Leveraging Deep Learning Models to Identify Structurally Adequate Drug Candidates for Parkinson's Disease

Garimella, Raghav (School: Evergreen Valley High School)

Parkinson's Disease (PD) is a neurodegenerative disease that causes gradual impairment of an individual through mutations in the SNCA (Synuclein Alpha) gene. Currently, there exist no widely available drugs for PD due to the long process of drug identification and FDA (Food and Drug Administration) certification which can take up to fifteen years, as well as the 300 million dollars often put into the creation and marketing of a new drug. This project aims to expedite the process and reduce cost by utilizing Machine Learning models to identify and repurpose already approved FDA drugs. The models constructed in this project will analyze Early Onset, Late Onset, and General PD utilizing a novel input combined matrix consisting of Simplified Molecular Input Line Entry System (SMILES) encoded drug representation vectors of drug molecular features and one-hot encoded disease representation vectors. Both Deep Learning and Regression models are evaluated to predict the efficacy of various drug-disease pairs. Considering the data representations of multiple FDA approved drugs, and utilizing a validation loss metric of Mean Squared Error, Valproic Acid emerges as a potential target for further research in its applications to PD. The best predictive model utilizes a Multilayer Perceptron Classifier and has an average validation loss of 0.0040. Valproic Acid is seen to inhibit HDAC6 enzyme production, potentially playing a role in preventing protein aggregation formations known as Lewy Bodies in PD. As a result, this project greatly accelerates the process of drug identification for different stages of PD, providing a foundation for future, targeted research into drug-PD associations.