Development of a Novel Machine Learning Algorithm in Biomolecules and Drugs for Measuring Molecular Surface Area: Applications in Long QT Syndrome and Neurodegenerative Diseases

Lokshin, Shawn (School: Staten Island Technical High School)
Ibrahim, Nouraldeen (School: Staten Island Technical High School)

This study involved developing a Machine Learning (ML) algorithm to create an efficient approach for predicting Molecular Surface Area (MSA). MSA is a fundamental property that models the amount of surface available in a molecule. However, since current MSA calculation approaches are demanding, time-consuming, and inefficient in large-scale computations, we coded a ML algorithm on Wolfram Mathematica to simplify MSA calculation. We retrieved descriptor values for 799 biomolecules and 682 drugs to find properties correlated with MSA and trained six different ML methods to utilize these properties to predict MSA. We assessed the accuracy of each, and the most accurate one was then applied to predict the MSA of two molecules from the dataset: the biomolecule Glutamine and the drug Citalopram. Neural Network evaluation retrieved the highest R^2 values of 0.997 in both classes, Root Mean Square Error values of 3.0 and 2.8 respectively, and a rapid evaluation time under 3 milliseconds/example. The MSA's of Glutamine and Citalopram were also predicted within 1 and 0.3 A^2, respectively. This is crucial because Glutamine's MSA corresponds with protein adsorption, which worsens the quality of pharmaceuticals, and aggregation, which causes neurodegenerative diseases. Likewise, Citalopram's MSA defines its channel blocking ability and its likelihood to induce Long QT Syndrome (LQTS). Determining if LQTS is a risk factor is essential for expeditious drug development against diseases like COVID-19 and cancer. Overall, this algorithm can be employed by researchers as an invaluable tool to greatly accelerate MSA calculation.

Awards Won:
YM American Academy: Third Award of $500.00