DISCRIMINATING PARKINSON'S DISEASE USING FUNCTIONAL CONNECTIVITY AND BRAIN NETWORK ANALYSIS

by

DANIEL GELLERUP

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN INDUSTRIAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2016

Acknowledgements

I would like to thank and acknowledge my thesis advisor, Dr. Wang. I would like to thank the other members of my thesis committee, Dr.Leboulluec, Dr. Chen, and Dr. Rosenberger for their time and support. Special thanks to my friend and research assistant Kin-Ming Puk for his support, advice, and constant help in making this endeavor possible.

April 22, 2016

Abstract

DISCRIMINATING PARKINSON'S DISEASE USING FUNCTIONAL CONNECTIVITY AND BRAIN NETWORK ANALYSIS

Daniel Gellerup, MS

The University of Texas at Arlington, 2016

Supervising Professor: Shouyi Wang

In this study, we explored the use of functional connectivity patterns in fMRI data to classify subjects on the basis of Parkinson's disease. We explore various brain networks and features. We partition our fMRI data in 5 filtered frequency ranges. We use a proximal support vector machine paired with a minimum-redundancy and maximum-relevance feature selection method on each frequency range. We use a majority voting ensemble classification method on the results of the proximal support vector machine classification results. We use a double 5-fold cross validation scheme for model validation. We achieve 84% accuracy 74% sensitivity, and 93% specificity. Our results indicate that the ensemble method is effective compared to a single broad frequency range, and that Bonferroni correction may enhance classification results. We produce brain graphs to illustrate the brain networks of Parkinson's and control subjects.

Table of Contents

Chapter 1

Introduction

Applying machine learning algorithms to neuroimaging data is a promising method of modern neuroscience. Brain signals or transformed brain signals from neuroimaging can be used as features in a machine learning classification algorithm. These features can be used by a classification algorithm to discriminate between brain states. However, there are difficulties with this approach due to the high dimensionality and low signal to noise ratio in fMRI data. Additionally, high variability between brains among individuals makes it difficult to compare them. Simply using brain signals as features may not yield desirable performance. More sophisticated features developed from the underlying brain signals may be useful in improving the learning approach in terms of accuracy, sensitivity, and specificity. Different network approaches for modeling the brain connections may improve classifier performance. Researchers apply methods to reduce noise in the data during a phase of fMRI analysis known as preprocessing. Filters may be applied on the data to focus on certain information bandwidths. Some frequency ranges of fMRI data may contain a better signal to noise ratio for classification than others. An ensemble method that applies majority voting on a range of frequency values may outperform a single frequency range. In this paper, we seek discriminative features and compare their merit with an ensemble support vector machine classification scheme and minimum-redundancy maximum-relevancy feature selection scheme. We perform supervised learning on subjects on the basis of Parkinson's disease, a neuro-degenerative disease that may affect the functional connectivity patterns in the brain. We explore various frequency bands in the preprocessing step. Our approach advances the understanding of classification of functional connectivity patterns in Parkinson's disease and the effect of ensemble classification with band-pass filters. We further contribute to

fMRI research on various correlation and network features used in brain state classification.

1.1 History of Modern Neuroscience

Modern neuroscience is a discipline interested in the scientific study of the nervous system. Neuroscientists study the properties of the brain, including the biological properties of the nervous system, the emergence of consciousness, memory, and cognitive processes. Some approaches used in the neuroscience are cognitive psychology, chemical-biological methods, computational neuroscience, and brain models. Neuroscience is a multi-disciplinary field that involves biologist, chemists, physicists, computer scientists, mathematicians, and related fields.

The central nervous system is a complex system of specific subsystems composes of two main classes of cells: neurons and glial cells. Neurons are the basic signaling units that transmit information throughout the nervous system. Neurons are capable of receiving information, making a decision with it using chemo-biological processes and transmitting to other neurons, using a process called neuronal signaling. Neurons transmit information via electrical impulses, and chemical molecules. Glial cells are non-neural cells that provide structural support, electrical insulation to neurons, and modulate neuronal activities. Neurons are connected in complex networks of dendrites, branches between neurons that transmit inputs between neurons. Synapses are the structures neurons use to send information to another neuron.

Neural communication in the brain involves complex patterns of connectivity involving sets of neurons that form networks of transmission called neural circuits. The structure and density of the network a neuron is contained in varies depending on the location of the neuron and what subsystem of the nervous system the neuron is a

member of. Neural circuits can combine with each other to make neural systems. Using anatomical criteria we can divide the brain into complex neural systems.

Cognitive Psychology explore the mental operations that are produced by the underlying biology of the brain. Cognitive psychologist seek to understand the higher order mental processing abilities of the brain. Cognitive psychology models the brain as input output machine that processes information with internal representations and transforms these representations via cognitive processes. Mental representations are the way information is encoded and processed in the brain. Transformations describe how sensory data is transformed into internal representations and can be further transformed into different mental representations depending on situational context or on the nature of the sensory inputs; for example, images being processed in the brain and prompting a memory of an event with similar images. Cognitive psychology explore the information processing abilities of the brain from a higher order perspective.

Chemical-biological methods explore the brain from its underlying biology. This ranges from the micro-foundations of the brain elements, the individual cells and molecules of the nervous system, or at the level of neural systems and brain systems. Methods include analysis of brain tissue, cells, and neural connections, analysis of brain systems using experiments, elucidating system properties from abnormal or damaged brains, and using imaging techniques to examine the brain. Changes in electrical impulses, fluctuations in blood flow, and shifts in utilization of oxygen and glucose are the changes in the brain that brain imaging exploit to view the inner functioning of the brain. Computed tomography, magnetic resonance imaging, diffusion tensor imaging, positron emission tomography, functional magnetic resonance imaging, and electroencephalography are all imaging techniques that measure various properties of the brain.

Computational neuroscience studies brain function in terms of the information processing properties of brain structures and units that compose the nervous system. It is a multi-disciplinary approach that uses methods from computer science, electrical engineering, and chemistry and biology to study these properties. Computational neuroscience studies all scales of the brain from single neurons, to cognition and consciousness. Brain imaging data is used in conjunction with computational methods such as pattern recognition, statistical analysis, and machine learning to understand the brain.

1.2 Motivation

The basic motivation behind research on supervised machine learning classification in the brain is to understand the functional properties of the brain and to determine mathematical models to determine brain states from underlying measurements of brain elements. One example of the utility of this research is diagnosing brain disorders based on underlying biological mechanisms in the brain, which may be more accurate than diagnosing based on symptoms which could be caused by multiple similar disorders, each with unique treatment options.

Another motivation of this research is to examine multiple features commonly used in fMRI supervised machine learning and comparing them for their predictive abilities. Similarly, filtering imaging data is a common practice, and the best filtering frequency range is an open question. Our research is based on the standard structure of supervised machine learning studies and uses standard features and classifiers. The research in this thesis explores these features, classifiers, and filtering frequencies. We graphically represent the functional connectivity data using brain graph software building on exploratory research of brain graphs of neuro-degenerative diseases.

1.3 Objective of Research

Following in the footsteps of prior research in fMRI brain state classification and research on Parkinson's disease, we seek network features in the brain to use for pattern recognition to diagnose Parkinson's patients. In particular, we explore a range of features from neuroscience theory to use in SVM classification to distinguish between Parkinson's patients and healthy individuals. Further, we are interested in the benefits of using an ensemble SVM method with a range of brain imaging frequency values to evaluate if this improves classification properties. We further seek to compare features and filter frequencies to evaluate their utility in a supervised machine learning scheme for diagnosing Parkinson's. In addition, we plot brain graphs of Parkinson's and healthy subjects to graphically display the functional connectivity patterns.

1.4 Organization of the Thesis

This thesis is organized as follows:

Chapter 2 summarizes the current state of the literature on fMRI classification and Parkinson's research.

Chapter 3 describes our research methodology.

Chapter 4 reports the results of our research.

Chapter 5 concludes the thesis and discusses limitations of the research and further research that is warranted.

Chapter 2

Literature Review

2.9 Diagnosis of Parkinson's and Challenges

Parkinson's is a progressive neurodegenerative disorder that manifests principally as resting tremor, rigidity, akinesia and postural instability, and the diagnosis is based on these symptoms [26]. Parkinson's results in cell death in the substantia nigra, a region of the brain. Parkinson's is associated with the degeneration of dopaminergic nigrostriatal neurons with consequent dysfunction of the cortico-striatal-thalamic loops [44]. It is irreversible once it reaches a mature stage and there are few therapy options. There are enormous costs associated with the disease. Studies have shown that multivariate pattern analysis is capable of extracting functional patterns from neuroimaging data and may be useful for identifying significant neuroimaging-based biomarkers [43]. Biomarkers to distinguish patients from healthy controls are desirable for Parkinson's disease. Biomarkers may be useful for identifying Parkinson's in early stages to be used for early diagnosis and possibly intervention. There may be a transition stage between normal cognition and Parkinson's.Functional connectivity studies on Parkinson's focus on certain brain regions such as the putamen, thalamus or SMA [26]. Certain patterns of functional connectivity associated with Parkinson's may exist as potential biomarkers that could provide additional information for the clinical diagnosis and treatment of this disease [43].

2.1 fMRI and Functional Connectivity

The fMRI technique measures changes in blood supply and blood oxygenation in activated brain regions. The power of this imaging technique is derived from the fact that cerebral blood flow and neuronal activity are related. Blood flow is correlated with brain

activity in particular regions of the brain. As regions of the brain transition from an inactive to an active state, blood flows to that area and the neural tissue may not be able to absorb the oxygen, resulting in a change in the proportion of oxygenated to deoxygenated hemoglobin in the blood. BOLD is the ratio of oxygenated to deoxygenated hemoglobin. The fMRI imaging yields a sequence of 3D brain images with measurements of Blood Oxygenation Level Dependent (BOLD) brain activations [1]. Researchers can identify brain regions activated during fMRI scans and test functional anatomy and functional connectivity [2]. Frequently a block study is employed with the experimental structure of presenting stimuli sequentially within a condition, alternating this with other moments known as epochs when a different condition is presented [3]. Resting state fMRI studies take fMRI images over a course of time when a subject is not exposed to stimuli. It is important to note that movement of blood to a region of the brain is preceded by the activation in the region from a stimulus. BOLD typically peaks approximately 6 to 10 seconds after stimulus [2].

Voxels are nodes of the brain that are measured over the course of an fMRI study. Voxels are a cube of brain tissue typically of length 2mm. Because of the temporal resolution of the fMRI, which requires approximately 2 seconds to scan, it is not possible to measure precisely when activity occurs in the brain because of the rapidity in which neurons can fire. [2]. Because of this, fMRI measures voxels over time and uses statistical analysis to analyze voxel activity [1].

Functional connectivity is brain connectivity defined by a functional connection but not necessarily a structural connection. Functional connectivity is a statistical concept. Two voxels are functionally connected if there are deviations from statistical independence between the time series of the voxels [2]. Therefore, voxels can be functionally related despite the fact that there do not exist structural anatomical brain links

between the two. Functional connectivity relationships can be inferred on functional connectivity in the brain by analyzing the relationships between voxels [4].

2.2 Brain Graphs

Representing functional connectivity in a graphical form using geometry and topology techniques is a useful method for analyzing the brain [5]. Nodes representing anatomical regions, such as voxels, can be used to produce graphs based off of imaging on these nodes. Brain graphs produced from fMRI data have revealed topological properties like small-world-ness, modularity, and heterogeneous distributions. They have also revealed efficient underlying biological mechanisms that the brain is based off of to minimize energy consumption [5].

Brain graphs are simple models of functional connectivity patterns in the brain. Nodes are connected by edges with values based on the strength on the functional connectivity between the two nodes. Some interesting properties of brain graphs are that they are generalizable to any scale of neuro-imaging data, exhibit small world-ness, modularity, and heterogeneity [5]. Small-worldness describes a network where most nodes are not connected by edges to each other, but most nodes can reach another node by a comparatively small sequence of connections between adjacent nodes relative to the overall size of the network [5]. A modular graph is one that can be decomposed into subsystems [5]. Heterogeneity describes graphs with central nodes that have many connections, also known as hubs [5]. Some studies have examined brain graphs of neurodegenerative diseases. For example, reductions in network efficiency have been associated with greater white matter lesion load in patients with multiple sclerosis [6], reductions in the average number of edges per nodes have been observed in patients with Alzheimer's disease [7], and Parkinson's may contribute to weaker connectivity strengths between brain regions compared to healthy individuals [8]

2.3 Preprocessing

2.3.1 Overview of Preprocessing

Preprocessing is a sequence of data transformations on the fMRI images to reduce noise and to make the data useful for statistical analysis [9]. There are several sources of noise during an fMRI reading and statistical corrections can be applied to remedy them. These sources include motion in the head, external body sources, and sources of noise generated by the fMRI scanner. In addition, preprocessing applies statistical techniques to adjust the brain images using normalization and other techniques to reduce noise and allow for statistical analysis. A brief discussion of various preprocessing techniques follows in the sequence they are typically performed [10].

2.3.2 Reconstruction from k-space Data

Reconstruction from k-space data is the first step. The raw signal from the scanner is received in a state known as k-state, which is a spatial frequency transformation of the actual image space. An inverse transform, typically the inverse Fourier transform, is performed to transform the data to image data. Often artefacts, corruptions in brain image, occur from data received from the brain scanner so artefact correction steps are often included in the reconstruction step [5].

2.3.3 Motion Correction

Motion in the head during a scan is a significant problem because the position of the voxels change over time and cause artefacts in data [11]. Data from subjects with too much motion may need to be discarded entirely and for those with less extreme amounts of movement it is desirable to apply realignment correction techniques to their data [11]. This amounts to finding a common orientation of images and performing operations such as translations and rotations to reorient the images to a common frame. Other sources of movement that occur from natural body operations such as the respiratory and cardiac

cycles also cause motion, however minor. Some researchers attempt to minimize the effects of these movements, but this is not a common technique [10].

2.3.4 Slice Timing Correction

Datasets are commonly measured using repeated 2D imaging methods, resulting in a temporal offset between slices. To compensate for this timing difference, slice-timing correction, or temporal data interpolation, has been used as a preprocessing step [12]. The typical method of temporal adjustment is to shift the time series values in the sequence to a reference time using interpolation [10].

2.3.5 Spatial Filtering

Spatial filtering applies an inter-subject averaging technique also referred to as blurring to the imaging data, as spatial normalization cannot perfectly align all structures. This increases signal to noise by removing noise from small scale changes in the image [13].

2.3.6 Normalization

Spatial normalization enables reporting of activations as coordinates within a known standard space [14]. Spatial normalization deforms human brain scans so one location in one subject's brain scan corresponds to the same location in another subject's brain scan [15]. Normalization is often achieved by scaling each voxels time series by a mean intensity over the course of the scan [10].

2.3.7 Temporal Filtering

Temporal filtering is used to filter out frequencies that are not salient to the analysis the researchers are performing. Physiological noise can be reduced by straightforward measures that include filters or more complicated corrections in k-space using navigator echoes or external monitoring and retrospective estimation [16]. High-pass filters can remove low-frequency drifts, which are thought to be caused by

physiological noise as well as by physical scanner-related noise [16]. Low-pass filters are used to reduce physiological high-frequency respiratory and cardiac noise [16]. Band-pass filters, which is both a low and high-pass filter, are frequently used and three frequency bands typically used in the analysis of fMRI data are: 0.01 Hz to 0.03 Hz, 0.03Hz to 0.07Hz, 0.01 to 0.10 Hz [17]. A research paper of particular interest uses frequency values of 0.23 to 0.45 Hz, 0.11 to 0.23 Hz, 0.06 to 0.11 Hz, 0.03 to 0.06 Hz [18]. A result from that paper finds that the most informative frequency ranges are below 0.11 Hz. The fast Fourier transform is a band-pass filter technique that can transform time series data to the frequency domain where it can be filtered, then an inverse Fourier transform can be applied to return to the time domain [19].

2.4 Functional fMRI Features

Common features for fMRI classification include voxel time series data, correlation matrices, topological features, and computational features. Time series from voxels can be effective features [20]. However, some studies have found that simply considering connections as features does not yield desirable predictions [8].

In general, by using domain knowledge to construct appropriate features, one can often improve a learning method that has only the raw features at its disposal [2]. Another approach examines voxels as clusters. ReHo values are a method to analyze voxels at the individual level using cluster assumptions of their neighboring voxel cluster [21]. Similarly, voxel time series can be computationally combined to improve signal to noise ratios. Regression techniques can reduce confounding voxels [22]. Correlation matrices between voxel time series, also known as functional connectivity matrices, are frequently used to represent the functional organization of the brain network across subjects and conditions. Partial correlations represent the strength of edges when all other time series have been regressed out, thereby putatively representing direct links between nodes [23].

Another type of correlation feature defines a feature vector whose elements are pairwise regional correlation coefficients [24]. However, previous studies have shown that correlations among network nodes have been discovered to change throughout a scan in a structured way [25]. Moreover, brain network patterns obtained from voxel-wise correlation features are often difficult to interpret in terms of the underlying neurobiology due to the limitation of feature selection techniques [26].

Topological features emphasize the network structure of the brain. The notions of brain connectivity naturally lend themselves to graphical representation and make use of modern graph techniques to describe and represent brain networks [27]. There is a wealth of previously defined metrics that can be used to characterize the topological architecture of the brain's anatomical or functional connectivity: degree and degree distribution, small-worldness and efficiency, modularity, and distance measures [5].

2.5 Modularity Analysis of Functional Connectivity Networks

An important property of a network is the existence of subsets of the network that form dense networks of their own and that are only weakly linked to the other nodes in the network at large, if at all. If such a subset of the network exists, it is called a community structure of the network. A metric that has been used to identify the community structures of a network is modularity. Modularity is, up to a multiplicative constant, the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random [28].

One common modularity method of identifying community structures is the Louvain method for community detection [29]. The Louvain method is an optimization method that optimizes the modularity of partitions of the network. The Louvain method first finds communities by optimizing modularity locally. Next, it aggregates nodes belonging to the same community and builds a new network whose nodes are the

communities. Then the algorithm iterates these two steps until a maximum value of modularity is achieved and the optimal communities are formed, if they exist [29].

2.6 Feature Selection

Feature selection removes features from the supervised learning algorithm with high noise levels and low signal levels that can reduce classifier performance [1]. Common feature selection methods for fMRI data are to identify specific regions of interest from neuroscience theory and to select active voxels [1]. One approach is to filter voxels based on neuroscience theory focusing only on which voxels are in a region of interest depending on the study [30]. An approach that can be used in conjunction with the region of interest approach is to use the fisher discriminant ratio to choose the most active voxels from the most active regions of interest as the most informative features [31]. Analysis of variance can be performed comparing mean values between voxels and identifying the most dissimilar voxels [1].

Another method to select active regions is independent component analysis (ICA). ICA can be used in fMRI modeling to understand the spatio-temporal structure of the signal [32]. Independent component analysis is frequently performed in conjunction with algorithms to identify and choose voxels with significant signal patterns in the brain activity [33].

An alternative feature selection algorithm is minimum redundancy maximum relevance (mRMR). This approach seeks features with maximum relevance measured in mutual information and minimum redundancy to select mutually exclusive features [34]. Some advantages of this approach are the reduction of noise to improve classification accuracy and more interpretable features of characteristics that can help identify and monitor the target diseases or function types [35]. This approach is common to genetic research and the similarities between fMRI and genetic datasets, such as large

dimensionality with many features that are not salient for the task being researched, make mRMR a useful approach for fMRI data as well [36]. Minimum redundancy eliminates features that are highly correlated with others leaving behind a sparse set of representative features. This is paired with a maximum relevance criteria such as maximum Euclidean distance, minimum pair-wise correlation, or mutual information [34].

2.7 Pattern Classification Models

Machine learning algorithms are well suited for the task of using relationships in fMRI data to decode cognitive states [1]. Supervised learning methods take features from labeled classes and produces a function to map new data to the classes [37]. fMRI brain images have been used in supervised learning algorithms to build decoders for mind states [38].

Pattern classification is a technique that identifies differences patterns and regularities in neural systems [39]. In supervised learning, pattern classification is trained on models with labeled data then applied to unlabeled data. Pattern classification takes the features as inputs and applies a mathematical model to map the data from the high dimensional feature domain to the low dimension class range with values representing classes.

Support vector machines are a prominent pattern recognition based classifier that produce nonlinear boundaries between non-separable classes by constructing linear boundaries in a transformed version of the feature space [37]. SVM selects a hyperplane in the feature space that maximizes its distance from the closes data point. The hyperplane partitions the feature space and assigns data points to one of two classes based on the margins created [35]. Support vector machines (SVM) have been widely used as an accurate and reliable method to decipher brain patterns from functional MRI data [20]. It is possible to augment SVMs with kernels to provide a computationally

efficient way to perform classification without explicitly using potentially long feature vectors [40].

Proximal classification is a computationally useful approach. It is a fast alternative to the standard SVM. Proximal SVM forms proximal planes around which points of the two classes cluster and which are located as far away as possible from each other. PSVM has similar classification performance properties to standard SVM classifiers [40].

Ensemble approaches apply various classification schemes on the data and apply a method to determine a final label value on each subject based on some voting mechanism to decide between the labels chosen by each classifier [37]. One ensemble approach that has been used on fMRI data is to partition the data into frequency bands, use a standard classification method on each sub-band, and use some voting mechanism to determine the final label chosen. This method has been shown to improve SVM classification properties [18].

Chapter 3

Methodology

3.1 Data Acquisition

The fMRI data were acquired using a Philips 3T Achieva MR System from University of Washington, Integrated Brain Imaging Center. A summary of the participants is summarized in Table 1. 45 Subjects were scanned. 24 with Parkinson's (17 male) and 21 control subjects (9 male). 264 functional areas (ROI) were selected for imaging based on the Power 264 template.

The Power template is based off of the combination of two network methods. The first used a large fMRI data set to identify voxels that were activated with a significant probability when research subjects performed specific tasks. The second method used a technique of mapping cortical areas. The combination of these resulted in 264 nodes representing an element of brain organization spanning the cerebral cortex, subcortical structures, and the cerebellum [45].

Descriptive statistics were calculated on the subjects and are represented in Table 1. The Hoehn and Yahr scale is a system for describing how Parkinson's disease progresses. A score of 1 describes a state of unilateral involvement only usually with minimal or no functional disability. A score of 2 describes a state of bilateral or midline involvement without impairment of balance. The number of right-handed subjects is reported because studies have shown evidence of functional differences between left-handers and right-handers that extends to declarative memory processes [46].

Table 3-1 Subject Data

Demographics	PD Patients	Controls	Total Participants
Number of Participants	24	21	45
Age at Scan	66.08(10.27)	61.90(10.00)	64.13(10.25)
Number of Male Participants	17(71%)	9(43%)	16.05(2.23)
Years of Education	16.17(2.12)	15.90(2.39)	16.05(2.23)
Hoehn and Yahr Scale (HY)	2.04(1-2.5)	-	-
Number of Right-handed	20	19	39

We note the similarities and dissimilarities between the patients and controls. The mean ages of Parkinson's patients and controls differ by 5.82 with standard deviations that differ be a value of .27. The years of education are also similar with a value of approximately 16 and standard deviations that differ by a value of .29.

3.2 Frequency Band Analysis of fMRI Data

fMRI data were preprocessed by the University of Washington. First fMRI images were corrected for within-scan acquisition time differences between slices and realigned to the first volume to adjust inter-scan head motions. Subjects with head movement greater than or equal to 2mm or rotation greater than or equal to 1 degree were discarded. Then motion-corrected functional volumes were spatially normalized to the Power 264 brain template and re-sampled into 2-mm isotropic voxels. Finally time-series for each voxel were passed through a temporal band-pass filter for low spatial resolution of .01 to 0.1Hz.

We imported the transformed fMRI time series data into matrices corresponding to each subject with nodes as rows, time as columns, and entries representing neural activity in a node at a specific time. These matrices are normalized with a function that rescales the matrix according to the following equation:

$$x_{ij} = \frac{(y_{ij} - \min(y_{i1,\dots,}y_{in}))}{(\max(y_{i1,\dots,}y_{in}) - \max(y_{i1,\dots,}y_{in}))}.$$

where,

x_{ij} is a normalized element in position (i,j) of the normalized matrix;

y_{ij} is an element in the time series matrix contained in row i of length n.

Next we pass the data through a Fast Fourier Transform (FFT) band-pass filter at a range of low and high filter values to limit our data to frequencies of interest. The Fast Fourier Transform's equation is:

$$X(k) = \sum_{j=1}^{N} x(j)\omega_N^{(j-1)(k-1)}.$$

The inverse fast Fourier transform's (IFFT) equation is:

$$x_j = (\tfrac{1}{N}) \sum_{k=1}^{N} X(k)\omega_N^{-(j-1)(k-1)}.$$

where,

ω_N is a root of unity;

$x(j)$ is an element of a vector of length N.

We compute the FFT, filter out values outside of our frequency range, then use the IFFT to return to our filtered time series. Most fMRI studies select one range of frequencies. Following in the approach of Richiardi's paper "Decoding Brain States from fMRI Connectivity Graphs", we select a range of values. The range of values we use, measured in hertz are as follows: 0.01 to 0.03; 0.03 to 0.06; 0.06 to 0.11; 0.11 to 0.23; and 0.23 to 0.45. These are the same values used by our reference paper, with the addition of the 0.01 to 0.03 that we add to capture signals from lower frequencies [18].

We apply different values for the filter in order to vary the information we receive and analyze how that affects our classification accuracy. We use the Fast Fourier Transform for its filtering properties.

3.3 Brain Network Statistical Measures

Following preprocessing, we initiate our pattern recognition study by calculating classical brain connectivity network matrices to represent functional connectivity in the brain. We use construct various brain networks using correlation and distance approaches. Our networks are constructed from the time series generated by the brain nodes. The brain networks we construct and a brief description of them follow.

Pearson correlation: computes a network matrix from Pearson correlation between all of the nodes:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{ns_x s_y},$$

where

x_i and y_i are a series of n measurements of two time series, X and Y;

s_x and s_y are the sample standard deviations of X and Y;

$\bar{x}$ and $\bar{y}$ are the sample means of X and Y.

Pearson correlation with Fisher's z-transformation: computes a matrix of Pearson correlation between node with the z-transformation to stabilize variance:

$$z = .5\ln\left(\frac{1+r}{1-r}\right).$$

Pearson correlation with a Bonferroni correction: computes a matrix with Pearson correlation p-values based off the null hypothesis that there does not exist a significant connection between nodes. This is computed between every node with a Bonferroni correction to correct the problem of multiple comparisons according to:

$$P_B = \frac{p}{c},$$

where

p is the p-value derived from the t-statistic;

c is the number of comparisons performed on the p-value.

Pearson Correlation with a Bonferroni Correction and a z-transformation: constructs a brain network matrix using a z-transformation on Pearson correlation with a Bonferroni correction as described above.

Tetrachoric correlation: computes a matrix with tetrachoric correlations between nodes, which estimates the correlation between two normally distributed variables to observed ordinal variables according to:

$$r_t = -\cos(2\pi p_{11}),$$

where

p_{11} is the number of indices in the time series of X and Y, with x_i and y_i greater than the median values of their respective time series.

t-stat distance: computes a matrix with a t-value assessing whether the means of the two nodes are statistically different from each other under the assumption that the paired differences are independent and identically normally distributed according to:

$$T_{xy} = \frac{|E[X]-E[Y]|}{\sigma_{xy}/n},$$

where

σ_{xy} is the sample standard deviation between time series X and Y.

Lagged correlation with maximum criterion: computes lagged cross correlation between nodes according to:

$$r = \frac{\sum_i[(x_i-\bar{x})*(y_{i-d}-\bar{y})]}{\sqrt{\sum_i(x_i-\bar{x})^2}\sqrt{\sum_i(y_{i-d}-\bar{y})^2}},$$

where

d is the delay.

Lagged correlation with optimum criterion: computes lagged cross correlation with an optimum lag criteria using cross correlation.

Euclidean distance: computes a matrix composed of the difference in distance between each brain node:

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + \cdots + (q_n - p_n)^2},$$

where

q, p are points with dimension n.

Standard Euclidean distance: computes a matrix composed of Euclidean distances where every element of a time series has been divided by the standard deviation of the time series.

Chebychev distance: computes the distance between each pair of nodes by the maximum distance along any of the three coordinate dimensions of the time series given by:

$$D_C(p,q) = \max\left(|p_i - q_i|\right).$$

City block distance: computes the minimum distance traveled along piecewise linear paths between nodes, whose constituent line segments are all parallel to one of the coordinate axes given by:

$$d_1(p,q) = \sum_{i=1}^{n} |p_i - q_i|.$$

3.4 Brain Connectivity Modularity

Next we apply an algorithm to search for modularity within each network computed using the Louvain method. The optimal community structure is a subdivision of the network into non-overlapping groups of nodes which maximizes the number of within-group edges, and minimizes the number of between-group edges. We find community structure by optimizing modularity defined as:

$$Q = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j),$$

where

A_{ij} represents the edge weight between nodes i and j;

k_i and k_j are the sum weights of the edges attached to nodes i and j respectively;

m is the sum of all edge weights in the graph;

c_i and c_j are the communities of the nodes;

$\delta(c_{i,}c_j) = 1$ if $c_i = c_j$ and 0 otherwise.

For each network, we retain our complete networks, calculate positive and negative only networks, submodule networks, and a network of only submodules. We select the five most modular submodules.

3.5 Feature Extraction

Next we extract features from each network. We extract both correlation features and topological features. Our correlation features take as inputs the networks and modulated networks computed by the correlation statistics described above. We take the elements of the matrices and translate them column-wise to a vector for all 45 subjects. Additionally, we investigated and evaluated a set of neuro-biologically meaningful topological brain network features calculated from the connectivity matrices. A brief description of these follow.

Clustering Coefficient: the fraction triangles around a node, where a triangle is a set of 3 nodes that are all connected to each other according to:

$$C = \frac{1}{n}\sum_{i=1}^{n}\frac{2t_i}{k_i(k_i-1)},$$

where

n is the number of nodes;

k_i is the degree of a node;

t_i is the number of triangles around a node.

Characteristic Path Length: calculates the average path length between nodes in the network according to:

$$L = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j\neq i} d_{ij}}{n-1},$$

where

d_{ij} is the shortest path length between nodes i and j.

Eccentricity: calculates the max value between every node and all other nodes, provided there is a connection between them according to:

$$ecc_i = \max\left(a_{i1,\ldots,}a_{ij}\right),$$

where

a_{ij} is row i and column j of the connectivity network.

Radius: the minimum eccentricity of the network

$$R = \min\left(ecc_1, \ldots, ecc_n\right).$$

Small world-ness: measures the network in terms of being more clustered than a random network, yet having same characteristic path length as random networks according to:

$$S = \frac{C/C_{rand}}{L/L_{rand}},$$

where

C_{rand} is the clustering coefficient of a randomly generated network with as many nodes as the network;

L_{rand} is the characteristic path length of a randomly generated network with as many nodes as the network;

Standard statistics: a set of standard statistical functions is applied on the network matrix: mean, standard deviation, quantiles, range, kurtosis, and skewness.

3.6 Classification Overview

We performed an ensemble classification scheme depicted in Figure 3-1.

We subdivide our data into 5 frequency ranges and perform mRMR feautes selection with PSVM classification on each band. We then perform ensemble classification on the results of each band.

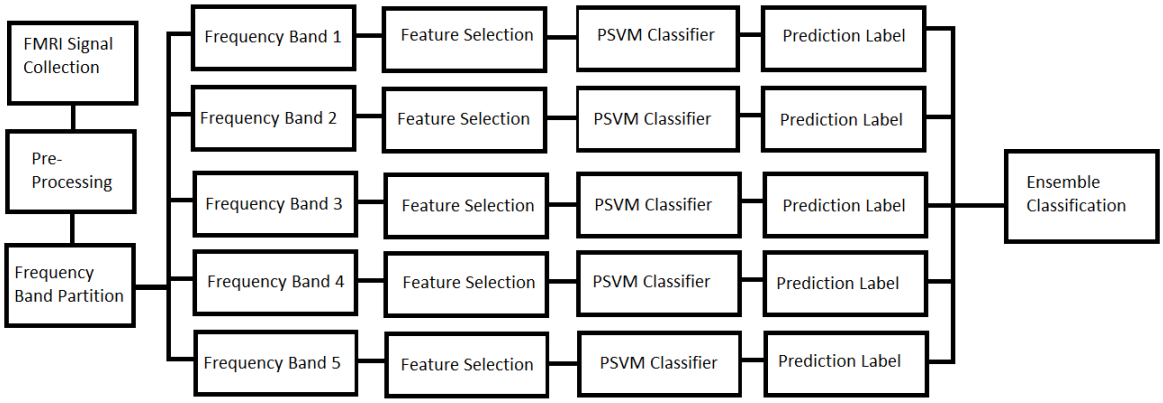

Figure 3-1 Classification Framework Overview

3.7 Feature Selection

Our feature extraction methods will produce many features depending on the number of modules computed by the Louvain algorithm. This has adverse consequences to classification. Therefore, we use an mRMR selection method to select the most informative features. This approach seeks the maximally relevant and minimally redundant features in our dataset. Our mRMR algorithm uses mutual information as the distance to compute feature to feature and feature to class labels similarities.

Mutual information is computed as:

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right),$$

where

X and Y are two features;

$p(x)$ and $p(y)$ are marginal probability functions;

$p(x, y)$ is the connected probability distribution;

An optimal subset of features can be obtained by minimizing the following objective function:

$$\varphi(Rd, Re) = \frac{1}{|S|^2}\sum_{i,j\varepsilon S} I(i,j) - \frac{1}{|S|}\sum_{i\epsilon S} I(h,j),$$

where

Rd is redundancy;

Re is relevance among features;

S is the set of features;

h is the vector of target class labels;

$I(h,j)$ is the amount of mutual information between features h and j;

3.8 PSVM Classification

Due to its computational efficiency with respect to time and classification properties we use proximal SVM as our classifier using the features selected by mRMR. The standard SVM with Linear Kernel is defined as:

$$\min ((\frac{1}{2})\|w\|^2 + \beta\sum_{i=2}^{n}\varepsilon_i,$$

$$such\ that\ z_i(w^t x_i + w_o) - 1 + \varepsilon_i \geq 0, \varepsilon_i \geq 0.$$

SVM creates hyperplanes used to classify classes by minimizing the margin to the other classes' hyperplane. PSVM showed the explicit nonnegativity constraint is not needed. PSVM formulation replaces the inequality constraint by an equality, which changes the nature of optimization problem, and one can obtain an explicit exact solution to the problem. PSVM uses the L2-norm error term which adds a stronger convexity condition to the objective function. With PSVM, the hyperplanes are not bounding planes anymore, but can be thought of as "proximal" planes, around which the points of each

class are clustered and which are pushed as far apart as possible. The PSVM formulation with linear kernel is defined as:

$$\min ((\tfrac{1}{2})\|w\|^2 + (\tfrac{1}{2})\beta \sum_{i=2}^{n} {\varepsilon_i}^2,$$

$$such\ that\ z_i(w^t x_i + w_o) - 1 + \varepsilon_i = 0.$$

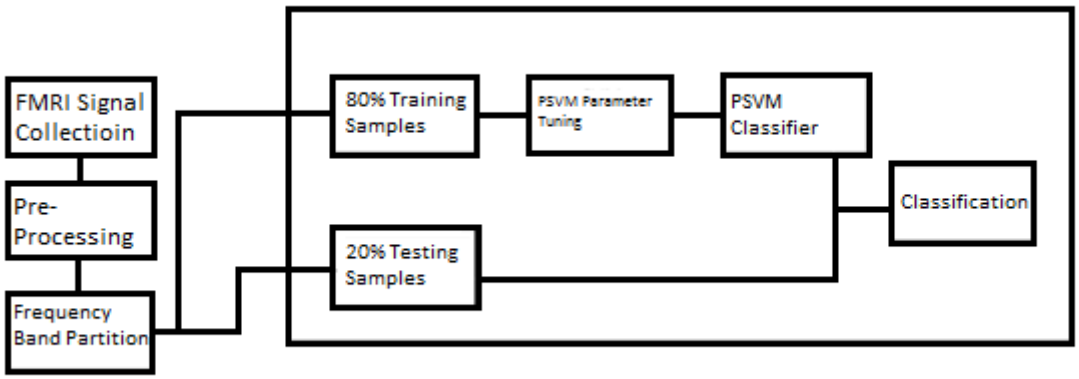

Figure 3-2 Proximal SVM Classification with 5-fold Cross-validation

3.9 Ensemble Classification

We then use an ensemble framework that uses majority voting on the labels assigned to each subject by each frequency range to determine a final label assigned to the subject. This label corresponds to Parkinson's or control. Our ensemble classification scheme is depicted by Figure 3-3.

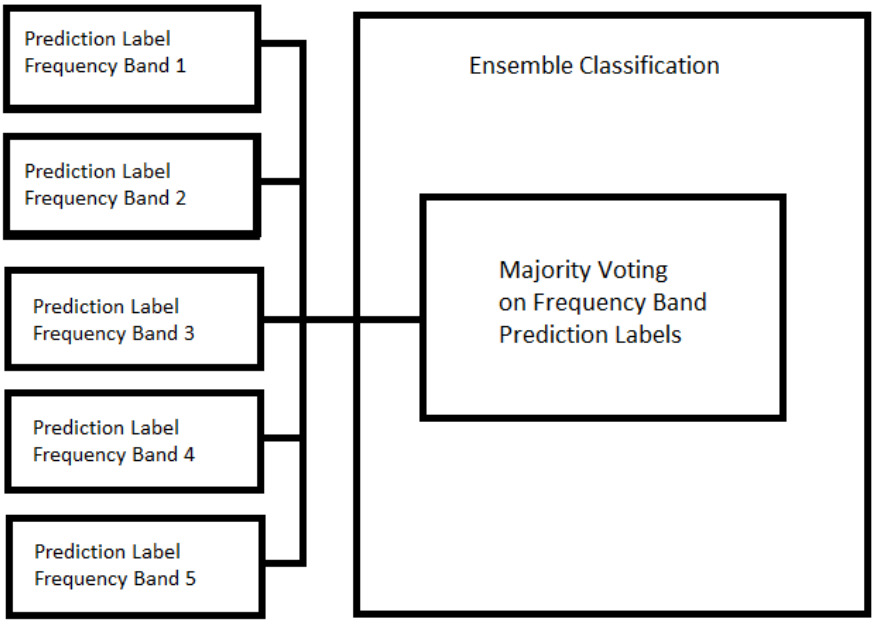

Figure 3-3 Ensemble Classification With Majority Voting

3.10 Classification Performance Evaluation

The N-fold cross-validation is an attractive model evaluation method when the sample size is small. It is capable of providing an almost unbiased estimate of the generalization ability of a classification model [41]. N-fold cross-validation divides the dataset into N approximately equal mutually exclusive and collectively exhaustive subsets. It then uses the N-1 subsets for training and the N subset that is omitted for testing. The process then cycles through all of the N subsets so each is omitted from training and used for testing once. In determining what value to use for N, it is important to consider prediction error and the size of the dataset. If the learning curve has a considerable slope at a given training set size, a larger N will have higher but lower variance compared to a lower N value [37]. Five-fold cross-validation is considered a good compromise [42]. We use 5-fold cross-validation for classification performance

evaluation. We evaluate performance in terms of sensitivity, specificity, and accuracy, defined as:

$$sensitivity = \frac{\#\ correctly\ classified\ type\ Parkinson's\ subjects}{Total\ number\ of\ type\ Parkinson's subjects},$$

$$specificity = \frac{\#\ correctly\ classified\ type\ control\ subjects}{Total\ number\ of\ type\ control\ subjects},$$

$$accuracy = \frac{\#\ correctly\ classified\ subjects}{Total\ number\ of\ subjects}.$$

For 5-fold cross-validation, we reserve 20% of the sample for testing with 80% for training and SVM parameter tuning. In particular, we divide the training data into 5 non-overlapping sets. We leave one of the sets out a perform feature selection and classification using the remaining nine subsets. The set left out is used to evaluate the performance of the trained classifier. Then we cycle the set that was left out back in and remove one of the 4 sets and repeat the method until each of the 5 subsets has been left out once and used for validation. We average sensitivity, specificity, and accuracy over the 5 subsets.

Chapter 4

Experimental Results

4.1 Modularity

The Louvain method yields the top 5 community structure modules for each of our correlation networks. In addition we have 3 other networks to examine. This results in a sum total of 8 networks listed below:

- 5 submodules from the Louvain method.
- Global network consisting of all of the values in the network.
- Global network positive values consisting of all of the positive values in the network.
- Global network negative values consisting of all of the negative values in the network.

4.2 Feature Selection

For each subject, we have 12 correlation network types, each with 9 networks, and each network with 6 feature types. This results in a large number of features for classification. We report the complete network (global network) statistics because of its superior classification performance. The number of features in submodules varies based on the size of the submodule, but is strictly less than the total number of features for the global network.

Table 4-1 Feature Type Sizes for Global Network Matrices

Feature Type	Rows	Columns
Correlation Value	45	69432
Clustering Coefficient	45	264
Characteristic Path Length, Eccentricity, Radius	45	266
Small-Worldness	45	264
Distance Statistics	45	9
All Features	45	70235

Because of the large number of features, we use mRMR as our feature selection method due to its ability to select the most informative features from the large set. We use two parameter for our mRMR scheme: choose the top 100 features or the top 50 features.

4.3 Classification

Our methods result in a total of 12,653 proximal SVM classifications on frequency bands. We perform classification on all of the combinations of networks, frequency ranges, and features. We focus our analysis on correlation features. However, we present the best performing submodule and topological feature combinations below on the Pearson network with ensemble classification.

Table 4-2 Classification with Submodules and Topological Features Pearson Network

Submodule	Feature	# Features	Accuracy	Specificity	Selectivity
4	Characteristic Path Length, Eccentricity, and Radius	32	0.7	0.68	0.71
4	Clustering Coefficient	10	0.67	0.71	0.63
4	Small-worldness	10	0.66	0.72	0.59
3	Standard Statistics	9	0.62	0.63	0.61
1	Small-worldness	50	0.61	0.64	0.57
2	Clustering Coefficient	39	0.6	0.6	0.6
5	Small-worldness	10	0.58	0.6	0.55

It is interesting to note that the best submodule for characteristic path length, eccentricity, and radius; clustering coefficient, and small world-ness is submodule 4. Submodule 3 has the best performance with the standard statistics. We also report the best features for submodules 1, 2, and 5.

Following the reference paper [18], we focused on the global positive network with correlation matrices as our features. The following tables present the results of proximal SVM on all networks and frequency bands.

Table 4-3 PSVM Frequency Band 0.01 – 0.03Hz PSVM

Network	Frequency Range	# Features	Accuracy	Sensitivity	Specificity
Pearson with Bonferroni	.01 - .03	100	0.81	0.88	0.73
Pearson	.01 - .03	100	0.74	0.73	0.74
Lagged Correlation Max	.01 - .03	50	0.70	0.75	0.64
Tetrachoric	.01 - .03	50	0.67	0.65	0.68
Pearson with Bonferroni and z	.01 - .03	100	0.63	0.6	0.65
Pearson with z	.01 - .03	50	0.62	0.64	0.6
Chebychev	.01 - .03	100	0.60	0.58	0.61
Lagged Correlation Optimum	.01 - .03	100	0.61	0.59	0.62
t-stat	.01 - .03	50	0.58	0.57	0.58
Euclidean	.01 - .03	100	0.53	0.55	0.5
City Block	.01 - .03	50	0.43	0.48	0.38
Standardized Euclidean	.01 - .03	100	0.44	0.48	0.39

Table 4-4 PSVM Frequency Band 0.03 – 0.06Hz PSVM

Pearson with z	.03 - .06	50	0.72	0.76	0.68
Pearson with Bonferroni and z	.03 - .06	100	0.69	0.68	0.7
Lagged Correlation Optimum	.03 - .06	100	0.67	0.67	0.67
Standardized Euclidean	.03 - .06	100	0.67	0.68	0.65
City Block	.03 - .06	50	0.65	0.68	0.61
Tetrachoric	.03 - .06	50	0.65	0.65	0.64
Pearson	.03 - .06	100	0.62	0.64	0.6
Pearson with Bonferroni	.03 - .06	100	0.61	0.65	0.56
Chebychev	.03 - .06	100	0.58	0.59	0.56
Euclidean	.03 - .06	100	0.53	0.55	0.5
Lagged Correlation Max	.03 - .06	50	0.48	0.48	0.48

t-stat	.03 - .06	50	0.47	0.5	0.43
Pearson with z	.03 - .06	50	0.72	0.76	0.68

Table 4-5 PSVM Frequency Band 0.06 – 0.11 PSVM

Network	Frequency Range	# Features	Accuracy	Sensitivity	Specificity
Lagged Correlation Optimum	.06 -.11	100	0.75	0.7	0.8
Pearson with z	.06 -.11	50	0.68	0.65	0.71
Pearson with Bonferroni	.06 -.11	100	0.65	0.64	0.65
City Block	.06 -.11	50	0.63	0.61	0.64
Lagged Correlation Max	.06 -.11	50	0.63	0.62	0.63
Pearson	.06 -.11	100	0.62	0.64	0.6
Pearson with Bonferroni and z	.06 -.11	100	0.63	0.62	0.63
Tetrachoric	.06 -.11	50	0.62	0.64	0.6
Euclidean	.06 -.11	100	0.55	0.54	0.56
Standardized Euclidean	.06 -.11	100	0.55	0.55	0.55
Chebychev	.06 -.11	100	0.51	0.54	0.47
t-stat	.06 -.11	50	0.51	0.54	0.48

Table 4-6 PSVM Frequency Band 0.11 – 0.23Hz PSVM

Network	Frequency Range	# Features	Accuracy	Sensitivity	Specificity
Pearson with Bonferroni	.11-.23	100	0.82	0.76	0.88
Pearson with Bonferroni and z	.11-.23	100	0.79	0.71	0.86
Pearson	.11-.23	100	0.69	0.71	0.67
Chebychev	.11-.23	100	0.67	0.67	0.67
Euclidean	.11-.23	100	0.65	0.63	0.67
Lagged Correlation Max	.11-.23	50	0.62	0.62	0.62
Pearson and z	.11-.23	50	0.62	0.65	0.59
t-stat	.11-.23	50	0.62	0.63	0.61
Standardized Euclidean	.11-.23	100	0.60	0.59	0.6

Tetrachoric	.11-.23	50	0.60	0.63	0.57
Lagged Correlation Optimum	.11-.23	100	0.58	0.6	0.55
City Block	.11-.23	50	0.58	0.56	0.59

Table 4-7 PSVM Frequency Band 0.23 – 0.45Hz PSVM

Network	**Frequency Range**	**# Features**	**Accuracy**	**Sensitivity**	**Specificity**
t-stat	.23 - .45	50	0.72	0.69	0.75
Pearson	.23 - .45	100	0.68	0.66	0.69
Pearson with z	.23 - .45	50	0.72	0.61	0.82
Tetrachoric	.23 - .45	50	0.68	0.63	0.73
Chebychev	.23 - .45	100	0.62	0.62	0.62
City Block	.23 - .45	50	0.63	0.61	0.64
Euclidean	.23 - .45	100	0.64	0.61	0.67
Pearson with Bonferroni and z	.23 - .45	100	0.60	0.58	0.61
Pearson with Bonferroni	.23 - .45	100	0.58	0.59	0.56
Lagged Correlation Max	.23 - .45	50	0.55	0.58	0.52
Lagged Correlation Optimum	.23 - .45	100	0.56	0.58	0.53
Standardized Euclidean	.23 - .45	100	0.52	0.52	0.52

In addition, we ran a full frequency range of .01-.45 in order to compare our ensemble result to non-filtered proximal SVM.

Table 4-8 PSVM Frequency Band 0.01 - 0.45Hz

Network	Frequency Range	# Features	Accuracy	Sensitivity	Specificity
Pearson and z	.01-.45	50	0.75	0.7	0.8
City Block	.01-.45	50	0.675	0.66	0.69
Pearson with Bonferroni and z	.01-.45	100	0.675	0.66	0.69
Tetrachoric	.01-.45	100	0.645	0.67	0.62
Lagged Correlation Optimum	.01-.45	50	0.62	0.63	0.61
Euclidean	.01-.45	100	0.615	0.59	0.64
Pearson with Bonferroni	.01-.45	100	0.6	0.61	0.59
t-stat	.01-.45	50	0.53	0.56	0.5
Pearson	.01-.45	100	0.53	0.56	0.5
Chebychev	.01-.45	100	0.505	0.54	0.47
Standardized Euclidean	.01-.45	100	0.475	0.52	0.43
Lagged Correlation Max	.01-.45	50	0.37	0.49	0.25

We applied majority ensemble voting on each set of 5 frequency ranges, with the following results:

Table 4-9 PSVM Ensemble Classification Global Network Positive Correlation

Network	Feature	# Features	Accuracy	Sensitivity	Specificity
Pearson with Bonferroni	Correlation	100	0.84	0.74	0.93
Lagged Correlation Optimum	Correlation	100	0.75	0.7	0.8
Pearson	Correlation	100	0.74	0.73	0.74
Pearson and z	Correlation	50	0.71	0.72	0.7
Tetrachoric	Correlation	50	0.71	0.74	0.68
Pearson with Bonferroni and z	Correlation	100	0.72	0.66	0.77
Lagged Correlation Max	Correlation	50	0.65	0.64	0.65
Chebychev	Correlation	100	0.62	0.59	0.64
Euclidean	Correlation	100	0.59	0.57	0.6
t-stat	Correlation	50	0.55	0.58	0.52
Standardized Euclidean	Correlation	100	0.53	0.55	0.5
City Block	Correlation	50	0.50	0.53	0.46

Our best performing ensemble data set was Pearson correlation with Bonferroni correction with an accuracy of 84%.

We present the best performing sub-band for comparison. It is interesting to note that the best performing sub-band in terms of accuracy varies between the different correlation networks.

Table 4-10 PSVM Most Accurate Sub-band per Network

Network	Frequency Range	# Features	Accuracy	Sensitivity	Specificity
Pearson with Bonferroni	.01 - .03	100	0.81	0.88	0.73
Pearson with Bonferroni and z	.11-.23	100	0.79	0.71	0.86
Lagged Correlation Optimum	.06 -.11	100	0.75	0.7	0.8
Pearson	.01 - .03	100	0.74	0.73	0.74
t-stat	.23 - .45	50	0.72	0.69	0.75
Pearson with z	.23 - .45	50	0.72	0.61	0.82
Lagged Correlation Max	.01 - .03	50	0.70	0.75	0.64
Tetrachoric	.23 - .45	50	0.68	0.63	0.73
Chebychev	.11-.23	100	0.67	0.67	0.67
Standardized Euclidean	.03 - .06	100	0.67	0.68	0.65
Euclidean	.11-.23	100	0.65	0.63	0.67
City Block	.03 - .06	50	0.65	0.68	0.61

We now compare the best frequency band to the ensemble method that uses the range of frequency bands to compare performance.

Table 4-11 Difference Between Ensemble and Best Sub-band Performance

Network	Delta Accuracy	Delta Sensitivity	Delta Specificity
Pearson with Bonferroni	0.06	0.03	0.07
Tetrachoric	0.03	0.11	-0.05
Pearson	0.01	0.00	0.00
Pearson with Bonferroni and z	0.01	0.05	-0.05
Lagged Correlation Optimum	0.00	0.00	0.00
Lagged Correlation Max	-0.05	-0.11	0.01
Chebychev	-0.05	-0.08	-0.03
Euclidean	-0.06	-0.06	-0.07
Pearson and z	-0.10	-0.16	-0.03
Standardized Euclidean	-0.14	-0.13	-0.15
City Block	-0.15	-0.15	-0.15
t-stat	-0.17	-0.11	-0.23

Table 11 reveals that the ensemble method never outperformed the best frequency sub-band in terms of accuracy, but did equal the same value for accuracy for 4 of the 12 Data sets. The bottom 7 best networks performed considerably worse than the best sub-band.

We also compare the ensemble method applied to the full frequency range 0.01 to 0.45. These differences are as follows:

Table 4-12 difference between ensemble and full frequency range

Network	Delta Accuracy	Delta Sensitivity	Delta Specificity
Lagged Correlation Max	0.28	0.15	0.40
Pearson with Bonferroni	0.24	0.13	0.34
Pearson	0.21	0.17	0.24
Lagged Correlation Optimum	0.13	0.07	0.19
Chebychev	0.12	0.05	0.17
Tetrachoric	0.06	0.07	0.06
Standardized Euclidean	0.06	0.03	0.07
Pearson with Bonferroni and z	0.04	0.00	0.08
t-stat	0.02	0.02	0.02
Euclidean	-0.03	-0.02	-0.04
Pearson and z	-0.04	0.02	-0.10
City Block	-0.18	-0.13	-0.23

We note that the Ensemble method is superior with respect to accuracy, for 9 of 12 networks with an average gain of .08, and that our most accurate ensemble classification scheme, Pearson with Bonferroni, is 24% more accurate than using the full frequency band.

4.4 Brain Graphs

We generate brain graphs of the networks of Parkinson's and healthy subjects using the BrainNet Viewer program to visualize the networks. This program illustrates human connectomes as ball-and-stick models displaying brain surface, nodes, and

edges. The brain graphs we provide are averaged over the groups to provide a representative graph for Parkinson's subjects and controls.

We plotted brain graphs of Pearson correlation for their representation of functional connectivity in the subject's brains. These results are an average of functional connectivity in the research subjects brains based on Parkinson's status. Figure 4-1 depicts the average functional connectivity patterns of the control subjects.

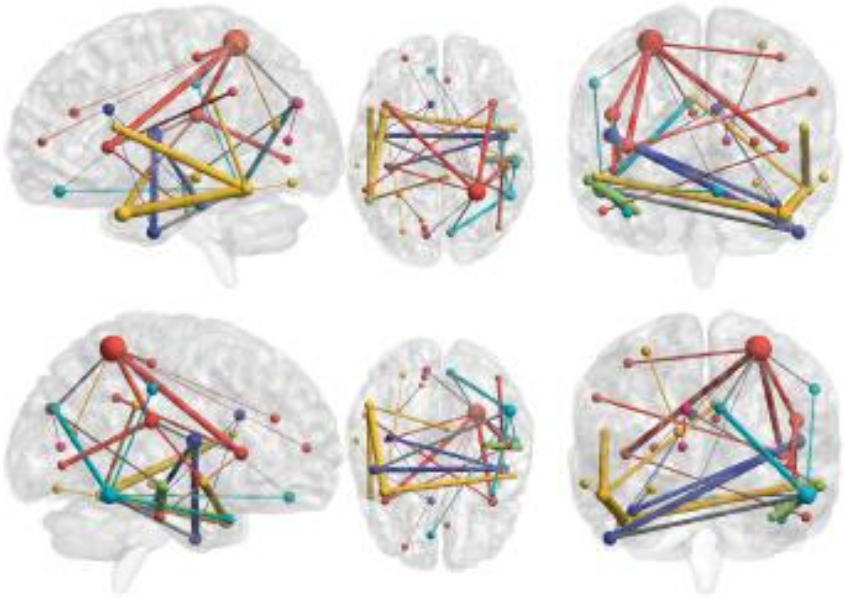

Figure 4-1 Pearson Correlation Control

Figure 4-2 depicts the average functional connectivity patterns of the Parkinson's subjects.

Figure 4-2 Pearson Correlation Parkinson's

Thicker bands correspond to stronger network connectivity. From visual inspection, we find similar functional connectivity patterns between the two classes in the Pearson correlation brain graphs. We note that control subjects have visibly stronger functional connection patterns than Parkinson's subjects.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This thesis has explored the utility of an ensemble classification method using majority voting on labels assigned by proximal SVM performed on frequency sub-bands. Our best ensemble performance is Pearson Correlation with a Bonferroni correction, which yields an accuracy of 80%, sensitivity of 74% and specificity of 93%. However, this performance is very similar to the best frequency sub-band of .11 to .23 Hz. We find that in general the best sub-band frequently outperforms or is similar to the ensemble method. We produce brain graphs based on average Pearson correlation of Parkinson's and control subjects but they do not yield major visual differences. We explored proximal SVM classification on a range of topological features. We used the Louvain method to determine community structures in the data and discovered 5 submodules of the full global network. We used the global network with positive values for our ensemble classification.

5.2 Future Work

The ensemble frequency approach applied to different fMRI studies may yield insights on its utility. The comparison between the ensemble approach and the best sub-band may suffer from bias by selecting the best sub-band ex-ante. An approach that tested ensemble and individual sub-bands could be warranted to better evaluate the ensemble method. We performed proximal classification using several topological features. A more in depth exploration of these features may be warranted with respect to Parkinson's data or other fMRI data with a full comparison between topological and correlation based features. A more in depth analysis of Parkinson's brain graphs using additional correlation types or topological methods coupled with numerical analysis may

yield insights. Similarly we explored several networks and subnetworks in our proximal classification scheme. These could warrant further exploration to understand their nature and function and possible relation to Parkinson's disease. The performance of Pearson correlation with Bonferroni may suggest further research on this correlation variant. Artefact removal in preprocessing may improve the signal to noise ratio.

References

1. *Machine Learning Approaches for Cognitive State Classification and Brain Activity Prediction: A Survey.* Parida, Shantipriya, Satchidananda Dehuri, and Sung-Bae Cho. 2015, Current Bioinformatics, pp. 344-359.

2. Gazzaniga, Michael, Richard Ivry, and G.R. Mangun. *Cognitive Neuroscience: The Biology of the Mind.* New York : W.W. Norton, 1998.

3. *Study Design in fMRI: Basic Principles.* Amaro, Edson, and Gareth Barker. 2006, Brain and Cognition, pp. 220-232.

4. *Connectomics and New Approaches for Analyzing Human Brain Functional Connectivity.* Craddock, R. Cameron, Rosalia L. Tungaraza, and Michael P. Milham. 2015, GigaScience, p. 13.

5. *Brain Graphs: Graphical Models of the Human Brain Connectome.* Bullmore, Edward, and Danielle Bassett. 2011, Annual Review of Clinical Psychology, pp. 113-140.

6. *Impaired Small-Worlk Efficiency In Structural Cortical Networks in Multiple Sclerosis Associated with White Matter Lesion Load.* He Y, Dagher A, Chen Z, Charil A, Zijdenbos A, et al. 2009, Brain, pp. 3366-3379.

7. *Cortical Hubs Revealed by Intrinsic Functional Connectivity: Mapping, Assessment of Stability, and Relation to Alzheimer's Disease.* Buckner RL, Sepulcre J, Taludkar T, Krienen FM, Liu H, el. 2009, Neurosci, pp. 1860-1873.

8. *Discriminating Parkinson's Disease (PD) Using Functional Connectivity and Brain Network Analysis.* Madhyastha, Tara, Shouyi Wang, Kinming Puk, Cao Xiao, Thomas Grabowski, W. Chaovalitwongse. 2016.

9. *Theoretical, Statistical, and Practical Perspectives on Pattern-based Classification Approaches to the Analysis of Functional Neuroimaging Data .* O'Toole A, Jiang F, Abdi H, Penard N, Dunlop J, Parent M. 2007, Cogn Neurosci, pp. 1735-1752.

10. *Pre-Processing of BOLD fMRI Data.* Jenkison, Mark, and Stephen Smith. 2006. Oxford Centre for Functional MRI of the Brain.

11. *Head Motion and Correction Methods in Resting-State Functional MRI.* Goto, Masami, et al. 2015, Magnetic Resonance in Medical Sciences.

12. *Slice-Timing Effects and their Correction in Functional MRI.* Sladky, Ronald, et al. 2011, Neuroimage, pp. 588-594.

13. *Adaptive Spatial Smoothing of fMRI Images.* Yue, Yu, Ji Loh, and Martin Lindquist. 2010, Statistics and its Interface, pp. 3-13.

14. *A Review of Geometric Transformations for Nonrigid Body Registration.* Holden, Mark. 2008, Medical Imaging, pp. 111-128.

15. *Correlations and Anti-correlations in Resting-State Functional Connectivity MRI: Aquantitative Comparison of Preprocessing Strategies.* Weissenbacher, A, Kasess, C, Gerstl, F, Lanzenberger, R, Moser, E, and Windischberger, C. 2009, Neuroimage, pp. 1408-1416.

16. *Comparison of Detrending Methods for Optimal fMRI Preprocessing.* Tanabe, Jody, et al. 2002, Neuroimage, pp. 902-907.

17. *Machine Learning with Brain Graphs: Predictive Modeling Approaches for Functional Imaging in Systems Neuroscience.* Richiardi, Jonas, et al. 2013, Signal Processing Magazine, pp. 58-70.

18. *Decoding Brain States from fMRI Connectivity Graphs.* Richiardi, Jonas, et al. 2011, Neuroimage, pp. 616-626.

19. *An Algorithm for the Machine Calculation of Complex Fourier Series.* Cooley, James, and John Tukey. 1965, Mathematics of Computation, pp. 297-301.

20. *Comparative Study of SVM Methods Combined with Voxel Selection for Object Category Classification on fMRI Data.* Song, Sutao, et al. 2011, PloS one.

21. *Altered Coritical and Subcortical Local Coherence in Obstructive Sleep Apnea: a Functional Magnetic Resonance Imaging Study.* Santarnecchi, Emiliano, et al. 2013, Journal of Sleep Research, pp. 337-347.

22. *Addressing Confounding in Predictive Models with an Application to Neuroimaging.* Linn, Kristin, et al. 2015, The International Journal of Biostatistics.

23. *Attentional Load Modulates Large-Scale Functional Brain Connectivity Beyond the Core Attention Networks.* Alnaes, dag, et al. 2015, Neuroimage, pp. 260-272.

24. *Classifying Minimally Disabled Multiple Sclerosis Patients from Resting State Functional Connectivity.* Richiardi, Jonas, et al. 2012, Neuroimage, pp. 2021-2033.

25. *A Tale of Two Dynamic Cities: Robust Dynamic Functional Connectivity in Healthy Elderly.* Madhyastha, T, S. Merillat, S. Hirsinger, M. Martin, and S. Willis. 2013. In: Presented at the 19th Annual Meeting of the Organization for Human Brain Mapping.

26. *Discriminative Analysis of Parkinson's Disease Based on Whole-Brain Functional Connectivity.* Chen, Yongbin, et al. 2015, PloS one.

27. *Cognitive Network Neuroscience.* Medaglia, John, Mary-ellen Lynall, and Danielle S. Bassett. 2015, Journal of Cognitive Neuroscience.

28. *Modularity and Community Structure in Networks.* Newman, Mark. 2006. Proceeding of the National Academy of Sciences. pp. 8577-8582.

29. *Fast Unfolding of Communities in Large Networks.* Blondel, Vincent, et al. 2008, Journal of Statistical Mechanics: Theory and Experiment.

30. *Region of Interest Analysis for fMRI.* Poldrack, Russell. 2007, Social Cognitive and Affective Neuroscience, pp. 67-70.

31. *A Multi-Voxel-Activity-Based Feature Selection Method for Human Cognitive States Classification by Functional Magnetic Resonacne Imaging Data.* Do, Luu-Ngoc, et al. 2015, Cluster Computing, pp. 199-208.

32. *ICA of Functional MRI Data: An Overview.* Calhoun, Vince, et al. Nara : s.n., 2003. 4th International Symposium on Independent Component Analysis and Blind Signal Separation. pp. 281-299.

33. *EEGLAB: an Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis.* Delorme, Arnaud, and Scott Makeig. 2004, Journal of Neuroscience Methods, pp. 9-21.

34. *Feature Selection Based on Mutual Information of Max-Dependency, Max-Relevance, and Min-Redundancy.* Peng, Hanchuan, Fuhui Long, and Chris Ding. 2005, Pattern Analysis and Machine Intelligence, pp. 1226-1238.

35. *Minimum Redundancy Feature Selection from Microarray Gene expression Data.* Ding, Chris, and Hanchuan Peng. 2005, Journal of Bioinformatics and Computational Biology, pp. 185-205.

36. *Selecting the Maximum Relevant Translated EEG Time-Frequency Features with a Minimum Redundancy Using the Mutual Information Measure for Newborn Seizure Detection and Classification.* Boubchir, Larbi, and Boualem Boashash. Qatar : s.n., 2012. Qatar Foundation Annual Research Forum.

37. Hastie, Trevor, Robert Tibshirani, and J. Friedman. *"The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* s.l. : Springer, 2001.

38. *Comparing EEG/ERP-Like and fMRI-Like Techniques for Reading Machine Thoughts.* Zanzotto, Fabio, and Danilo Croce. 2010, Brain Informatics, pp. 133-144.

39. *fMRI Pattern Classification Using Neuroanatomically Constrained Boosting.* Martinez-Ramon, Manvel, V. Koltchinsii, G. Heileman, and S. Posse. 2006, Neuroimage, pp. 1129-1141.

40. *Proximal Support Vector Machine Classifiers.* Mangasarian, O., and E. Wild. 2001, Knowledge Discovery and Data Mining, pp. 77-86.

41. *Cross-Validatory Choice and Assessment of Statistical Predictions.* Stone, M. 1974, Journal of the Royal Statistical Society: Series B, pp. 1271-1278.

42. *Submodel Selection and Evaluation in Regression. The X-Random Case.* Breiman, Leo, and Philip Spector. 1992, International Statistical Review, pp. 291-319.

43. *Machine Learning Classifiers and fMRI: a Tutorial Overview.* Pereira, Francisco, Tom Mitchell, and Matthew Botvinick. 2009, Neuroimage, pp. 199-209.

44. *The Substatnia Nigra of the Human Brain.* Dameir, P., et al. 1999, Brain, pp. 1421-1436.

45. *Functional Network Organization of the Human Brain.* Power, J., A. Cohen, S. Nelson, A. Wig, K. Barnes, J. Church, A. Vogel, T. Laumann, F. Miezin, B. Schlaggar, and S. Peterson. 2011, Neuron, pp. 665-678.

46. *Effect of Handedness on fMRI Activation in the Medial Temporal Lobe During an Auditory Verbal Memory Task.* Cuzzocreo, Jennifer, et al. 2009, Human Brain Mapping, pp. 1271-1278.

Biographical Information

I am graduating with a Master's degree in Industrial Engineering. I am employed as a manufacturing engineer. I am interested in math, economics, and engineering.