A Novel Machine Learning Method and Drug for the Diagnosis and Treatment of Depression

Lakkaraju, Ananya (School: Hamilton High School)

Major Depressive Disorder negatively affects millions worldwide by decreasing productivity, degrading physical well-being, and increasing the risk of suicide by 60%. Moreover, miscalculating the risk of suicide after rehabilitation, by inaccurately diagnosing depression, accounts for many suicide deaths. Studies find that the gene expression levels significantly affect the development of depression. However, current antidepressants increase serotonin and dopamine levels in the brain, not addressing the root cause of depression. This study provides a statistical analysis using machine learning to diagnose depression and identify a target gene to produce a drug that combats it. A Convolutional Neural Network (CNN) was used to identify patterns between an NIH RNA-sequencing dataset and depression labels. Next, Principal Component Analysis (PCA) was used to categorize the various differentially expressed genes (DEGs) by their influence on depression. To increase the model's accuracy, SMOTE was used to augment samples of the minority class for the training data, while original samples were used for testing. Using these methods, a Convolutional Neural Network (CNN) model achieved an accuracy of 93.24%. Next, PCA identified the top genes contributing to a depression diagnosis by assessing differential expression, identifying MT-ND4 as the top contributor. Protein transduction pathway analysis revealed that inhibiting 7UDQ activated the production of ND4, the protein made by MT-ND4, mitigating depression. The chemical composition of a drug that inactivates 7UDQ with a lower binding energy (-8.5 kcal/mol) than the original ligand (-8.2 kcal/mol) resulted from molecular docking simulations of several experimental ligands.

Awards Won: