**Introduction**

The neurological (neurology?) community has been slow to accept computer-based classification of neurological disorders due to lack of conclusive evidence and generalizability (Lord & Jones, 2012). With the ever increasing computational power, advanced classification techniques are rapidly released that provide researchers tools that may change the diagnosis process. The challenge then lies in identifying reliable and reproducible biomarkers for these neurological disorders.

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by a wide range of symptoms and levels of impairment or disability in children and adults. Current diagnosis levels are about 1 in 68 children, an increase from roughly 1 in 2000 in the 1960s through the 1980s (Kogan et al., 2009). Although the prevalence of ASD in children has significantly increased, researchers still have no clear biological cause. According to the National Institute of Mental Health, the current diagnosis process focuses on clinical interactions between a child and medical professionals. The interactions are monitored for behavioral and social cues for diagnosis. Manual diagnosis methods risk high variance of diagnosis (Zijdenbos et al., 2002) and misdiagnosis (Nylander et al., 2013). The use of automated classifiers could alleviate the problems encountered with manual diagnosis as well as decrease the diagnosis time.

Magnetic resonance imaging (MRI) has been used in several previous ASD studies. Although it has been proposed as a complement to behavioral diagnosis, the research is relatively recent and must be refined before implementing in a clinical setting. Investigating ASD using MRI can provide insight into biomarkers of ASD patients to use in a classifying algorithm. Previous studies utilizing MRI have focused on structural imaging of the brain (Ecker et al., 2010), whole-brain analysis (Calderoni et al., 2012), and connectivity (Anderson et al., 2011, Tyszka et al., 2014) but are limited by a small number of included patients. While these studies may enjoy high classification accuracy, a clinical applicable classification algorithm must be robust to accurately diagnose patients from various demographics and from different imaging sites.

The release of the Autism Brain Imaging Data Exchange (ABIDE) database was the first large, open-access database with functional MRI (fMRI) data on over 1,000 ASD and control subjects. The ABIDE database is a grassroots effort “dedicated to aggregating and sharing previously collected resting-state fMRI (R-fMRI) datasets from individuals with ASD” (Di Martino et al., 2014). The database is part of the International Data-sharing Initiative’s 1000 Functional Connectomes Project, which aims at collecting and sharing high quality neuroimaging data of several different neurological disorders. The database is an aggregate collection of 1,112 subjects from 20 different studies at 16 international sites. The available data includes raw fMRI as well as processed MP RAGE images.

The release of the data provides a wide-range of researchers access to data they may not be able to acquire on their own. This includes researchers in fields not normally prevalent in the medical or neurological fields such as computer science and statistics.

**Methodology**

**2.1 Data**

The raw fMRI data was aggregated and disseminated through the Functional Connectomes Project in support of their grassroots effort to provide high quality ASD fMRI data. The raw data must be preprocessed to permit quality analysis. This includes attempting to standardize the data from across all 20 studies to reduce any variance. Preprocessing the data is an essential step to constructing a robust classification model due to the inherent variability among different fMRI capture sites. A goal of this research is to construct a classifier that is not only accurate, but generalizes well enough to accommodate small variability within the data.

In support of the Preprocessed Connectomes Project (PCP), five teams used four different common neuroimagery preprocessing pipelines on the data from the ABIDE database. The PCP was developed to provide quality, open-source, preprocessed MRI data to researchers. Due to discrepancies in literature, the PCP conducts preprocessing through a variety of different methods to accommodate the needs of the researcher (Craddock & Bellec, 2015). This research utilizes the data preprocessed using the Configurable Pipeline for the Analysis of Connectomes (CPAC) tool (Craddock et al., 2015). All fMRI and corresponding phenotype data included in the ABIDE database has been scrubbed of protected personal identifying information in accordance with HIPPA guidelines. Every image was acquired with informed consent according to the human subjects research boards at each study’s respective institution. Details about each study’s guidelines can be found at *http://fcon\_1000.projects.nitrc.org/indi/abide/*.

**2.2 Functional Magnetic Resonance Imaging**

Functional Magnetic Resonance Imaging (fMRI) detects changes in blood perfusion, volume, or oxygenation that are thought to accompany neurological activity throughout the brain (Matthews & Jezzard, 2004). Acquired through sequential two-dimensional images, or slices, the entire brain is mapped through several repetitions as the machine moves throughout the brain region. The ABIDE fMRI data utilizes blood oxygenation level dependent (BOLD) contrasts to identify brain activity within the subjects. It is suggested that increased blood flow to specific regions of the brain is caused by neurotransmitter action, identifying local signaling (Matthews & Jezzard, 2004) as oxygen is extracted from the bloodstream (Ogawa et al., 1990). This research utilizes this to search for activity patterns across the subjects.

How the raw fMRI data is cleansed and preprocessed…

Utilizing fMRI presents another problem for classification; the size of the data can be several gigabytes and most classification algorithms cannot use the common fMRI filetype.

**2.3 Functional Network Connectivity**

Functional network connectivity (FNC), a technique that estimates “connectivity” levels between brain regions, accommodates the limitations presented by fMRI data with classification techniques. FNC is defined as the observed temporal correlation of neuroactivity between different parts in the brain (Friston et al., 1993).