XAI-Driven Biomarker Discovery for Early Alzheimer's Detection

Verma, Agrima (School: Pascack Hills High School)

Dementia affects 55 million people globally, projected to reach 139 million by 2050. Although deep learning models achieve 90–98% accuracy for dementia classification, clinical adoption is hindered by their "black box" nature. Current explainability methods, like Grad-CAM, lack sufficient anatomical detail for clinical use. This study introduces the first end-to-end pipeline to translate 3D convolutional neural network (3D CNN) decisions into anatomically specific, subject-level explanations. Our 3D CNN achieved 93% accuracy in classifying cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) using: (i) regional intensity-weighted saliency, (ii) affine-aware spatial registration aligning segmentations with Grad-CAM, and (iii) z-score normalization against CN distributions. Applied to 150 OASIS-3 subjects, this pipeline produced neuroradiologist-validated reports ranking brain regions by diagnostic contribution. Population-level analysis computed AD-to-CN saliency frequency ratios for 103 regions. Canonical biomarkers (hippocampus, entorhinal cortex, amygdala) were identified at 2.8–2.9 times the frequency, confirming biological validity. Non-canonical subcortical structures (caudate, thalamus, putamen, brainstem) showed similar ratios, with monotonic progression from CN to MCI to AD. These findings suggest the model detects structural correlates of tau spread, cholinergic denervation, and mixed pathology not captured by current frameworks. Overall, the study shows that structural MRI holds more biological information than currently used, and explainability tools can turn black-box models into actionable hypotheses.

Awards Won: