Predicting progression of Alzheimer's Disease with clinical and genotype data

Josh Tycko, Spencer Penn, Juan Jose Lopez Delgado

Project significance: Alzheimers disease (AD) is predicted to affect 1 in 85 people globally by 2050, causing dementia and eventual death. Care in the US costs $100 billion annually, and the available drugs can only help relieve some symptoms. It is currently difficult to predict the progression of AD, and it often progresses undiagnosed for years. Machine learning algorithms have the potential to assist doctors and patients by accurately predicting disease progression based on clinical and genetic data. We are using data collected over 2 years by the Alzheimers Disease Neuroimaging Initiative (ADNI) in 767 patients, including mental examinations and genotype in order to predict the progression of AD over time. If successful, this could lead to a new approach for early diagnosis of AD and earlier disease treatment.

Traditional methods: Since the causes of AD are currently unknown and there are no laboratory tests that can accurately perform a diagnosis, AD progression is quantified with psychological tests like the mini-mental state examination (MMSE) - a questionnaire used to measure cognitive impairment. This set of 30 questions was developed in 1975 and remains the standard.

Prior work: Clinical and genotype data may be used to aid in diagnosing the disease at an early stage, as AD is a brain disorder. Machine learning algorithms have been used on ADNI data with varying success to predict the change in MMSE. Interestingly, no single algorithm has been shown to be superior across all AD datasets, particularly when progression is measured up to varying time points.

Novel approach: We aim to develop an algorithm that is robustly accurate across data sets, by creating an ensemble model of the top models tried previously (simple logistic regression, random forests, and Bayesian nets). By weighting our ensemble with boosting, we will try to create an ensemble model that is superior in accuracy to any of the constituent models. There is also an opportunity to exceed the accuracy of previous algorithms by incorporating additional data attributes and more granularity in time from the AddNeuroMed Study. There is also great potential in extracting features from MRI image data to complement the clinical and genotype data.

Standardized metric of success: Our goal is to predict the change in MMSE compared to baseline (delta MMSE). We are going to check our success by calculating both the Pearson and Spearman correlation between our predicted delta MMSE and the observed delta MMSE in a testing set of AD patients from a different study (with all the same attributes except the final MMSE score). We can do this by submitting our predictions to synapse.org, where there is an online challenge to develop algorithms.

Data source: Data used in preparation of this proposal were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). We are considering supplementing our data set with the AddNeuroMed Studys data collected in 409 AD patients over 1.5 years, which includes significantly more clinical attributes.