Machine Learning-Based Algorithm to Predict Alzheimer's Disease Risk Using Inflammatory and Genetic Biomarkers in Caribbean Hispanic Cardiovascular Patients

Duconge, Ana (School: University Gardens High School)

Latin Americans are the ethnic group with the second highest prevalence of Alzheimer's disease cases. However, most studies on this condition have been conducted in European populations. To address this health disparity, a Machine Learning prediction algorithm was developed using clinical, demographic, genetic, and inflammatory data to predict the occurrence of this illness in a high-risk population of Caribbean Hispanic cardiovascular patients, given that heart disease is known to increase Alzheimer's risk by ~40%. Clinical and demographic data were collected through questionnaires conducted during recruitments, while genetic and inflammatory profiles were obtained using TaqMan RT-PCR and ELISA, respectively, to identify the presence of pre-selected genetic variants and circulating cytokines. Of the 7 Machine Learning models tested, Gradient Boosting Classifier (GBD) obtained an AUC ROC of 0.98 and an accuracy of 0.87. The features of the most crucial value in the final model's performance were IL-12p70 (~23%), TREM2 (~12%), APOE, and smoker status (4% each). By exceeding standard metrics and an AUC ROC of 0.8, the efficiency of this model in predicting the occurrence of the disease within the study population was confirmed. The development of a tool that assists clinicians in the detection of individuals at high risk of Alzheimer's disease within a Caribbean Hispanic cohort may help establish better strategies to identify cases at an early stage of onset, hence preventing disease progression. Consequently, patient care in this minority group can be improved, lessening the burden on the healthcare system and the economy within the studied regions and globally.

Awards Won: