Exploring Genetic Similarities With AI: A Novel Python Algorithm That Finds HOG Descriptors of GWAS LD Patterns

Cohen, Romy (School: iPrep Academy North)
Cohen, Ruby (School: iPrep Academy North)

This project looked at two neurodevelopmental disorders, Autism Spectrum Disorder and Dyslexia, that both contain a hereditary component and frequently co-occur. The purpose of this project was to create an artificial intelligence device using a Python-based program to find the genetic overlaps between the diseases. We chose this approach due to the extensive data available for both neurodevelopmental conditions. My goal was to find out if my original Python code algorithm based on the linkage disequilibrium (LD) patterns of the GWAS can determine the extent of these two disorders being genetically similar. To collect this data, an original Python code was made to extract all the available data for each of the 50 variants used for this study: 25 of the lowest p-value variants for dyslexia and 25 of the lowest p-value variants for Autism Spectrum Disorder which told me the Linkage Disequilibrium (LD) Pattern for each of the 50 variants used. Then a code was created to rank the most similar LD patterns between the two disorders based on disequilibrium values. This was repeated instead it was the rank of HOG, Histogram of Oriented Patterns. The results were compared to identify similarities based on the Euclidean distance. In conclusion, this project presented a novel dissimilarity-based learning model for similar structured but diverse chart classification.

Awards Won: