A Novel CNN-Based Method Using Genetic Data for Early Prediction of Alzheimer's Disease

Wang, Angela (School: University School of Milwaukee)

Alzheimer’s disease affects over 55 million people worldwide, making early prediction crucial for timely interventions to slow disease progression and improve quality of life. Single Nucleotide Polymorphisms (SNPs) within the APOE gene are strong indicators of disease susceptibility. However, two major challenges persist: (1) the genetic complexity of the disease and (2) limited access to individual-level SNP data due to privacy regulations such as HIPAA. CNNs have been used for image processing, but very few have been extended to genetic data. This study leverages CNNs for Alzheimer’s prediction using genetic data. Three models were developed: (1) one-dimensional CNN trained on APOE SNP sequences (2) two-dimensional CNN trained on the same SNP data reshaped into a 2D format (3) two-dimensional CNN trained on Linkage Disequilibrium (LD) matrices constructed from population-level APOE SNP data. Each model used the Adam optimizer and fine-tuned hyperparameters to prevent overshooting and overfitting. Model performance was evaluated using (1) mean squared error (MSE) for regression prediction of hippocampal volume, a key biomarker for Alzheimer’s, and (2) confusion matrix to assess classification accuracy for dementia status. The 2D CNN applied to LD matrices achieved a regression accuracy of 82.53% and a classification accuracy of 85.17%, demonstrating not only the strong predictive capability of CNNs on genetic data but also the ability of LD matrices to preserve key genetic patterns while protecting individual privacy. This study presents the first usage of 2D CNNs on LD matrices for Alzheimer’s prediction, introducing a privacy-preserving approach that overcomes data access barriers and paves the way for AI-driven, large-scale genetic research and disease prevention.

Awards Won: