**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). 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.