CALIFORNIA STATE SCIENCE FAIR
2017 PROJECT SUMMARY

Name(s)

Pravin Ravishanker

Project Number

S0827

Project Title

ALZCan: Machine Learning Based Multimodal Neuroimaging Genetics Framework to Predict Future Onset of Alzheimer's Disease

Abstract

Objectives/Goals

To create an accurate diagnostic and prognostic machine learning based methodology for Alzheimer's early detection by using polygenic risk scores constructed from single nucleotide polymorphism (SNP) genotype data; cerebrospinal fluid (CSF) levels; cognitive tests; demographic info; 3 neuroimaging modalities (678 Resting State Functional MRI (rs-fMRI), 427 Fluorodeoxyglucose PET (FDG-PET), 407 Florbetapir PET brain scans).

Methods/Materials

Using R programming language and data from Alzheimer's Disease Neuroimaging Initiative (ADNI), I considered 23 neuroimaging and CSF phenotypes linked to AD pathology: 3 CSF protein levels; FDG-PET cerebral metabolic activity in 5 brain regions; Florbetapir-PET beta-amyloid plaque levels in 7 regions; and 8 rs-fMRI functional network connectivity (FNC) measures (3 ROI-to-Voxel Ratio + 5 Graph-theory based). Graph-theoretic FNC metrics were extracted from each rs-fMRI scan using ICA, signal Cross-Correlation, and graph-rendering algorithms.

Using GLMs, Genome-Wide Association Analyses evaluated 14 million+ interactions between 608,586 SNP genetic variants (393 participant genotypes) and the aforesaid 23 AD phenotypes. 23 polygenic risk scores per subject were computed by combining weighted additive effect of each subject's multiple SNP variants, with association results determining each SNP's weight in risk scores.

Results

Based on a subject's 23 polygenic risk scores, demographic info, and cognitive scores; gradient boosting machine learning could differentiate between Healthy Control, Mild Cognitive Impairment (MCI), Alzheimer's (AD) with diagnostic accuracy of 98.10%; and predict onset of AD and MCI 6, 12, 24, 36 months into future with prognostic accuracies of 91.89%, 91.72%, 85.38%, 70.67% respectively. Polygenic risk scores constructed from top genetic variants affecting 1) Posterior Cingulate/Left Angular Gyrus metabolic activity; 2) Precuneus beta-amyloid plaque levels; 3) rs-fMRI FNC as measured through Default Mode Network ROI-to-Voxel ratio and Transitivity; 4) CSF phosphorylated tau levels were most capable of predicting MCI and AD status in the future.

Conclusions/Discussion

With power of polygenic risk scores for risk prediction/intervention/personalized medicine and machine learning's promise to discover relationships amongst massive genomic/neuroimaging datasets, we can precisely pinpoint Alzheimer's future onset and prevent irreversible brain damage.

Summary Statement

I created a novel machine learning framework for Alzheimer's and Mild Cognitive Impairment's early diagnosis and prognosis, fusing signal processing, graph theory, neuroimaging, genome-wide association analyses, and polygenic risk scoring.

Help Received

I thank the Alzheimer's Disease Neuroimaging Initiative (international effort tracking AD biomarkers) and its investigators for the vital data repository. I am grateful to all who have supported me throughout my 4-year progressive computational neuroscience research on Alzheimer's prediction.