Optimizing Random Forest Algorithms for Autism Prediction: Enhancing Diagnostic Accuracy Through Machine Learning

Madhar, Tanish (School: Fort Myers High School)
Cintron, Gabriel (School: Fort Myers High School)

The purpose of this study is to pursue utilization of the most advanced machine learning models in order to improve the accuracy and reliability of Autism Spectrum Disorder (ASD) diagnoses. Behavioral biomarkers such as responses from the Q-CHAT-10 questionnaire and individual behavioral metrics (e.g., A1, A2, A5, and A8) were incorporated into Random Forest and Neural Network models that were trained with exceptional performance for predicting ASD traits. Both models had perfect evaluation metrics (accuracy, precision, recall, F1) scores and a confusion matrix that confirms the zero misclassifications. The Q-CHAT 10 score was revealed as a significant contributor to model predictions, as it contributed over 75% to model predictions. Although individual features had inferior predictive power singly, their combination made the model more robust and generalizable over wide variety of situations. Diagnostic accuracy was similar as class imbalances in the dataset and preprocessing steps, such as normalization, categorical encoding were addressed. This paper proposes a data driven, subjectively less subjective, more interpretational approach to ASD diagnostic tool, employing feature selection and visualization techniques. Additionally, this research demonstrates the feasibility of using behavioral data aggregation to enhance understanding of ASD associated traits. In future work we would like to integrate environmental and genetic factors in a much broader scope of ASD diagnostics with data privacy and ethical considerations in mind. In this study, we show that machine learning is a viable means to accelerate ASD diagnoses, enabling early intervention and better outcomes for persons with ASD and their families.

Awards Won: