Name(s)
Jack D. Albright

Project Number
J1501

Project Title
Predicting the Future: Using Machine Learning to Forecast the Progression of Alzheimer's Disease

Abstract

Objectives/Goals
Alzheimer's disease (AD) is the most common neurodegenerative disease in older people. Despite considerable efforts to find a cure for AD, there is a 99.6% failure rate of clinical trials for AD drugs, likely because AD patients cannot easily be identified at early stages. AD research would benefit from the ability to use current patient data to predict the clinical state of patients in future years. This project investigated machine learning approaches to enable such predictions.

Methods/Materials
Clinical data from 1737 patients was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and was preprocessed using a novel methodology involving the comparison of all possible pairs of temporal data points for each patient. This data was then used to train machine learning models implemented with the Python library scikit-learn. Models were first evaluated using k-fold cross-validation on the training dataset, and then confirmed using data from a separate testing dataset (110 patients).

Results
Several machine learning-based classifiers, including support vector machines and neural networks, were effective (86.6% MAUC score) at predicting the progression of AD, both in patients who were initially cognitively normal and in patients suffering from mild cognitive impairment.

Conclusions/Discussion
This project developed a machine learning model that can correlate clinical data obtained from patients at one time point with the month-by-month progression of AD in the future. Such a model could be used to identify patients having high AD risk before they become symptomatic and who are therefore good candidates for clinical trials for AD therapeutics.

Summary Statement
I developed a machine learning model and novel preprocessing technique that uses clinical data to identify patients at risk for Alzheimer's disease and predict the future progression of their clinical diagnoses on a month-by-month basis.

Help Received
I received guidance from my computer science teacher Jen Selby and my science teacher Rachel Dragos. I also received feedback on my machine learning model from Alex Abbas, a bioinformatics researcher at Genentech, and Lisa Bell, a regulatory consultant experienced in neurological disease clinical trials.