Decoding Genetics of Aging: A Neural Network Interpretation on Age-Associated Biomarker Data and Diseases

Guan, Hannah (School: BASIS San Antonio Shavano Campus)

Aging studies can sustain longer lives with reduced periods of disability, and genetics is a powerful tool for understanding the underlying mechanisms of aging. As part of the complicated aging mechanisms, epigenetic aging is related to specific patterns and changes in the epigenome, e.g., DNA methylation. Multiple epigenetic aging clocks have been developed, but little research has focused on their interpretations. Neural networks are powerful prediction models which can model higher-order interactions between biomarkers. This research addressed the challenging overfitting problem caused by high-dimensionality and low sample size in DNA methylation data, by developing the Correlation Pre-Filtered Neural Network (CPFNN) model. CPFNN uses Spearman Correlation to reduce the feature space before inputting them into neural networks. It outperformed 11 other aging formulas by at least 1 year with a Mean Absolute Error of 2.7 years. To interpret the CPFNN aging model, I analyzed its weight parameters in comparison to that of LASSO Neural Networks and Elastic Net Neural Networks. I concluded that, for a large number of candidate features, such as genome-wide DNA methylation data, a key factor in improving prediction accuracy is to appropriately weight features that are highly correlated with the outcome of interest. CPFNN is one of the first neural network models for age prediction and age acceleration research. It can be widely used in large-scale biomarker datasets for prediction and feature selection, ultimately improving our understanding of the aging process and benefiting public health.

Awards Won:

Fourth Award of $500

Fourth Award of $500