. The NaN values seemed to cluster around specific ROIs but they also only occurred in a few subjects. If a subject registered a NaN value, there were usually a large number (> 10) rather than an isolated case. Due to this, any subject with NaN connectivity values was scrubbed from the data. There were 375 autistic and 430 control subjects remaining after converting to FNC values.

While some studies excluded subjects based on handedness, sex, or age, we decided to keep all subjects in the data. One of the motivations of this study was to investigate if generalization is possible for ASD. Several subsequent models were run with reduced

subjects restricted by age to compare results to full inclusion.

3.4 Data Preparation

This section discusses how the data was prepared before it was run through the classifier. Several different methods were used to investigate which produced the most accurate classification model. These include dimensionality reduction, including phenotype data such as gender, or running a full model with interactions and polynomials. Each of these methods was optional and could be included or excluded based on the desire of the researcher.

Principal Component Analysis.

A common method of dimensionality reduction, principal component analysis projects high dimension data into orthogonal components to explain the maximum amount of variance. While useful, PCA does not allow for easily interpretable data. The features created by the algorithm cannot be easily mapped to the originals which is useful when trying to explain where in the brain the connections are most important. However, PCA is a quick and readily available algorithm that can be used to compare different dimensionality reduction techniques.

Scikit-learn provided the PCA algorithm in this research. This algorithm allows for researchers to specify how much variance the components should explain. The parameter could be optimized but with regularization, unnecessary features would be stripped anyways. We ran a rough grid search between 60% and 95% at 5 percentage point intervals.

The PCA algorithm was fit using the training set of data. The training, validation, and test data were then transformed before running the model. These data partitions are explained in Section 3.5.

Artificial Noise.

Introducing an artificial noise vector can be an effective way to gauge how discriminatory feature selection performs. Since regularization is not a random process, there is not a great way to select the most important features over all of the experiments. Including a noise feature introduces a threshold that can be used to filter only those features selected more than randomly. Guyon & Elisseeff (2003) explain that the addition of noise can be useful when the data has redundant variables to test the stability of the models.

For example, say a dataset has one hundred features and one hundred experiments are run, selecting only a subset of the features. Each feature would be included in a certain percentage of the models. If we introduce an artificial noise variable to the data and rerun the experiments, we could select only the features that were included more often than the noise variable for our final model.

This method is useful because of the stochastic nature of the partitioning the dataset into train, validate, and test data. While the most important features will be included in the subset for a large percentage of experiments, those lesser important features identified by the model should also be included. The noise vector provides this ability.

This vector was developed as a random Gaussian vector with $x \sim N(0, 1)$. Since FNC data is correlation values, every observation in the vector was scaled between [-1, 1] by,

$$\begin{aligned} x_{i,std} &= \frac{x_i - x_{min}}{x_{max} - x_{min}} \\ x_{i,scale} &= x_{i,std} * (2) - 1 \end{aligned} \tag{19}$$

Gender.

Autism seems to be more prevalent in males than females. There has yet to be any concrete evidence for why this happens. To accompany possible differences between male and female brains, the models could include a gender variable. The gender of each subject was selected from the accompanying phenotype data and included in the FNC data as an external variable.

To test for the significance of including a gender variable, three experiments were set up. The first uses the FNC data without the subject's gender. The second includes the subject's gender. The third includes both the subject's gender and an artificial noise variable to test whether the gender variable was being used by the model. The program was seeded (seed=41) to replicate training, validation, and testing partitions over the three experiments.

Age Restricted Models.

Several studies in Chapter 2 noted how only including certain age subjects increased model accuracy. Other studies suggest that connectivity values in the brain switch from hyperconnectivity to hypoconnectivity as a subject increases in age. If this is true, then using both young and old subjects in the same model could cause serious accuracy problems.

To test these hypotheses, the data has the option to be scrubbed of subjects outside a prescribed age range. A maximum or minimum age could be used to limit the data. The total number of subjects included in the model drops when controlling for age as seen in Table 6. This trade-off between accuracy and model generalization may be necessary to create a useful model.

Table 6. Comparison of the number of subjects remaining in study after controlling for age

Age	ASD	Control	Total
No restriction	375	430	805
< 19	283	318	601
12 - 18	177	202	379

Interaction and Polynomial Terms.

Although the theory of sparse effects in design of experiments offers that low order effects (single variables) are more likely to be more important than higher order effects (interactions between and powers of variables), the complexity of neuroprocessing may lend itself to a higher order model.

The initial attempt to create a full model with every interaction term and poly–nomial failed. The input vector of this model is,

$$\mathbf{x}_{full} = \begin{bmatrix} \mathbf{x} \\ \mathbf{x}_i\mathbf{x}_j \\ \mathbf{x}^2 \end{bmatrix}^T \quad \text{for: } i \in \{1, ..., k-1\},\ j \in \{i, ..., k\} \tag{20}$$

where k is the number of features in the first order model.

The total number of features included in a full model is,

$$N = \left(\frac{n^2 - n}{2}\right)\left(\frac{n^2 - n}{4} + 1\right) \tag{21}$$

where n is the number of ROIs. A full model with the Talairach atlas and 97 ROIs would be 10,843,824 features. Eight hundred and five observations of each feature was simply too large for our machine to handle.

Although there was not enough memory to run a full model with every interac–tion, a pseudo-interaction model could potentially give desired results. This model

was constrained by the memory and the number of interactions was limited. Initial experiments output columns included in the models by the regularization. Using the artificial noise vector, only 384 features were included more often than the noise. These features were used to create the full model.

The data for the full model was constructed by combining the original features and their squares with the interactions of each of the 384 selected features. The number of features in this model is 82,848.

3.5 Cross Validation

Machine learning algorithms create highly flexible models meaning they can fit a variety of data extremely well. A problem with these techniques is overfitting the data. It is possible to create a model that fits the data perfectly so there is no error. This is caused by the model trying to fit for every small variation in the data. In this case, the model is probably fitting the noise rather than the true signal in the data. It is highly likely that the model will make a lousy prediction when new data is presented to the model.

Cross validation (CV) is a common technique used to maintain model generalization. Hyperparamters which control model behavior must be optimized to correctly fit the model but keep it general enough for use with external data. These hyperparamters can be easily optimized for perfect performance on the data only to fail to make good predictions later. Cross validation selects hyperparameters that perform well over a variety of data with the aim of preserving model generalization.

To perform cross validation, the data is randomly split into separate partitions: training, validation, and testing data. The training data is the data used to fit a model. This is the only data that is actually used by the optimization techniques. The validation data is used in the hyperparameter search to test for generalization; this

data is only used to predict responses and to calculate the classification error. Finally, the testing data is a hold-out set that is only used after the hyperparameter search to test the accuracy of the model. It remains unused until the entire hyperparameter optimization process is complete. We used 50% of the data for the training set, 25% for the validation set, and 25% for the testing set.

Error is defined as mean square error,

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \tag{22}$$

where n is the number of observations, $\hat{Y}_i$ is the predicted classification, and Y_i is the actual classification. With binary classification, the prediction is either correct or not. Models with the ability to output probabilistic classification such as logistic regression could also define the error as how far the probability is from the threshold.

As model complexity increases, the training error approaches zero. The validation error also decreases until a point where the model begins to overfit the data. The hyperparameter that results in the minimum validation error is considered the "optimal" model. The model then classifies the test data and the resulting accuracy is the model's performance.

Figure 6 illustrates how increasing model complexity (increasing the model regularization parameter, C) affects the error of each data partition. As the parameter increases, the model becomes more complex and the training error decreases to nearly zero. Near the start of the search when the model is very general ($C \approx 0$), the model does not show good results. The minimum validation error occurs around $C \approx 0.5$ and begins increase afterwards. This parameter is considered the optimized hyperparameter for the model given our data.

Due to the stochastic nature of partitioning the data, the accuracy of the model is expected to vary with different training sets. Multiple experiments with different

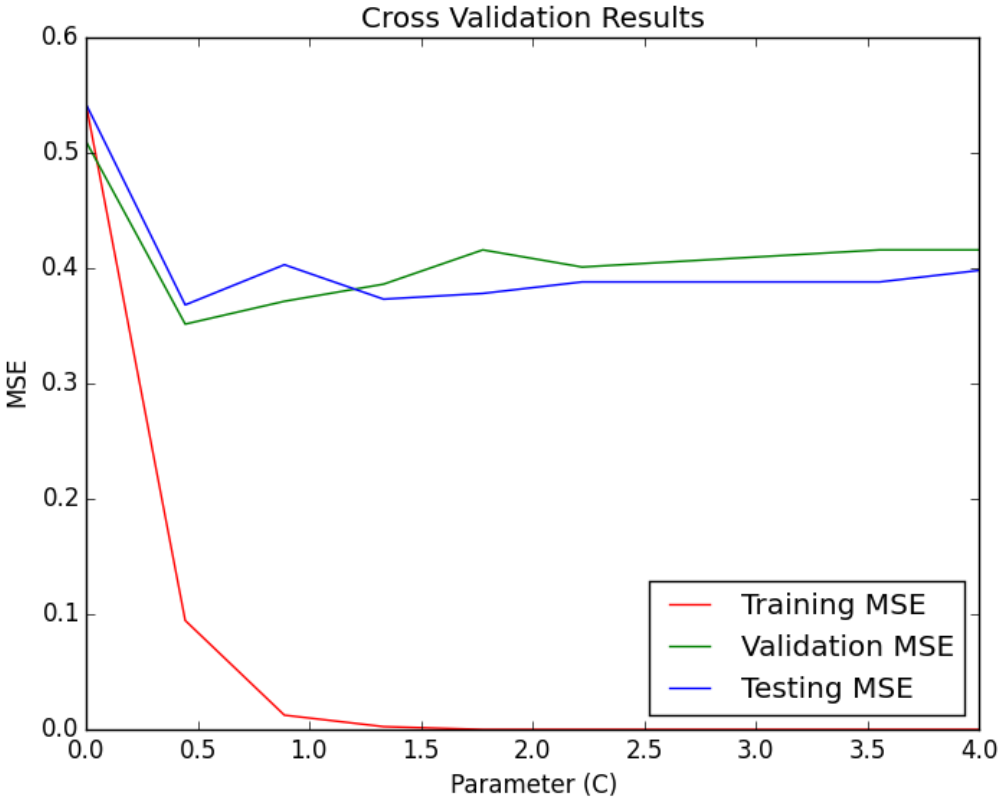


Figure 6. Model error comparison of training, validation, and holdout test data.

randomized partitions is a way to estimate the classification accuracy. The accuracy is averaged over all experiments and presented as the best estimate for the model's performance. The majority of models developed in this thesis were run 1,000 times. That is, 1,000 distinct training, validation, and testing partitions and hyperparameter searches.

3.6 Restricted Isometery Property

As stated in Chapter 2, the solution to the compressed sensing quadratic program will be unique if the restricted isometery property holds for a given sparsity. However, guaranteeing RIP holds for the data is proven NP-hard (Bandeira *et al.* , 2012). Since we cannot check all possible combinations in a given sparse set without full

enumeration, we rely on experimentation.

Simulated data is useful to illustrate how δ_s depends on the size of the basis. Our simulated data was a 805×4,656 matrix with each column as a Gaussian distribution, $N(0, 1)$. The data was standardized between [-1, 1] as the FNC data can only lie within this range. This represents the optimal conditions of data for successfully satisfying the restricted isometery property.

The algorithm for these experiments is in Algorithm 2 below.

Algorithm 2 Restricted isometery property experiments

initialize run = True, $s = 1$, r_{max}, s_{max}
standardize every column in $\mathbf{X}$ to unit norm
while run **and** $s \leq s_{max}$ **do**
    $r = 1$
    while $r \leq r_{max}$ **do**
        $\beta_i = \frac{1}{\sqrt{s}}$ for $i = [1, \ldots, s]$
        $\mathbf{z} \leftarrow b$ random columns from $\mathbf{X}$
        $\delta_{i,s} = |\sum_{i=1}^{s} (\beta_i \cdot \mathbf{z}_i)^2 - 1|$
        $r = r + 1$
    end while
    Extract 10^{th}, 50^{th}, and 90^{th} percentile δ_s
    if 10^{th} percentile $\delta_s > 1$ **then**
        run = False
    end if
    $s = s + 1$
end while
return 10^{th}, 50^{th}, and 90^{th} percentile δ_s for $s = 1, \ldots, s_{max}$

The experiments were set for a maximum sparsity, $s_{max} = 100$, and 10,000 runs per sparsity, s. Figure 7 displays how δ_s changes as the number of coefficients kept in the basis increases for the simulated data. As you can see, a basis with 100 coefficients is still far less than the maximum, $\delta_s = 1$. This is expected because the simulated data is designed to be well conditioned for RIP.

Replacing the simulated data with the FNC data provides insight into whether RIP may hold for our data. If $\delta_s \leq 1$ for relatively large values of s, the regularization

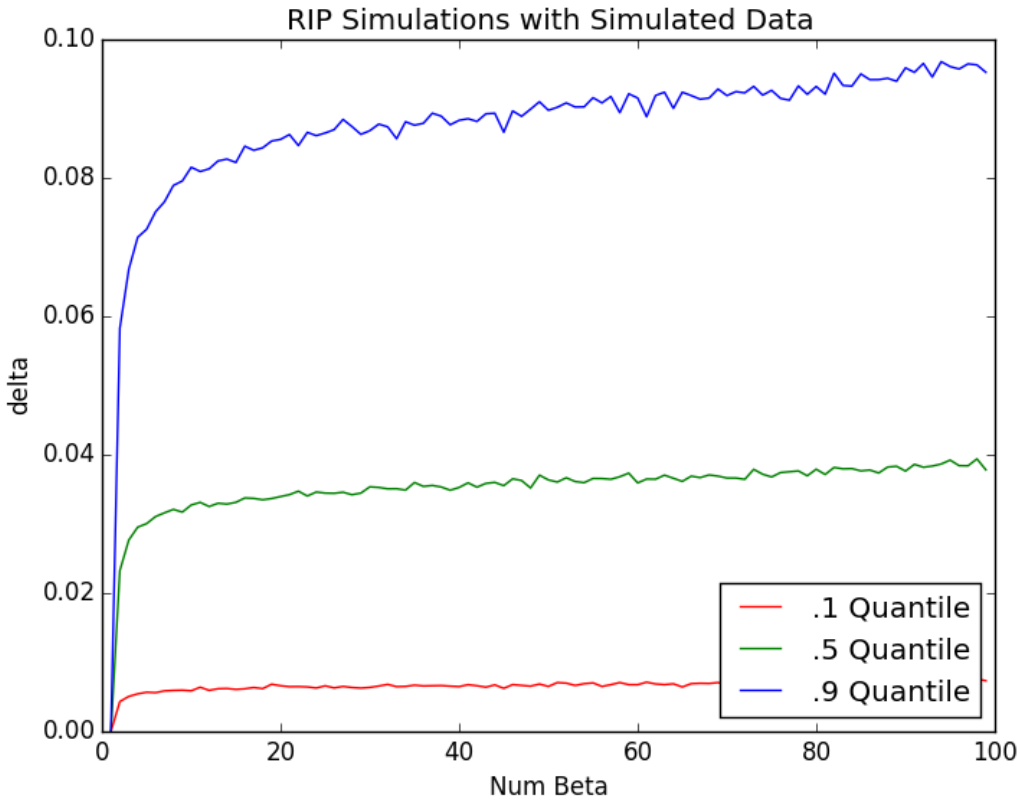


Figure 7. 10%, 50%, and 90% quantiles of δ_s for a given sparsity, s from the simulated, standardized data with a Gaussian ($\mu = 0$, $\sigma^2 = 1$) distribution.

may be capturing the unique solution to the problem in Equation 17.

3.7 Algorithms

Coding in Python allowed for a multitude of experiments with different machine learning algorithms. Running different models only required relatively simple changes to the code. Although the research was focused on using logistic regression to classify ASD, a few other techniques named in chapter two were also investigated.

Random Forest.

A popular first approach in many machine learning problems, a random forest was noted in Chen *et al.* (2015) for high accuracy when restricted to only a few features. This algorithm might provide an excellent fit because the data does not have to be linearly separable to be accurate.

The random forest classifier provided by scikit-learn fits a number of decision tree classifiers on samples of the data and averages to improve accuracy and reduce overfitting. We searched for the optimal number of trees in the forest, `n_estimators`, using a coarse grid-search from 1 to 91 at 15 unit intervals. The number of features to consider when looking for the best split was $\sqrt{N}$, where N was the total number of features in the data. Bootstrapping was enabled in the algorithm, meaning that the training data was sampled with replacement while fitting the model.

Although bootstrapping means the entire dataset can be used to construct a model without cross-validation, using a holdout test set provides insight on the model's generalization. We kept the test data separate and use it to report the final accuracy of the random forest model.

AdaBoost.

Although it did not show up in any previous autism spectrum disorder studies, the AdaBoost algorithm may provide a flexible model for this type of data. The algorithm creates a complex decision boundary but is slow to overfit the data.

The AdaBoost algorithm in scikit-learn builds a classifier with a maximum of `n_estimators` before the boosting is terminated. The algorithm terminates early if the model perfectly fits the data. While the number of estimators could be set to a large value, the time to train the model also increases. This is the tradeoff between learning rate and accuracy.

The maximum number of estimators was set to 1,000 as the algorithm would end in case of a perfect fit. The algorithm in scikit-learn uses a decision tree classifier.

Support Vector Machine.

The linear SVM utilizes a subset of the data, the support vectors, to determine a linear boundary separating the two classes. These models are popular in literature but had better accuracy with smaller rather than larger datasets. A downside is that SVMs are not probabilistic which could cause problems depending on the research.

The SVM with a linear kernel in scikit-learn takes a hyperparameter, `C`, the penalty parameter of the error term. This was optimized through a grid search from 0.0001 to 5 with 50 intermediate steps.

The radial basis function SVM operates the same as the linear SVM, except the kernel does not force a linear boundary. The RBF kernel is defined by Murphy (2012) as,

$$\kappa(\mathbf{x}, \mathbf{x}') = \exp\left(-\gamma||\mathbf{x} - \mathbf{x}'||^2\right) \tag{23}$$

where γ defaults to $1/n$ observations in the scikit-learn toolkit. The RBF kernel's output depends on the observation's distance from the origin producing different results than a linear kernel. The same technique as the linear SVM was used to optimize the penalty parameter.

Scikit-learn's SVM algorithm fits in time $> O(n^4)$, which makes it one of the slowest algorithms tested (Pedregosa *et al.* , 2011b).

Logistic Regression with ℓ_1 Regularization.

Logistic regression is a common tool for binary classification. The model predic–tions are probabilities matched against a threshold value. Probabilities greater than the threshold are classified as a one and vice versa.

The logistic regression coefficients are estimated through maximum likelihood estimation (MLE). The negative log likelihood for logistic regression is,

$$\text{NLL}(\mathbf{w}) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i)) \tag{24}$$

The MLE cannot be written in closed form and requires an optimization algorithm (Murphy, 2012).

The logistic regression classifier in scikit-learn with ℓ_1 regularization fits a model using a sparse subset of the input vector. The ℓ_1 regularization was discussed in Section 2.8. The regularized logistic regression solves the following problem,

$$\min_{w,c} ||w||_1 + C \sum_{i=1}^{n} \log(1 + \exp(-y_i(\mathbf{w}^T \mathbf{x}_i + c)) \tag{25}$$

where C is the penalty parameter provided by the user and c is the intercept. Scikit-learn uses the liblinear solver from Fan *et al.* (2008) to optimize the problem.

The penalty parameter enforces the tradeoff between enforcing sparsity and minimizing the negative log likelihood. As C increases, sparsity decreases. A fine grid search between [0.0001, 4] with 50 steps was used to select the penalty parameter for the TT atlas data. The search was expanded to [0.0001, 5] for the CC200 atlas to accommodate the possibility that more columns could be included in the basis due to the larger number of ROIs in this dataset. The classification threshold remained at 0.50 for every experiment, although modifying this could be investigated in the future.

3.8 Output

Retaining information from the models fit during each experiment provides insight about how and why the models fit as they do. The program outputs the coefficient

vectors from each optimized model as well as the accuracy, hyperparameter value, and number of non-zero coefficients.

Summary

In this chapter we discussed the methodology behind the research. The data was manipulated to test several different models accounting for gender and age. We also used several different machine learning algorithms noted in literature to compare their performance against our regularized logistic regression, although they were not the focus of this research. The next chapter will discuss the results of our research.

IV. Results

This chapter covers the results from the research. In addition to model results, we present insight into the models and detail some findings from the model. A significance level, $\alpha = 5\%$ is assumed unless otherwise stated.

The baseline accuracy for all 805 subjects is 53.4%. Any classifier that classified each observation as a control would get this result. The baseline is calculated by,

$$\text{acc} = \frac{\max(\#\text{ ASD}, \#\text{ Control})}{\text{Total}} \tag{26}$$

Baseline accuracy for models with a reduced number of subjects are explained in their respective sections.

4.1 Restricted Isometery Property

Section 3.6 introduced the restricted isometery property experiment with sim–ulated data. The simulated data provides the best case scenario for RIP to hold. Remember that proving exact RIP is NP-hard, but with experiments we can at least predict whether RIP may hold. The closer δ_s is to zero, the better.

Figure 8 displays the results from the RIP experiments on the TT FNC data. δ_s is less than one only for very sparse bases. The 90^{th} quantile surpasses $\delta_s = 1$ when 4 columns are included in the basis. The 10^{th} quantile included sparsity up to six columns before the program quit.

This result concludes that the restricted isometery property does not hold for our data except for a basis that is extremely sparse. Using the best case scenario when $s = 6$, this represents only 0.131% of the original 4,656 columns in the TT data. These six columns would have to be extremely important to the classifier for regularization to only select such a small amount of the data.

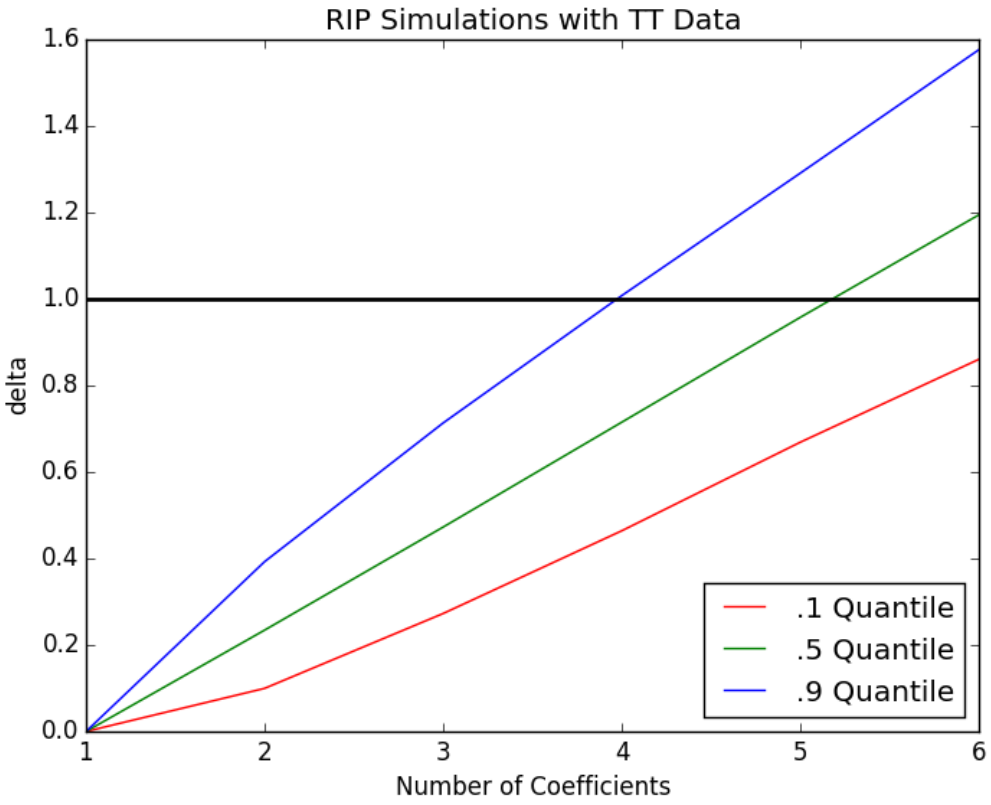


Figure 8. 10%, 50%, and 90% quantiles of δ_s for a given sparsity, s from the TT FNC data. Values above δ_s violate RIP.

Due to the failure of the RIP experiments, we cannot guarantee that the ℓ_1 regu–larization technique employed throughout this research provides the unique solution to the regularization problem. Since RIP ensures a convex model, ℓ_1 regulariza–tion may unfortunately return a local optima. However, Loh & Wainwright (2011) demonstrated that even without RIP, gradient descent algorithms converge with high probability to a coefficient vector close to global optimum. We use this technique to enforce sparsity in our models in combination with the artificial noise vector to assess the stability of our sparse coefficient vectors.

4.2 Preprocessing Methods

While the Preprocessed Connectomes Project provided four different preprocessing pipelines, our data was preprocessed through the CPAC tool (see Table 1 in Section 2.3). The PCP also provided four different data filtering and regression methods. Most of the controversy in literature was focused on global signal regression. The PCP's data allowed for comparison of these methods.

Each preprocessing combination was run through the logistic regression classifier one thousand times. The accuracy, number of columns used in the model, s, and the optimal hyperparameter, C, were recorded and exported upon completion. Table 7 displays the results of these experiments.

Table 7. Mean results for each filtering and regression strategy of TT FNC data. These values are the means over 1,000 experimental runs.

Name	Band-Pass Filtering	Signal Regression	Accuracy (%)	Number of Coefficients (s)	Parameter (C)
`filt_global`	Yes	Yes	58.53	83.34	0.79486
`filt_noglobal`	Yes	No	62.79	147.22	0.75420
`nofilt_global`	No	Yes	62.45	130.08	0.61796
`nofilt_noglobal`	No	No	62.57	144.54	0.73592

The combination of band-pass filtering and signal regression (`filt_noglobal`) showed the greatest accuracy of the three other combinations. While the accuracy of `filt_noglobal` was not significantly different from `nofilt_global` or `nofilt_noglobal`, analysis of variance (ANOVA) highlighted a significant difference between groups ($p < 0.0001$). It is intuitive to the casual observer that something about the `filt_global` data is different than the remaining strategies.[1] It could be possible that band-pass filtering exaggerates the problems encountered with global signal regression as the `nofilt_global` data does not significantly differ from the best accuracy.

[1]Subsequent experiments on the same data resulted in similar results. The data does not seem to be corrupted in any way.

Due to the controversy of global signal regression and the lack of significant accu–racy difference, `filt_noglobal` is the strategy that is used for all other models unless otherwise specified. Figure 9 displays the number of non-zero coefficients in all 1,000 experimental models on this data. The impact of the regularization parameter, C, on the size of the basis is very evident from this figure. The smaller the parameter, the less columns were selected for the basis. There was also a slight accuracy decrease as the size of the basis increases. Perhaps this is due to the model trying to fit too much of the noise rather than the signal itself.

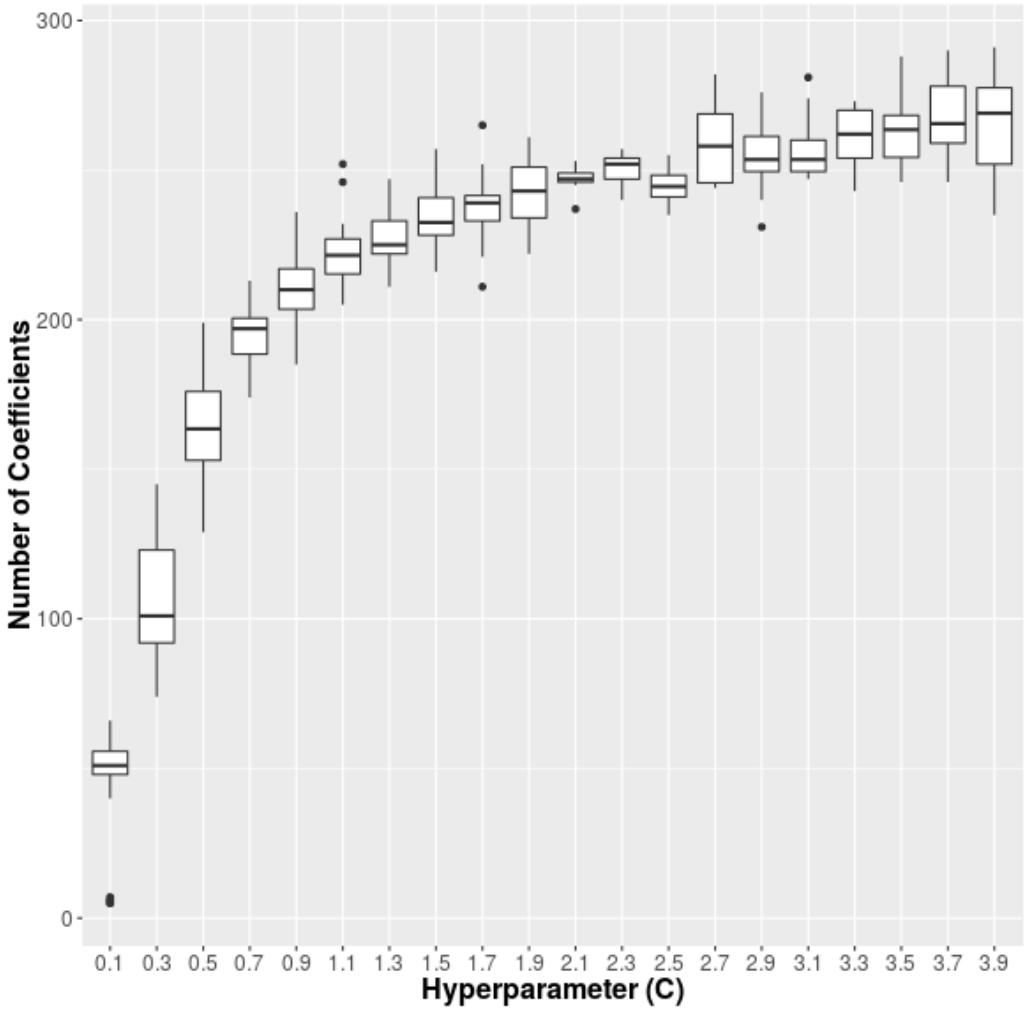


Figure 9. Logistic regression experimental results on the TT FNC data with band-pass filtering and without global signal regression. This plot illustrates the clear increase in number of coefficients in the model as the hyperparameter also increases.

The results of this model suggest we are within the limits of the variance in the data. The accuracy remains relatively stable within the hyperparameter range (0, 0.75] indicating that while the optimal parameter for our model lies somewhere within the range, the accuracy will probably only slightly increase at best.

While the mean selected hyperparameter for the `filt_noglobal` data was $C = 0.75420$, Figure 10 shows how the data is skewed right. 65.5% of the selected parameters were $C < 0.5$ and 76% were $C < 0.75$. Evidently the model performs best with a regularization parameter somewhere in this range meaning the model favors a sparse coefficient vector.

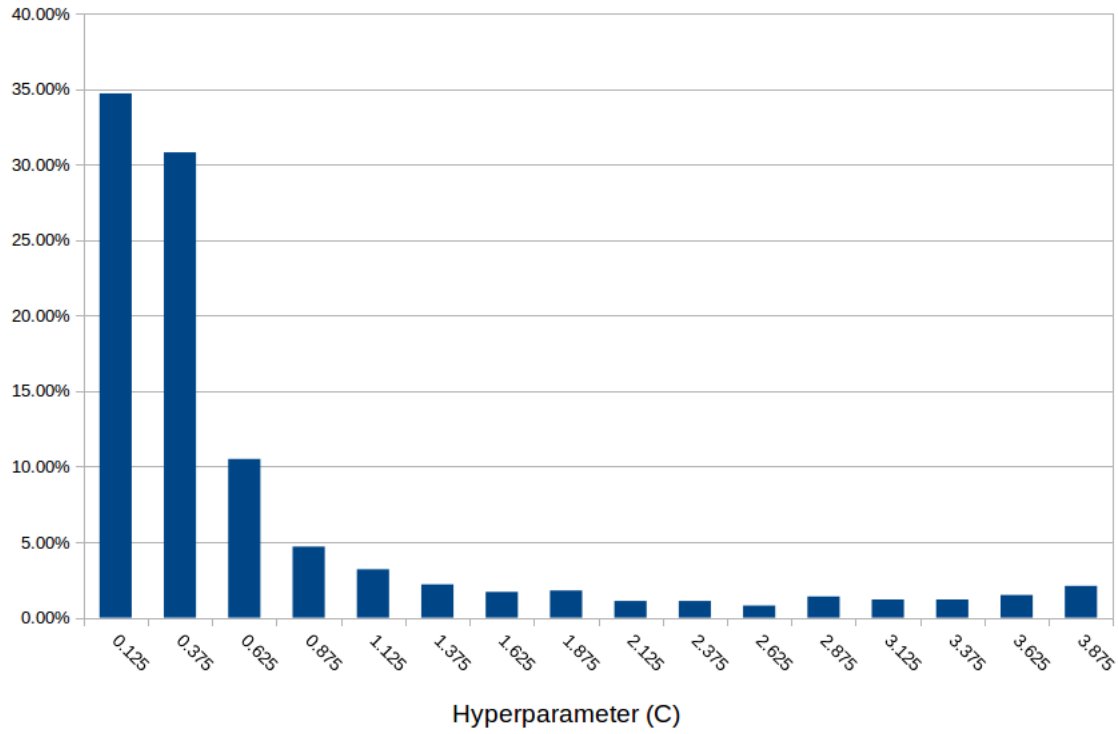


Figure 10. A histogram of the occurrences of optimal hyperparameters as reported by the model on the TT `filt_noglobal` data. Most of the hyperparameters selected are under $C = 1$, suggesting the model favors sparse coefficients.

4.3 Other Machine Learning Algorithms

Random Forest.

We attempted to create a random forest classifier to compare against several studies that found good performance with such model. The random forest classifier produced only $57.9 \pm 0.2\%$ accuracy on the hold-out data, barely above the 53.4% baseline. The mean number of trees used to create the optimal classifier was 62.65 ± 4.78.

The accuracy in our model with a hold-out set and 805 subjects was similar to the reported results in Chen *et al.* (2015). This model was not the primary focus of the research and could possibly be improved by restricting the input to only the most important vectors as was done in other research.

AdaBoost.

Initially it was hypothesized that AdaBoost might provide a model that can pro–vide a great fit for complex neuroimaging data. The AdaBoost model delivered an average accuracy of $56.9 \pm 0.2\%$, less than the random forest. There is not much dif–ference between the accuracy of this model and the baseline accuracy. This suggests that the model had a difficult time distinguishing between ASD and controls.

The difficulty could stem from the theory behind boosting algorithms. Since autism spectrum disorder comes in a variety of different severity, highly functional autistic subjects or subjects with asperger syndrome might be difficult to correctly classify. Long & Servedio (2010) demonstrated that boosting algorithms with an exponential error function perform no better than a random classifier when using noisy data. They define noisy as a mislabeled observation. As the boosting algorithm iteratively weights misclassified observations, the algorithm may be focused too much on these difficult subjects rather than all of the subjects.

The mean number of estimators used to create the model was 56.0 ± 5.8.

Support Vector Machine.

The support vector machine is a commonly used algorithm in the machine learning discipline. Its use in Autism Spectrum Disorder studies was discussed in Chapter 2. Two SVMs were used in our research; a linear and radial basis function (RBF) kernel SVM.

The linear SVM is perhaps the simplest of all SVM. The performance of the linear kernel highly depends on the linear separability of the data. Using all 4,656 features, the linear SVM averaged $61.0 \pm 0.2\%$, better than both the random forest and AdaBoost models.

Interestingly, the optimal error penalty parameter, C, was very stable. The ex–periments selected $C = 0.2042$ 96.0% of the time. This was much more stable than the logistic regression, random forest, or the AdaBoost parameter searches.

An RBF kernel SVM was also tested for comparison. The RBF kernel was hy–pothesized to potentially increase the classification accuracy due to the complexity of the data. The model significantly increased the accuracy over the linear SVM. Accuracy of the RBF SVM was $62.4 \pm 0.2\%$.

The error penalty parameter, however, was significantly different from the linear SVM. The mean selected parameter was 7.0735 ± 0.4020, higher than the linear SVM's usual selection of $C = 0.2042$. Figure 11 below illustrates the parameter experimental results. If the linear SVM's results were included, every response would be in the 0-1 column of the histogram.

SVM's provided a model comparable in accuracy to those seen in Section 4.2 but with much greater variance. The basic SVM models from these experiments suggest that focusing on optimizing the hyperparameters and preparing the data specifically

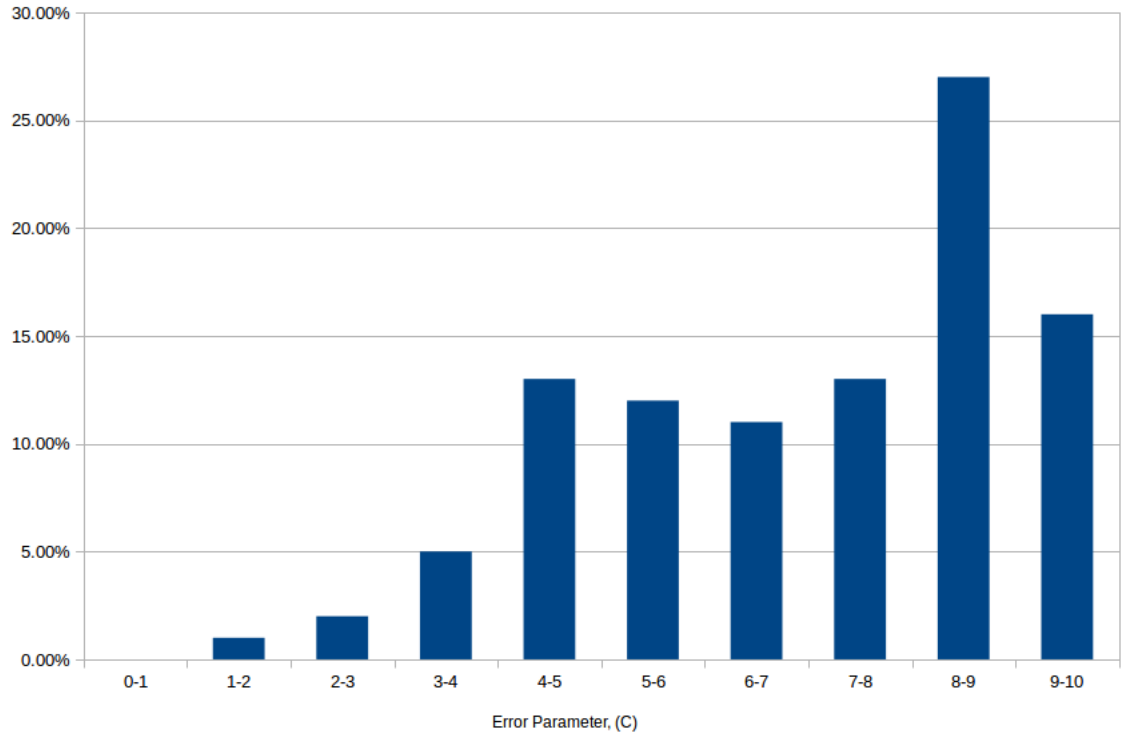


Figure 11. Histogram of error penalty parameter, C, in the RBF SVM experiments.

for use by such models may provide reasonably high accuracy. The difficulty lies in preparing the data in such a way that it increases the SVM's accuracy while remaining interpretable. Finally, the SVM models fit much slower than the logistic regression models, limiting the number of experiments that could be run in a reasonable amount of time.

4.4 Gender Models

The inclusion of the gender indicator variable did significantly improve accuracy of the logistic regression models. The accuracy can be seen in Table 8. The data used was collected through three different experiments using the `filt_noglobal` FNC values. The program was seeded (seed=41) so the data partitions remained constant for each experiment.

Table 8. Gender inclusion in the model significantly increased model accuracy, although the increase is small.

Gender	Accuracy
No	62.38 ± 0.20
Yes	62.79 ± 0.20

Gender was included almost four times as often as the noise variable, indicating that the model generally found that gender was an important feature. 97.4% of the gender coefficients were positive suggesting that female subjects were more likely to classify as a control rather than ASD. The average positive coefficient value was 0.1578 for an odds ratio of 1.171, meaning the probability that a female was classified as a control was 4 percentage points more often than a male, ceteris paribus.

4.5 Full Model

The inclusion of the polynomial and selected interaction terms did not significantly change the accuracy from the single factor model. Table 9 displays the results of the two different models.

Table 9. Comparison of results of the interaction and polynomial terms versus the single order `filt_noglobal` models

Model	Accuracy
Single Order	62.79 ± 0.20
Full	62.65 ± 0.20

Including the interactions of the 384 features selected more often than the noise variable created a much larger dataset. The experimental results indicate that the regularization parameter increased to $C = 1.17 \pm .07$, significantly larger than that of the single factor model. This is not unexpected as an increase in total features usually corresponds with more features selected in the basis.

Finally, such a large model was not particularly stable when selecting the relevant features. 98.9% of all features in the data were included in less that 5% of the

experimental models. 3.7% of the first order features were never included in any model compared to 68% of the polynomial and 94.5% of the interaction terms. 2.1% of first order features were selected in more than 25% of the models. 0.2% of the polynomial features and none of the interaction terms were included more than 25%. Figure 12 provides a visual of the inclusion rates among the different features.

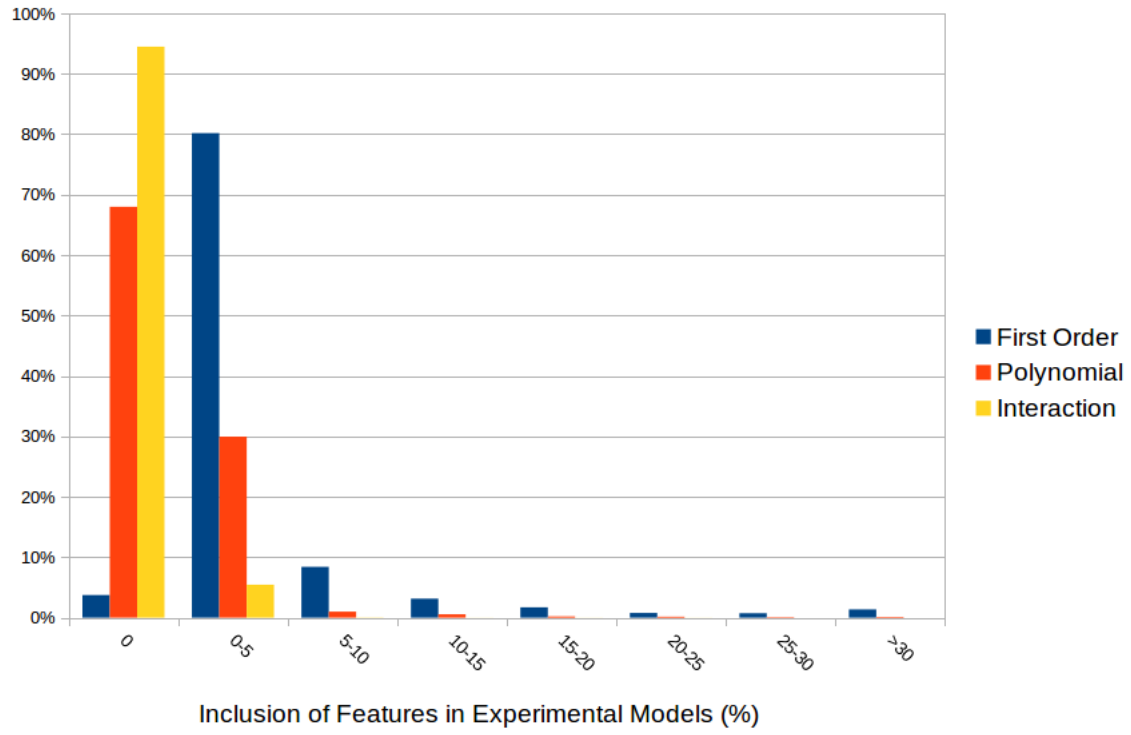


Figure 12. A histogram comparing inclusion rates among different variables in the full model.

Interestingly, of all of the single order features included in more that 50% of the models, only two of their polynomials were selected more than 2% of the time. However, when the polynomial term was included in the model, the first order term was also present in 97.9% of the models for the first feature and in 87.5% for the second feature.

Although the classification accuracy of the full models did not significantly differ from the first order models, the feature selection stability was much less. The time it

took to run this model was also eight to nine times longer than the first order models. It seems that this type of model introduces more noise than signal and therefore does not positively contribute to building a better classifier.

4.6 Comparison of Talairach and Tournoux and Craddock 200 Atlases

Different atlases provide different levels of resolution throughout the brain volume. Atlases with fewer ROIs use average signals from regions of the brain rather than individual voxel activity. This can be important to reduce the noise but could also restrict the ability to identify important activity patterns in smaller regions than some ROI atlases provide.

This section investigates the similarities and differences between the common Talairach and Tournoux (TT) atlas and the newer Craddock 200 (CC200) atlas. The TT atlas was important to this research as it has the least amount of ROIs, providing a simple atlas with not too large data that allowed for relatively fast experiments. The CC200 was developed specifically for FNC data and provides smaller regions that may capture brain activity closer to individual voxels than the TT atlas. This atlas has more than twice the number of ROIs, which can significantly slow some algorithms. In fact, while a single 1,000 run logistic regression experiment took about three hours for the TT data, the CC200 took between eight and fourteen hours.

The data from each atlas was run including the gender variable and same filtering strategies to remain consistent. The grid search for the CC200 regularization parameter was expanded from [0.0001, 4] to [0.0001, 5] to account for the increased number of ROIs in the atlas. Table 10 presents the information about the two atlases. Accuracy results from the experiments did not significantly differ between the atlases. The parameter and size of the basis did increase for the CC200 atlas, although that did not come as a surprise. The basis was less than twice the size of the TT atlas

even as the number of features in the CC200 atlas more than quadrupled.

Table 10. Experimental results of Talairarch and Tournoux and Craddock 200 atlases.

Atlas	Accuracy (%)	Parameter (C)	Number of Coefficients (s)
TT	62.79	0.7542	147.2
CC200	63.05	1.9375	266.5

Figure 13 illustrates how the size of the basis differed between the two atlases. The difference is quite distinct among the two atlases. While the Talairach atlas tends to center around the 100 to 150 basis, the CC200 is almost distributed uniformly across a wide range of values.

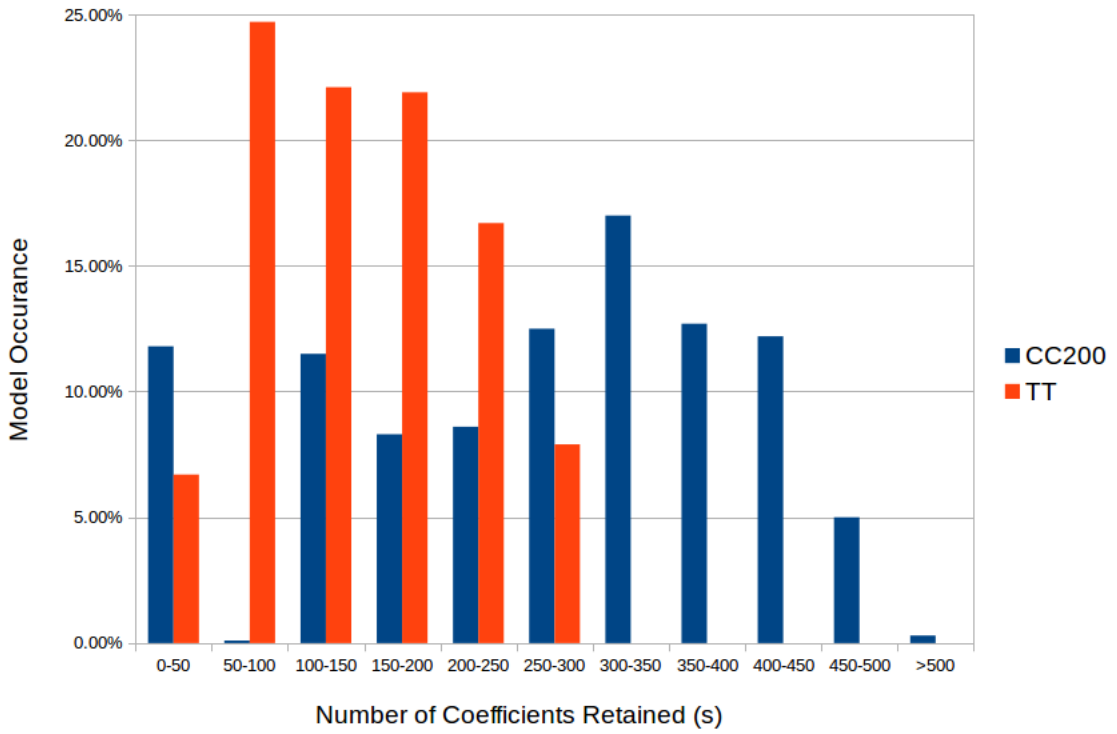


Figure 13. Experimental basis sizes for the Talairach and Tournoux and Craddock 200 atlases.

Although there was no evidence supporting that the CC200 atlas significantly outperformed the TT atlas at classification, information provided by the PCP allows

for direct interpretation of the model features. The CC200 remains a valuable atlas and is used later in the research.

4.7 Age Restricted Models

Several studies have experimented with restricting subjects to a certain age range with excellent results. It is hypothesized that brains of older ASD subjects may have different activity patterns than younger subjects. Two different age models were experimented with and their results are below.

The child model included all subjects under the age of nineteen. There were 283 ASD and 318 control subjects for a baseline accuracy of 52.91%. Both the TT and CC200 atlases were tested. The results from the TT atlas were disappointing. Accuracy dropped to $60.43 \pm 0.24\%$, less than 8 percentage points better than baseline as opposed to over 9 percentage points better with all subjects. The average number of columns in the basis was 135.5, similar to the other model, but the regularization parameter increased to 2.04.

The child model with CC200 increased accuracy to $63.19 \pm 0.24\%$, 10.3 percentage points greater than baseline. The basis and parameter were consistent with the full subject model. These results can be seen in Table 11.

The second model included adolescent subjects, defined as those whose ages were between 12 and 18 years. This model included 177 ASD and 202 control subjects and a baseline accuracy of 53.3%. Once again, the TT atlas model significantly worsened. The experiments reported an accuracy of only $60.80 \pm 0.32\%$, almost the same amount better than baseline as the child model. The average size of the basis did decrease though, as only 81.1 features were used in the model.

The CC200 adolescent model did not significantly outperform the model that included all subjects. The number of coefficients in the reduced basis was significantly

less than the other models. The CC200 model significantly outperforms the TT basis for every age group.

Table 11. Results of the child and adolescent models on the TT and CC200 data. Accuracy above baseline is the percentage point increase over each model's baseline accuracy.

Atlas	Subject Age	Accuracy (%)	Hyperparameter (C)	Number of Coefficients (s)	Accuracy Above Baseline
TT	All	62.79	0.7542	147.2	9.37
	≤ 18	60.43	2.0366	135.5	7.52
	12-18	60.80	2.0058	81.1	7.50
CC200	All	63.05	1.9375	266.5	9.63
	≤ 18	63.19	2.2959	226.3	10.28
	12-18	63.29	2.2400	144.8	9.99

The TT model results are somewhat surprising as literature suggested that separating subjects by age could boost accuracy within the model. The accuracy decrease with the TT atlas model could be caused by the larger ROIs failing to capture the relevant signal for the younger subjects. The CC200 atlas consistently outperforms the TT atlas in model accuracy, supporting the hypothesis that atlases play a major role in FNC classification.

4.8 Principal Component Analysis

Principal component analysis was used as a comparison to regularization as a feature selection method. The difficulty with PCA is the interpretability of the principal components. Regularization is superior in this aspect as the features selected in the model can be directly interpreted as a region in the brain whereas the principal components are combinations of these features. Regardless, PCA can be a useful tool.

Creating components to explain 60% of the variance in the data performed significantly better than any other PCA model for the TT atlas. Its accuracy was almost one percentage point higher than the baseline TT model. There was no significant

accuracy difference between the four other reported models. Table 12 displays the results of the experiments. Below 60% explained variance, the accuracy severely decreases.

Table 12. Experimental results of different principal component analysis based on the desired amount of explained variance.

Explained Variance (%)	Accuracy (%)	Components	Number of Coefficients (s)
60	63.76	43.2	27.6
75	63.16	80.9	65.1
80	63.05	106.6	81.7
85	63.19	141.1	102.5
90	63.00	188.9	115.1

RBF and sigmoid PCA kernels were also tried but failed to produce significant results (62.7% and 61.8%, respectively). The other machine learning algorithms were also run using PCA and while many significantly increased in accuracy, none failed to surpass our baseline 62.79% accuracy. These results are in Table 17 in Appendix 1.1.

Using PCA on the CC200 data also significantly improves the model performance as seen in Table 13. Principal components explaining 85% of the variance in the CC200 data improved the model accuracy from 63.05% to 65.52%. As the CC200 atlas contains smaller volume ROIs, the principal components could essentially be combining ROIs into a single feature that transforms the FNC data to something between the TT and CC200 atlas. Perhaps some regions in the brain can be effectively summarized over large volumes while other regions require display more significant activity near the voxel level.

The 65.5% accuracy was the highest average accuracy achieved among any at–tempted models, although several experimental runs reported holdout accuracy of almost 80%.

Table 13. Significant PCA results on the CC200 atlas data based on the desired amount of variance

Explained Variance (%)	Accuracy (%)	Components	Number of Coefficients
60	64.95	73.7	54
75	65.38	125.9	105.4
85	65.52	199.1	145.1

Adolescent Model.

While PCA had meager effects on the TT data for all subjects, it significantly increased the accuracy of the models that only include the adolescent subjects. Without PCA, this model reported 60.8% accuracy. With PCA explaining 75% of the variance of the data, this accuracy increased to $64.5 \pm 0.3\%$. In fact, each of the experimented variance levels above 60% increased significantly increased accuracy. The results of the other experiments are in Table 18 in Appendix 1.1.

Using PCA on the CC200 adolescent model significantly increased its accuracy as compared to the non-PCA model (64.2% versus 62.6%) but was more than one percentage point less accurate than using all subjects. The difference between the two atlases and their adolescent and all subject models may lie in where the important signals are extracted. The CC200 atlas may effectively capture the required signal to accurately classify regardless of age while the TT model may not for the younger subjects.

Regardless of the differences between the atlases, the results of using PCA to reduce the dimensionality of the data are promising. Principal component analysis increased the accuracy of every model, implying that the FNC data may be reducible to a lower dimension. PCA is a simple linear combination of the features but more advanced manifold learning techniques may be able to provide a better reduction while retaining interpretability.

4.9 Brain Abnormalities

The connectivity of specific regions in brains of those with autism spectrum disorder could be a significant factor in the diagnosis process. Vidaurre *et al.* (2013) reported that a majority of studies concluded that adult subjects with ASD showed a decreased connectivity (hypoconnectivity). This section presents our findings gained by the CC200 models. The first subsection discusses overall connectivity in the brains as suggested by the experiments while the second section discusses specific regions highlighted as important in the models. Connectivity is defined in this section as the average FNC value for a given feature.

Connectivity Levels.

The experiments on the CC200 model including all 805 subjects selected 692 features more often than the random noise. We refer to the top 227 (the average number of coefficients in the experimental models) of these as the selected features. The subject's gender was included in the selected features but removed for this analysis. The average connectivity of subjects with ASD was slightly, but not significantly, less than the average connectivity of the control subjects. Those with autism spectrum disorder had an average connectivity of 0.307 ± 0.011 versus the average connectivity of 0.314 ± 0.011 for the control subjects.

Table 14. Average functional network connectivity of the selected features suggests that ASD subjects have slightly below average connectivity.

Subject	Average Connectivity
Control	0.314
Autism	0.307

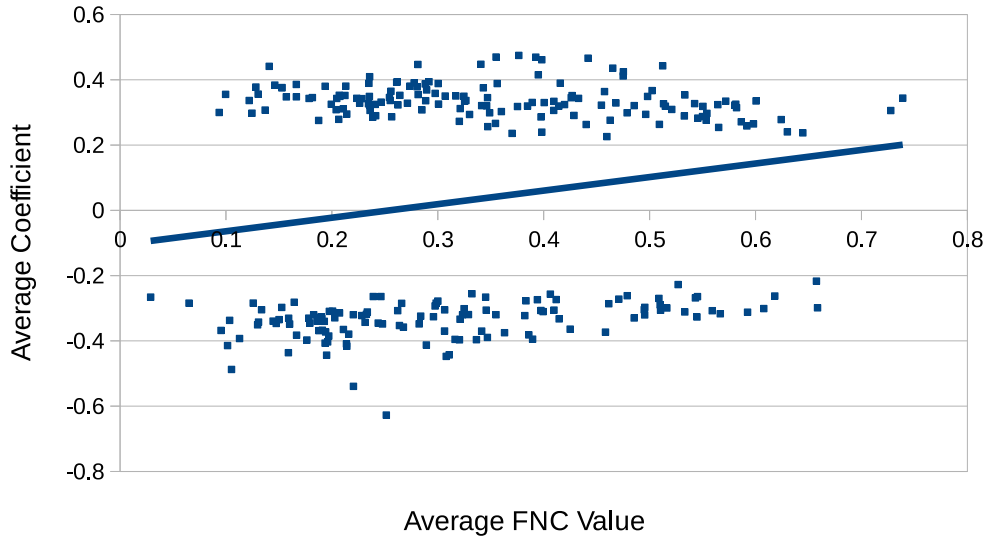
However, investigating how the average connectivity influences the model coefficients presents a clearer result. Figure 14 illustrates the differences between the

two diagnoses. ASD data is represented in orange and the controls in blue. There is a clear separation of positive and negative coefficients, indicating that the coefficients of selected features were stable. Most of the data is clustered around the mean connectivity and lower.

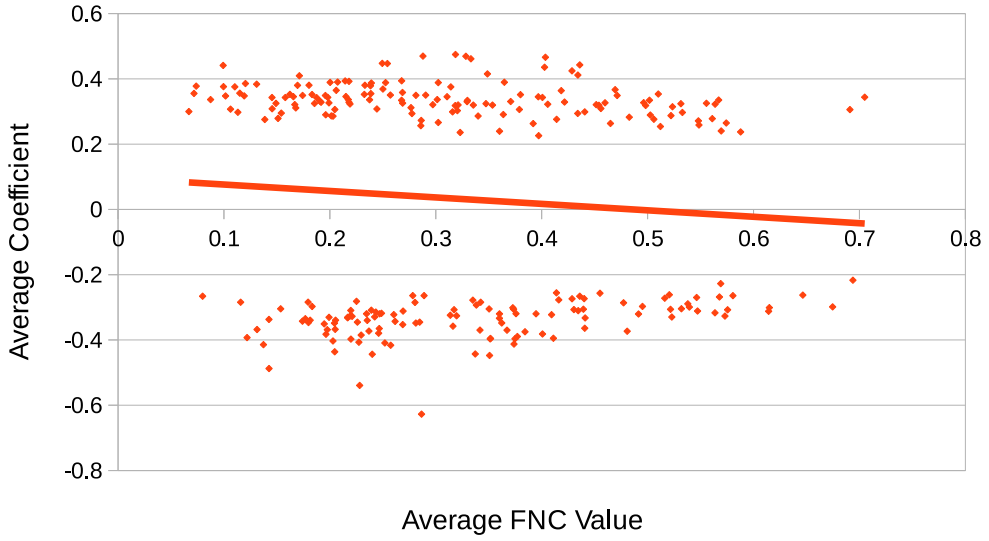
Two linear models are overlaid on the figure; their color corresponds to their respective diagnosis group. These models intersect at 0.311, the average connectivity of all subjects. The positive slope of the control model indicates that an increasing connectivity correlates with a more positive coefficient. This suggests that the model generally finds greater connectivity to be an indicator of a control subject. The negative slope of the ASD model indicates the same: increasing connectivity corresponds with a more negative coefficient.

There are, of course, features in which this is not the case. Many coefficients indicate a higher connectivity is more likely to be classified as an autistic subject. The cluster of points under the horizontal axis in Figure 14 displays this effect. Most of these negative coefficients are with a greater FNC of autistic subjects than control subjects.

The top 60 most selected features reveals that the model coefficients are positive about 65% of the time. This, combined with the evidence of the linear models, supports the theory that subjects with ASD display hypoconnectivity in comparison to control subjects. We did not find evidence that younger ASD subjects displayed hyperconnectivity. Figure 19 in Appendix A displays the results from the models only including subjects eighteen years old and younger. The younger control subjects display a more positive linear regression slope than all subjects ($\beta_{<18} = 0.548$ vs. $\beta_{all} = 0.425$) indicating an increased hypoconnectivity compared to the model with all subjects.



(a) Control Subjects



(b) Autistic Subjects

Figure 14. The two figures above display how the average subject connectivity affects the model coefficients. Figure (a) with the control subjects displays a positive linear relationship ($p = 0.002$), suggesting that the model classifies higher connectivity values as controls. While the ASD subjects in Figure (b) shows a slight negative relationship, the trend is not statistically significant ($p = 0.226$). This is evidence that the model attributes lower connectivity values with ASD.

Regional Findings.

In this section, model occurrence refers to how often a feature is included in the reduced basis of the regularized model over the 1,000 experimental runs. Due to the

restricted isometery property failing to hold, we are not guaranteed a unique reduced basis for our data. Therefore, the inclusion of features varies over each experimental run. Here is where the artificial variable is useful, as we can compare how often it is selected versus the features in the data.

Model occurrence is not the only way to determine the relative importance of a feature. Coefficient magnitude is another successful strategy. Larger magnitudes indicate greater importance in the individual models. These two measures are cor–related ($\rho = 0.605$); coefficients with a greater magnitude are also more likely to also occur in more of the experimental runs. Figure 15 displays how a greater magnitude also usually means a higher rate of occurrence. This analysis focuses on the features with the greatest rates of occurrence.

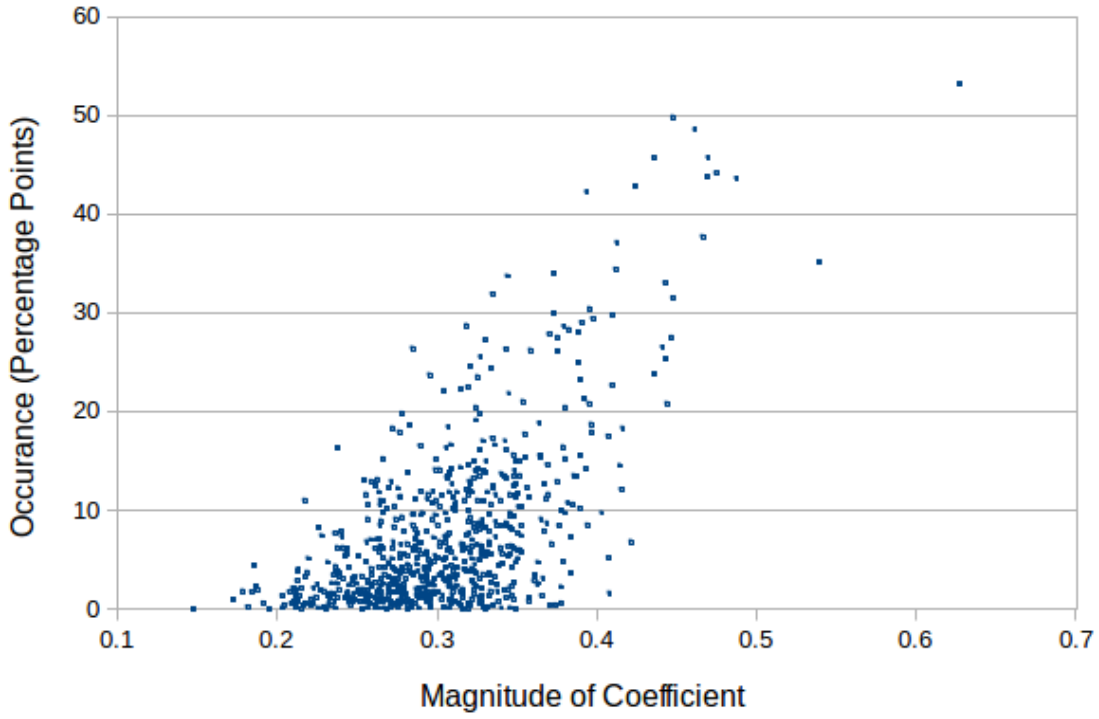


Figure 15. The magnitude of the model coefficient impacts the rate of occurrence in the experimental models. A larger magnitude correlates to an increased occurrence rate, displayed here in percentage points greater than the artificial noise variable.

Although the CC200 atlas does not segment the brain volume by "functional" regions, the location of the center of mass of the CC200 ROIs can determine the corresponding regions of other atlases. The Talairach and Tournoux atlas provides our interpretations. Due to differences from crafting the different atlases, most of the ROIs from the CC200 atlas overlap multiple regions of the TT atlas. The TT region that includes the most volume of the CC200 ROI is reported.

Table 15 shows the ten connections most often selected for the model. Eight of these ten features display hypoconnectivity among the ASD subjects, with matching positive coefficients. Immediately these results stand out as the left and right temporal gyrii are prevalent throughout these connections. A visual representation of these results is in Figure 16. The size of the coefficient is represented by the color saturation

of the nodes while the edges are the center of mass of the individual regions of interest of the feature.

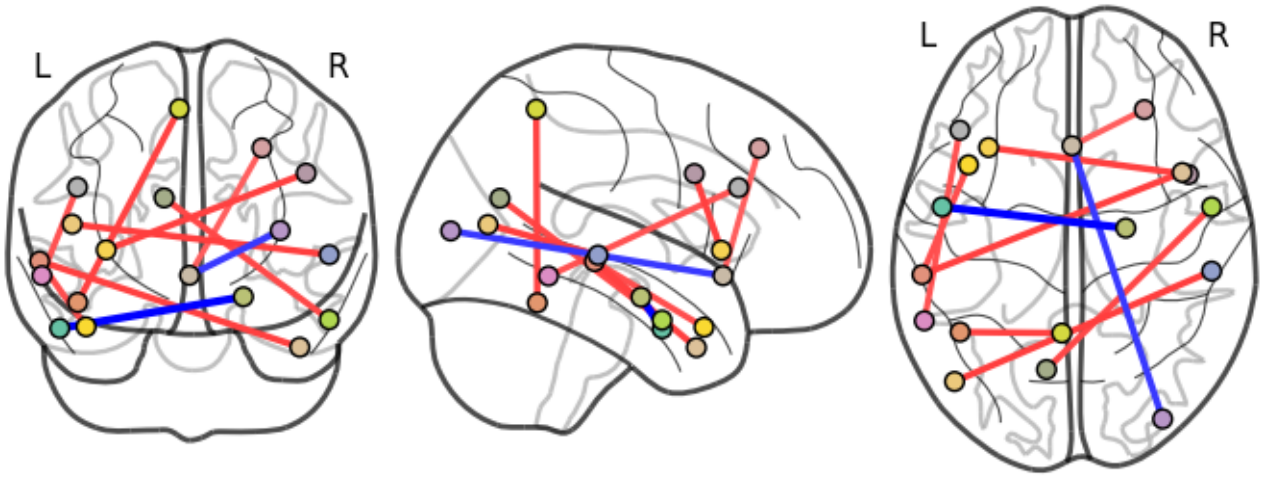


Figure 16. The size of the model coefficient is represented by the color saturation. Red edges signify a positive coefficient while blue represent a negative coefficient. Each node is the center of mass of the ROIs from the selected features in Table 15.

Table 15. The ten features with the highest rate of occurrence in the experimental runs. Each connectivity value corresponds to the correlation between two regions in the brain. These regions are below, along with their average model coefficient and the connectivity values for both ASD and control subjects.

Region of Interest		Average Coefficient	Average Connectivity	
			Autism	Control
Left Inferior Temporal Gyrus	Right Parahippocampal Gyrus	-0.627	0.287	0.251
Left Precuneus	Right Middle/Inferior Temporal Gyrus	0.448	0.249	0.341
Left Fusiform Gyrus	Left Precuneus	0.461	0.333	0.398
Left Middle Temporal Gyrus	Left Superior Temporal Gyrus	0.436	0.403	0.465
Right Middle/Inferior Frontal Gyrus	Left Insula	0.470	0.288	0.355
Left Middle Temporal Gyrus	Right Middle/Superior Temporal Gyrus	0.475	0.319	0.377
Left Middle Temporal Gyrus	Right Middle/Superior Temporal Gyrus	0.469	0.328	0.393
Right Middle Occipital Gyrus	Anterior Cingulate Cortex	-0.487	0.143	0.105
Left Middle Temporal Gyrus	Left Middle/Inferior Frontal Gyrus	0.424	0.429	0.475
Anterior Cingulate Cortex	Right Middle/Superior Frontal Gyrus	0.394	0.214	0.262

The temporal lobe is one of the four major lobes of the human brain. It is responsible for memory storage and retrieval (Squire & Zola-Morgan, 1991), as well as visual stimuli, including facial recognition (Baylis *et al.* , 1987). The left middle temporal gyrus is highlighted in Figure 17 to provide the reader a location reference. The left and right middle temporal gyrus were selected in half of the top connections.

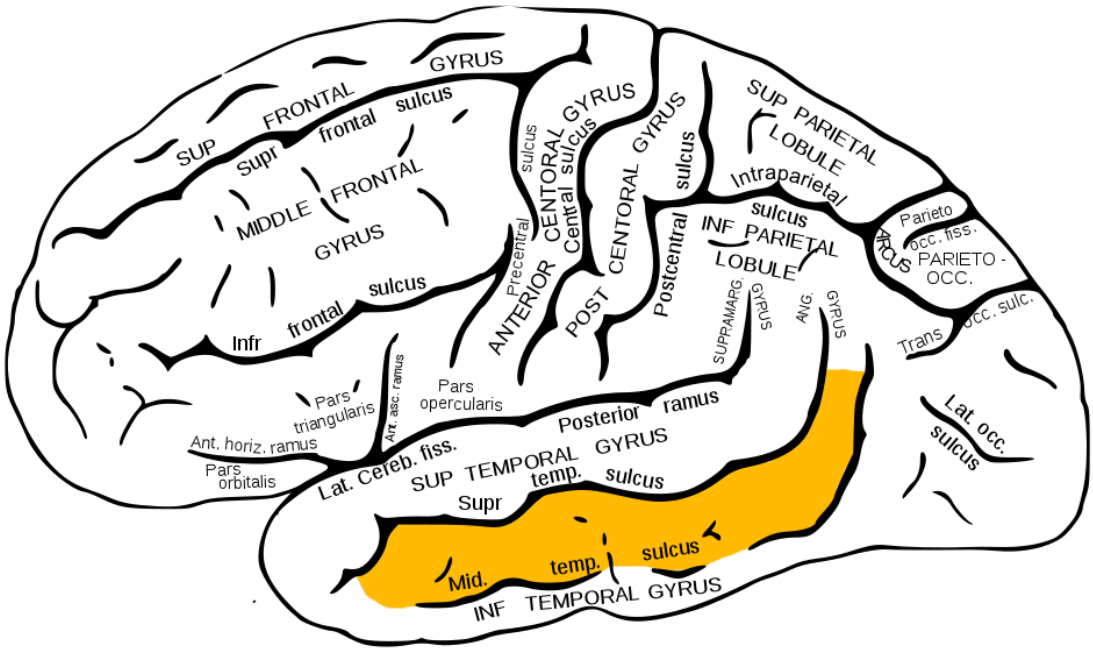


Figure 17. The left middle temporal gyrus is highlighted in yellow.
From: `(Gray, 1918), Mysid and was_a_bee / Wikipedia Commons / Public Domain`

The most significant feature is the connectivity between the left inferior temporal gyrus and the right parahippocampal gyrus. The large negative coefficient signifies hyperconnectivity between the regions in autism. The left inferior temporal gyrus is thought to be partly responsible for facial recognition (Haxby *et al.* , 2000), numer–ical perception (Shum *et al.* , 2013), and spatial awareness, although still debated (Karnath *et al.* , 2001). The parahippocampal gyrus is thought to be responsible for perceiving the surrounding visual environment (Epstein & Kanwisher, 1998) as well as understanding social cues (Rankin *et al.* , 2009).

There is evidence of significant hypoconnectivity in ASD between the left middle

temporal gyrus and the right middle and superior temporal gyrus. Furthermore, there is also evidence of hypoconnectivity within the left temporal lobe, as identified by the fourth coefficient in Table 15. These findings suggest a possible hypoconnectivity within the entire temporal lobe.

While not exactly known, there are several studies that suggest the functions of middle temporal gyrus. Acheson & Hagoort (2013) proposes that the posterior middle temporal gyrus is used for language storage and retrieval. They found that the combination of the middle temporal gyrus and the inferior frontal gyrus may be very important for reading comprehension.

Haxby *et al.* (2000) noted that while facial recognition sparks activity all throughout the brain, the middle temporal gyrus is prevalent throughout the process. They noted that the superior temporal gyrus, in the same region as the middle, was responsible for understanding the "changeable aspects" of faces, such as eye gaze, expression, and lip movement. This could play a role in how humans understand other humans' emotions.

Vandenberghe *et al.* (1996) also found that the middle temporal gyrus was partially responsible for processing visual stimuli. By comparing comprehension of images of objects such as a squirrel versus an image of the word squirrel, they found a semantic network was present through the left middle temporal gyrus.

The frontal lobe was also identified in multiple connections. This part of the brain is responsible for voluntary action, such as movement and understanding future consequences (Miyake *et al.* , 2000) as well as the dopamine system. Figure 18 shows the left inferior frontal gyrus, with the middle and superior frontal gyrus above it.

There is no evidence in our top selected features of abnormal connectivity within the frontal lobe, although the connectivity among several gyrus' in the lobe and regions outside do show decreased connectivity in ASD subjects. The following is a

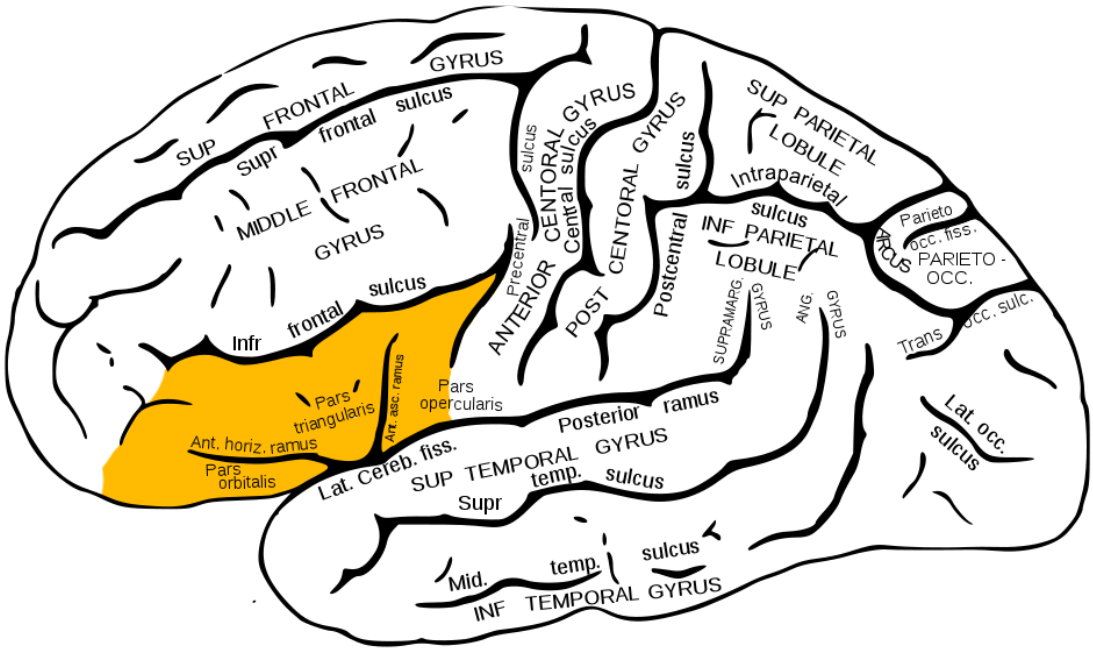


Figure 18. The left inferior frontal gyrus is highlighted in yellow. The middle frontal gyrus is directly above it.
From: (Gray, 1918), Mysid and was_a_bee / Wikipedia Commons / Public Domain

brief discussion on the functions of the significant gyrus within the frontal lobe.

Pedersen *et al.* (1998) consistently found that activity within the left middle frontal gyrus was detected up between 900 and 200 ms before the subject moved a finger. The left and right middle and inferior frontal gyrus are also associated with inhibition. Aron *et al.* (2004) found that subjects with damaged frontal lobes performed significantly worse on inhibition tests. They also note that this region might be essential for the inhibition of memory retrieval (such as trying to forget a bad memory).

Hampshire *et al.* (2010) goes further, explaining that the right inferior frontal gyrus is not just about inhibition. They hypothesize that this region is also respon–sible for attention control, adapting to new stimuli before proceeding with inhibition signals.

Finally, the anterior cingulate cortex (ACC) was identified twice in the top con–

nectivity. The ACC is part of the limbic system in the brain and is thought to be involved with cognition and emotional response. The cortex influences activity in the other brain regions to regulate cognitive, motor, endocrine, and visceral responses (Bush *et al.* , 2000). It is located at the rear of the brain.

This region had evidence of hyperconnectivity among autistic subjects with the right middle occipital gyrus, which is mainly responsible for processing visuals (Van–denberghe *et al.* , 1996). There was evidence of hypoconnectivity between the ACC and the right middle and superior frontal gyrus. The function of that region was discussed above.

These ten connectivities are similar to findings in other ABIDE studies. Nielsen *et al.* (2013) reported the highest accuracy from the parahippocampal and fusiform gyri, insula, prefrontal cortex, posterior cingulate cortex, and superior temporal gyrus (reported as the Wernicke Area), all the same or close regions to our results. They also reported the intraparietal sulcus, which was not included in our results.

Chen *et al.* (2015) reported that the parietal and left occipital lobes had the most significant connections, followed by the frontal and the right occipital lobe. Our results agree with the frontal and somewhat with the occipital, but failed to replicate those connections in the parietal lobe.

Tyszka *et al.* (2014) found that the left frontal gyrus and left middle and inferior temporal gyrus exhibited significant differences between ASD and control subjects.

Summary

This chapter explained the results from our research. While the restricted isome–tery property failed to hold for our data, there are still techniques that result in a close to global optimal sparse coefficient basis. We also saw how different variables affected the models and found that the Craddock 200 atlas and principal component

analysis combined to create the most accurate model. The results of these different models are in Table 16

Finally, we reported on several regions highlighted as significantly impacting our model. While these regions traversed the brain volume, several were clustered in the temporal lobe. Interestingly, the functions of the regions reported with abnormal connectivity values are all associated with behaviors and processes often deficient in autism. While we are not medically trained nor experts in the field of autism, several of these findings may prove useful to researchers in their quest to determine the cause of autism spectrum disorder.

Table 16. A summary of different model accuracies from our research.

Model	Accuracy
TT Logistic Regression	62.79
TT Interaction Logistic Regression	62.65
TT Logistic Regression with PCA	63.76
CC200 Logistic Regression	63.05
CC200 Logistic Regression with PCA	**65.52**
Random Forest	57.9
AdaBoost	56.9
Linear SVM	61.0
RBF SVM	62.4

V. Conclusions and Recommendations

Autism spectrum disorder can be a debilitating condition for those diagnosed. With annual costs commonly in excess of $60,000 for extra care, autism can be a significant burden on affected families (AutismSpeaks). Diagnosis still takes several months of behavioral observation and the diagnosis can vary among professionals. While several studies on fMRI and ASD classification reported good results, these studies were all small scale with only a few subjects. The release of the Autism Brain Imaging Data Exchange, with data on over 1,100 subjects, allows for the first time research on a large data sample. The goal of this research was to attempt to create a logistic regression classifier that could take fMRI data and accurately distinguish between autism spectrum disorder and healthy patients.

Classifiers

We used several different machine learning algorithms to compare how the accuracy differed with different techniques. Several algorithms, such as random forest and support vector machine, were reported in literature to provide highly accurate classification for ASD. These algorithms failed to outperform our logistic classifier, although they could possibly be specifically tuned for our data.

5.1 Logistic Regression Classifier Results

Our logistic regression classifier with ℓ_1 regularization effectively reduced the dimension of the data from almost 20,000 features with the CC200 atlas to only a couple hundred in the reduced basis. The CC200 data was classified with 63.05% accuracy, not exactly groundbreaking but about the same as other studies have reported. This result and other research on the ABIDE data suggest that large scale

model generalization is difficult with ASD classification.

Principal component analysis was also used as an attempt to reduce the dimension of the data before fitting the model. PCA has been known to increase the accuracy of models as it can extract hidden dimensions that could explain the data better. We found that PCA does significantly increase the accuracy of the models, boosting the CC200 model to 65.52%. We lose the interpretability however, limiting the practicability of using the technique.

5.2 Connectivity Abnormalities

Although the model did not provide evidence of volume-wide hypoconnectivity in ASD subjects, over 60% of the most important features and 8 of the top 10 displayed lower connectivity levels. In general, decreased connectivity increased the odds of classifying as ASD.

Abnormal connectivities were reported in several regions of the brain. The temporal lobe displayed hypoconnectivity within the entire lobe and the left temporal gyrii were less connected to outside regions in subjects with ASD as compared to the controls.

Several gyrii within the frontal lobe also displayed lower connectivity values in ASD subjects. Two connections were hyperconnected in ASD subjects, although both of these connectivity values were significantly less than average for both the control and ASD subjects.

The significance of the parahippocampal gyrus, insula, prefrontal cortex, cingulate cortex, and temporal gyrii are also reported in other autism studies. The abnormalities in the left precuneus and right middle occipital gyrus were previously unreported in the autism studies and could be subject to further research.

5.3 Recommended Future Work

The differences between the TT and CC200 atlas and the increased accuracy with principal component analysis suggest that a novel atlas could be effective for extracting FNC values relevant to ASD. A new database, the ABIDE 2, is due to release soon and could be used to compare the performance of this new atlas over an even larger collection of subjects.

Principal component analysis was also effective in reducing the experimental run times. PCA reduced CC200 run times from around nine to only one and a half hours. Data interpretability suffers after PCA which suggests that a more advanced dimen–sionality reduction technique could be useful for this neuroimaging data. Manifold learning techniques could possibly reduce the size of the data while a higher degree of interpretability.

Of course, if interpretability is not a goal of research, deep convolutional neural networks have shown promise of high classification accuracy albeit more complex and slower to fit than a logistic classification.

Finally, although our classifier performed well with respect to literature, a 65% classification accuracy is too low for clinical value. Due to the early age of typical di–agnosis, there may also be problems acquiring the data. This data does not represent the typical diagnosis age (only a few years old) for ASD. Truly diagnosing ASD is not possible unless new models are developed using fMRI data from young subjects.

Appendix A.

1.1 Other Model PCA Results

Table 17. Experimental results on PCA with other machine learning algorithms. Reported is the average model accuracy and the change in accuracy of the model without using PCA. Each model was run with 85% variance explained.

Algorithm	Accuracy (%)	Δ%
Random Forest	56.8	-1.1
ADABoost	57.2	0.3
Linear SVM	62.1	1.1
RBF SVM	62.2	-0.1

Table 18. PCA results on the adolescent TT data indicates a significant increase in classification accuracy above 65% explained variance.

Explained Variance (%)	Accuracy (%)
30	51.3
50	59.6
65	64.2
70	63.8
75	64.5
85	63.2
90	63.1
95	63.3

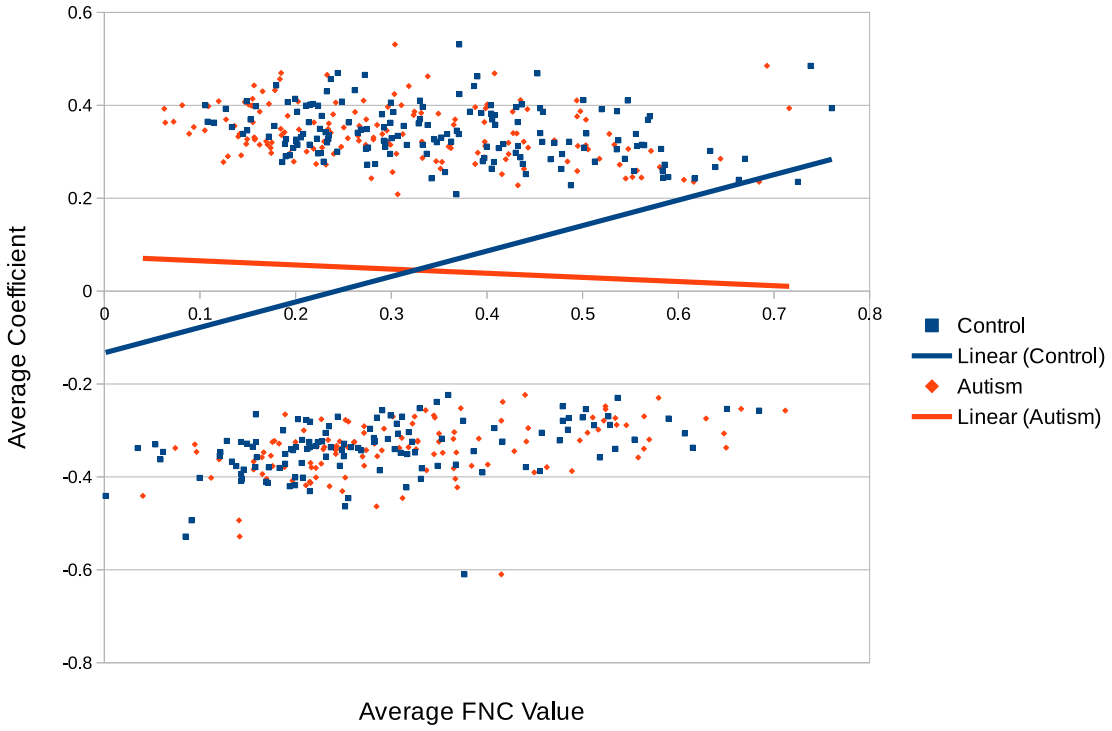


Figure 19. Coefficients versus average connectivity of subjects aged 18 years and younger. The increasing linear model for the control subjects again provides evidence of hypoconnectivity among autistic subjects. The Control subjects have a positive slope ($\beta = 0.548, \quad p < 0.0001$) while the ASD linear model is not statistically different from zero ($p = 0.530$)

Bibliography

Abraham, Alexandre, Pedregosa, Fabian, Eickenberg, Michael, Gervais, Philippe, Mueller, Andreas, Kossaifi, Jean, Gramfort, Alexandre, Thirion, Bertrand, & Varoquaux, Gaël. 2014. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, **8**.

Acheson, Daniel J, & Hagoort, Peter. 2013. Stimulating the brain's language network: syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus. *Journal of cognitive neuroscience*, **25**(10), 1664–1677.

American Psychiatric Association and others, year=2003, publisher=ManMag. *Diagnostic and Statistical Manual of Mental Disorders:: DSM-5*.

Anderson, Jeffrey S, Nielsen, Jared A, Froehlich, Alyson L, DuBray, Molly B, Druzgal, T Jason, Cariello, Annahir N, Cooperrider, Jason R, Zielinski, Brandon A, Ravichandran, Caitlin, Fletcher, P Thomas, *et al.* . 2011. Functional connectivity magnetic resonance imaging classification of autism. *Brain*, **134**(12), 3742–3754.

Aron, Adam R, Robbins, Trevor W, & Poldrack, Russell A. 2004. Inhibition and the right inferior frontal cortex. *Trends in cognitive sciences*, **8**(4), 170–177.

Bandeira, A, Dobriban, E., Mixon, D., & Sawin, W. 2012. Certifying the restricted isometry property is hard. *IEEE Transactions on Information Theory*, **59**, 3448–3450.

Baylis, Gordon C, Rolls, Edmund T, & Leonard, CM. 1987. Functional subdivisions of the temporal lobe neocortex. *The Journal of Neuroscience*, **7**(2), 330–342.

Birn, Rasmus M, Molloy, Erin K, Patriat, Rémi, Parker, Taurean, Meier, Timothy B, Kirk, Gregory R, Nair, Veena A, Meyerand, M Elizabeth, & Prabhakaran, Vivek. 2013. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage*, **83**, 550–558.

Biswal, Bharat B, Kylen, Joel Van, & Hyde, James S. 1997. Simultaneous assessment of flow and BOLD signals in resting-state functional connectivity maps. *NMR in Biomedicine*, **10**(45), 165–170.

Bush, George, Luu, Phan, & Posner, Michael I. 2000. Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences*, **4**(6), 215–222.

Calderoni, Sara, Retico, Alessandra, Biagi, Laura, Tancredi, Raffaella, Muratori, Filippo, & Tosetti, Michela. 2012. Female children with autism spectrum disorder: an insight from mass-univariate and pattern classification analyses. *Neuroimage*, **59**(2), 1013–1022.

Candes, E., & Tao, T. 2005. The Dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, **35**(6), 2313–2351.

Candes, E., & Tao, T. 2006. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, **52**(12), 5406–5425.

Candes, Emmanuel J, & Tao, Terence. 2005. Decoding by linear programming. *Information Theory, IEEE Transactions on*, **51**(12), 4203–4215.

Candès, Emmanuel J, & Wakin, Michael B. 2008. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, **25**(2), 21–30.

Caruana, Rich, & Niculescu-Mizil, Alexandru. 2006. An empirical comparison of supervised learning algorithms. *Pages 161–168 of: Proceedings of the 23rd international conference on Machine learning*. ACM.

Cetin, Mustafa S, Christensen, Fletcher, Abbott, Christopher C, Stephen, Julia M, Mayer, Andrew R, Canive, José M, Bustillo, Juan R, Pearlson, Godfrey D, & Calhoun, Vince D. 2014. Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia. *NeuroImage*, **97**, 117–126.

Chen, Colleen P, Keown, Christopher L, Jahedi, Afrooz, Nair, Aarti, Pflieger, Mark E, Bailey, Barbara A, & Müller, Ralph-Axel. 2015. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *NeuroImage: Clinical*, **8**, 238–245.

Chen, Heng, Duan, Xujun, Liu, Feng, Lu, Fengmei, Ma, Xujing, Zhang, Youxue, Ud–din, Lucina Q, & Chen, Huafu. 2016. Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivitya multi-center study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, **64**, 1–9.

Cortes, Corinna, & Vapnik, Vladimir. 1995. Support-vector networks. *Machine learning*, **20**(3), 273–297.

Craddock, Cameron, Milham, Michael, Giavasis, Steven, Clark, Daniel, Froehlich, Carol, & Pellman, John. 2015 (Apr). *Configurable Pipeline for the Analysis of Connectomes*. `https://fcp-indi.github.io/docs/user/`.

Craddock, R. Cameron, & Bellec, Pierre. 2015. *Preprocessed Connectomes Project: ABIDE*. `https://preprocessed-connectomes-project.github.io/abide/index.html`.

Craddock, R Cameron, James, G Andrew, Holtzheimer, Paul E, Hu, Xiaoping P, & Mayberg, Helen S. 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping*, **33**(8), 1914–1928.

Dagli, Mandeep S, Ingeholm, John E, & Haxby, James V. 1999. Localization of cardiac-induced signal change in fMRI. *Neuroimage*, **9**(4), 407–415.

Della-Maggiore, Valeria, Chau, Wilkin, Peres-Neto, Pedro R, & McIntosh, Anthony R. 2002. An Empirical Comparison of SPM Preprocessing Parameters to the Analysis of fMRI Data. *NeuroImage*, **17**, 19–28.

Di Martino, Adriana, Yan, Chao-Gan, Li, Qingyang, Denio, Erin, Castellanos, Francisco X, Alaerts, Kaat, Anderson, Jeffrey S, Assaf, Michal, Bookheimer, Susan Y, Dapretto, Mirella, *et al.* . 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, **19**(6), 659–667.

Dillon, William R, & Goldstein, Matthew. 1984. *Multivariate analysis: Methods and applications*.

Ecker, Christine, Marquand, Andre, Mourão-Miranda, Janaina, Johnston, Patrick, Daly, Eileen M, Brammer, Michael J, Maltezos, Stefanos, Murphy, Clodagh M, Robertson, Dene, Williams, Steven C, *et al.* . 2010a. Describing the brain in autism in five dimensions magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. *The Journal of Neuroscience*, **30**(32), 10612–10623.

Ecker, Christine, Rocha-Rego, Vanessa, Johnston, Patrick, Mourao-Miranda, Janaina, Marquand, Andre, Daly, Eileen M, Brammer, Michael J, Murphy, Clodagh, Murphy, Declan G, Consortium, MRC AIMS, *et al.* . 2010b. Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. *Neuroimage*, **49**(1), 44–56.

Epstein, Russell, & Kanwisher, Nancy. 1998. A cortical representation of the local visual environment. *Nature*, **392**(6676), 598–601.

Fan, Rong-En, Chang, Kai-Wei, Hsieh, Cho-Jui, Wang, Xiang-Rui, & Lin, Chih-Jen. 2008. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, **9**, 1871–1874.

Fox, Michael D, Zhang, Dongyang, Snyder, Abraham Z, & Raichle, Marcus E. 2009. The global signal and observed anticorrelated resting state brain networks. *Journal of neurophysiology*, **101**(6), 3270–3283.

Friston, KJ, Frith, CD, Liddle, PF, & Frackowiak, RSJ. 1993. Functional connectivity: the principal-component analysis of large (PET) data sets. *Journal of cerebral blood flow and metabolism*, **13**, 5–5.

Ghiassian, Sina, Greiner, Russell, Jin, Ping, & Brown, Matthew RG. 2013. *Learning to Classify Psychiatric Disorders based on fMR Images: Autism vs Healthy and ADHD vs Healthy*.

Gray, Henry. 1918. *Anatomy of the human body.* Lea & Febiger. Plate 726, Accessed: Wikipedia Commons / Public Domain.

Guyon, Isabelle, & Elisseeff, André. 2003. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, **3**, 1157–1182.

Hampshire, Adam, Chamberlain, Samuel R, Monti, Martin M, Duncan, John, & Owen, Adrian M. 2010. The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage*, **50**(3), 1313–1319.

Haxby, James V, Hoffman, Elizabeth A, & Gobbini, M Ida. 2000. The distributed hu–man neural system for face perception. *Trends in cognitive sciences*, **4**(6), 223–233.

Iidaka, Tetsuya. 2015. Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex*, **63**, 55–67.

Jiao, Y, Chen, R, Ke, X, Cheng, L, Chu, K, Lu, Z, & Herskovits, EH. 2011. Pre–dictive models for subtypes of autism spectrum disorder based on single-nucleotide polymorphisms and magnetic resonance imaging. *Advances in medical sciences*, **56**(2), 334.

Just, Marcel Adam, Cherkassky, Vladimir L, Keller, Timothy A, Kana, Rajesh K, & Minshew, Nancy J. 2007. Functional and anatomical cortical underconnectivity in autism: evidence from an FMRI study of an executive function task and corpus callosum morphometry. *Cerebral cortex*, **17**(4), 951–961.

Karnath, Hans-Otto, Ferber, Susanne, & Himmelbach, Marc. 2001. Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, **411**(6840), 950–953.

Kogan, Michael D, Blumberg, Stephen J, Schieve, Laura A, Boyle, Coleen A, Perrin, James M, Ghandour, Reem M, Singh, Gopal K, Strickland, Bonnie B, Trevathan, Edwin, & van Dyck, Peter C. 2009. Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics*, **124**(5), 1395–1403.

Lebedev, Alexander V. 2014. MLSP 2014 Schizophrenia Classification Challenge.

Loh, Po-Ling, & Wainwright, Martin J. 2011. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Pages 2726–2734 of: Advances in Neural Information Processing Systems.*

Long, Philip M, & Servedio, Rocco A. 2010. Random classification noise defeats all convex potential boosters. *Machine Learning*, **78**(3), 287–304.

Lord, Catherine, & Jones, Rebecca M. 2012. Annual Research Review: Re-thinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, **53**(5), 490–509.

Lord, Catherine, Cook, Edwin H, Leventhal, Bennett L, & Amaral, David G. 2000. Autism spectrum disorders. *Neuron*, **28**(2), 355–363.

Matthews, PM, & Jezzard, P. 2004. Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, **75**(1), 6–12.

Miyake, Akira, Friedman, Naomi P, Emerson, Michael J, Witzki, Alexander H, How–erter, Amy, & Wager, Tor D. 2000. The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive psychology*, **41**(1), 49–100.

Montgomery, Douglas C, Peck, Elizabeth A, & Vining, G Geoffrey. 2012. *Introduction to linear regression analysis*. Vol. 821. John Wiley & Sons.

Mugler, John P, & Brookeman, James R. 1990. Three-dimensional magnetiza–tion-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, **15**(1), 152–157.

Murphy, Kevin, Birn, Rasmus M, Handwerker, Daniel A, Jones, Tyler B, & Bandet–tini, Peter A. 2009. The impact of global signal regression on resting state corre–lations: are anti-correlated networks introduced? *Neuroimage*, **44**(3), 893–905.

Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective*. MIT press.

Narayan, Manjari, & Allen, Genevera I. 2015. Mixed Effects Models to Find Differ–ences in Multi-Subject Functional Connectivity. *bioRxiv*, 027516.

Natarajan, B. K. 1995. Sparse Approximate Solutions to Linear Systems. *SIAM J. Comput.*, **24**, 227–234.

National Institute of Mental Health. 2015. *Autism Spectrum Disorder*. `http://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml`. Accessed: 2015-05-18.

Nielsen, Jared A, Zielinski, Brandon A, Fletcher, P Thomas, Alexander, Andrew L, Lange, Nicholas, Bigler, Erin D, Lainhart, Janet E, & Anderson, Jeffrey S. 2013. Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in human neuroscience*, **7**.

Nielsen, Jared A, Zielinski, Brandon A, Fletcher, P Thomas, Alexander, Andrew L, Lange, Nicholas, Bigler, Erin D, Lainhart, Janet E, & Anderson, Jeffrey S. 2014. Abnormal lateralization of functional connectivity between language and default mode regions in autism. *Molecular autism*, **5**(8).

Ogawa, Seiji, Lee, Tso-Ming, Kay, Alan R, & Tank, David W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, **87**(24), 9868–9872.

Orrù, Graziella, Pettersson-Yeo, William, Marquand, Andre F, Sartori, Giuseppe, & Mechelli, Andrea. 2012. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*, **36**(4), 1140–1152.

Pedersen, Jane R, Johannsen, Peter, Bak, Christen K, Kofoed, Bent, Saermark, Knud, & Gjedde, Albert. 1998. Origin of human motor readiness field linked to left middle frontal gyrus by MEG and PET. *Neuroimage*, **8**(2), 214–220.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. 2011a. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, *et al.* . 2011b. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, **12**, 2825–2830.

Rankin, Katherine P, Salazar, Andrea, Gorno-Tempini, Maria Luisa, Sollberger, Marc, Wilson, Stephen M, Pavlic, Danijela, Stanley, Christine M, Glenn, Shenly, Weiner, Michael W, & Miller, Bruce L. 2009. Detecting sarcasm from paralinguistic cues: anatomic and cognitive correlates in neurodegenerative disease. *Neuroimage*, **47**(4), 2005–2015.

Rauhut, Holger. 2010. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, **9**, 1–92.

Schapire, Robert E, & Freund, Yoav. 2012. *Boosting: Foundations and algorithms*. MIT press.

Shum, Jennifer, Hermes, Dora, Foster, Brett L, Dastjerdi, Mohammad, Rangarajan, Vinitha, Winawer, Jonathan, Miller, Kai J, & Parvizi, Josef. 2013. A brain area for visual numerals. *The Journal of Neuroscience*, **33**(16), 6709–6715.

Sladky, Ronald, Friston, Karl J, Tröstl, Jasmin, Cunnington, Ross, Moser, Ewald, & Windischberger, Christian. 2011. Slice-timing effects and their correction in functional MRI. *Neuroimage*, **58**(2), 588–594.

Squire, Larry R, & Zola-Morgan, Stuart. 1991. The medial temporal lobe memory system. *Science*, **253**(5026), 1380–1386.

Talairach, Jean, & Tournoux, Pierre. 1988. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging*. Stuttgart: Thieme Classics.

Tyszka, J Michael, Kennedy, Daniel P, Paul, Lynn K, & Adolphs, Ralph. 2014. Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cerebral Cortex*, **24**(7), 1894–1905.

Van Der Walt, Stefan, Colbert, S Chris, & Varoquaux, Gael. 2011. The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, **13**(2), 22–30.

Van Essen, David C. 1997. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *NATURE-LONDON-*, 313–318.

Vandenberghe, Rik, Price, Cathy, Wise, R, Josephs, O, & Frackowiak, RS. 1996. Functional anatomy of a common semantic system for words and pictures. *Nature*, **383**(6597), 254–6.

Vidaurre, Diego, Bielza, Concha, & Larrañaga, Pedro. 2013. A survey of L1 regression. *International Statistical Review*, **81**(3), 361–387.

Vigneshwaran, S, Mahanand, BS, Suresh, S, & Sundararajan, N. 2015. Using regional homogeneity from functional MRI for diagnosis of ASD among males. *Pages 1–8 of: Neural Networks (IJCNN), 2015 International Joint Conference on*. IEEE.

Zhu, Ji, Zou, Hui, Rosset, Saharon, & Hastie, Trevor. 2009. Multi-class adaboost. *Statistics and its Interface*, **2**(3), 349–360.

Zijdenbos, Alex P, Forghani, Reza, & Evans, Alan C. 2002. Automatic" pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *Medical Imaging, IEEE Transactions on*, **21**(10), 1280–1291.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
24–03–2016	Master's Thesis	SEP 2014 — MAR 2016

4. TITLE AND SUBTITLE

Diagnosing Autism Spectrum Disorder through Brain Functional Magnetic Resonance Imaging

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Palko, Kyle A., Second Lieutenant, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENC-MS-16-M-123

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A.
Approved for Public Release; distribution unlimited.

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

Autism spectrum disorder (ASD) is a neurodevelopmental condition that can be debilitating to social functioning. Previous functional Magnetic Resonance Imaging (fMRI) classification studies have included only small subject sample sizes ($n < 50$) and have seen high classification accuracy. The recent release of the Autism Brain Imaging Data Exchange (ABIDE) provides fMRI data for over 1,100 subjects. In our research, we derive a subject's functional network connectivity (FNC) from their fMRI data and develop a regularized logistic classifier to determine whether a subject has autism. We obtained up to 65% classification accuracy, similar to other studies using the ABIDE dataset, suggesting that generalizing a classifier over a large number of subjects is much more difficult than smaller studies. The connectivity among several brain regions of ASD subjects were highlighted in the model as abnormal compared to the control subjects which potentially warrants future investigations about how these regions affect ASD. Although the classification accuracy was lower than what could be considered as clinically applicable, this research contributes to the continuing development of an automated classifier for diagnosing autism.

15. SUBJECT TERMS

autism spectrum disorder, functional network connectivity, logistic regression, regularization

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lt Col RD Kappedal, AFIT/ENC
U	U	U	U	102	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4539 ryan.kappedal@afit.edu