IntelliGenome: Using Artificial Intelligence to Diagnose Genetic Diseases on the Genomic Level

Farid, Michael

The purpose of this project was to create a novel artificial intelligence system to diagnose genetic disease based solely on genetic data. The first half of the project involved creating an artificial neural network to distinguish between neutral and disease-causing Single Nucleotide Variations (SNVs) in genes associated with genetic diseases and the second half of the project involved diagnosing complex genetic diseases using microarray gene expression data. First, datasets of the both types of SNVs were downloaded from various genetic databases. Macros were programmed and run to calculate 5 crucial properties about each of the 2000 protein with the SNV. The five properties were pI, molecular weight, hydropathicity, instability index, and aliphatic index.. Then, an s-back-propagation neural network was designed with Matlab with 500 hidden layers and data was inputted for training. After training, the neural network demonstrated an accuracy of 91.2% which was incredibly high. A second improved network was made using 43940 SNVs and had 93.4% accuracy. The network was then used to classify the online unknown SNVs that were found in human genes but were not yet classified. The neural network was later deployed into a user interface for easy usage. For the second half, microarray data for lung and breast cancer were downloaded. Then, biomarkers were found using various bioinformatics algorithms. A neural network was made to diagnose based on biomarker data but it failed. Linear regression on the biomarkers values and surprisingly rendered an accuracy of 98.7% and 99.3%. Overall, the project was extremely successful.

Awards Won: