Title
Functional Analysis of Magnetic Resonance Spectroscopy Signals: Applications to Major Depression

Author
Patel, Karina Shanti

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Functional Analysis of Magnetic Resonance Spectroscopy Signals: Applications to Major Depression

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Bioinformatics

by

Karina Shanti Patel

2018

ABSTRACT OF THE THESIS

Functional Analysis of

Magnetic Resonance Spectroscopy Signals:

Applications to Major Depression

by

Karina Shanti Patel

Master of Science in Bioinformatics

University of California, Los Angeles, 2018

Professor Shantanu Hemachandra Joshi, Co-Chair

Professor Van Maurice Savage, Co-Chair

Magnetic Resonance Spectroscopy (MRS) is an in-vivo, non-invasive technique to measure biochemical metabolite concentrations in the brain. Here, we propose novel tools for processing and analysis of MRS data. We apply these tools to classify disease status in major depressive disorder (MDD). Our tools enable the manipulation and formatting of MRS data, and allow for the application of several machine-learning classifiers used to predict disease status. We test and derive several representations of MRS data including the conventional metabolite concentration levels, full MRS spectral function data points, elastic functional distances from the mean, and principal component analysis (PCA) components of the MRS spectra. We apply a number of

different machine learning algorithms including logistic regression, random forest classification, and support vector machines (SVMs). The highest performance of all these classifiers resulted from the use of random forest classification on full spectral data points with a precision of 0.698 and a recall of 0.695. This predictive approach to disease status classification can be used to allow possible depression cases to be caught early on as well as for confirmation of diagnosis.

The thesis of Karina Shanti Patel is approved.

Matteo Pellegrini

Katherine L Narr

Shantanu Hemachandra Joshi, Committee Co-Chair

Van Maurice Savage, Committee Co-Chair

University of California, Los Angeles

2018

To my supportive family…

thank you for always

encouraging me to pursue my passions,

both in and out of academia

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge all of the individuals in my life who have made it possible to complete my thesis. I am extremely grateful for my committee members Professor Shantanu Joshi, Professor Van Savage, and Professor Matteo Pellegrini. These individuals have helped shape my academic experience and have provided invaluable guidance throughout my research. I also would like to thank Professor Katherine Narr and those working in the UCLA Brain Mapping Center for providing me the opportunity to work with data that I would not have had access to otherwise. I would also like to acknowledge Stephanie Njau for preprocessing this data and for the valuable insights on metabolite selection. The collection of data used in this study was made possible thanks to the National Institute of Mental Health (grant numbers R01MH092301 and K24MH102743). I am additionally appreciative for all my professors, teaching assistants, fellow friends, and classmates who have helped me grow as a student and individual. I learned so much in and out of the classroom throughout my time at UCLA, all of which I will take with me in my future endeavors.

CHAPTER 1

Introduction

Major depressive disorder (MDD) is a prevalent mental disorder impacting the lives of over 350 million individuals around the world.[1] MDD is the fourth leading cause of disability worldwide and takes over one million lives due to suicide each year, yet due to the number of different symptoms that are associated with depression, it is often difficult to diagnose.[1,2,3] Factors that contribute to the development of MDD include genetics, hormones, brain chemistry, and life experiences, which helps explain why diagnosis indicators significantly differ depending on the individual studied.[3] Initial diagnosis of Major Depressive Disorder involves a distinct mood change often associated with inability to sleep, withdrawal from society, appetite loss, and/or suicidal thoughts.[4,5]

Following diagnosis, standard antidepressant medication and psychotherapy are often administered. For those patients resistant to standard approaches after two or more treatment trials, electroconvulsive therapy (ECT) is a highly effective treatment option. While the exact biological mechanism of the therapy is largely unknown, there are several symptomatic manifestations including change in transcriptional factors, metabolite concentrations, and gene

expression throughout ECT.[6] With an average remission rate of 87%, electroconvulsive therapy is the most effective and widely studied treatment for resistant patients.[7,8] ECT involves the use of anesthesia and muscle relaxants while simultaneously sending small electric currents to the brain through strategically placed probes.[9,10] The electric current induces a small seizure which alters the chemistry of the brain and can be observed through the acquisition of Magnetic Resonance Spectroscopy (MRS) signals.[8]

Magnetic Resonance Spectroscopy provides a non-invasive diagnostic tool to track biochemical changes in the brain.[11] Proton MRS (^{1}H-MRS) utilizes signals from hydrogen protons to determine the concentrations of a number of different brain metabolites including N-acetyl aspartate (NAA), Myo-inositol (Ins), Creatine (Cr), and glutamate (Glu).[11,12] In comparison to Magnetic Resonance Imaging (MRI), which utilizes the protons in water molecules to generate the major signals, ^{1}H-MRS captures excitations of protons attached to molecules other than water.[2] ^{1}H-MRS excites the small and highest concentrated molecules, which explains why we are only able to measure key metabolites such as NAA, Ins, Cr, and Glu.[13] This common spectroscopic strategy involves acquisition of a signal from a single region of the brain through single voxel spectroscopy. In this methodology, a small volume (30 x 18 x 12 mm) of the brain known as an aggregate voxel is specified prior to MRS acquisition and results in the output of a Free Induction Decay (FID) signal. There are a number of different voxels each associated with different regions of the brain. The left hippocampus, right hippocampus, dorsal cingulate, and subgenual cingulate regions are the four voxels that we were interested in studying due to their crucial roles in emotion and mood. For the purpose of this study, we focused solely on the left

hippocampus voxel.[2] The FID signal is analyzed either in the time domain or the spectral domain using a wide variety of techniques including TARQUIN, jMRUI, and VeSPA.[14,15,16,17]

TARQUIN analyzes spectroscopic data in the time-domain using a non-linear fitting model. It utilizes a basis set constructed from metabolite signals, macromolecules, and lipids while calculating amplitudes with a non-negative least squares projection.[14] Another software package often used is jMRUI, a Java graphical user interface software that also allows time-domain analysis of MRS signals.[15] This package performs removal of residual water, signal-to-noise enhancement, and uses a variety of fitting algorithms to quantify metabolites including the Hankel Lanczos Squares Singular Values Decomposition (HLSVD), Hankel Lanczos Total Least Squares (HLTLS), and the Linear Prediction Singular Value Decomposition (LPSVD). VeSPA is a graphical user interface offering a number of tools for MRS data analysis.[16,17] VeSPA is capable of handling high dimensional data and optimizing data acquisition sequences for spectral analysis. Additionally, VeSPA allows for the exploration of MRS data acquisition and improve MRS data processing Siemens scanners.

Most methods initialize the signal by setting a reference metabolite concentration for unsuppressed water spectrum. Line shaping is performed assuming a Gaussian or a Lorentzian profile to account for spectral broadening and shift. Singular value decomposition (SVD) is then used for either filtering water signals or to even quantify metabolites. The SVD step is often followed by a least-squares or a non-linear least squares fitting procedure that may or may not be followed by corrections of misalignment in the spectrum.[18] The last step is metabolite

quantitation, which again uses several SVD based techniques such as HLSVD, HLTLS, or non-linear least squares algorithm such as AMARES.[19]

After performing metabolite quantitation, researchers typically analyze either absolute metabolite concentrations or ratios of concentrations (in reference to NAA or in reference to Cr + PCr) for a population. This thesis addresses the analysis of absolute metabolite concentrations while also proposing a new approach for analyzing the full spectral function of the MRS signal by analyzing its shape. We provide evidence for differences seen in the entire signal function as well as changes in individual peaks when comparing depressed patients to healthy controls. Our approach also allows us to calculate elastic distances between functions and perform principal component analysis for reducing the infinite dimensionality of the spectral functions. Finally, we use the individual metabolites, full spectral functions, reduced dimensional functional features from Principal Component Analysis (PCA), and elastic MRS spectral distances to draw distinction between individuals with depression and healthy controls. We introduce an R package `mrspecr` that encapsulates our MRS function analysis methods in a reproducible manner.

For visualization purposes, the MRS signal is shown as a spectrum of resonance peaks corresponding to metabolite concentrations in parts per million (ppm). While there does exist age-related and regional variations in the metabolite concentrations, these spikes are usually seen in early years of life and mainly reflect myelination patterns.[11] Disregarding possible spikes from myelination, we are able to match peaks to different core metabolite concentrations. Through the analysis of these metabolic peaks, we are able to identify changes to metabolite concentrations that help differentiate healthy controls and depressed individuals. We utilize machine-learning

prediction to assist with diagnosis confirmation by identifying patients or controls based on metabolite concentrations or full signals of each subject. Identifying how possible neurometabolite levels associated with ECT treatment response could help with diagnosis and tracking of treatment progress, which are highly desirable as we push towards more effective and individualized treatment.

This thesis generates novel ideas for representation and analysis of MRS signals with the following contributions:

1. We proposed an idea of matching the shapes of MRS spectral functions rather than using metabolite peaks. This is a new idea that has not been studied to our knowledge.
2. We introduced the notion of an elastic MRS spectral distance as a univariate feature that allows for shape analysis and classification based solely on the shape of the signal; the concept of utilizing MRS function shape is a novel concept that has not been applied to the field of MRS analysis. Further, using dimensionality reduction of full spectral functions to extract principal component analysis features allowed us to find key indicators for classification.
3. Finally, we developed and implemented an R package `mrspecr` to allow other researchers to perform similar MRS analysis and preprocessing quickly and effectively.

This thesis is organized as follows. We begin by discussing participant recruitment and the process of ECT administration to the patient population that was studied in this research. Then we introduce our novel ideas of MRS signal representation and matching. Next, we focus on data preprocessing, formatting through the creation of the R package `mrspecr`, a package created

for MRS analysis. We elaborate on results from utilizing several different data sets in machine learning classification algorithms through R. Finally, we discuss how these results compare to each other and outline other avenues of future research based on this thesis.

CHAPTER 2

Methods

2.1 Participant Recruitment and ECT Administration

Patients for this study were recruited from the University of California, Los Angeles (UCLA) Resnick Neuropsychiatric Hospital. Individuals set to receive ECT due to depression had been diagnosed by a board of certified psychiatrists (R.E.) through the DSM-IV-Revised criteria as well as confirmation from the Mini-International Neuropsychiatric Interview (M.I.N.I.).[20,21] In order to qualify for this study, patients must have completed two or more standard medication trials without response. Of the patients planning on receiving ECT, we excluded those individuals with diagnosis of post-traumatic stress disorder, bipolar mania, schizophrenia, attention deficit/hyperactivity disorder, and/or any primary anxiety disorders. Additionally, patients qualifying for the study must have already experienced at least two major depressive episodes. The patients who fit the specifications of this study were then gradually taken off their antidepressant medication over a span of two to three days prior to ECT treatment.

Of the 107 adult subjects who participated in the study, 46 of the individuals (24 men) were patients suffering from MDD based on the DSM-IV diagnosis. The ages of the individuals ranged from 20 to 61 years with a mean ± standard deviation (SD) of 39.9 ± 12.1 years. We additionally had 61 healthy individuals (23 men) ranging from 20 to 63 years of age. The control group was recruited using local advertising and the mean ± standard deviation (SD) of the group was 37.7 ± 10.5 years.

Data was collected from patients at three time points, but only the data collected for patients and controls at baseline was utilized for analysis and classification. T1 was taken the day prior to the initial ECT treatment. The patient was then treated twice in a span of 48 hours before recording T2 (Figure 2.1). Following T2 data collection, the patient continued to receive ECT three times a week for a four-week period prior to collection of T3 data. The Hamilton Rating Scale for Depression (HAMD) test was used at each of the 3 points of data collection for all patients.[2,21] The controls were chosen based on similar age, sex, and geographic similarity to our patients. Additional screens were performed using M.I.N.I. to remove those who had past experiences with depression, antidepressants, or medical illness.[22] We scanned these controls for MRS data collection 2 times: the first time at baseline (C1) and an additional scan four weeks later for a follow-up (C2). UCLA Institutional Review Board approved this study protocol and all patients and controls provided written consent to participate (Figure 2.1).

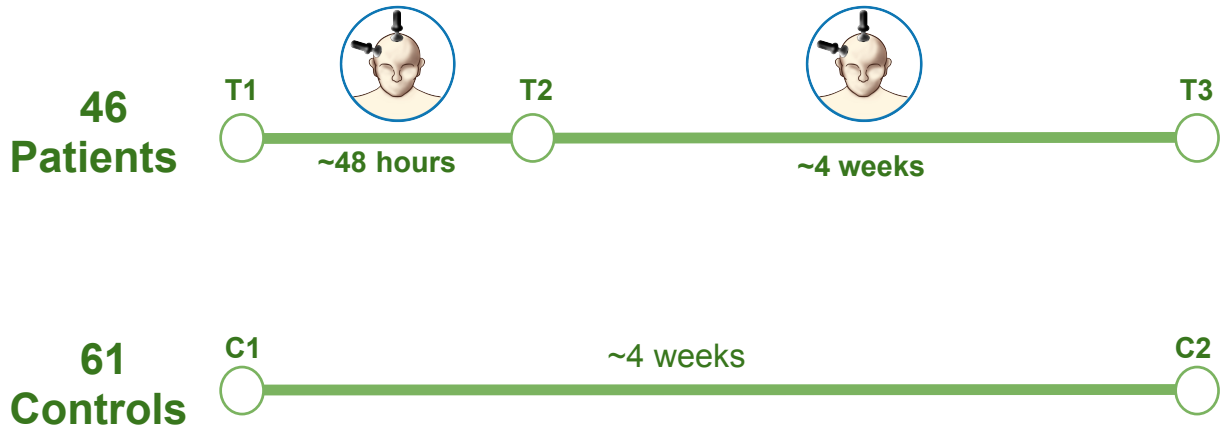


Figure 2.1: ECT Design Study at UCLA: The data collection involved 107 individuals, 46 of them being depressed patients undergoing ECT while the remaining 61 were healthy controls participating for normalization confirmation.

ECT was performed in combination with paralysis using 1mg/kg of succinylcholine and anesthesia using 1mg/kg of methohexital. Clinical response was recorded using the HAMD-17, allowing us to determine MRS marker association.

2.2 Data Acquisition

Using the Siemens 3 T Allegra system with and without water suppression (128/1 averages) we were able to acquire single-voxel point resolved spectroscopy sequences.[17] Repetition time was 2200 ms, echo time was 30 ms, and spectral width was 2000 Hz, 1024 samples. Using a volumetric navigator in real time, we were able to correct any motion or B0 noise. With the Siemens 3 T Allegra system, we were also able to acquire motion-correlated multi-echo MPRAGE images. For these images, the conditions used included echo times of 1.74, 3.6, 5.46, and 7.32 ms. The repetition and inversion time for the MPRAGE images were set to 2530 ms and 1260 ms, respectively. Additionally, the images were captured at 192 different sagittal slices

at a flip angle of 7°, a field of view of 256 x 256 mm, and a resolution of 1.3 x 1.0 x 1.0 mm^3. These images were then resliced to position the ^{1}H-MRS in the left and right hippocampus as well as the midsagittal dorsal anterior cingulate (dACC) and sub-genual anterior cingulate (sgACC).

After utilizing Siemens 3 T Allegra System for the spectroscopy, we next processed the data.[21] Upon receiving these signals, data preprocessing was necessary to improve signal-to-noise ratio and eliminate overall interference from patient movement throughout the test. For each MRS signal, a Gauss-Seidel iterative update scheme was utilized for denoising.[23] Denoising allowed for cleaner and overall smoother signals to remove as much noise as possible prior to inputting the data into our machine-learning framework.

After denoising was complete, LCModel software was utilized to compute water-referenced metabolite concentrations.[24,25] LCModel additionally performs fitting procedures, which involve a singular value decomposition followed by metabolite fitting using regression.[24] We can visualize the output of LCModel in a plot across different chemical shifts (Figure 2.2). To minimize noise in the data, we select the subjects whose metabolites show a noise standard deviation of less than 20%.

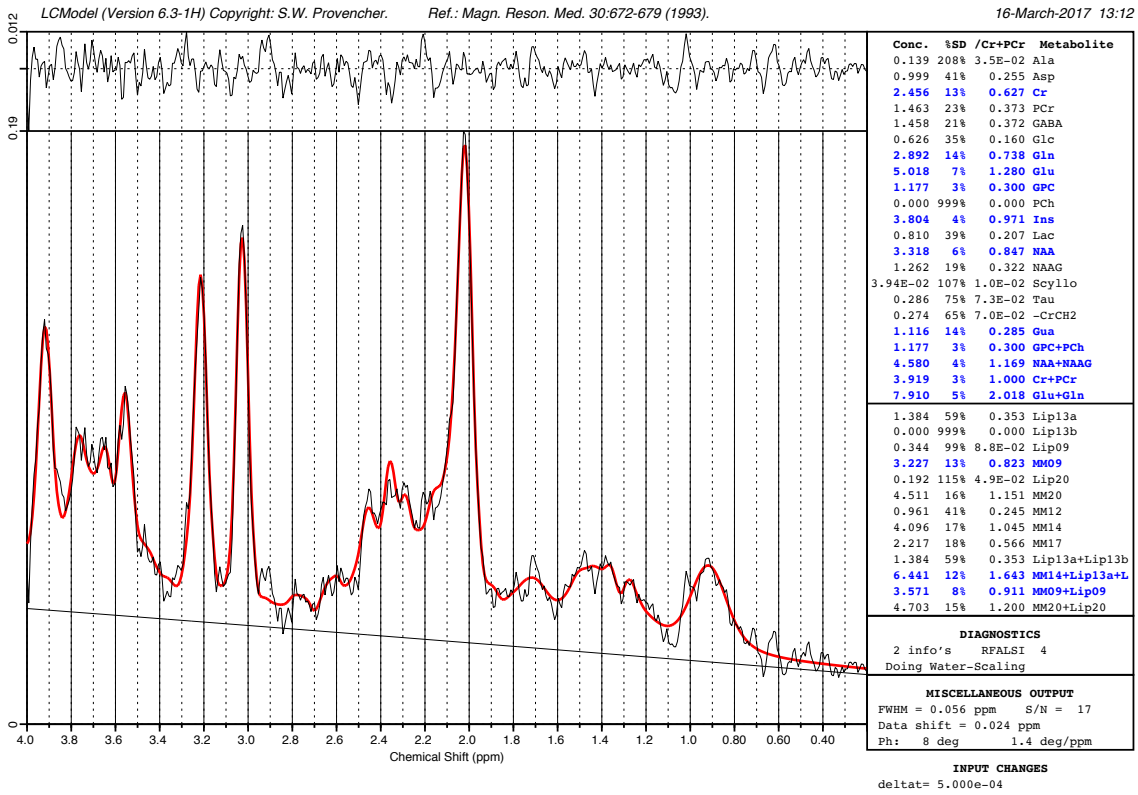


Figure 2.2: Example of an output from LCModel. The left-hand side shows the LCModel generated spectrum for the dorsal cingulate voxel. The right-hand side lists the metabolite concentrations (measured in ppm) and their corresponding standard deviations.

2.3 Functional Shape Analysis of MRS Signals

In this section, we outline our new approach for analyzing the full spectrum using a functional approach. We used a previously proposed functional shape representation approach and adapt it for representing MRS signals here. [26,27,28,29] Given that many magnetic resonance analyses have low detection of parameters below our threshold of significance, we perform a shape based alignment of these signals through elastic functional registration.[26] The elastic time series approach allows us to construct a template of the shape for optimal temporal alignment. Utilizing square root velocity field (SRVF) parameterization, we are able to define a time course signal

$f : I \equiv [0,1] \rightarrow \mathbb{R}$ by a square-root velocity field map q given by $q : [0,1] \rightarrow q : [0,1] \longrightarrow \mathbb{R}, q(t) = \frac{\dot{f}(t)}{\sqrt{|\dot{f}(t)|}}$. In this case, $\dot{f}(t) = \frac{df}{dt}$ corresponds to the time course velocity while $|\dot{f}(t)|$ corresponds to the magnitude.[27,28,29] Given the absolute continuity of f, the SRVF transformation allows the mapping q to be square integrable with the set of SRVFs given by the Hilbert space $\mathbb{L}^2([0,1], \mathbb{R})$. We are then able to recover the time domain fMRI signal using $f(t) = f(0) + \int_0^t q(\tau)|q(\tau)|d\tau$. At time t = 0 we presume $f(0) = 0$ and the SRVF mapping is invertible up to the given $f(0)$. To get the unit length constraint on this q function we can take $\frac{q}{||q||}$ and which ensures q will lie on a Hilbert sphere $\mathbb{S}_f$, where $\mathbb{S}_f \equiv \{q \in \mathbb{L}^2 \left| \int_0^1 [\![q(s), q(s))_{\mathbb{R}^2} ds]\!] = 1, q(s): [0,1] \rightarrow \mathbb{R}^2\}$.[26]

Time reparameterization allows us to better align the response shapes with the conditions by shifting the signals to match peaks and troughs more closely. Representing the time reparameterization function with diffeomorphic function $\gamma : I \rightarrow I$ such that $\dot{\gamma} > 0 \; \forall t$, we can alter the time reparameterization using a composite $f \circ \gamma$. In the case of the SRVF space, we use $q \cdot \gamma = \frac{(\dot{f} \circ \gamma)\dot{\gamma}}{\sqrt{|(\dot{f} \circ \gamma)\dot{\gamma}|}} = (q \circ \gamma)\sqrt{\dot{\gamma}}$. We next compute an unbiased optimal template of the functional response, ultimately using the geometry of the time course representation space through the space of the q functions. More specifically, we assume elastic form of the tangent space Riemannian metric $T_q(\mathbb{S}_f)$ as $\langle q_1, q_2 \rangle = \int_0^1 q_1 q_2 dt$. The closed form solution for geodesics between q_1 and q_2 has a geometry described by $\chi_t(q_1; v) = \cos(t\cos^{-1}\langle q_1, q_2 \rangle)\, q_1 + \sin(t\cos^{-1}\langle q_1, q_2 \rangle)\, v$. In this case, t $\in$ [0,1] and v $\in T_{q1}(\mathcal{Q})$ is described using $v = q2 - \langle q_1, q_2 \rangle q_1$. The geodesic distance between these 2 shapes q_1 and q_2 on $\mathbb{S}_f$ is calculated as

$d(q_1, q_2) = \int_0^1 \sqrt{\langle \dot{\chi}_t, \dot{\chi}_t \rangle}\, dt$ which we are then able to use in calculating the optimal temporal alignment as ($\hat{\gamma} = argmin_{\gamma \in \Gamma} d(q_1, q_2 \cdot \gamma)$ Using gradient descent and dynamic programming, we are able to achieve the temporal alignment for two time course signals. Calculating the template for unbiased within-condition time series, we can align all time courses to the mean to achieve temporal alignment.

CHAPTER 3

An R Package for MRS Preprocessing: `mrspecr`

3.1 Overview

Here we introduce a new R package `mrspecr` that we developed specifically for MRS analysis and signal processing. The R package contains a number of data input-output and statistical tools to help extract necessary data from the output of an MRS test after being sent through LCModel. The processed FID signal through LCModel produces several output files containing information on metabolite concentrations, noise standard deviations, full FID spectra, and other diagnostic measures generated during the LCModel analysis.

Our package includes functions to read MRS Digital Imaging and Communications in Medicine (DICOM) acquisitions, denoise the FID signals, extract coordinates and metabolites, and align the MRS signals (Figure 3.1). Each of these features ultimately automates the preprocessing of batch subject folders by parsing the text files for the data that we need for feature matrices that are inputted into a machine learning classifier.

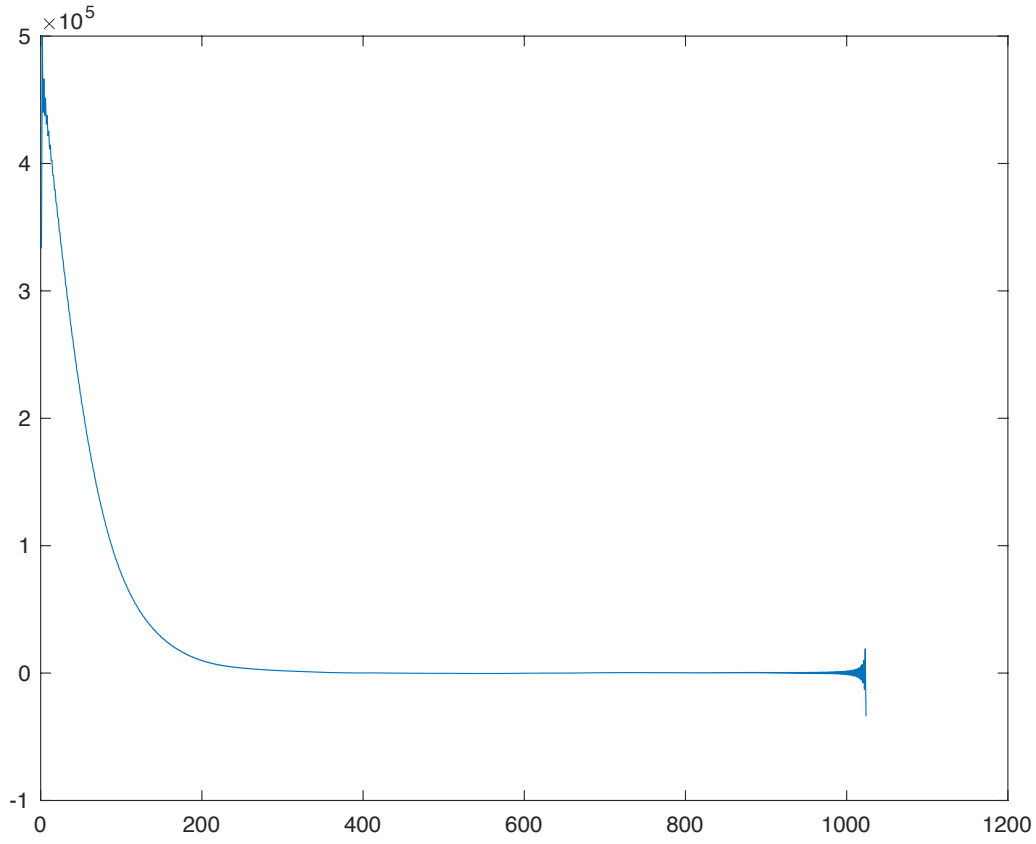


Figure 3.1: The function read_siemens_dicom takes in a DICOM image and reads it to produce the FID signal envelope of time versus intensity as shown. This FID signal in the time domain is then transformed to the MRS spectra in the frequency domain using the LCModel software.

3.2 Extracting Coordinates and Metabolites

The file output by LCModel contains processed data to be extracted. Since these files mainly are stored in an unstructured format, we implemented a text parser to extract the desired signal and metabolite information. The coordinate text file contains information regarding each of the x and y axes data coordinates that are needed to reconstruct the signal from LCModel. For all subjects, the x coordinates of this axis will be a constant set of 480 data points representing chemical shift

in ppm. For each of these ppm values 1 to 480, we have corresponding amplitude or peak. These 480 y values for our group of patients and controls are the values that we are interested in pulling into a matrix as these values can be used to differentiate patients versus controls. We parse through these text files for not only the coordinates of the function but the metabolite peaks as well.

Metabolite concentrations correspond to the peaks in the plot from LCModel. These values can be found in a comma separated value (csv) file that must be parsed for the concentration levels along with their standard deviations. Looking at these metabolite concentration levels is what we have seen in previous studies as a measure to possibly differentiate between patients and controls, which is why these values are of interest.[30,31] In order to extract these concentration levels, we must parse for each metabolite name and pull the corresponding concentrations into a data frame in R. The coordinate y values and metabolite concentrations are the foundation for our analysis as they provide the critical data for classifying patients versus controls.

3.3 Batch Alignment

Once we have extracted the desired coordinate data using the coordinate function in `mrspecr`, certain preprocessing steps must be performed to eliminate interference and noise from the MRS imaging tests. During MRS acquisition, patient movement and other artifacts may cause signal shifts and peaks to vary in ppm. Because we are aware of the metabolite peaks occurring at the same time points (ppm), we can utilize functional alignment to structure and straighten our data based off a mean function of all the signals.

The align function is built into the R package `fdasrvf`, a package built to perform alignment and PCA for single and multidimensional functions.[32] This package implements geometric methods for shape matching of curves and functions.[28,29] These capabilities utilize the square-root velocity framework and provide elastic analysis opportunity.[26] The functional data analysis is based off of separation between phase and amplitude of the signal to preserve the structure of the observed data.[33] We utilize the commonality of phase to a shift the signal and better align the amplitude for analysis.

Extracting and accounting for the amplitude variability in batch subject cases involves utilization of the coordinate matrix to vertically and horizontally shift data to more closely follow the mean (Figure 3.2). The batch alignment function returns a list including the original functions for all subjects along with the mean, the aligned functions, and a number of additional features associated with PCA. The batch alignment function also contains plotting capabilities, which allow a comparison between the original functions and the aligned functions. The aligned functions are more closely correlated to the actual peaks where metabolite concentrations are known to occur, removing variation where possible (Figure 3.2). We are able to see how the mean computed for the aligned functions differs from the standard Euclidean mean when plotting the two along the same time axis (Figure 3.3).

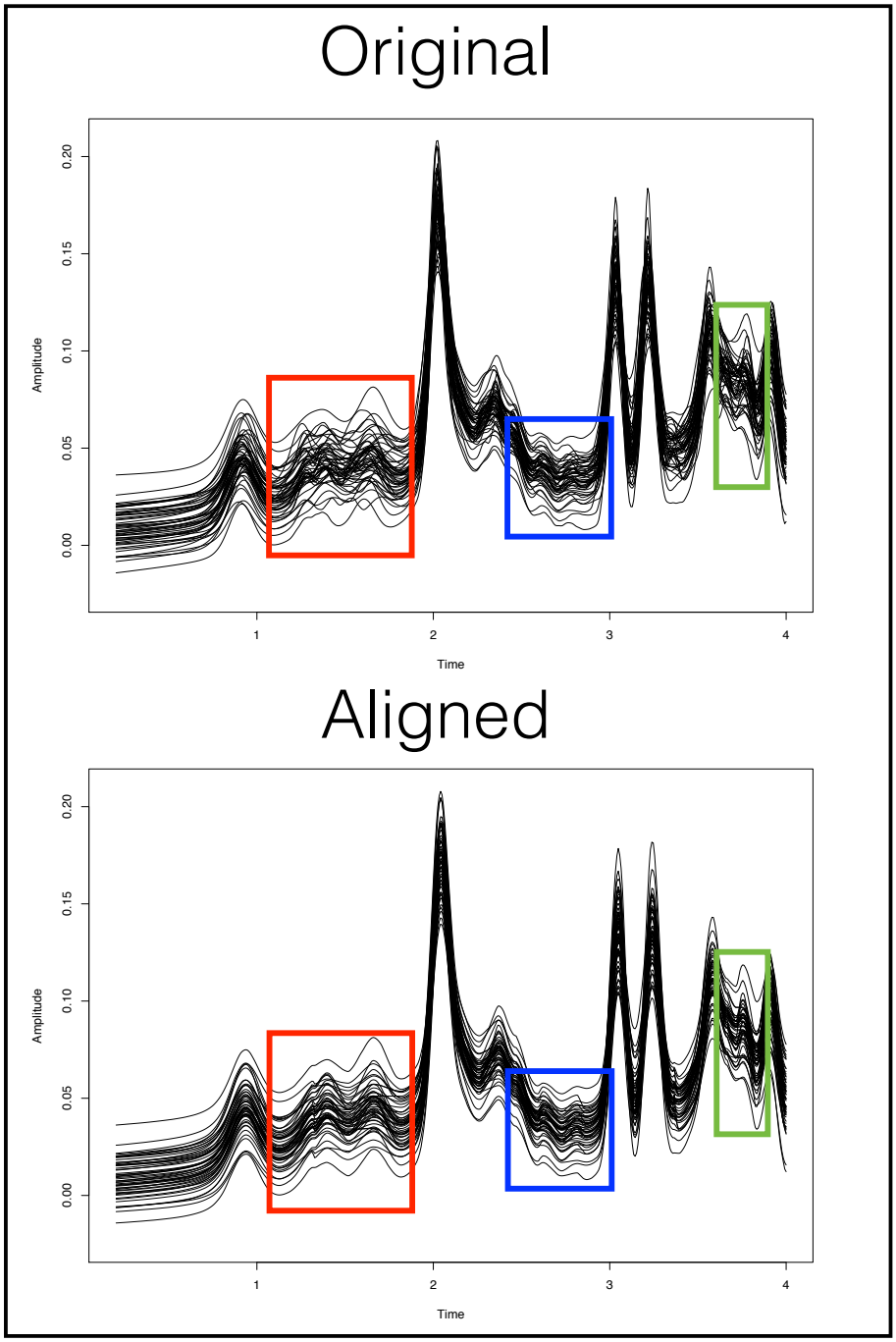


Figure 3.2: Plotting the original functions (top) and the aligned functions (bottom) for all 59 subjects, we are able to visualize 3 main areas of the curve in which alignments is helping smooth and eliminate noise. Each of the three areas of interest is boxed in a unique color in both the original and aligned plots.

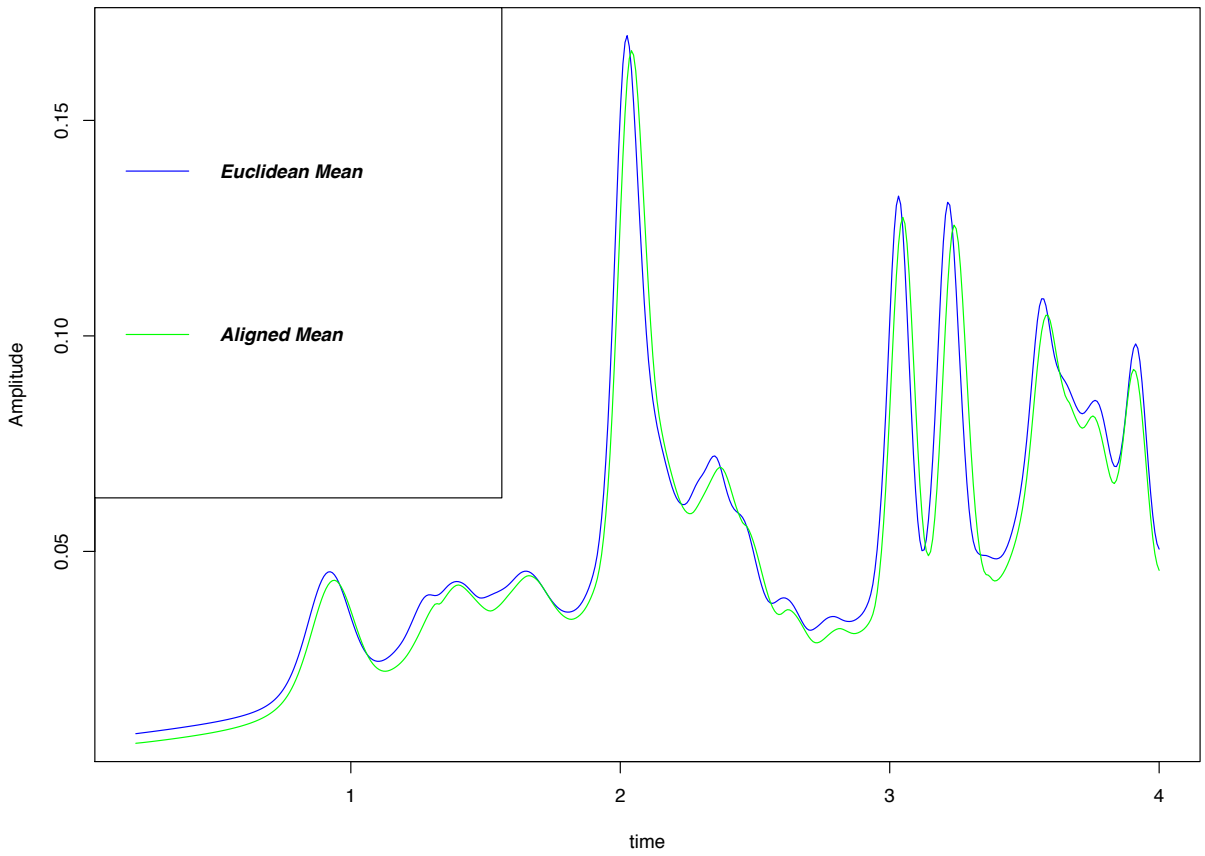


Figure 3.3: Comparing the plot of the mean for all original functions (blue) to the mean of all aligned functions for the 59 subjects (green) we are able to see the shift in the aligned mean. There are a number of reasons for alignment but two of the most important reasons are noise and peak alignment.

3.4 Data Matrices of MRS Features

After extracting the relevant functional and metabolite information from output datasets from LCModel, we are now ready to formulate this information into feature matrices for classification. There are multiple different data sets that we are looking at to use to independently classify patients versus controls. The input data that must go into the classification matrix function for coordinates is the classification vector indicating patient or control, the subject ID, and the

matrix of all coordinates for the given set of subjects. In the case of the metabolite matrix function, we must input the classification vector and the subject ID again, but instead of the coordinates we feed in the metabolite matrix. Additional matrices that are formed include the distance matrix and the PCA matrix. Each of these two matrices has the outputted feature of distance or PCA components along with the classification vector. In any case, the classification of each subject is binary, 1 being a depressed patient and 0 corresponding to a healthy control.

CHAPTER 4

Results

4.1 Feature Selection

The voxel of interest was narrowed down to the Left Hippocampus due to its significant volume differences between MDD patients and healthy controls.[22] For each voxel, LCModel estimated 35 metabolite concentrations. Using our previous findings, we selected 4 metabolites that exhibited most significant patterns of change throughout ECT. These metabolites include N-acetyl aspartate (NAA), Creatine (Cre), Myo-inositol (Ins), and glutamate (Glu).[22]

4.2 Subject Selection

Using the subset of data from the feature selection, we next cleaned the data to ensure that the input data into our machine learning classifier was accurate and contained minimal outlier subjects. In order to do so, the standard deviations associated with each metabolite were compared per subject to an upper bound of 20%. We removed 48 subjects through this processing and were able to narrow the data set to 59 total subjects. Of these 59 subjects, 27 were identified as patients with depression and the remaining 32 subjects were healthy controls. Given

that this data was selected based on the metabolite concentrations, we needed to pull the selected data for these subjects for each of the classification matrices. Using these restrictions, we re-selected the coordinates according to the 59 subjects for the full function matrix. We additionally used these coordinates to calculate distances between patients and controls to the mean and for PCA. Given the filtering of the data, the demographics of the new data set may contain slight differences from the original 107 subjects.

4.3 Classification Overview

After the pre-processing, denoising, and filtering of these signals, we then moved to the prediction of diagnosis classification as patients or controls based on each of these feature matrices using a number of different machine learning classifiers. We used a combination of scikit-learn in Python and Weka to verify the machine learning model results we built in R.[34,35] From the LCModel output data, we generated four different derived data sets to train a model that aimed to predict whether the individual is an MDD patient (depressed) or a control (healthy). A number of different classifiers were tested on each of these data sets including random forest, logistic regression, and support vector machines (SVMs) based on precision, recall, and accuracy metrics. For each of these classifiers, we were solely testing its ability to predict the diagnosis of an individual through binary classification. The logistic regression and SVM models utilize the `caret` package in R while the random forest model utilizes the `randomForest` package in R.[36,37] To train and test the models, we used repeated 10-fold cross validation to evaluate the performance of these models. This involves splitting the dataset of 59 subjects into 10 sets. We choose 9 of these sets to train on and then test on the remaining 10th dataset, iterating through each of the random permutations of training and testing until all 10 sets were tested. Cross

validation provides a more robust performance evaluation metric than a single split validation. Due to the fairly balanced classes of diagnosis for depressed patients and healthy controls, accuracy was a good metric for performance. We ran these tests over 300 iterations, recording the accuracy and computing the mean after the 300 iterations to calculate the balanced accuracy. These results were reported in the form of a histogram and boxplot of the balanced accuracy among all iterations. In addition, we also collected the confusion matrix for the model and plotted threshold ROC curves for each data set showing true positives vs. false positives for patient classification at different threshold values. After running each of these classifiers using 10-fold cross validation, we were able compare model performance for each data set.

4.4 Classification using Metabolites Concentrations

Using the data extracted from LCModel and guided by previous studies, a straightforward idea for classification was to use the metabolite concentrations as features.[22] Using the metabolites NAA, Cr, Ins, and Glu as selected above as well as the 59 subjects, which were selected based on standard deviation limits, we tested a number of machine learning algorithms; each of these metabolites is a single feature in the classification matrix. The highest performing algorithm on this data was logistic regression with a precision of 0.599 and recall of 0.600 (Table 4.1). The Receiver Operating Characteristic (ROC) of the logistic regression performance was 0.583 (Figure 4.1). This metabolite classifier tells us that of the 27 depressed patients, 15 of them were correctly classified as depressed from our model. Additionally, of the 32 healthy controls, 21 of these were correctly classified as healthy.

The overall accuracy of this model was calculated through balanced accuracy (Figure 4.2A). We see that our model successfully predicts 15 of the actual 27 patients correctly given the patient recall of 0.555, but fails to predict 12/27 of the patients as depressed. The accuracy from our confusion matrix here is 0.600, which is similar to the average value of 0.601 we were able to compute in R using the balanced accuracy model (Figure 4.2B). For this data set, random forest classification performed the most poorly with a precision of 0.556 and recall of 0.555.

Table 4.1: Summary of the results from logistic regression for classification. Using the metabolite concentrations as features, we find that logistic regression has the strongest results out of all the different classifiers tested with a precision of 0.599 and a recall of 0.600.

	True Positive Rate	False Positive Rate	Precision	Recall	F-Measure	ROC Area
Patient	0.536	0.344	0.577	0.536	0.556	0.583
Control	0.656	0.464	0.618	0.656	0.636	0.583
Weighted Average	0.600	0.408	0.599	0.600	0.599	0.583

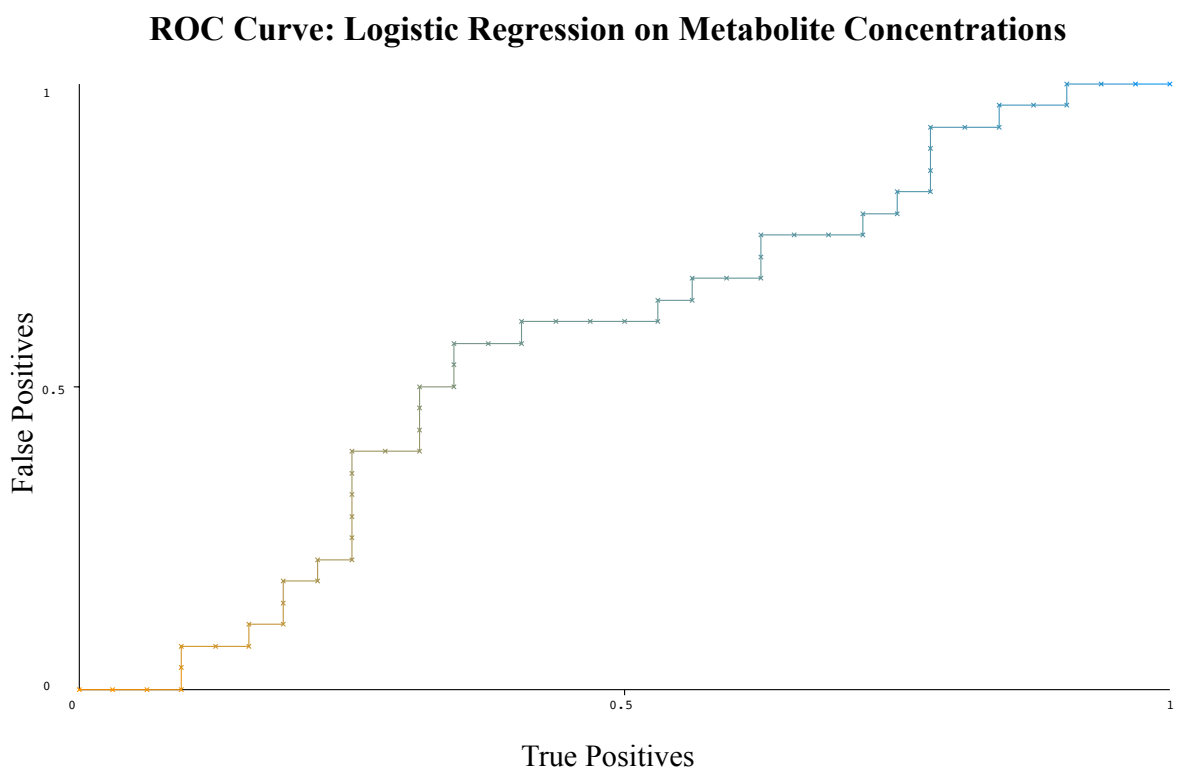
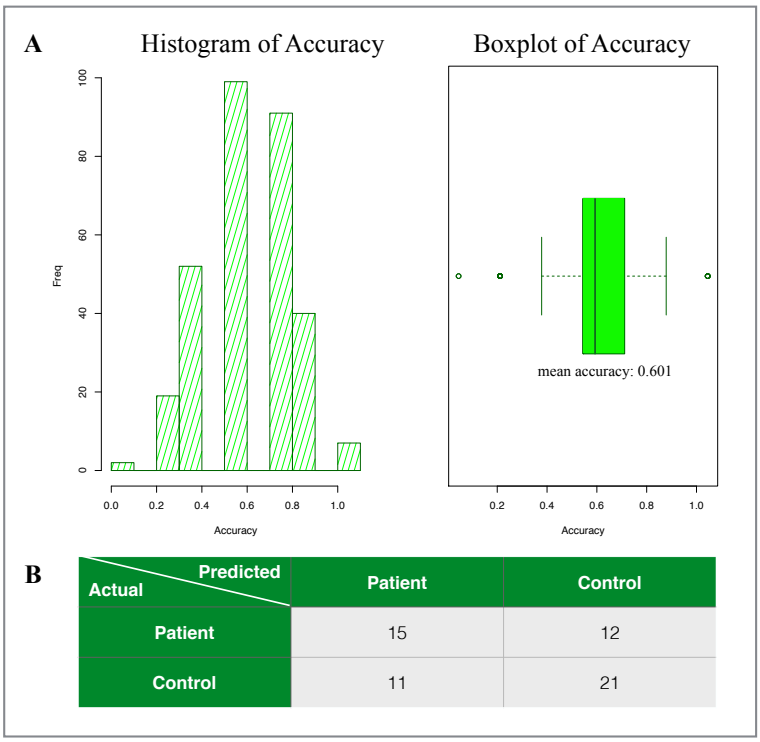


Figure 4.1: Logistic regression ROC Curve for metabolite concentrations plots true positives to false positives.



Actual \ Predicted	Patient	Control
Patient	15	12
Control	11	21

Figure 4.2: A: Histogram and boxplot for classification accuracy for logistic regression. Balanced accuracy was estimated to be 0.601. **B:** The logistic regression confusion matrix shows the predictions of the model compared to the actual classification of individuals.

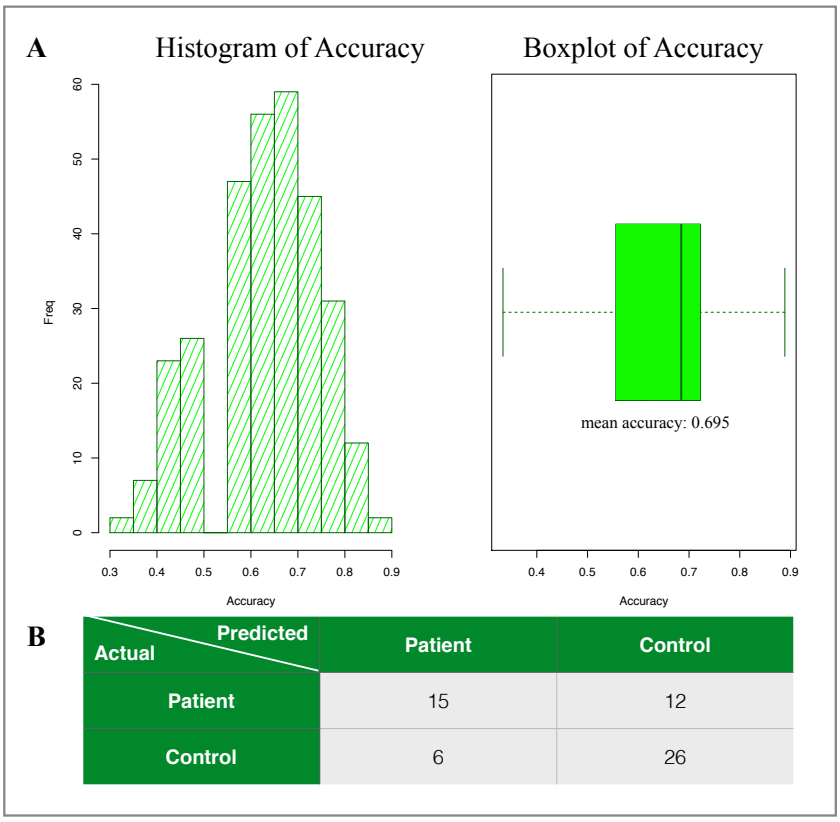
4.5 Classification using Full Spectral Functions

Next, we wanted to explore if the full spectral function, as a classification feature would improve classification accuracy. By default, LCModel discretizes the spectral function into 480 samples. Thus, we have 480 different discrete features representing a single data point or subject. The underlying thought process behind using every feature was that there could be significant areas of the function that vary greatly between patients and controls but may not correspond to one of the metabolite peaks used in the metabolite classification. We tested each of the classification algorithms with this data to compare performance and found different results than what we had

seen in the original case of metabolite peaks. This time, the highest performance was obtained using random forest classification; this model gave us a precision of 0.698 and a recall of 0.695 (Table 4.2). From these results we generated a confusion matrix and an ROC curve to visualize the performance of our classifier (Figure 4.3 and 4.4).

Table 4.2: Summary of the results from random forest classification. Using the full spectral functions as features, we find that the random forest model has the strongest results out of all the different classifiers tested with a precision of 0.698 and a recall of 0.695.

	True Positive Rate	False Positive Rate	Precision	Recall	F-Measure	ROC Area
Patient	0.556	0.188	0.714	0.556	0.625	0.730
Control	0.813	0.444	0.684	0.813	0.743	0.730
Weighted Average	0.695	0.327	0.698	0.695	0.689	0.730



Actual \ Predicted	Patient	Control
Patient	15	12
Control	6	26

Figure 4.3: A: Random forest classification shows a balanced accuracy of 0.695 in the histogram and boxplot of the accuracy. **B:** The random forest classification confusion matrix shows the predictions of the model compared to the actual classification of individuals with a high true positive and low false positive.

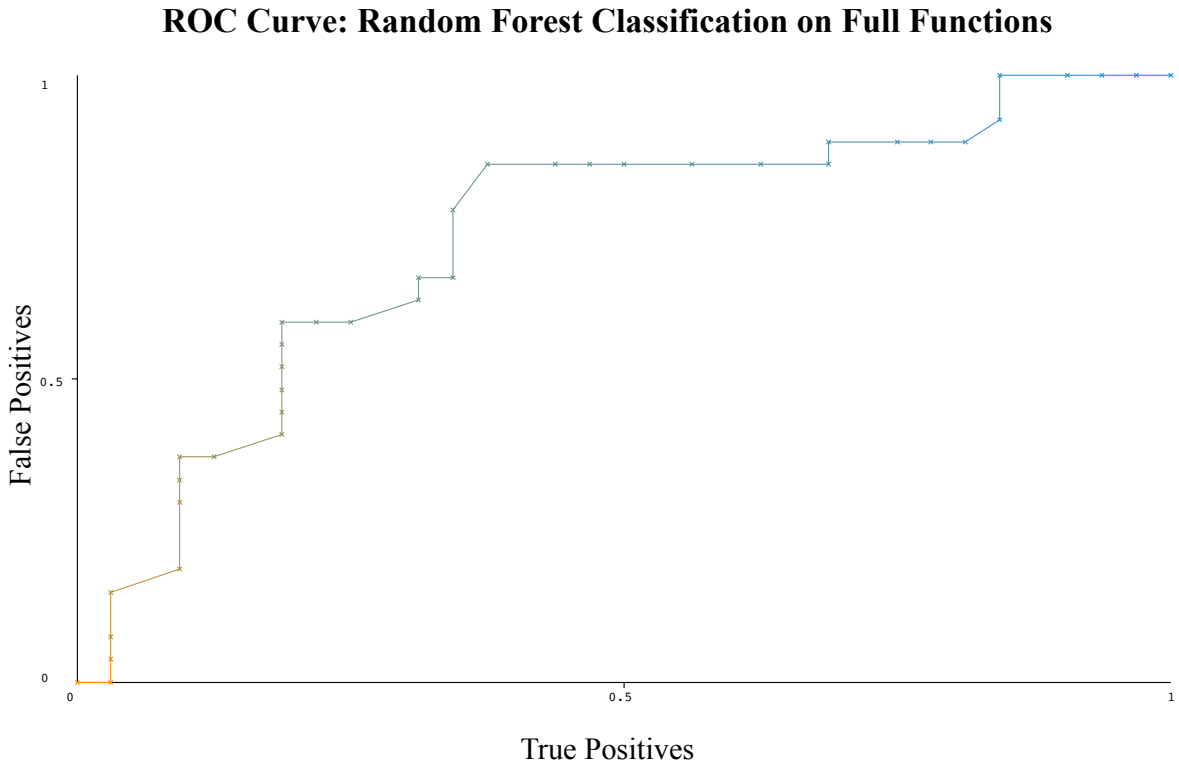


Figure 4.4: Random forest ROC Curve for full spectral functions showing the comparison of true positives to false positives at different threshold values along the x-y axis. A prediction model that performs as unbiased chance would be visualized as a linear plot of y=x meaning that the numbers closer to the top left corner of the graph denote better performing models.

We note that the function classification results originate from the same subjects (N=59) used in the metabolite classification experiment. The results table allows us to see the breakdown of how patients and controls are performing in our predictive model (Table 4.2). Looking at the controls, our model correctly predicts control subjects as healthy 81.3% of the time. Additionally, the ROC curve allows us to see the high performance of this model along different thresholds with an ROC value of 0.730 (Figure 4.4). Only 6 controls were incorrectly classified as patients, which is far lower than the 11 controls that were classified as patients in our previous model. The

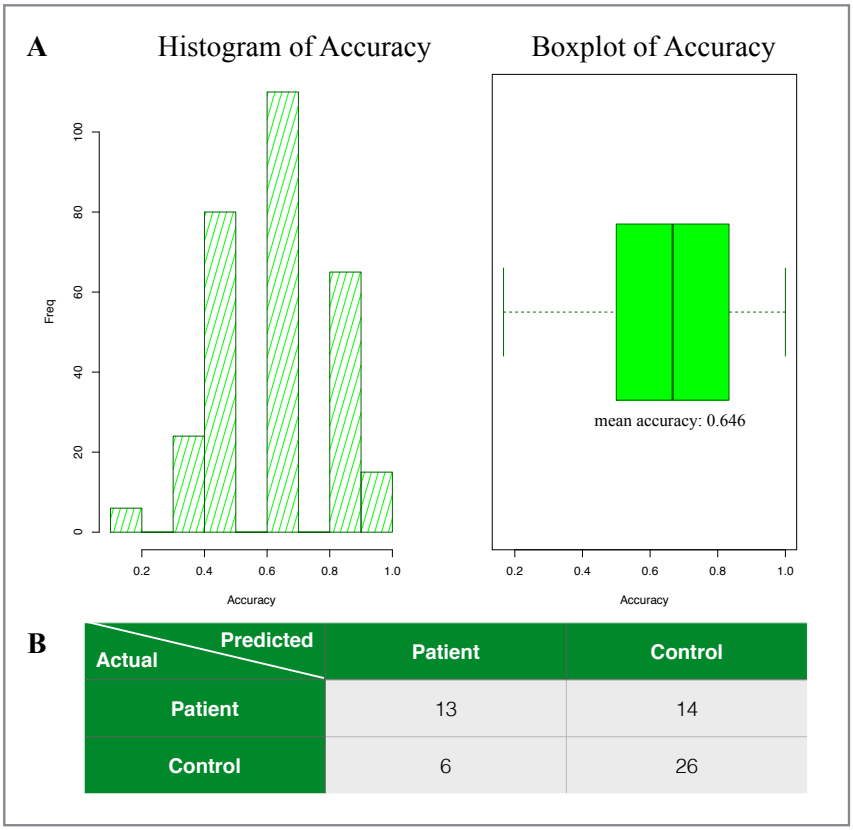
balanced accuracy of 0.695 reflects the results seen in the confusion matrix given that the balanced accuracy computes a mean over 300 iterations of this random forest model (Figure 4.3).

4.6 Elastic Shape Distances Between Spectral Functions

Following the full functional classification, we began to discuss the possibility of using the shape distance of the MRS spectral function with respect to a standard template as a feature in classification. The framework described in Section 2.3 allows us to compute an intrinsic (only depending upon the geometry of the shape space) mean shape of all the spectra using elastic matching.[26,27,28] This yields a single univariate (scalar) measure for each subject. We expect that the distances of the elastic shape to the population mean should be different when comparing patients and controls given that the metabolite concentrations are different in each of the two. Thus we hypothesize that if we were to compute elastic distances for each subject and compare that distance to the mean to get a baseline distance, these values could distinguish patients from controls. We used the R package `fdasrvf` to compute elastic distances between pairs of functions.[28] First we computed the mean function after aligning individual spectra from all patients and controls. This mean function served as the template for the whole population. We then computed elastic distances from each subject's spectral function to the template. These elastic distances were used as the feature for that subject entry in the data matrix. In doing so, our feature matrix shrunk significantly from the previous data used from containing 480 features to now just 1 single feature.

We now utilize the distance as a sole feature in classification and find that the classifier performance rankings are stronger than those seen in the metabolite classification case. The

highest performing model when using distance as the feature was logistic regression. Logistic regression performed with a precision of 0.666 and a recall of 0.661 while having an ROC of 0.669 (Figure 4.5b and 4.6).



B

Actual \ Predicted	Patient	Control
Patient	13	14
Control	6	26

Figure 4.5: A: Balanced accuracy for the elastic function distance feature is shown to be 0.646 along with the histogram and boxplot of the accuracy as shown. **B:** The logistic regression confusion matrix shows the predictions of the model compared to the actual classification of individuals with a high true positive and low false positive as we saw in the random forest of full functions. Our model correctly predicts 13/27 of patients correctly but fails to predict 14/27 of the patients as depressed.

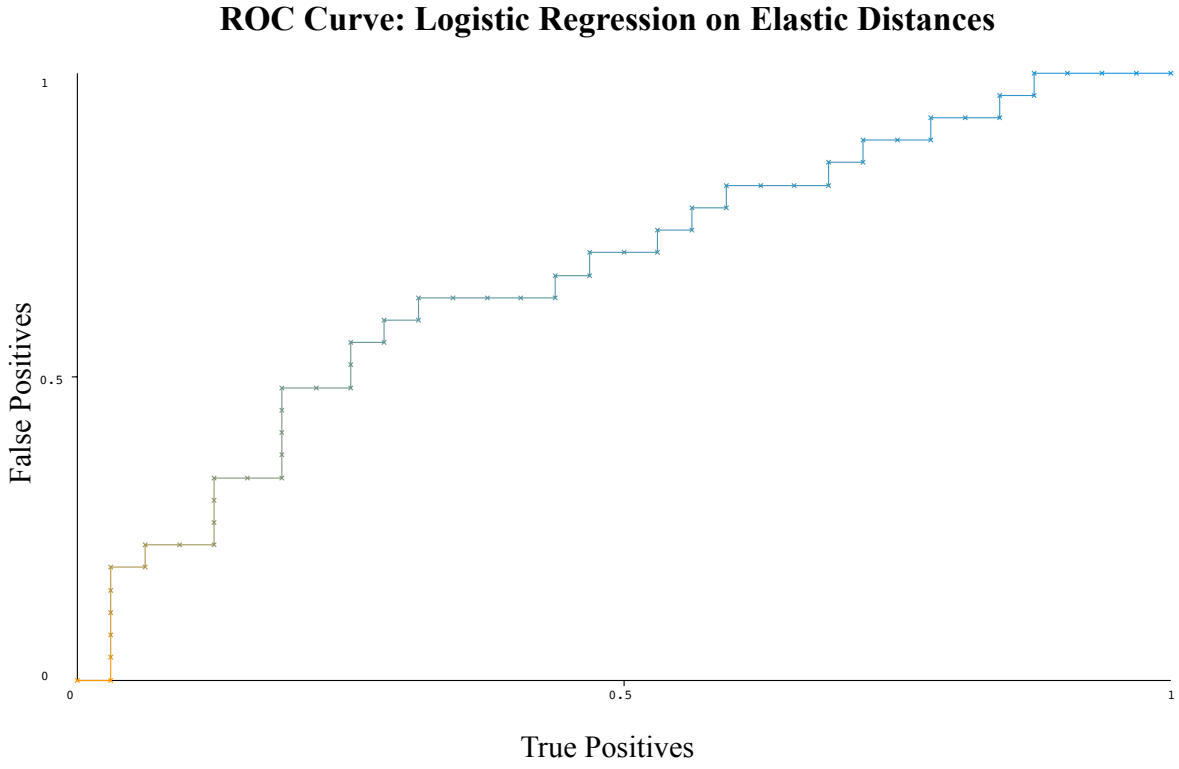


Figure 4.6: Logistic regression ROC Curve for elastic distances showing the comparison of true positives to false positives.

The performance of this model was unexpected given that we decreased the number of features from 480 to just one single feature but testing these results over 300 iterations allows us verify our results. The balanced mean accuracy for this logistic regression model falls at 0.646 (Figure 4.5a). We found the random forest model to be the lowest performer with a precision of 0.522 and a recall of 0.488 (Table 4.3). Comparing these results to the rest of the results from classification, we see that this was the lowest precision, recall, and accuracy set of results between all data sets that were inputs into the classification algorithms thus far. It is interesting to see how random forest performs strongly with full spectral functions (480 data points) while performing poorly with a single feature of elastic distance.

Table 4.3: Summary of the results from logistic regression for classification. Using the elastic distance as a univariate feature, we find that logistic regression has the strongest results out of all the different classifiers tested with a precision of 0.666 and a recall of 0.661.

	True Positive Rate	False Positive Rate	Precision	Recall	F-Measure	ROC Area
Patient	0.481	0.188	0.684	0.481	0.565	0.669
Control	0.813	0.519	0.650	0.813	0.722	0.669
Weighted Average	0.661	0.367	0.666	0.661	0.650	0.669

4.7 Principal Component Analysis

After utilizing metabolites, full functions, and distances, we compared the results. Looking at the results allowed for the exploration of new possibilities for manipulating the data according to where our models were performing best. The full spectral data matrix had the highest performance compared to metabolites and distances. Because the metabolites represent a very small subset of the full function matrix, we began to investigate if there were certain important features involved in the full function classification that were the key driving factors in performance. We utilized PCA with the full functions at different number of features to find an optimum set of features that explained the highest variance in the set.

Comparing the results for PCA with 3 features, 10 features, and 20 features, we found that 10 features had the highest performance of the three. It was interesting to see which models performed well in each of the different cases; we ultimately found that the highest performance came from PCA with 10 features using support vector machines (SVMs) at an ROC value of 0.643 (Figure 4.7). The precision of this model was 0.645 and the recall was 0.644, which are not

the strongest performers overall but still were stronger results than the 4 metabolites used initially (Table 4.1 and Table 4.4). In this model, we see that the number of controls classified as patients and patients classified as controls are relatively equal (Figure 4.8b). These result patterns are comparable to what was seen in the metabolite classification case. Additionally, the balanced accuracy of this model is 0.657, which is comparable to the results seen in the confusion matrix with an accuracy of 0.644 (Figure 4.8a).

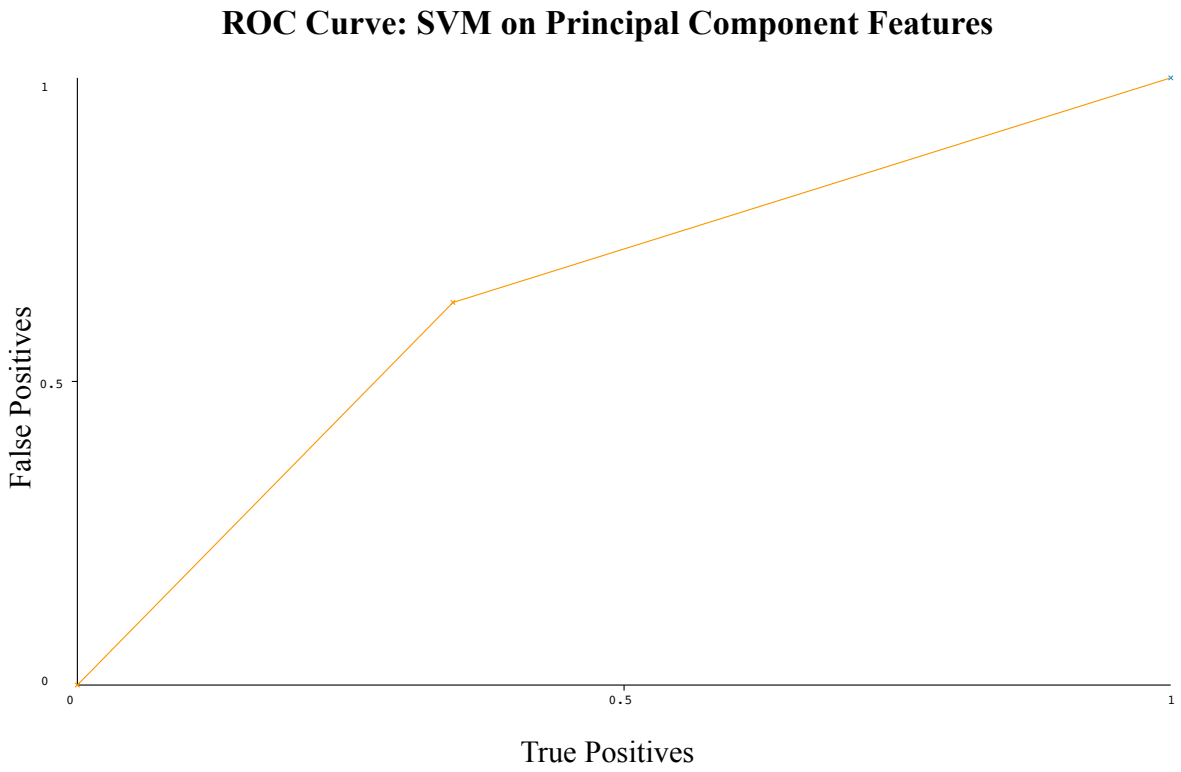
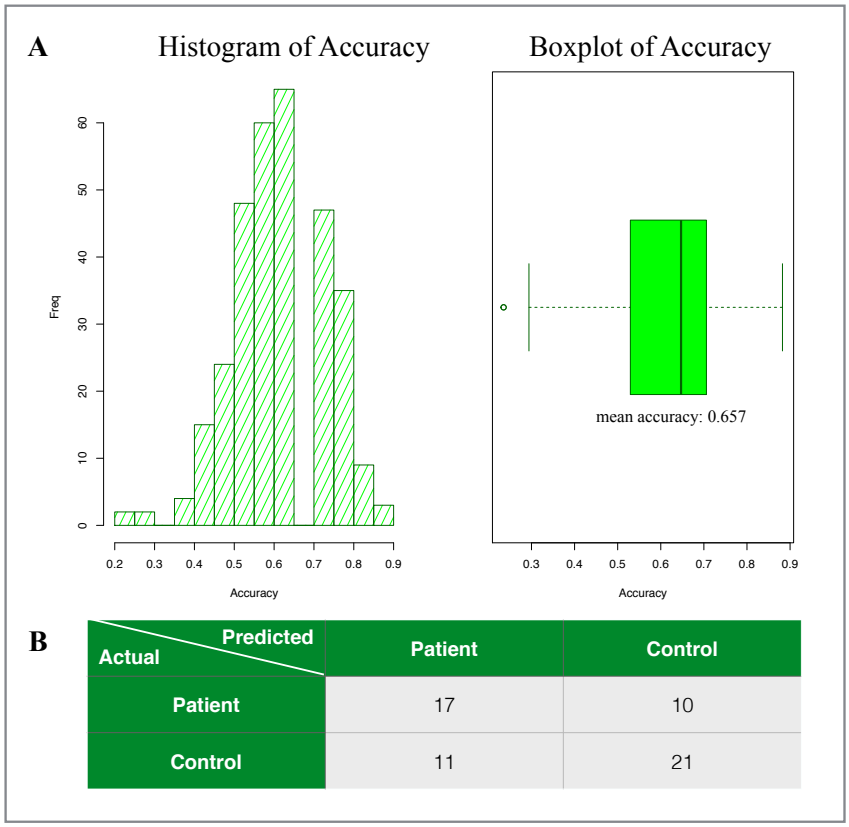


Figure 4.7: SVM ROC Curve for principal component features showing the comparison of true positives to false positives.

Table 4.4: Summary of the results from SVMs for classification. Using PCA to extract components as features, we find that SVMs has the strongest results out of all the different classifiers tested with a precision of 0.645 and a recall of 0.644.

	True Positive Rate	False Positive Rate	Precision	Recall	F-Measure	ROC Area
Patient	0.630	0.344	0.607	0.630	0.618	0.643
Control	0.656	0.370	0.677	0.656	0.667	0.643
Weighted Average	0.644	0.358	0.645	0.644	0.644	0.643



Actual \ Predicted	Patient	Control
Patient	17	10
Control	11	21

Figure 4.8: A: Balanced accuracy for PCA features of the spectral functions is obtained as 0.657 along with the histogram and boxplot of the accuracy as shown. **B:** The SVM confusion matrix shows the predictions of the model compared to the actual classification of individuals with a lower performance than the results seen in the random forest of full functions. Our model

correctly predicts 17/27 of patients correctly, which is higher than the previous models but also fails to predict 10/27 of the patients as depressed.

CHAPTER 5

Discussion

5.1 R Package

We have utilized and created novel tools to analyze and preprocess magnetic resonance spectroscopy signals. Specifically, we have developed an R package known as `mrspecr` that can take in LCModel output data and perform data formatting and processing. The denoising and batch alignment functions allow the data to be manipulated and provide a cleaner format of the data. The coordinate and metabolite extraction functions allow the data necessary for classification to be extracted through file text parsing. Lastly, the classification matrices can be formulated using this `mrspecr` package to allow for machine learning classification on the terms of classifying whether a subject is a depressed patient or a healthy control. As of now, this package specifically has been designed to work with data outputted from LCModel, meaning that each voxel has a corresponding folder with text files containing coordinate data points, metabolite concentrations, and signal to noise ratios that are utilized as inputs to the functions contained in `mrspecr`.

5.2 Machine Learning and Classification

After testing a number of different classification algorithms utilizing scikit-learn, R, and Weka, we found that a single classification model did not work on all types of features. Results diverged depending upon the different types of models used; more specifically, some features that biologically made sense (metabolite concentrations), and that have showed statistical effects when comparing patients and controls, did not perform as strongly as expected. We see that our strongest performing model according to balanced accuracy is the random forest classification using full spectral functions as features (Table 5.1).

Table 5.1: Balanced accuracies of top performing models for each feature set. We see that using random forest classification on full spectral function data points yields the strongest performance.

Data Set:	Metabolites	Distance	PCA	Functions
Classifier:	Logistic	Logistic	SVM	Random Forest
Balanced Accuracy:	0.601	0.646	0.657	0.695

Based on previous literature, we expected that the logistic regression model using the metabolite peaks as features would yield the strongest performance. In testing a number of different classification models on these features, we did not obtain good performance. At a precision of 0.599 and a recall of 0.600, this model was not performing significantly better than a random guess. Likewise, random forest had even less robust results, performing with an average precision of 0.556 and a recall of 0.555 These two classifiers may have better performance upon

future exploration of metabolite concentrations to find the optimal subset of features for input but compared to the other results we have, this did not yield strong results.

Overall, the random forest classification model applied to full spectral data points yielded superior performance, achieving a precision rate upwards of 0.698 with cross validation. We note that the idea of using full functions for classification has not been proposed before. The recall rate of this prediction model was also strong with a value of 0.695. The analysis of the ROC curve for this model allowed us to measure performance along different values for a threshold, and demonstrates the strong predictive capabilities of this model early on (Figure 4.4). We expect a strong prediction curve will follow the top left corner of the plot closely as this would signify minimum false positives and maximum true positives over time with respect to the threshold.

We saw that the confusion matrix points out specific limitations in our model in which we can improve upon, specifically in prediction of false controls. Our model predicts 12 of our 59 subjects as controls when these individuals are actually depressed patients (Figure 4.3). This is the main source of concern because we want to minimize the number of individuals who are not being diagnosed as depressed when they truly do need treatment. The more patients going untreated equates to a larger percent of the population suffering from MDD; this is highly undesirable given the fact that false diagnosis is what we are trying to avoid with these analyses. On the other hand, we only have 6/59 subjects being classified as patients when they are truly healthy controls. This is a small fraction of the population and raises less concern given that this predictive model is meant to assist diagnosis, not replace it. In a model where we desire to

properly diagnose individuals, these results would support majority of diagnoses from mental status examinations and would allow doctors to more confidently diagnose and treat patients with MDD.

Our elastic shape distance approach to classification performed much stronger than initially expected. This is because the shape distance compresses the full function into a single univariate measure and achieves drastic feature reduction. The underlying idea behind the distance calculation was to incorporate the distances between shapes functions for classification. To our knowledge, this idea has not been applied in the context of MRS classification even across any disease modality. Using the `fdasrvf` elastic distance calculation for each subject, we determined a baseline distance difference by comparing the mean function to each of the elastic distances. Using this curve metric for classification, the model performed second to the full function metabolites. These results may have been due to the differences seen in the patient curves when compared to controls. We know that the curves differ because the metabolite concentrations are different, but what we were most interested in seeing is if the distance calculations could outperform the model that involved every single data point as a feature in classification. Using SVMs, the distance matrix performed with a precision of 0.645 and a recall of 0.644 (Figure 4.5).

The final feature set used for classification was principal component analysis (PCA). We expected PCA to outperform the metabolites because rather than choosing the metabolites for classification on our own, we let the PCA choose the top 3, 10 and 20 features that best explained the variance of the data. The PCA with 10 features performed best and the PCA with 3

features performed worst. The reason for poor performance with just 3 features could be explained due to these 3 features not being able to differentiate patients and controls enough to successfully predict new subjects.

Overall, the strongest performance came from random forest classification using the entire 480 data points as features in our classification. Our experiments generated valuable data that suggests innovative approaches and opportunities to grow and expand this line of research (functional analysis of MR spectroscopy) to produce stronger results with more meaningful biological implications.

CHAPTER 6

Future Directions and Conclusion

6.1 Future Directions

An improvement to incorporate in the future would allow for different formats of MRS data to run through `mrspecr` for increased flexibility and utilization. Given this expanded application, more researchers would be provided with the tools to format and process MRS data for machine learning analysis.

On the topic of expanding the algorithms used, there are many avenues for altering or expanding on current practice models. One idea is to tweak random forest to incorporate the elastic distances calculated as one of the bifurcating factors. We already have the algorithm to calculate the elastic distance so utilizing this information in the random forest may boost performance. Another option would be to integrate these models by using the distance as a feature in combination with multiple other features. Combining the distance feature with features from other data sets such as the PCA components as well as core metabolite concentrations could allow for the classification algorithms to integrate this data together and discover distinguishing factors that may not have been clear from using each of the features independently.

With the estimation of PCA features, a future topic of interest would involve mapping those PCA features back to the original LCModel signal. To do this, each feature would need to be decomposed into the eigenvector components and combined with the data from all voxels to utilize metabolite data at a hierarchical scale. This overlay of different levels of data would allow us to see which features from PCA, if any, correspond to a metabolite peak from our metabolite classification matrix. Performing this analysis could allow us to find new metabolites that may be of significant importance and would allow us to reevaluate our choices for the original metabolites (NAA, Cr, Ins, and Glu) in our classification.

6.2 Concluding Summary

In summary, the presented work has showcased the development and utility of the designed R package in processing MRS signals. Additionally, classification models were used on this data to identify key driving features or the signal. In the future, these techniques and algorithms can ultimately assist in the tracking of treatment progress throughout electroconvulsive therapy as well as the diagnosis of patients who may be at risk of MDD.

REFERENCES

1. Marcus, Marina, et al. Depression: A Global Public Health Concern. *PsycEXTRA Dataset*, 2012, doi:10.1037/e517532013-004.
2. Drago, Thomas, et al. "A Comprehensive Regional Neurochemical Theory in Depression: A Protocol for the Systematic Review and Meta-Analysis of 1H-MRS Studies in Major Depressive Disorder." *Systematic Reviews*, vol. 7, no. 1, 2018, doi:10.1186/s13643-018-0830-6.
3. National Institute of Mental Health, U.S. Department of Health and Human Services. Depression. Available: www.nimh.nih.gov/health/topics/depression/index.shtml.
4. Belmaker, R. H. & Agam, G. Major Depressive Disorder. *N. Engl. J. Med*. 358, 55–68 (2008).
5. Fava, Maurizio, et al. "Major Depressive Disorder." *Neuron*, vol. 28, no. 2, 2000, pp. 335–341. doi:10.1016/s0896-6273(00)00112-4.
6. Duman, R.S, et al. A molecular and cellular theory of depression. *Arch. Gen. Psychiatry* 54, 597–606.
7. Petrides, G., et al. ECT Remission Rates In Psychotic Versus Nonpsychotic Depressed Patients: A Report From *CORE. J. ECT* 17, 244–53 (2001).
8. National Institute of Mental Health, U.S. Department of Health and Human Services. Brain Stimulation Therapies. June 2016. Available: www.nimh.nih.gov/health/topics/brain-stimulation-therapies/brain-stimulation-therapies.shtml.

9. Njau, Stephanie, et al. Variations in myo-inositol in fronto-limbic regions and clinical response to electroconvulsive therapy in major depression. *Journal of Psychiatric Research*, 80, 45-51. doi:10.1016/j.jpsychires.2016.05.012

10. Pirnia, T, et al. Electroconvulsive therapy and structural neuroplasticity in neocortical, limbic and paralimbic cortex. *Translational Psychiatry*, 6(6). doi:10.1038/tp.2016.102

11. Gujar, Sachin, et al. "Magnetic Resonance Spectroscopy." *Journal of Neuro-Ophthalmology.* 1 Sept. 2005 Available: insights.ovid.com/crossref?an=00041327-200509000-00015.

12. Duncan, John S. Magnetic Resonance Spectroscopy. *Epilepsia*, vol. 37, no. 7, 1996, pp. 598–605. doi:10.1111/j.1528-1157.1996.tb00622.x.

13. Shirayama, Yukihiko, et al. "Myo-Inositol, Glutamate, and Glutamine in the Prefrontal Cortex, Hippocampus, and Amygdala in Major Depression." Biological Psychiatry: *Cognitive Neuroscience and Neuroimaging,* vol. 2, no. 2, 2017, pp. 196–204. doi:10.1016/j.bpsc.2016.11.006.

14. Kossowski, Bartosz, et al. "Follow-up Analyses on the Effects of Long-Term Use of High Fat Diet on Hippocampal Metabolite Concentrations in Wistar Rats: Comparing Tarquin Quantification of 7.0T Rat Metabolites to LCModel." *Biology, Engineering and Medicine, vol. 2,* no. 4, 2017, doi:10.15761/bem.1000129.

15. Stefan, D, et al. "Quantitation of Magnetic Resonance Spectroscopy Signals: the JMRUI Software Package." *Measurement Science and Technology, vol. 20, no. 10,* Apr. 2009, p. 104035., doi:10.1088/0957-0233/20/10/104035.

16. Simpson, Robin, et al. “Advanced Processing and Simulation of MRS Data Using the FID Appliance (FID-A)-An Open Source, MATLAB-Based Toolkit.” *Magnetic Resonance in Medicine, vol. 77*, no. 1, 2015, pp. 23–33., doi:10.1002/mrm.26091.

17. VeSPA: Integrated Applications for RF Pulse Design cds.ismrm.org/protected/11MProceedings/files/1410.pdf.

18. Near, Jamie, et al. “Frequency and Phase Drift Correction of Magnetic Resonance Spectroscopy Data by Spectral Registration in the Time Domain.” *Magnetic Resonance in Medicine, vol. 73, no. 1*, 2014, pp. 44–50., doi:10.1002/mrm.25094.

19. Vanhamme, Leentje, et al. “Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge.” *Journal of Magnetic Resonance, vol. 129*, no. 1, 1997, pp. 35–43., doi:10.1006/jmre.1997.1244.

20. Blazer, D.G., et. al. The prevalence and distribution of major depression in a national community sample: the National Comorbidity Survey. *Am. J. Psychiatry* 151, 1994, 979–986.

21. Godlewska, Beata R., et al. “Neurochemistry of Major Depression: a Study Using Magnetic Resonance Spectroscopy.” *Psychopharmacology*, vol. 232, no. 3, 2014, pp. 501–507. doi:10.1007/s00213-014-3687-y.

22. Njau, Stephanie, et al. “Neurochemical Correlates of Rapid Treatment Response to Electroconvulsive Therapy in Patients with Major Depression.” *Journal of Psychiatry & Neuroscience*, vol. 42, no. 1, 2017, pp. 6–16. doi:10.1503/jpn.150177.

23. Joshi, Shantanu et al. Denoising of MR spectroscopy signals using total variation and iterative Gauss-Siedel gradient updates. *IEEE International Symposium on Biomedical Imagi*ng; 2015 Apr 16 -19; New York, New York.

24. Provencher, Stephen W. “Automatic Quantitation of Localized in vivo 1H Spectra with LCModel.” *NMR in Biomedicine*, vol. 14, no. 4, 2001, pp. 260–264., doi:10.1002/nbm.698.

25. Provencher, Stephen. LCModel User's Manual. s-provencher.com/lcm-manual.shtml.

26. Lee, David S., et al. “Elastic Registration of Single Subject Task Based FMRI Signals.” *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018 Lecture Notes in Computer Science,* 2018, pp. 154–162., doi:10.1007/978-3-030-00931-1_18.

27. Vasavada, M, et al. Short- and Long-term Cognitive Outcomes in Patients With Major Depression Treated With Electroconvulsive Therapy. *The Journal of ECT*, 33(4), 278-285. doi:10.1097/yct.0000000000000426

28. Joshi, Shantanu, et al. A novel representation for Riemannian analysis of elastic curves in R^n. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7. IEEE, 2007.

29. Joshi, Shantanu, et al. Removing shape-preserving transformations in square-root elastic (SRE) framework for shape analysis of curves. *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR),* pages 387–398, 2007.

30. Jorgensen, A, et al. Regional brain volumes, diffusivity, and metabolite changes after electroconvulsive therapy for severe depression. *Acta Psychiatr Scand*. 2016; 133:154–164. doi: 10.1111/acps.12462.

31. Knudsen, M, et al. Magnetic resonance (MR) spectroscopic measurement of γ-aminobutyric acid (GABA) in major depression before and after electroconvulsive therapy. *Acta Neuropsychiatrica,* 1-10. doi:10.1017/neu.2018.22

32. Tucker, J. Derek, et al, fdasrvf: Elastic Functional Data Analysis. R package version 1.8.3. Available: https://CRAN.R-project.org/package=fdasrvf

33. Tucker, J. Derek, et al. "Generative Models for Functional Data Using Phase and Amplitude Separation." *Computational Statistics & Data Analysis*, vol. 61, 2013, pp. 50–66. doi:10.1016/j.csda.2012.12.001.

34. Pedregosa et al. Scikit-learn: Machine Learning in Python. *JMLR 12*, pp. 2825-2830, 2011.

35. Eibe, Frank, et al. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016

36. Kumar, Ajeet. "Pre-Processing and Modelling Using Caret Package in R." *International Journal of Computer Applications,* vol. 181, no. 6, 2018, pp. 39–42., doi:10.5120/ijca2018917530.

37. Svetnik, Vladimir, et al. "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling." Journal of Chemical Information and Computer Sciences, vol. 43, no. 6, 2003, pp. 1947–1958., doi:10.1021/ci034160g.