# NEUROIMAGING-BASED STATISTICAL MACHINE LEARNING CLASSIFICATION OF SCHIZOPHRENIA

Kevin Lee Chen

A senior honors thesis submitted to the faculty of the University of North Carolina at Chapel Hill in fulfillment of the requirements for the Honors Carolina Senior Thesis in the Department of Statistics and Operations Research.

Chapel Hill
2018

Approved by:

Eran Dayan, Thesis Advisor

Nicolas Fraiman, Thesis Advisor

Sayan Banerjee

# Abstract

Schizophrenia is a chronic mental disorder that affects millions in the US and tens of millions globally. It is largely believed to be caused by structural and functional differences in the brain, but its exact cause is unknown. Due to the complicated structure of the human brain, its functional connections are often represented by networks. In this thesis, we utilize brain networks generated by functional magnetic resonance imaging (fMRI) data to develop machine learning classification models that can accurately make inferences on single subjects to predict the diagnosis of schizophrenia. We look at a number of local and global connectivity measures derived from correlation-based functional connectivity matrices to do so, using a dataset provided by the National Institute of Health Center of Biomedical Research Excellence (COBRE) and 1000 Functional Connectomes project. Preprocessing and analysis of data is done through the CONN functional connectivity toolbox and MATLAB. Using a subset of subjects and global metrics, we first conduct a preliminary group comparison to determine the existence of a significant difference between patient and control groups with respect to the selected metrics. Then, we investigate machine learning classifiers using *k*-nearest neighbors and support vector machine models on the full dataset using an expanded set of metrics. Using these models, we observe classification accuracy rates of up to approximately 84% on testing sets using 10-fold cross validation, with sensitivity of approximately 91% and specificity of 77% using a polynomial kernel. This rate is fairly consistent with that of other studies, which generally report classification accuracies of 60-90%. As such, the models we have developed demonstrate the potential of networks in determining the nature of schizophrenia and the uses of statistical learning in the diagnosis of neuropsychiatric disorders.

# Acknowledgements

First, I would like to especially thank Dr. Eran Dayan in the UNC School of Medicine Department of Radiology for his continued support and guidance over the last two years as both a research and thesis advisor. He has also given invaluable advice about my development as a student and researcher and has been a crucial resource in my applications for graduate study. His unwavering support in all aspects of my life have made a huge impact on my academic career and personal growth and the extent of my thanks to him cannot be expressed properly with words alone.

I would also like to express my gratitude towards Dr. Nicolas Fraiman in the UNC Department of Statistics and Operations Research for taking on this thesis project with me nearly a year ago and for his continued investment in me as a research student.

# Table of Contents

# Chapter 1

# Introduction and Preliminaries

Schizophrenia is a chronic mental disorder that affects millions in the US and tens of millions globally. While relatively rare in relation to other psychological disorders like phobia and alcohol abuse, schizophrenia has a high disability rating and is known to be significantly disruptive socially (Eaton et al., 2009). Furthermore, treatment is essentially required for patients with schizophrenia in order to avoid psychotic breaks, and in many cases is mandated by law for those exhibiting symptoms of psychosis. In addition to the social and personal burden, there is also a significant economic cost as a result of the disease. Schizophrenia is estimated to cost over $70 billion per year in the United States alone, and the burden continues to grow as the world population increases (Eaton et al., 2009). As with many neuropsychiatric disorders, the exact cause of schizophrenia is unknown, making diagnoses often difficult and complicating treatment processes. Such diagnoses are primarily clinical decisions often based on the ruling-out of other diseases, as the ability of laboratory studies and other testing procedures to make such decisions has not yet been demonstrated (Astrachan et al., 1972). As a result, the development of innovative methods of diagnosing the disorder is highly desired. In addition, these methods could also potentially be used to find prognostic markers to identify the presence of disease in otherwise asymptomatic patients or predict deterioration in previously diagnosed patients.

In order to study the brain, functional magnetic resonance imaging (fMRI) is widely used as a non-invasive imaging technique to measure neuronal activity in different areas in the brain (Ma & Xu, 2016). Over a period of time, scans of the brain are taken with measurements of brain activity, which can then be processed and used in analysis. One such method of analysis is based

in networks, where areas of the brain are represented by nodes and connections are represented by edges. Together, these nodes and edges create what is called a graph network which is extremely useful in analyzing and understanding neuroimaging data.

Networks have extensive uses in modeling complex data, including those in neuroscience. The human brain is frequently modeled as networks depicting either structural or functional connections between various regions of the brain. These networks are crucial in the study of neurological disease such as schizophrenia, where researchers have hypothesized various structural and functional differences in the brains of schizophrenic patients and healthy controls (Demirci et al., 2009; Brown et al., 1986). The usage of networks has the advantage of retaining a richer and more nuanced set of features and information. However, the key difficulty in analyzing brain networks is the determination of useful predictors of disease given datasets of tremendous magnitude. Analysis of datasets with hundreds or thousands of predictors is practically impossible for human analysts, who could take months or years to find useful data that is not readily evident to the human eye. The usage of machine learning offers an approach for detecting subtle and otherwise unapparent patterns in such datasets.

The motivation of this thesis is to explore classification methods using the computational advantage of machine learning in an effort to introduce automation through a data-driven process to accurately determine a diagnosis of schizophrenia in a patient. Our goal is to utilize existing well-studied machine learning methodologies combined with network science to develop new models for brain network classification. Based on data in which category membership is already known, these models can be used to predict the category in which a new observation with unknown membership belongs. We propose that methods of discrimination between schizophrenic and normal brain networks can benefit from the utilization of network metrics as

parameters in machine learning classifiers. Previous neuroimaging studies have typically studied differences between patients and healthy controls at a group level, leaving individual decisions to the professional discretion of the physician (Yue, Li, & Hao, 2003). Machine learning allows us to make inferences at an individual level, which boosts the accuracy and efficiency of diagnostic and prognostic practices in neuropsychiatry. This thesis aims to bring insight into the application of network science in statistical machine learning analysis of neuropsychiatric disorders.

The remainder of Chapter 1 will focus on an introduction to statistical machine learning, *k*-nearest neighbors and support vector machine classification, as well as give a brief background of neuroimaging and brain networks.

## 1.1 Overview of Statistical Machine Learning

Statistical machine learning techniques are widely used in the analysis of datasets which have a much greater number of predictor variables than samples, otherwise known as high-dimensional datasets. Predictor variables can also be referred to as "features" and samples as "observations". In these high-dimensional datasets, machine learning algorithms can filter out unimportant variables ("noise") and determine the best methods for predicting outcomes. This is especially useful in brain networks, where there can be thousands of predictors and millions of connections. Due to the highly connected nature of such networks, many of the observations in fMRI data are highly correlated, and are redundant as a result. This fact combined with the sheer number of predictors leads easily to overfitting, where the model is highly accurate in classifying the given training data but performs poorly on new data. This phenomenon is known as the curse of dimensionality and is a common problem when dealing with high-dimensional datasets. The methods we use to alleviate this issue are explained in Chapter 3.

Two of the most commonly used categories of machine learning are supervised and unsupervised learning. Supervised learning is a subgroup of machine learning methods based on using observed data (i.e., predictor variables) to make predictions of a response variable (e.g., schizophrenia vs. healthy). Unsupervised learning, on the other hand, does not use response variables and uses only the observed data to make inferences on patterns or associations present naturally in the data. While unsupervised learning allows for exploratory analysis of a dataset without prior knowledge or assumptions, supervised learning gives us the ability to measure model performance and make substantive predictions from new data based on the learned optimal model.

### 1.1.1 Support Vector Machines

Our primary focus is on support vector machines (SVM), a type of supervised learning task which uses a known training dataset with both predictor and response values to build a classification model that captures the relationship between the predictors and categorization label. This optimal model can then be tested on a different set of data, known as the test set, under the assumption that a model which truly captures the relationship between the features and responses would also predict accurately on data outside of the training set (Pereira et al., 2009). SVM is a class of discriminative classifiers that separate different groups by a separating hyperplane, which aims to maximize the margin between groups (Vapnik, 1995 as cited in Orrù, Pettersson-Yeo, Marquand, Sartori, & Mechelli, 2012). Figure 1.1 demonstrates a hypothetical example based on two features, where schizophrenic patients are denoted by purple stars and healthy controls are denoted by red stars. Each star represents the fMRI data of a single subject. As seen on the left, each green line presents a possible separating hyperplane for the data.

However, we wish to find a hyperplane that maximizes the margin between the two classes to ensure maximal separation. SVM accomplishes this by focusing on the points closest to the hyperplane, known as support vectors. The right plot demonstrates the usage of support vectors (circled) in SVM, with the solid line being the optimal separating hyperplane and the dotted lines being the resulting margin.

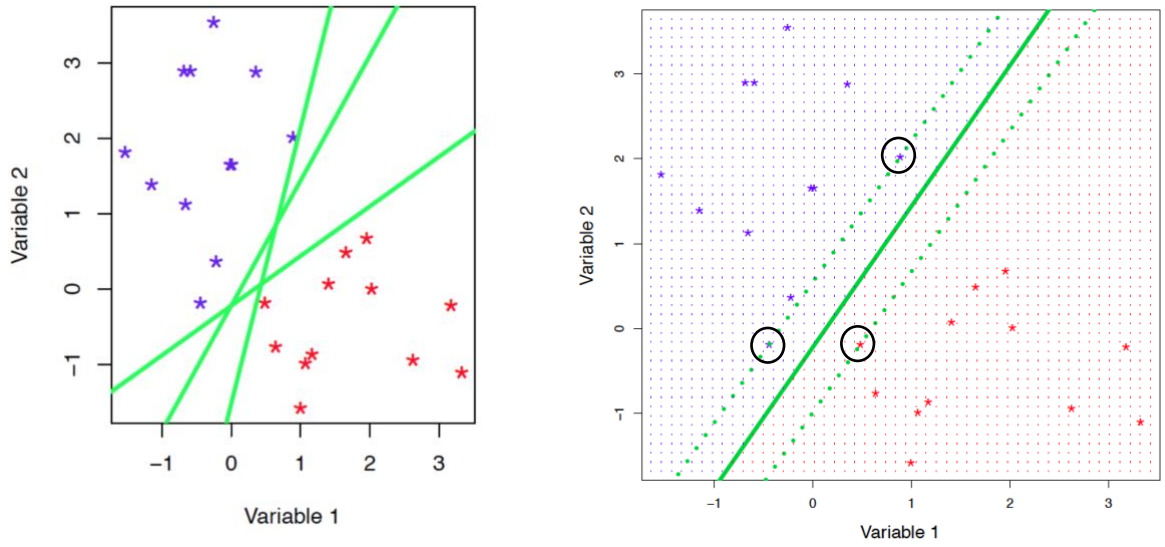


*Figure 1.1. Hypothetical example of an SVM classification problem, with each data point representing fMRI data of either schizophrenic patients (purple stars) or healthy controls (red stars). (Left) Each of the green lines represents a possible separating hyperplane that differentiates the two classes. (Right) The optimal separating hyperplane is represented by the solid green line based on the support vectors, which are circled. The green dotted lines represent the margin maximized by the SVM algorithm. Reprinted from Classification Methods II [PowerPoint slides] by Y. Liu, 2018, Chapel Hill, NC. Retrieved from the University of North Carolina at Chapel Hill STOR 565.*

The example shown in Figure 1.1 is an example of a linear classification problem, where the classes can be separated by a linear hyperplane. Mathematically, the goal in the perfectly separable case is as follows, for a vector *w* and scalar *b*:

$$Minimize\ \frac{1}{2}\|w\|^2 \tag{1}$$

$$such\ that:\ y_i(\langle w, x_i\rangle + b) \geq 1, \qquad i = 1, 2, \ldots, n$$

The width of the margin is given by $\frac{2}{\|w\|}$. In general, however, the classes will not be perfectly separable. In this case, a tuning parameter must be added in order to balance the misclassification rate and margin width. The tuning parameter controls the amount of weight placed on these factors. Thus, in the general case for SVM, the optimal hyperplane is given by:

$$\min_{b,w,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \tag{2}$$

$$such\ that: y_i(x_i^T w + b) + \xi_i \geq 1, \quad \xi_i \geq 0, \ \ i = 1, 2, \ldots, n$$

where *C* is the tuning parameter.

A large value of *C* puts more weight on misclassification rate, meaning that misclassified data receives a higher penalty. A small value of *C* puts more weight on having a large margin, allowing for more misclassification. As the choice of tuning parameters in these models significantly affects their prediction ability and applicability to new observations, we use cross-validation to choose an optimal parameter.

In the case where the data is not linearly separable, we can use kernels to train non-linear boundaries. A kernel is used to transform $x_i$ from the input space into a higher dimensional feature space, where a linear separating hyperplane exists. The idea: what is linear in the feature space is nonlinear in the input space. The kernel trick allows us to create complex decision surfaces in a relatively simple manner. Examples of the usage of nonlinear kernels is demonstrated in Figure 1.2.

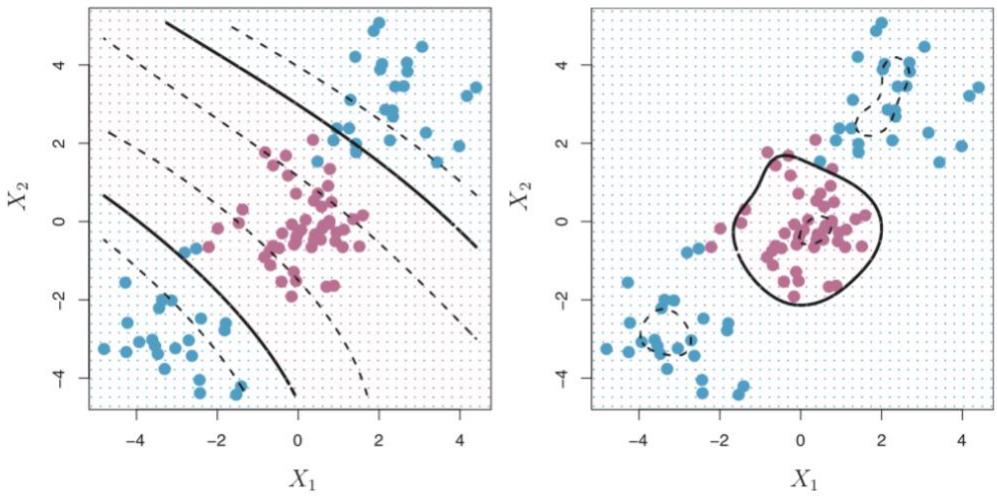


*Figure 1.2. Examples of SVM with nonlinear kernels applied to nonlinear data. (Left) SVM with a polynomial kernel of degree 3. (Right) SVM with a radial basis kernel. Reprinted from An Introduction to Statistical Learning with Applications in R (p. 353), by G. James, D. Witten, T. Hastie, and R. Tibshirani, 2013, New York, NY: Springer. Copyright 2013 by Springer.*

It is important to note that in problems where the number of feature dimensions exceeds the number of samples (as in many machine learning problems), it is always possible to find a linear decision boundary– therefore researchers in neuroimaging problems generally prefer using linear kernels (Orrù et al., 2012). For the sake of comparison, we will look at three commonly-used kernels in Chapter 3– linear, polynomial and radial basis kernels.

### 1.1.2 *k*-Nearest Neighbors

In addition to SVM, we look at the well-studied *k*-nearest neighbors algorithm (KNN), which is a relatively simple nonparametric classification method. A nonparametric method is one that does not make assumptions on the underlying distribution of the data, making it useful when we are unsure if the data follows the theoretical assumptions normally made for other methods. KNN is referred to as a lazy-learning algorithm, meaning that it does not do any learning with the training set; instead, all of the computation is done at the time of classification (James,

Witten, Hastie, & Tibshirani, 2017). Rather than training a model that is then used to make predictions on new observations, KNN makes predictions using the raw training data in a single step. Depending on the value of *k*, which is chosen by cross-validation in practice, KNN determines the *k* training observations that are closest to the test observation. While there are many different ways of determining the closeness of an object, a widely-used measure (and the one that is used in this study) is the Euclidean distance, which is given by:

$$d(a, b) = \sqrt{(a - b)^2}$$

for objects *a* and *b* (Wasule & Sonar, 2017).

The KNN algorithm will then classify the test observation based simply on majority vote of these *k* training observations. An example is shown in Figure 1.3, where the black *x* is a test observation among training observations denoted by colored circles.

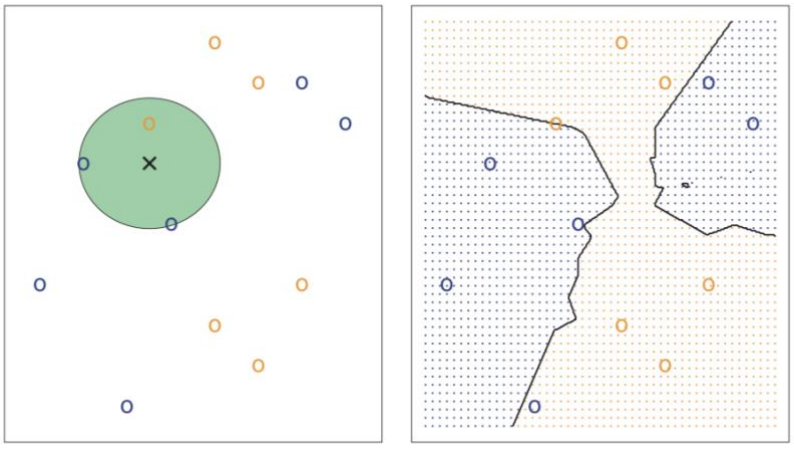

*Figure 1.3. An example of KNN. (Left) A test observation (black x) among training observations (colored circles). The neighborhood of the test observation is shown by the green circle using K=3. (Right) The KNN decision boundary for this example given by the black solid line, with regions of classification shaded in respective colors. Reprinted from An Introduction to Statistical Learning with Applications in R (p. 40), by G. James, D. Witten, T. Hastie, and R. Tibshirani, 2013, New York, NY: Springer. Copyright 2013 by Springer.*

This example uses *K=3* as shown on the left, where the test observation is classified as blue due to there being two blue observations as opposed to one orange observation in its neighborhood. On the right, the decision boundary for KNN is shown. A test observation falling into the orange shaded region would be classified as orange, and an observation falling into the blue shaded region would be classified as orange (James et al., 2017).

## 1.2 Background of fMRI

Functional magnetic resonance imaging (fMRI) is a method of measuring activation in various areas of the brain over time. It is based on the same technology as MRI, but measures fluctuations in cerebral blood oxygenation to detect areas of activity in the brain (Fornito, Zalesky, & Bullmore, 2016). The scanner creates image slices of the brain over time, which creates a 3-dimensional model of the brain that highlights areas of high oxygenation as possible neuronal activity based on the differing magnetic properties of hemoglobin in its oxygenated and deoxygenated states. Well-oxygenated areas (those with increased blood flow) have higher magnetic resonance signal intensities than those with decreased blood flow due to the faster magnetic resonance signal decay rate of deoxyhemoglobin (Cohen & Bookheimer, 1994). Therefore, neural activity can be measured based on magnetic resonance signal intensity under the assumption that increased blood flow corresponds to increased neuronal activity (Kim et al., 1999 as cited in Anderson & Cohen, 2013). This technique is known as blood-oxygen-level dependent (BOLD) contrast imaging and indirectly relates to the level of neuronal activity in the area (Fornito et al., 2016). These measures of neuronal activity are then temporally correlated between regions, leading to the identification of (potentially spatially remote) functionally correlated regions that are modeled as functional brain networks (Patel, Aggarwal, & Gupta,

2016). As such, fMRI can be used in the discovery of regions specific to certain tasks such as language processing or facial recognition and in the study of functional network structures of patients with neuropsychiatric disorders (Bookheimer, 2002; Gauthier et al., 1999; Ford et al., 2003; Anderson et al., 2010 as cited in Anderson & Cohen, 2013).

Historically, fMRI has been used to measure neuronal activity while performing actions, such as squeezing a ball or looking at a visual stimulus. However, in the past few decades, researchers have discovered that by taking fMRI scans while patients are not performing an explicit task, spontaneous fluctuations in oxygenation levels can be measured and applied in the study of various neuropsychiatric disorders, such as Parkinson's disease, attention deficit hyperactivity disorder, Alzheimer's disease, and schizophrenia (Lee et al., 2013). This method is universally applicable given its lack of constraints; however, researchers have found that functional patterns observed during rest are exceptionally robust (Fox & Raichle, 2007; Biswal et al., 2010; Kalcher et al., 2012). Therefore, resting-state fMRI has become a strong candidate for a standard scan protocol in neuroimaging that can be generalized across different studies and environments (Huf et al., 2014).

## 1.3 Functional Brain Networks

Complex brain networks can be modeled using graph theoretical methods. In this study, we will model nodes as various regions of the brain and edges as functional connections between two of these nodes. It is important to note the difference between structural connectivity and functional connectivity– while structural connectivity refers to the physical connection of different brain regions, functional connectivity refers only to a statistical dependency or correlation between brain regions that may or may not have a direct physical connection (Friston,

2011). We use an atlas consisting of cortical and subcortical Regions of Interest (ROIs) from the *Harvard-Oxford Atlas* (Makris et al., 2006; Frazier et al., 2005; Desikan et al., 2006; Goldstein et al., 2007) and cerebellar parcellation from the *AAL Atlas* (Tzourio-Mazoyer et al., 2002) to divide the human brain into 132 regions, each of which are represented by a single node in the network. We model functional connections in correlation-based matrices, which represent all pairwise connections between the nodes that we have defined above. In these matrices, elements depict the correlation between the functional properties of two brain regions, measured by their blood oxygenation levels. Using these matrices, we can calculate a number of other metrics that can be used to characterize the structure of the network in quantitative terms. Furthermore, these metrics can be classified as *local* or *global*, with local measures providing information about individual elements in the network and global measures quantifying the structure of the network as a whole (Braun et al., 2015). In this study, we will first focus on four global measures in our initial exploratory analysis as detailed in Chapter 3: mean node strength, mean clustering coefficient, global efficiency, and characteristic path length. We then expand this analysis in Chapter 4 using a set of 14 global and local metrics, which are used as parameters to build our schizophrenia classification model.

## 1.4 Comparison with Other Studies

SVM and KNN are well-studied and widely used in research relating to the diagnosis of disease. Previous studies on machine learning classification of schizophrenia were able to achieve up to 90% classification accuracy (Anderson & Cohen, 2013). However, these studies often use leave-one-out cross validation, which is known to have high variance in tradeoff with lower bias and thus can lead to unreliable predictions and reproducibility (Efron, 1983 as cited in

Kohavi, 1995). In addition, due to the high cost and difficulty in obtaining fMRI data, such studies often employ extremely small samples ($n \approx 20$) and thus create models that often do not validate across different populations or patient groups (Anderson & Cohen, 2013). Therefore, these results run into issues with reproducibility with criticism that such high classification rates could be due to chance or model mining, where so many models are created that one is bound to suggest an anatomical or functional difference that does not exist in reality (Anderson & Cohen, 2013).

Past methods in the study of schizophrenia have largely been focused on conventional ROI and voxel-based analysis of group differences (SBIA). However, these methods are insufficient for single-subject classification, leading to a push for the development of machine learning methods that can accurately perform such differentiation. Current state-of-the-art classification methods are largely based on pattern analysis of white and gray matter or other structural biomarkers in the brain. Our study aims to further utilize functional neuroimaging techniques and network statistics to develop new methods of classification that alleviate the reproducibility problems found in previous studies. In addition, we investigate the usage of global versus local metrics in an attempt to find meaningful differences that can be capitalized upon in future work. Our goal is to improve upon the prediction accuracy and robustness of classification techniques in schizophrenia by employing the utility of both graph theory and statistical machine learning.

# Chapter 2

# Network Metrics and Exploratory Group Comparisons

In this section, we detail the methodologies used for analysis of the fMRI data, computation of network metrics, and group comparisons using a subset of the resulting metrics. The COBRE dataset includes structural and functional MRI scans in NIfTI format for 72 schizophrenia patients and 74 controls. Subjects were excluded if they were found to have a history of: neurological disorder, mental retardation, severe head trauma with more than 5 minutes loss of consciousness, or substance abuse or dependence within the last 12 months. We preprocess the raw fMRI and anatomical MRI data and obtain functional connectivity correlation-based matrices using the CONN functional connectivity toolbox based on MATLAB. Details of the preprocessing pipeline and calculation of connectivity matrices are given in the following section.

## 2.1 Connectivity Matrices

The CONN functional connectivity toolbox (http://www.nitrc.org/projects/conn), created by members of the Gabrieli Lab at MIT, was used to preprocess, denoise, and obtain functional connectivity matrices for each of the subjects. The preprocessing pipeline consisted of the following: functional realignment and unwarping, functional centering, functional slice-timing correction, functional outlier detection, functional direct segmentation and normalization, structural centering, structural segmentation and normalization, and functional smoothing. These steps are done in order to correct for issues in the raw data such as head movement, discrepancies

caused by the acquisition of different image slices at different times, and detection of outliers in the time course. During this stage, both structural and functional images are centered, segmented into nodes, and normalized. It is also necessary for the structural images of a subject to be aligned with their functional images to properly determine regions of interest. For each subject, CONN produces a two-dimensional square matrix, where the *ij*th matrix element corresponds to the Pearson's correlation between some node *i* and node *j* in the matrix. These matrices are undirected and weighted, with weights ranging from -1 (perfect negative correlation between two time courses) to 1 (perfect positive correlation). Figure 2.1 displays visualizations of two connectivity matrices from a healthy control and a schizophrenic patient.

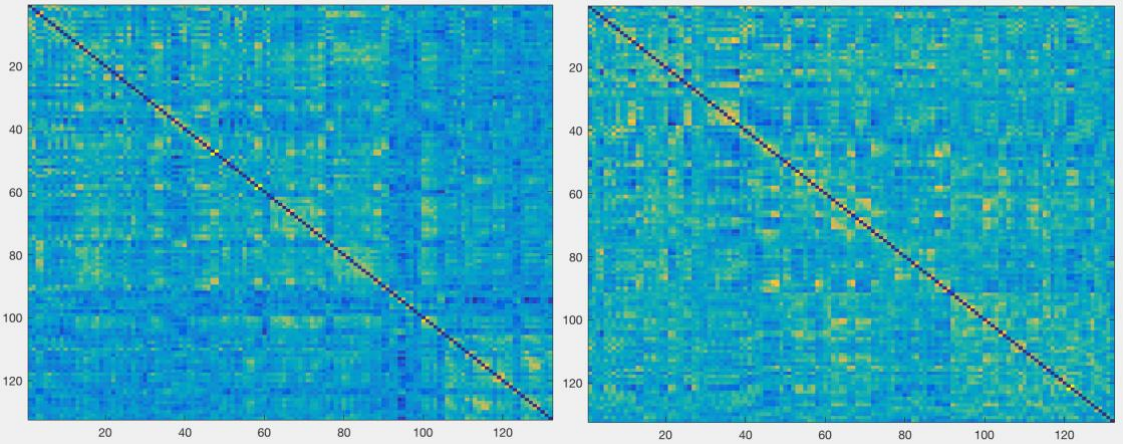


*Figure 2.1. Connectivity matrix visualized for an individual control (left) compared to that of a schizophrenic patient (right). Different colors signify strengths of functional connectivity between two nodes.*

## 2.2 List of Network Metrics

We calculate fourteen different network metrics, including eight local and six global metrics, using the Brain Connectivity Toolbox for MATLAB (Rubinov & Sporns, 2009). The

fourteen metrics used are listed and briefly described below, with items 1-6 being global metrics and items 10-18 being local metrics:

**1. Diameter:** The maximum shortest path in the network– otherwise known as the largest distance between any two nodes in the network (Barabási & Pósfai, 2017). Diameter is equal to the maximum eccentricity, which is explained in number #10 on this list (Hage & Harary, 1995).

**2. Radius:** The minimum eccentricity in the network (refer to #10 on this list).

**3. Global Efficiency:** The average inverse shortest path length observed in the network. Networks with high efficiency may have shorter path lengths, leading to a faster transfer of information. It is defined as:

$$E_{glob} = \frac{\sum_{i \neq j} \epsilon(n_i, n_j)}{N(N-1)}$$

where *N* is the number of nodes, and $n_i$ and $n$ refer to node *i* and node *j*, respectively (Ek et al., 2015).

**4. Characteristic Path Length:** The average shortest path length in the network, and is inversely related to the global efficiency:

$$\lambda = \frac{2 \sum_{i,j} d_{ij}}{N(N-1)}$$

where *N* is the number of nodes and $d_{ij}$ is the shortest path length between node *i* and node *j* (Ioannis & Eleni, 2007).

**5. Transitivity:** The ratio of triangles to connected triplets in the network, where a connected triplet is defined as a single vertex with edges leading to an unordered pair of vertices (Newman, 2003). It can be defined as:

$$C = \frac{3 * number\ of\ triangles\ in\ the\ network}{number\ of\ connected\ triplets\ of\ vertices}$$

In other words, it is the mean probability that two vertices that are individually connected to another vertex will also be connected to each other (Newman, 2003). Transitivity is often used as an alternative to mean clustering coefficient, which is normalized individually and thus may be disproportionally affected by vertices with low degree (Rubinov & Sporns, 2009). Transitivity solves this problem by normalizing vertices collectively (Newman, 2003).

**6. Assortativity Coefficient:** A measure of resilience that is a correlation coefficient between the degrees of all nodes on two opposite ends of a link. A positive assortativity coefficient generally means that nodes will link with nodes that have a similar degree (Rubinov & Sporns, 2009).

**7. Betweenness Centrality:** The fraction of all shortest paths in the network that contain a given node. High values of betweenness centrality indicate that the node participates in a large number of shortest paths in the network (Rubinov & Sporns, 2009). Betweenness centrality is calculated as such:

$$b_i = \frac{1}{(n-1)(n-2)} \sum_{\substack{h,j \in N \\ h \neq j, h \neq i, j \neq i}} \frac{\rho_{hj}(i)}{\rho_{hj}}$$

for a node *i*, where $\rho_{hj}$ is the number of shortest paths between nodes *h* and *j*, and $\rho_{hj}(i)$ is the number of shortest paths between nodes *h* and *j* that pass through node *i*.

**8. Clustering Coefficient:** The fraction of triangles around a node:

$$C_i = \frac{n_i}{k_i(k_i-1)} = \frac{\sum_{j,k} a_{ij} a_{jk} a_{ki}}{k_i(k_i-1)}, k_i \neq 0,1$$

where $\frac{n_i}{2} = \frac{\sum_{j,k} a_{ij} a_{jk} a_{ki}}{2}$ is the actual number of triangles in which node *i* participates, and $\frac{k_i(k_i-1)}{2}$ is the maximum possible number of links in a fully connected subgraph around node *i*

(Ioannis & Eleni, 2007). The clustering coefficient can be used as a measure of connectedness around a node.

**9. Degree:** The number of edges connected to a node:

$$k_i = \sum_{j \in \Pi(i)} a_{ij}$$

where $a_{ij}$ are the elements of the adjacency matrix ($N$x$N$ matrix where the element $a_{ij}$ is 1 if there exists an edge between $i$ and $j$, and 0 if such edge does not exist, for $i,j \in 1, 2, \ldots, N$), and $\Pi(i)$ is the neighborhood of the node $i$ (Ioannis & Eleni, 2007).

**10. Eccentricity:** The maximal shortest path length between some node $i$ and all other nodes in the network.

**11. Eigenvector Centrality:** A centrality measure of the importance of a node in the network. High eigenvector centrality means that a node is connected to many other nodes with high eigenvector centrality, and is determined as such:

Each node $i$ is assigned a weight $x_i$ that is proportional to the sum of weights of all nodes connected to node $i$, such that $\boldsymbol{A}x = \lambda x$ for $\lambda > 0$, where $\boldsymbol{A}$ is the adjacency matrix and $x$ is the vector of $x_i$ values.

The result is that higher values of eigenvector centrality will correspond to higher influence in the network.

**12. Local Efficiency:** The global efficiency computed on node neighborhoods (subgraphs around a node). Local efficiency is related to the network's connectedness at a local level and is computed by:

$$E_{loc} = \frac{1}{N} \sum_{i \epsilon G} E_{glob}(G_i)$$

where $G_i$ is the subgraph composed of the neighbors of node *i*. It is clear from this formula that the local efficiency is equivalent to the average of the global efficiencies of the subgraphs around node *i* (Strang, Haynes, Cahill, & Narayan, 2017).

**13. Modularity:** A measure of the degree to which a network is divided into a community structure. Rubinov & Sporns defines the dense interconnectedness of groups in a network as a community structure in which the network is subdivided into nonoverlapping groups of nodes such that the number of within-group links is maximized and the number of between-group links is minimized (2009). As such, many networks with high modularity will have small-world properties; that is, the strong within-group connectivity results in high clustering coefficient while weak within-group connectivity results in low characteristic path length (Fornito et al., 2016). An example of such a modular network is shown in Figure 2.2, where different groups are represented by different colors. Strong within-group connections are represented by colored edges, while weak between-group links are represented by gray-colored edges.

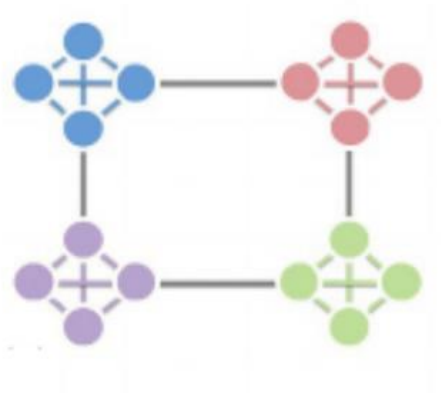

*Figure 2.2. An example of a modular network, with each "module" represented by a different color. Within groups there is high connectivity (represented by colored edges) while there are fewer long-range links between groups (represented by gray edges). Reprinted from Fundamentals of Brain Analysis (p. 304), by A. Fornito, A. Zalesky, and E. Bullmore, 2016: Copyright 2016 by Elsevier Inc.*

**14. Node Strength:** The sum of the weights of links connected to a node *i*, given by:

$$s_i = \sum_{j \in \Pi(i)} w_{ij}$$

where $w_{ij}$ is weight of the edge between nodes $i$ and $j$ and is equal to 0 if the two nodes are not connected (Ioannis & Eleni, 2007).

## 2.3 fMRI Analysis and Preliminary Group Comparisons

For ease of computation in the preliminary analysis portion of the study, we first select a subset of the COBRE dataset– 30 schizophrenia patients and 30 controls were selected by simple random sample from the full dataset of 72 schizophrenia patients and 74 controls. The connectivity matrices obtained from CONN are concatenated into two separate 132x132x30 matrices, one for schizophrenia patients and one for controls. These matrices are transformed by an inverse Fisher's z-transformation from Pearson's r-values to normally-distributed z-values. As CONN automatically fills diagonal elements with NaN (not-a-number) values, we then convert all NaN values to 1 in order to denote perfect correlation between a node and itself. A threshold of a minuscule non-zero number is then applied, removing negative and zero weights. From the Brain Connectivity Toolbox, we use a subset of four measures from the original fourteen metrics calculated, summing the node strengths to compute a global network strength and averaging the clustering coefficients to compute a global average clustering coefficient. Observed differences between patients and controls are then calculated and stored for each measure.

We then perform a permutation test in order to build a sampling distribution based on our null hypothesis of no difference between the patient and control groups. By resampling the observed data, we do not need to make assumptions about the data distribution as in a normal t-test. To perform the permutation tests, we randomly divide the 60 subjects into two groups. We

then calculate the means of each randomly assigned group for each network measure, and their difference is stored in a separate array. This process is repeated 10,000 times to ensure we sample a null distribution. We then compare our observed differences with that of the resampled data.

## 2.4 Group Comparison Results

Initial results show slightly higher mean clustering coefficients, characteristic path lengths, global efficiency, and network strengths for the schizophrenia patients than the healthy controls (results shown in Figure 2.3). Figure 2.4 shows the plots of the differences in each measure.

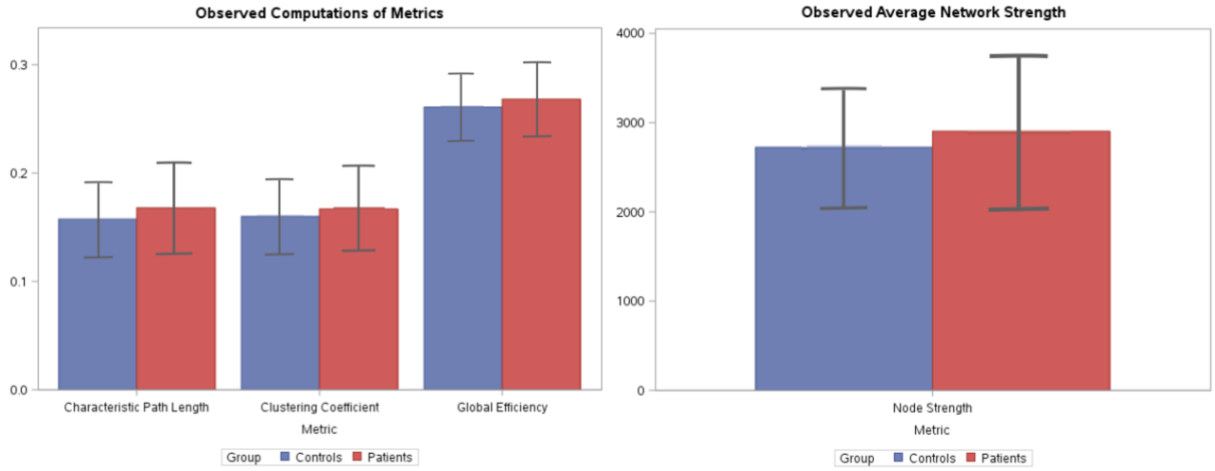


*Figure 2.3. Observed mean values of measures with standard errors. Controls depicted in blue, patients depicted in red. Units for node strength are given in weight of links. All other units are arbitrary.*

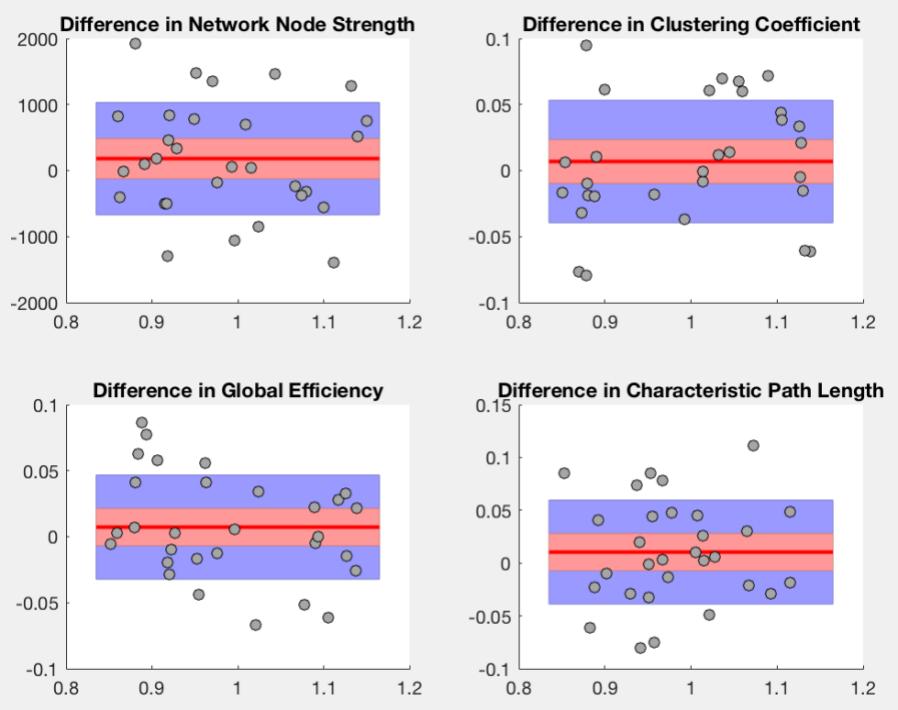


*Figure 2.4. Plots of the differences in each measure between schizophrenic patients and healthy controls (controls subtracted from schizophrenic patients). Red lines show the mean difference, with the red highlight showing 1 standard deviation and blue highlight showing 95% confidence intervals for the mean difference. Gray dots represent individual values.*

The plots above graphically depict the mean differences in addition to a 95% confidence interval for each measure. However, the permutation tests show that this difference is *not* significant at the $p < 0.05$ level. Figure 2.5 displays the null distributions sampled by the permutation tests along with the mean observed differences (marked by a red line on the graph). The percentages of samples from the permutation test that saw smaller mean differences than our observed values are shown in Table 2.1 as follows:

*Table 2.1. Percentages of samples with smaller mean differences than observed values from permutation tests.*

| Measure | Percent |
|---|---|
| Node Strength | 87.04% |
| Clustering Coefficient | 77.84% |
| Global Efficiency | 83.23% |
| Char. Path Length | 87.32% |

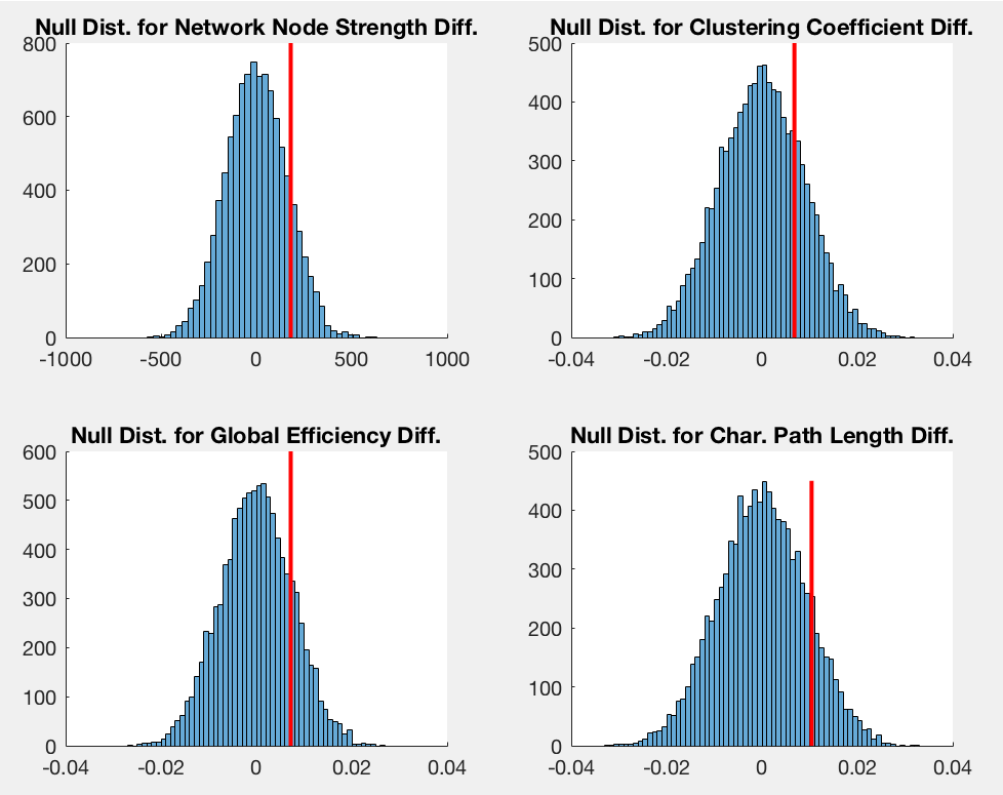


*Figure 2.5. Null distributions of mean metric differences, with observed mean differences marked by red lines.*

Clearly, the metrics used in this analysis cannot account for the difference between the networks in schizophrenic and non-schizophrenic brains. From these results, we can now expand our efforts to working with all subjects and full set of graph-theoretic metrics. The next chapter details the methods used for statistical learning and model-building for classification using this larger dataset.

# Chapter 3

# Machine Learning Classification

In this chapter, we describe the methods of creating classification models using our previously calculated network data. SVM is decided as the primary technique of choice for this study due to its usefulness in high-dimensional spaces (as is necessary with brain networks), as a result of its ability to control model complexity independently of data dimensionality (Clarke et al., 2008). As such, support vector machines are generally less susceptible to the curse of dimensionality. We also utilize the nonparametric *k*-nearest neighbors algorithm as a comparison point. The full COBRE dataset includes anatomical and functional MRI data with 72 schizophrenia patients and 74 healthy controls. We perform classification with four general sets of features: using only global metrics, using only local metrics (averaged across regions in order to achieve fair comparison with global metrics), and using all metrics (global combined with local) without feature selection, and using all metrics with feature selection. All data subsets are centered and divided into training and testing sets of 70 and 30 percent of the full dataset, respectively. Each model is trained using 10-fold cross validation to choose optimal tuning parameters and the classifier is built with linear, polynomial, and radial basis kernels. Predictions are then made on the testing set using each model, and confusion matrices of predicted and observed classifications are created, allowing us to calculate accuracy rate, sensitivity, and specificity values. We primarily use the 'caret' package in R to build the classifiers and assess model performance. The general procedure for classification is outlined in Figure 3.1.

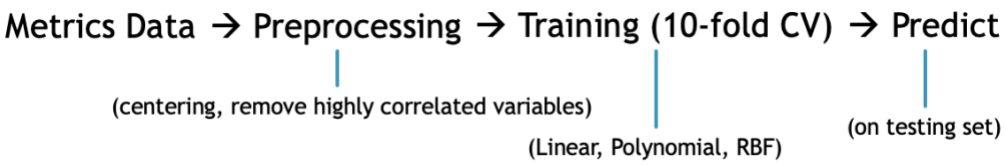


*Figure 3.1. Outline of general classification procedure. Note that the 146 observations are divided into training and testing sets of 102 and 44 observations (approximately 70/30 split), respectively.*

## 3.1 KNN & SVM Classification on Global vs. Local Feature Sets

In this section, we conduct a brief comparison of the efficacy of using global vs. local measures in classification. We perform classification with KNN and SVM models using all three aforementioned kernels on the two datasets and look at the difference in prediction accuracy. In order to achieve relative fairness with respect to number of features, we average local metrics across the 132 brain regions so there is only one averaged measure for each local metric per subject.

### 3.1.1 Global Feature Set

To begin, we use only global metrics as features. In this case as well as with averaged local metrics, it is important to note that the number of samples is significantly greater than the number of features. Therefore, we do not perform feature selection for these data subsets as the risk of overfitting is reduced. Using only global metrics as features, we find that SVM with a linear kernel achieves the best classification accuracy of the four models, although the rate is considerably lower than ideal at 61.4%. Table 3.1 shows the classification accuracy, 95% confidence interval for the accuracy rate, sensitivity, and specificity of the global metrics-only model. Figure 3.2 shows variable importance scores for each of the variables used in the SVM with linear kernel model. As SVM does not have a built-in importance score in the 'caret'

package, the importance is obtained by calculating the area under the Receiver Operating Characteristic (ROC) curve, which is a graph that measures classification performance by plotting true positive rate (i.e., sensitivity) against false positive rate (i.e., 1 – specificity).

*Table 3.1. Performance characterization for models using the data subset consisting of only global metrics. Shown are classification accuracies, 95% confidence intervals for these accuracy rates, sensitivity, and specificity values based on predictions made on the testing set. SVM with linear kernel is shown to be the model with highest accuracy.*

| | Accuracy | 95% CI | Sensitivity | Specificity |
|---|---|---|---|---|
| K-Nearest Neighbors | 52.3% | (36.7%, 67.5%) | 59.1% | 45.5% |
| SVM w/ Linear Kernel | 61.4% | (45.5%, 75.6%) | 59.1% | 63.6% |
| SVM w/ Polynomial Kernel | 59.1% | (43.3%, 73.7%) | 63.6% | 54.6% |
| SVM w/ Radial Basis Kernel | 47.7% | (32.5%, 63.3%) | 36.4% | 59.1% |

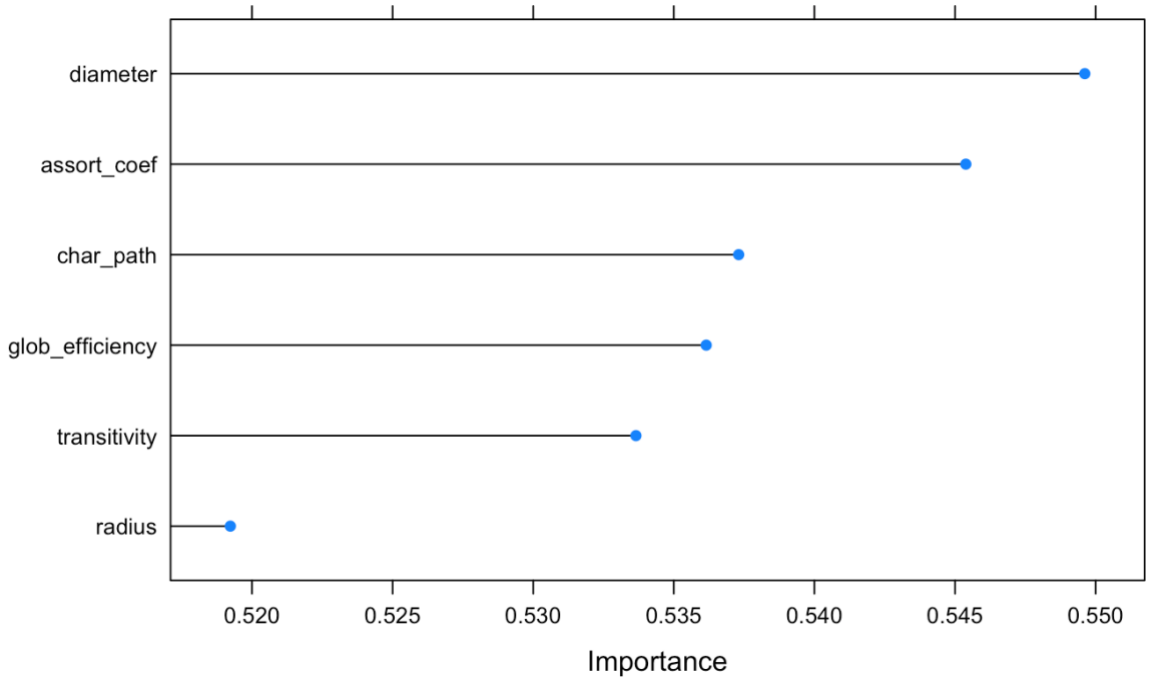


*Figure 3.2. Variable importance plot of SVM classifier with linear kernel for the global metric feature set.*

### 3.1.2 Local Feature Set

We now build the same classifiers using only the subset of local metrics. In order to compare with our global metric classifier, we compute averages of each local metric across regions to create a network metric. For example, we originally have local efficiency values for each of the 132 brain regions previously parcellated. In order to obtain a "global" metric, we calculate the average of the local efficiency over the 132 regions. This process is repeated for each of the eight local metrics. As with the global metric feature set in section 3.1.1, we do not perform feature selection due to the limited number of features as compared with the number of observations. Table 3.2 shows the classification performance measures for this feature set. SVM with linear kernel also has the highest performance in this case with an accuracy rate of 59.1%. Figure 3.3 is the variable importance plot for the local features, again using the SVM model with linear kernel.

*Table 3.2. Performance characterization for models using the data subset consisting of only averaged local metrics. Shown are classification accuracies, 95% confidence intervals for these accuracy rates, sensitivity, and specificity values based on predictions made on the testing set. SVM with linear kernel is shown to be the model with highest accuracy.*

| | Accuracy | 95% CI | Sensitivity | Specificity |
|---|---|---|---|---|
| K-Nearest Neighbors | 56.8% | (41.0%, 71.7%) | 63.6% | 50.0% |
| SVM w/ Linear Kernel | 59.1% | (43.3%, 73.7%) | 68.2% | 50.0% |
| SVM w/ Polynomial Kernel | 47.7% | (32.5%, 63.3%) | 40.9% | 54.6% |
| SVM w/ Radial Basis Kernel | 50.0% | (32.5%, 65.4%) | 100% | 0% |

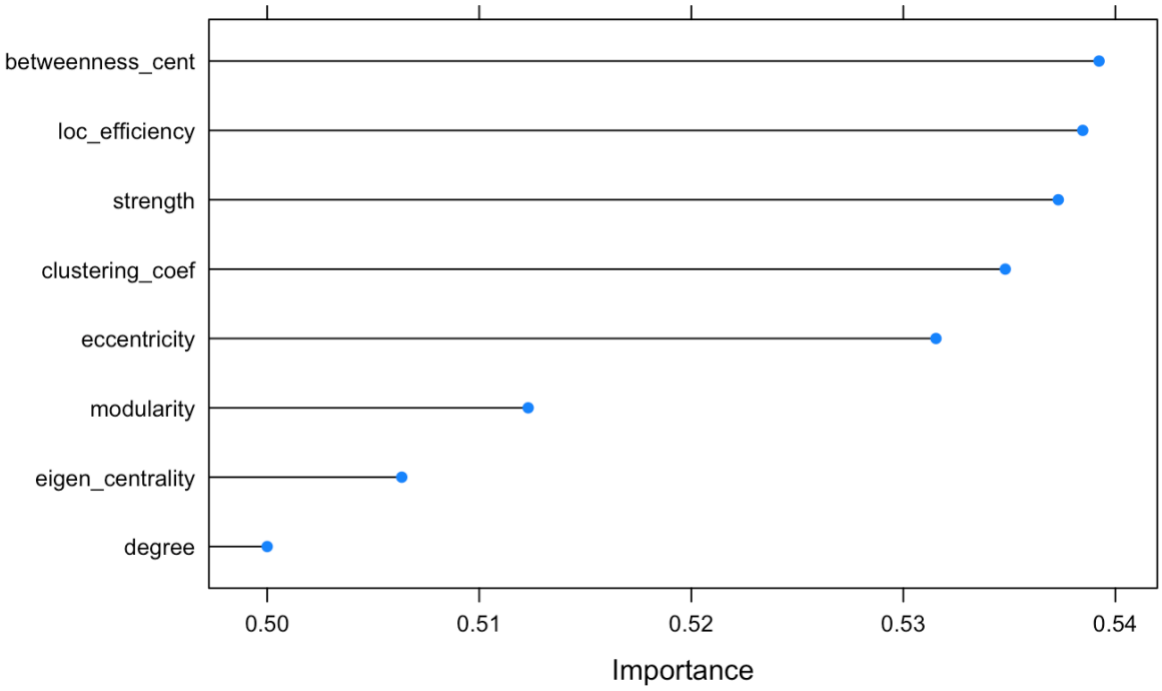


*Figure 3.3. Variable importance plot of SVM classifier with linear kernel for the averaged local metric feature set.*

While somewhat lower than in the global metric feature set, the difference in classification accuracy is not compelling enough to arrive at a conclusion about the nature of global versus local metrics for any of the classifiers. In the next section, we use the full feature set to perform classification in order to improve accuracy.

## 3.2 KNN & SVM Classification on Full Feature Set

A brief sample of the dataset using the full feature set is shown in Figure 3.4, separated into the matrix of predictors and the response vector.

| assort_coef | char_path | diameter | glob_efficier | radius | transitivity |
|---|---|---|---|---|---|
| 0.19515 | 0.1463 | 1.5063 | 0.26593 | 0.20192 | 0.18249 |
| 0.047961 | 0.25816 | 1.7483 | 0.4142 | 0.65725 | 0.35523 |
| 0.078552 | 0.14611 | 1.4612 | 0.2731 | 0.48119 | 0.17565 |
| 0.20298 | 0.13331 | 1.605 | 0.26461 | 0.41272 | 0.15744 |
| 0.10467 | 0.13163 | 1.5859 | 0.24973 | 0.30796 | 0.16084 |
| 0.050542 | 0.12968 | 1.7068 | 0.25152 | 0.45209 | 0.15091 |
| 0.13676 | 0.15341 | 1.8206 | 0.28078 | 0.43811 | 0.1872 |
| 0.19571 | 0.15059 | 1.7468 | 0.2815 | 0.39691 | 0.18465 |
| 0.073525 | 0.17913 | 1.66 | 0.31313 | 0.45226 | 0.22769 |

• • •

146x1194

| schizo |
|---|
| 0 |
| 0 |
| 0 |
| 0 |
| 0 |
| 0 |
| 0 |
| 0 |
| 0 |

146x1

*Figure 3.4. Sample of dataset with all features (pre-feature selection). There are 146 observations in total, with the predictor matrix on the left and response vector on the right. Elements of the response vector are separated into two classes: 1 for schizophrenia patient and 0 for healthy control.*

### 3.2.1 Without Feature Selection

Initially, no feature selection is performed. We find that with this dataset consisting of more information, the classification accuracy rate increases significantly across all models into the 65-80% range with sensitivity and specificity increasing to 63-82%. The best model selected based on accuracy rate for this dataset is SVM with radial basis kernel with a classification accuracy of 79.6%. Table 3.3 shows the full performance statistics along with a 95% confidence interval of the accuracy rate.

*Table 3.3. Performance characterization for models using the dataset of all features without feature selection. Shown are classification accuracies, 95% confidence intervals for these accuracy rates, sensitivity, and specificity values based on predictions made on the testing set. SVM with radial basis kernel is shown to be the model with highest accuracy.*

| | Accuracy | 95% CI | Sensitivity | Specificity |
|---|---|---|---|---|
| K-Nearest Neighbors | 65.9% | (50.1%, 79.5%) | 63.6% | 68.2% |
| SVM w/ Linear Kernel | 68.2% | (52.4%, 81.4%) | 68.2% | 68.2% |
| SVM w/ Polynomial Kernel | 72.7% | (57.2%, 85.0%) | 77.3% | 68.2% |
| SVM w/ Radial Basis Kernel | 79.6% | (64.7%, 90.2%) | 81.8% | 77.3% |

Figure 3.5 shows the variable importance plot for this feature set of the SVM with radial basis kernel model, showing only the top 10 most important predictors. It is important to note the difficulty of interpretation of this plot– for example, **degree_62** and **degree_64** both refer to the degree of divisons of the left parahippocampal gyrus, but what does this mean in terms of the graph structure as a whole (Whitfield-Gabrieli & Nieto-Castanon, 2012)? The parahippocampal gyrus has been implicated as a possible factor in schizophrenia by Brown et al., but further investigation would be necessary to make stronger claims (1986).

We can observe, however, that **eccentricity_119**, **strength_119**, and **eigen_centrality_119** are all in the top 10 important predictors, representing the eccentricity, node strength, and eigenvector centrality of the right cerebellum 8 region, respectively (Whitfield-Gabrieli & Nieto-Castanon, 2012). This could indicate some significance about this brain region, which is corroborated in Andreasen & Pierson's study of cerebellar anomalies in schizophrenia, which speaks to the growing evidence for the influence of the cerebellum in the presence of the disorder (2008).

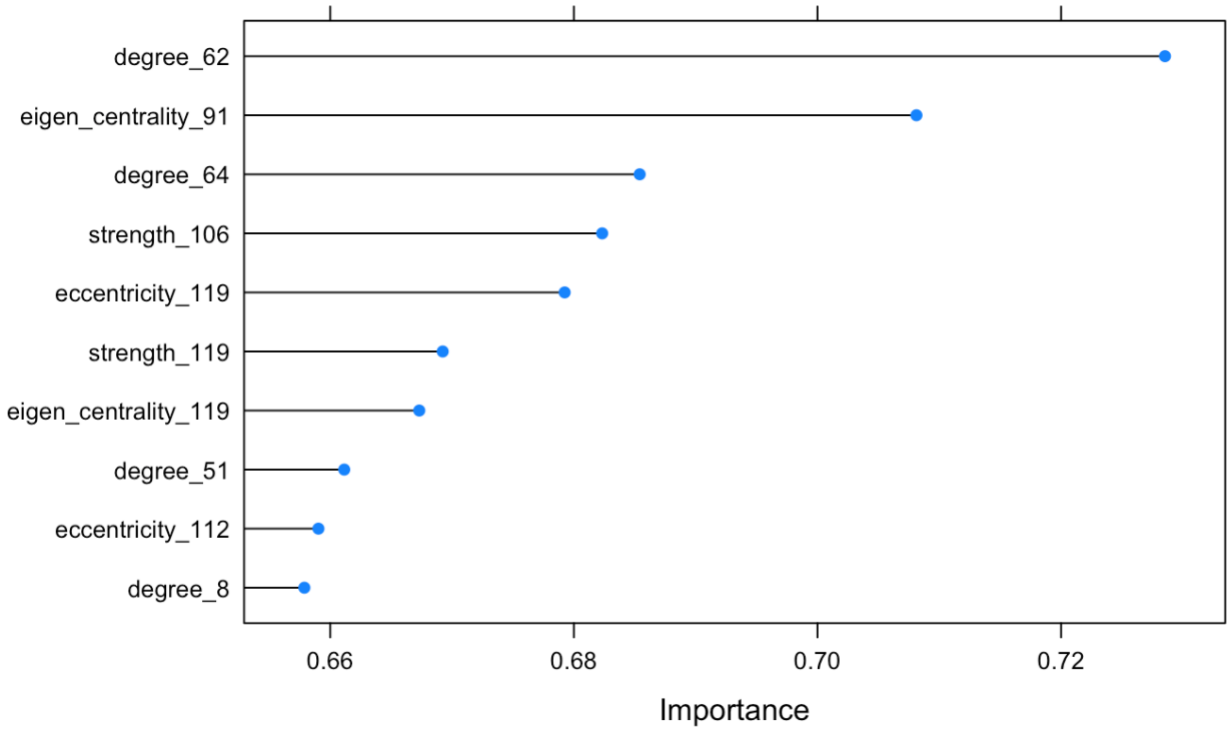


*Figure 3.5. Variable importance plot of SVM with radial basis kernel on full feature set. Note that indexing begins at 0, therefore for example,* ***eccentricity_112*** *refers to the eccentricity of the 113th node (or brain region). In context, the 113th node refers to the left cerebellum 4 5 left region (Whitfield-Gabrieli & Nieto-Castanon, 2012).*

### 3.2.2 With Feature Selection

We now perform feature selection under the assumption that the removal of redundant features will reduce overfitting and improve overall model accuracy. We first compute a correlation matrix from the predictor matrix and consider the absolute pairwise correlations between predictors. If two predictors are found to be highly correlated (using a set cutoff of 0.75), the variable with the largest mean absolute correlation is removed. The resulting dataset includes 567 remaining predictors after removing 627 highly correlated variables. Results of model creation and testing are shown in Table 3.4.

*Table 3.4. Performance characterization for models using the dataset of all features with feature selection (removal of highly correlated variables). Shown are classification accuracies, 95% confidence intervals for these accuracy rates, sensitivity, and specificity values based on predictions made on the testing set. SVM with polynomial kernel is shown to have highest classification accuracy.*

| | Accuracy | 95% CI | Sensitivity | Specificity |
|---|---|---|---|---|
| K-Nearest Neighbors | 68.2% | (52.4%, 81.4%) | 81.8% | 54.6% |
| SVM w/ Linear Kernel | 77.3% | (62.2%, 88.5%) | 72.7% | 81.8% |
| SVM w/ Polynomial Kernel | 84.1% | (69.9%, 93.4%) | 90.9% | 77.3% |
| SVM w/ Radial Basis Kernel | 77.3% | (62.2%, 88.5%) | 95.5% | 59.1% |

Overall, classification accuracy improves or remains constant with the exception of the SVM with radial basis kernel classifier. SVM with linear and polynomial kernels as well as the KNN classifier make fairly significant improvements with feature selection. The SVM with polynomial kernel achieves the highest accuracy of the models created with a classification accuracy rate of 84.1% with sensitivity of 90.9% and specificity of 77.3% on this dataset. The variable importance plot for SVM with polynomial kernel is shown in Figure 3.6, which is noted to be the same as the variable importance plot for the feature set without feature selection. This is not an unexpected result as it is reasonable to assume that feature selection would not change the importance of the most useful predictors.

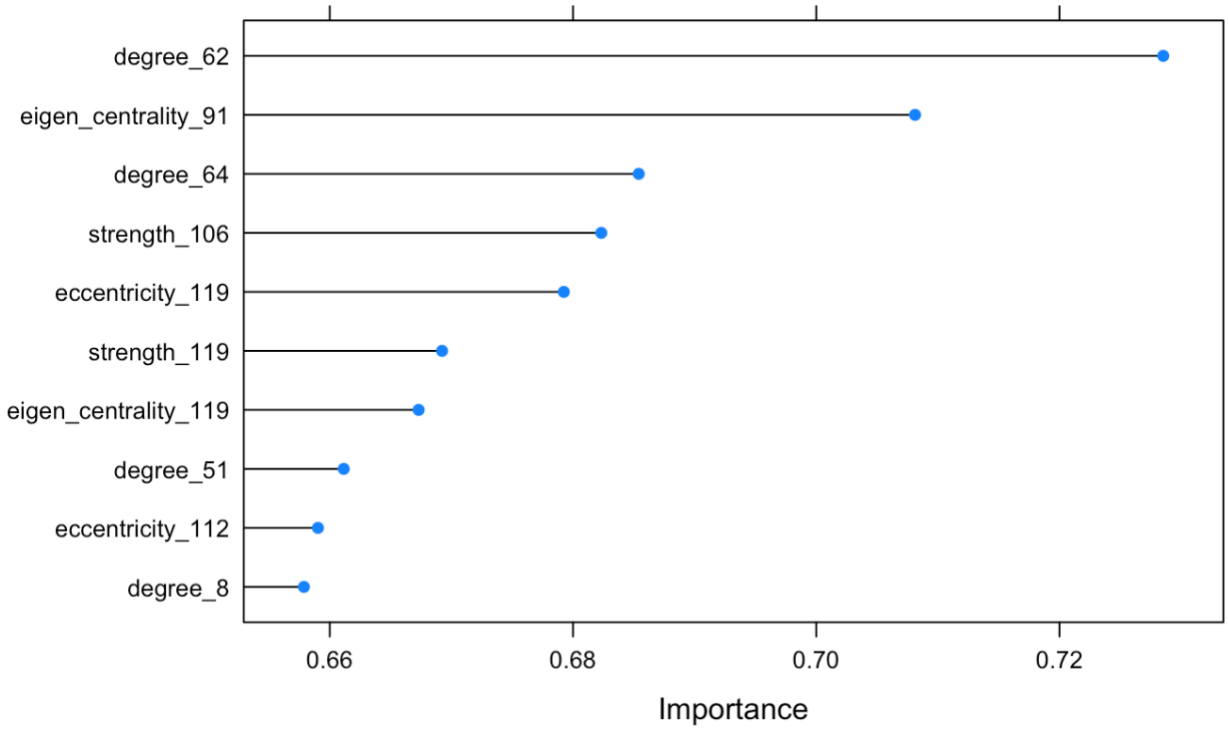


*Figure 3.6. Variable importance plot of most important variables using SVM with polynomial kernel on feature set with highly correlated variables removed. Note that this gives the same plot as in the feature set without feature selection.*

## 3.3 Classification Results

Table 3.5 gives a summary of the models with highest classification accuracy for each feature set. The highest accuracy obtained is SVM with polynomial kernel using the combined feature set with feature selection (i.e., removing highly correlated variables), which is bolded in Table 3.5. A one-sample t-test finds that a significant difference in both the SVM with radial basis kernel model on the full feature set and the SVM with polynomial kernel on the feature-selected feature set compared with random chance.

*Table 3.5. Summary of results. The best classifier for each feature set is given based on testing classification accuracy. One-sample t-tests were also performed for each model and these p-values are given in the last column. The model with highest accuracy is bolded.*

| Feature Set | Model (Kernel) | Accuracy | Sensitivity | Specificity | Significance |
|---|---|---|---|---|---|
| Global Metrics | SVM (linear) | 61.4% | 59.1% | 63.6% | .0871 |
| Local Metrics | SVM (linear) | 59.1% | 68.2% | 50.0% | .2257 |
| All Metrics, no feature sel. | SVM (radial) | 79.6% | 88.7% | 77.3% | .0002*** |
| **All Metrics w/ feature sel.** | **SVM (polynomial)** | **84.1%** | **90.9%** | **77.3%** | **<.0001*** ** |

* $p$ <0.05
** $p$ <0.01
*** $p$ <0.001

The bolded model in Table 3.5 achieves an 84.1% classification accuracy, with cross validation choosing a polynomial kernel of degree 3 and tuning parameter values of 0.001 and 0.5 for gamma and C (regularization parameter of the error), respectively. Compared to random chance, these results are statistically significant with $p < .0001$. The accuracy rates are consistent with, if not somewhat more compelling, than that of previous schizophrenia machine learning classification studies, which generally report classification accuracies of 70-80%, with some studies reporting figures as low as 60% or as high as 90%. These results are discussed further in Chapter 4.

# Chapter 4

# Discussion and Conclusion

In this study, we have discussed a number of statistical machine learning methods designed to accurately predict the presence of schizophrenia in a patient through the analysis of functional magnetic resonance imaging data. We proposed models using the well-established KNN and SVM algorithms on multiple feature subsets based on graph theoretical network measures. These network measures are well-studied and include information about node strength, eccentricity, modularity, path length, etc. They can be grouped into two general categories: global metrics, which provide information about the network as a whole, and local metrics, which provide information on the node level. The models were fit on training sets using four feature subsets: global metrics, averaged local metrics, combined metrics without feature selection, and combined metrics with highly-correlated variables removed. The combined feature subsets contained both global metrics and metrics calculated on a node level, resulting in a large feature space. Therefore, fitting is done both on the full feature space as well as a reduced feature space.

The efficacy of classifiers trained on global metrics and those trained on averaged local metrics was compared, finding no convincing evidence that one classifier was more accurate than the other. We then fit models on the combined feature sets with an overall improvement in classification results. Thus, we find that the increase in information on a node-level has a positive impact on the discriminatory power of our SVM classifier. On the feature-selected dataset, the testing accuracy of our parameter-tuned SVM classifier using a polynomial kernel is 84.1%, which is found to be significant when compared to random chance ($p < .0001$). A

comparison between the results of this study and that of previous studies on classification of schizophrenia using support vector machine classifiers is shown in Table 4.1.

*Table 4.1. Comparison of studies on classification of schizophrenia using support vector machine classifiers. Information on inputs is given along with sample size and classification accuracy results. Findings of the optimal model in this thesis are in bold. * Studies using COBRE dataset.*

| Authors | Parameters (Inputs) | Sample Size | Accuracy |
|---|---|---|---|
| Gould et al., 2014 | Gray- and white-matter volume data | Patients: 126<br>Controls: 134 | 72% |
| Iwabuchi et al., 2013 | Gray- and white-matter volume data | Patients: 19<br>Controls: 20 | 77% |
| Schnack et al, 2014 | Gray-matter volume data | Patients: 66<br>Controls: 66 | 90% |
| Yang et al., 2010 | Single neucleotide polymorphism (SNP) data<br>fMRI voxel data<br>SNP-Voxel combination | Patients: 20<br>Controls: 20 | 74%<br>82%<br>87 |
| Anderson & Cohen, 2013* | Functional connectivity measures | Patients: 72<br>Controls: 74 | 65% |
| Savio & Graa, 2015* | Functional connectivity measures | Patients: 72<br>Controls: 74 | 80% |
| Chyzhyk, Savio, & Graa, 2015* | Functional connectivity measures | Patients: 72<br>Controls: 74 | 91% |
| Hsieh, Sun, & Liang, 2014* | Functional connectivity measures | Patients: 69<br>Controls: 72 | 72% |
| Zeng et al., 2018* | Functional connectivity measures | Patients: 71<br>Controls: 74 | 73% |
| **Proposed Method*** | Functional connectivity measures | Patients: 72<br>Controls: 74 | 84% |

While the results of this study were largely on-par with previous studies on the subject and with current state-of-the-art methods, there are a number of limitations worth addressing. First, the COBRE dataset used was not large enough for complete out-of-sample testing. Therefore, future work could include the collection of out-of-sample data to verify the results obtained in this study. In addition, overfitting is still a concern given the number of features remaining after feature selection. Further feature selection not limited to correlated variables should be performed with feature extraction as an additional consideration. Another limitation

stems from the current gold standard for schizophrenia diagnosis, which is by clinical interview. The models created in this study do not necessarily add simple usage for physicians given the preprocessing and computation required. Therefore, the added benefit for a doctor attempting to diagnose the disease may be limited.

Future directions will also be aimed at maximizing information usage in an effort to improve the classifier. While we used a set of 14 global and local metrics as provided by the Brain Connectivity Toolbox for MATLAB, there is no shortage of additional measures inspired by graph theory that could be used as predictors for a model. We also plan to conduct more study into statistical methods using the raw graphs provided by CONN rather than summary statistics in order to determine whether there is information that is lost in the computation of network metrics. These methods include centroid-based clustering and may provide more differentiation power as compared to the methods developed in this thesis, given evidence from Chapter 3 that suggests a possible loss of useful data as a result of metric summarization.

From this thesis, we have shown the likely presence of differences in resting-state functional connectivity between schizophrenic patients and normal brains. In addition, network measures have proven useful in the discrimination of these groups, which has been demonstrated in the models created in this study using statistical machine learning methods. With further study, machine learning and graph theoretical methods can be utilized to improve classification accuracy and contribute to automated diagnosis of schizophrenia and potential development of prognostic markers for at-risk populations.

# References


Anderson, A., & Cohen, M. S. (2013). Decreased small-world functional network connectivity and clustering across resting state networks in schizophrenia: An fMRI classification tutorial. *Frontiers in Human Neuroscience, 7*. doi:10.3389/fnhum.2013.00520

Andreasen, N. C., & Pierson, R. (2008). The Role of the Cerebellum in Schizophrenia. *Biological Psychiatry, 64*(2), 81-88. doi:10.1016/j.biopsych.2008.01.003

Astrachan, B.M., Harrow, M., Adler, D., Brauer, L., Schwartz, A., Schwartz, C., & Tucker, G. (1972). A Checklist for the Diagnosis of Schizophrenia. *British Journal of Psychiatry, 121*(564), 529-539. doi:10.1192/bjp.121.5.529

Barabási, A., & Pósfai, M. (2017). *Network science*. Cambridge: Cambridge University Press.

Biswal B.B., Mennes M., Zuo X.N., Gohel S., Kelly C., Smith S.M., et al. (2010). Toward discovery science of human brain function. *Proc Natl Acad Sci USA. 107*:4734–9. 10.1073/pnas.0911855107

Braun, U., Muldoon, S. F., & Bassett, D. S. (2015). On human brain networks in health and disease. *eLS*.

Brown R., Colter N., Corsellis J.A.N., et al. (1986). Postmortem Evidence of Structural Brain Changes in Schizophrenia:Differences in Brain Weight, Temporal Horn Area, and Parahippocampal Gyrus Compared With Affective Disorder. *Arch Gen Psychiatry, 43*(1):36–42. doi:10.1001/archpsyc.1986.01800010038005

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience*, *10*(3), 186.

Chyzhyk, D., Savio, A., & Graña, M. (2015). Computer aided diagnosis of schizophrenia on resting state fMRI data by ensembles of ELM. *Neural Networks, 68*, 23-33. doi:10.1016/j.neunet.2015.04.002

Clarke, R., Ressom, H.W., Wang, A., Xuan, J., Liu, M.C., Gehan, E.A., & Wang, Y. (2008). The properties of high-dimensional data spaces: Implications for exploring gene and protein expression data. *Nature Reviews Cancer, 8*(1), 37-49. doi:10.1038/nrc2294

Cohen, M., & Bookheimer, S. (1994). Localization of brain function using magnetic resonance imaging. *Trends in Neurosciences, 17*(7), 268-277. doi:10.1016/0166-2236(94)90055-8

Demirci O., Stevens M.C., Andreasen N.C., Michael A., Liu J., White T., Pearlson G.D., Clark V.P., Calhoun V.D. (2009). Investigation of relationships between fMRI brain networks in the spectral domain using ICA and Granger causality reveals distinct differences between schizophrenia patients and healthy controls. *Neuroimage, 46*(2), pp. 419-431.

Desikan R.S., Ségonne F., Fischl B., Quinn B.T., Dickerson B.C., Blacker D., Buckner R.L., Dale A.M., Maguire R.P., Hyman B.T., Albert M.S., & Killiany R.J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage 31*(3):968-980

Eaton, W.W., Alexandre, P., Bienvenu, O.J., Clarke, D., Martins, S.S., Nestadt, G., & Zablotsky, B. (2012). The Burden of Mental Disorders. *Public Mental Health, 2-30*. doi:10.1093/acprof:oso/9780195390445.003.0001

Ek, B., VerSchneider, C., & Narayan, D.A. (2015). Global efficiency of graphs. *AKCE International Journal of Graphs and Combinatorics, 12*(1), 1-13.

Fornito, A., Zalesky, A., & Bullmore, E. (2016). Fundamentals of brain network analysis.

Fox, M.D., & Raichle, M.E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience, 8*(9), 700-711. doi:10.1038/nrn2201

Frazier J.A., Chiu S., Breeze J.L., Makris N., Lange N., Kennedy D.N., Herbert M.R., Bent E.K., Koneru V.K., Dieterich M.E., Hodge S.M., Rauch S.L., Grant P.E., Cohen B.M., Seidman L.J., Caviness V.S., & Biederman J. (2005). Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *Am J Psychiatry 162*(7):1256-1265

Friston, K.J. (2011). Functional and Effective Connectivity: A Review. *Brain Connectivity, 1*(1), 13-36. https://doi.org/10.1089/brain.2011.0008

Goldstein J.M., Seidman L.J., Makris N., Ahern T., O'Brien L.M., Caviness V.S. Jr., Kennedy D.N., Faraone S.V., & Tsuang M.T. (2007). Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biol Psychiatry 61*(8):935-945

Gould, I.C., Shepherd, A.M., Laurens, K.R., Cairns, M.J., Carr, V.J., & Green, M.J. (2014). Multivariate neuroanatomical classification of cognitive subtypes in schizophrenia: A support vector machine learning approach. *NeuroImage: Clinical, 6*, 229-236. doi:10.1016/j.nicl.2014.09.009

Hage, P., & Harary, F. (1995). Eccentricity and centrality in networks. *Social Networks, 17*(1), 57-63. doi:10.1016/0378-8733(94)00248-9

He, Y., & Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Current opinion in neurology, 23*(4), 341-350.

Hsieh T.H., Sun M.J., Liang S.F. (2014). Diagnosis of Schizophrenia Patients Based on Brain Network Complexity Analysis of Resting-State fMRI. In: Goh J. (eds) The 15th International Conference on Biomedical Engineering. IFMBE Proceedings, vol 43. Springer, Cham

Huf, W., Kalcher, K., Boubela, R. N., Rath, G., Vecsei, A., Filzmoser, P., & Moser, E. (2014). On the generalizability of resting-state fMRI machine learning classifiers. *Frontiers in Human Neuroscience, 8*. doi:10.3389/fnhum.2014.00502

Ioannis, A., Eleni, T. (2007). Statistical analysis of weighted networks. *arXiv preprint arXiv:0704.0686*.

Iwabuchi, S.J., Liddle, P.F., & Palaniyappan, L. (2013). Clinical Utility of Machine-Learning Approaches in Schizophrenia: Improving Diagnostic Confidence for Translational Neuroimaging. *Frontiers in Psychiatry, 4*. doi:10.3389/fpsyt.2013.00095

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning with applications in R*. New York: Springer.

Kalcher, K., Huf, W., Boubela, R.N., Filzmoser, P., Pezawas, L., Biswal, B., et al. (2012). Fully exploratory network independent component analysis of the 1000 functional connectomes database. *Frontiers in Human Neuroscience, 6*. doi:10.3389/fnhum.2012.00301

Kohavi, R., A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the 14th international joint conference on Artificial intelligence, p.1137-1143, August 20-25, 1995, Montreal, Quebec, Canada

Lee, M. H., Smyser, C. D., & Shimony, J. S. (2013). Resting state fMRI: A review of methods and clinical applications. *AJNR. American Journal of Neuroradiology*, *34*(10), 1866–1872. http://doi.org/10.3174/ajnr.A3263

Lynall, M. E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., & Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience*, *30*(28), 9477-9487.

Ma, F., & Xu, J. (2016). The Application of kNN and SVM in the Decoding of fMRI Data. *Proceedings of the 2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)*. doi:10.2991/aiie-16.2016.78

Makris N., Goldstein J.M., Kennedy D., Hodge S.M., Caviness V.S., Faraone S.V., Tsuang M.T., & Seidman L.J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr Res. 83*(2-3):155-171. [doi 10.1016/j.schres.2005.11.020.

Newman, M.E. (2003). The Structure and Function of Complex Networks. *SIAM Review*, *45*(2), 167-256. doi:10.1137/s003614450342480

Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., & Mechelli, A. (2012). Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience & Biobehavioral Reviews, 36*(4), 1140-1152. doi:10.1016/j.neubiorev.2012.01.004

Patel, P., Aggarwal, P., & Gupta, A. (2016). Classification of Schizophrenia versus normal subjects using deep learning. *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing - ICVGIP 16*. doi:10.1145/3009977.3010050

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage, 45*(1). doi:10.1016/j.neuroimage.2008.11.007

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage, 52*(3), 1059-1069. doi:10.1016/j.neuroimage.2009.10.003

Savio, A. & Graña, M.. (2015). Local activity features for computer aided diagnosis of schizophrenia on resting-state fMRI. Neurocomputing. 164. 10.1016/j.neucom.2015.01.079.

Schnack, H.G., Nieuwenhuis, M., Haren, N.E., Abramovic, L., Scheewe, T.W., Brouwer, R.M. et al. (2014). Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *NeuroImage, 84*, 299-306. doi:10.1016/j.neuroimage.2013.08.053

Section for Biomedical Image Analysis (SBIA). (n.d.). Retrieved from https://www.med.upenn.edu/sbia/schizophrenia.html

Strang, A., Haynes, O., Cahill, N.D., & Narayan, D.A. (2017). Relationships Between Characteristic Path Length, Efficiency, Clustering Coefficients, and Graph Density. *arXiv preprint arXiv:1702.02621*.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage 15*: 273-289

Wang, J. (2010). Graph-based network analysis of resting-state functional MRI. *Frontiers in Systems Neuroscience*. doi:10.3389/fnsys.2010.00016

Wasule, V., & Sonar, P. (2017). Classification of brain MRI using SVM and KNN classifier. *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS)*. doi:10.1109/ssps.2017.8071594

Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: A functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity*. doi:10.1089/brain.2012.0073

Yang, H., Liu, J., Sui, J., Pearlson, G., & Calhoun, V. D. (2010). A Hybrid Machine Learning Method for Fusing fMRI and Genetic Data: Combining both Improves Classification of Schizophrenia. *Frontiers in Human Neuroscience, 4*. doi:10.3389/fnhum.2010.00192

Yue, S., Li, P. & Hao, P. Appl. Math. Chin. Univ. (2003). 18: 332. https://doi.org/10.1007/s11766-003-0059-5

Zeng, L.L. & Huaning, Wang & Hu, Panpan & Yang, Bo & Pu, Weidan & Shen, Hui & Chen, Xingui & Liu, Zhening & Yin, Hong & Tan, Qingrong & Wang, Kai & Hu, Dewen. (2018). Multi-Site Diagnostic Classification of Schizophrenia Using Discriminant Deep Learning with Functional Connectivity MRI. *EBioMedicine. 30*. 10.1016/j.ebiom.2018.03.017