



Early Detection of Alzheimer's Dementia Using Multimodal Clinical and MRI Biomarkers: A Machine Learning Analysis of the OASIS-2 Longitudinal Cohort



MS Health Data Science, Saint Louis University
 Vudem Shruthi Reddy (Msc. Health Data Science)
 Instructors: Divya S. Subramaniam, PhD, MPH & Dipti P. Subramaniam, PhD, MPH
 Saint Louis University School of Medicine , Department of Health & Clinical outcomes Research

Background

- Alzheimer's disease (AD) is the most common cause of dementia and produces measurable changes in cognition and brain structure several years before clinical diagnosis.
- Early identification of "at-risk" individuals enables timely intervention, improved care planning, and better monitoring of disease progression.
- However, combining clinical + MRI features in a unified predictive model remains limited in real-world datasets because of small sample sizes, collinearity, and class imbalance.
- The OASIS-2 dataset provides longitudinal MRI and cognitive assessments, enabling evaluation of prediction models to classify individuals as demented vs. nondemented at baseline.

Goals of Study

- Characterize baseline differences in age, cognition, and brain volume between demented and nondemented individuals.
- Develop and evaluate machine-learning models (Logistic Regression, Random Forest) for baseline dementia classification.
- Compare fixed vs. ROC-optimized thresholds for realistic model performance.
- Identify key biomarkers using SHAP explainability to support early risk stratification.

Analysis

Statistical & ML Analysis

- Exploratory analysis: group distributions, missing data, correlations.
- Models:**
 - Logistic Regression (baseline linear model)
 - Random Forest (nonlinear, interaction-aware)
- Evaluation: 80/20 split, ROC-AUC, PR-AUC, accuracy, sensitivity, specificity, PPV.
- Thresholds:**
 - Fixed 0.50
 - ROC-optimal (Youden Index)
- Cross-validation: 5-fold CV on Random Forest.
- Explainability: SHAP summary & interaction plots.

