A Predictive Diagnose for Parkinson's Disease Through Machine Learning

Ray, Shounak (School: Webber Academy)

Currently, there is no fully reliable or accurate way to determine whether an individual has Parkinson’s disease. Moreover, 90% of clinically-confirmed Parkinson’s disease cases are idiopathic, suggesting the extent to which a more accurate and reliable diagnosis would be beneficial. This neurological disorder proffers a tremendous financial burden to the patient in terms of initial diagnostic and treatment costs. In this study, human demographic, movement, and speech data were analyzed to determine if an individual has Parkinson’s disease, thus resulting in a binary classification problem. Two data were used: one for demographics and movement and another for speech data. Machine learning and statistical testing were conducted on the two data sets individually. Over 30 different machine learning models, from lazy-based to tree-based, were analyzed through visualizations, model metric analysis, and external statistical testing. Upon in-depth exploration of the data set and the multiple models, an Android application was created in order to prove the merits of each machine learning model. The application extracts users’ demographic, movement, and speech data – both through manual input and artificial intelligence components, such as automatic speech recognition and model optimization. Ultimately, after rigorous statistical testing procedures, the locally weighted learning (LWL) was the best demographic-movement model (accuracy = 98.8%) and an ensemble model was the best human speech model (accuracy = 95.5%). Cumulatively, the machine learning framework founded on demographic, movement, and speech data suggests a more accurate, time-efficient, and cost-effective gold standard for Parkinson’s disease diagnosis.

Awards Won: