Final Report

By Alekhya Vittalam, Kirpa Kaur, Guantong Zhang, and Bingcheng Wang

Problem Statement

This project explores two key questions:

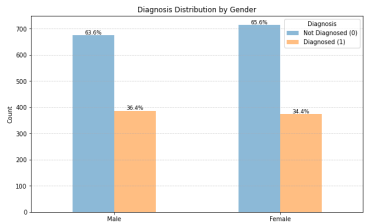
1. Can an optimized predictive model accurately diagnose Alzheimer's disease?
   Our goal is to leverage machine learning techniques to build a predictive model capable of accurately diagnosing Alzheimer's based on predictors such as demographics, lifestyle factors, medical history, and cognitive assessments. This tool aims to assist healthcare professionals in evaluating the likelihood of Alzheimer's early in its progression.
2. Can correlations between patient attributes and diagnostic outcomes inform recommendations for tests?
   We analyzed relationships between demographic, lifestyle, medical, and cognitive measurements and assessments to identify meaningful patterns. Insights from these correlations could guide diagnostic test prioritization based on patient profiles.

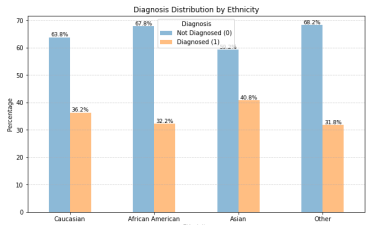
Background

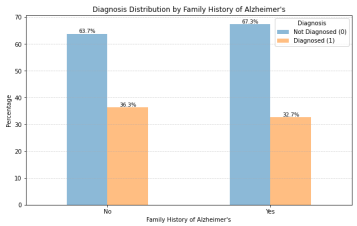
Alzheimer's disease is a progressive neurological disorder and a leading cause of disability in the elderly. Early diagnosis is crucial for timely intervention and a better quality of life. However, its complexity—requiring clinical, imaging, and laboratory assessments—makes diagnosis challenging. Machine learning offers a promising solution by analyzing multidimensional datasets to uncover patterns traditional methods might miss, enabling earlier and more accurate detection.

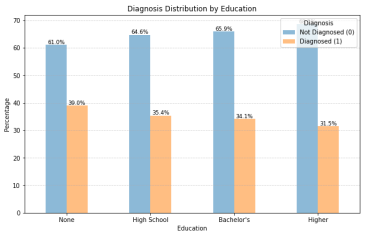
Data

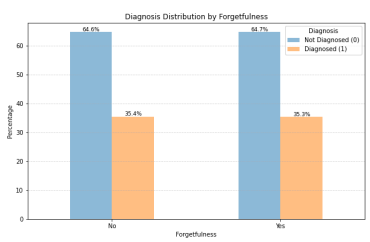
Our dataset includes health information for 2,149 patients with 33 predictors spanning demographics, lifestyle factors, medical history, cognitive assessments, and a binary Alzheimer's diagnosis response variable. Key findings from exploratory analysis:

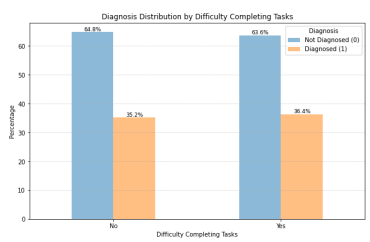


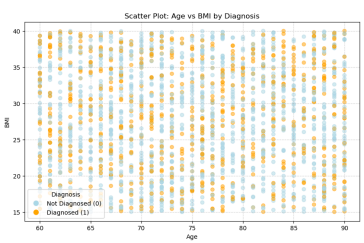


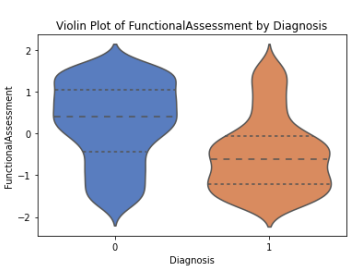










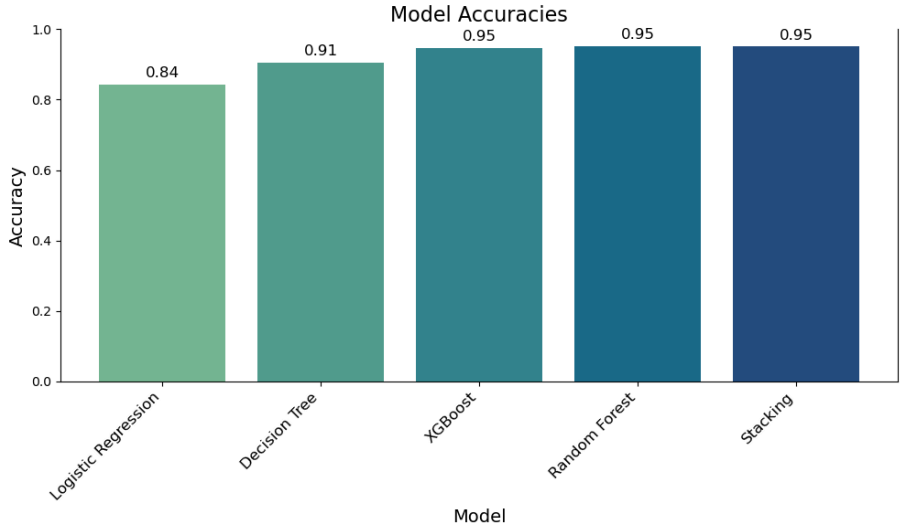


- Gender: Minimal predictive value with similar proportions between groups.
- Education & Ethnicity: Higher diagnosis rates for individuals with lower education and Asian ethnicity.
- Family History: Surprisingly, individuals without a family history of Alzheimer's had a higher diagnosis rate (36% vs. 32%).
- Symptoms: No significant differences in traditionally associated symptoms (e.g., Difficulty in Completing Tasks, Forgetfulness).
- Age & BMI: Not significant predictors.
- Cognitive Assessments: Functional assessment scores were notably higher in non-diagnosed patients, highlighting their predictive potential.

Given these findings, we prioritized building a machine-learning model to improve diagnostic accuracy.

Modeling

We addressed the imbalance in diagnostic groups and low predictive value of many variables by applying K-best feature selection to identify the 10 most impactful features. Supervised learning methods—Logistic Regression, Decision Tree, Random Forest, and XGBoost—were evaluated, with Random Forest achieving the highest overall accuracy (95.35%). We also tested stacking as an ensemble method for enhanced performance.



To dive deeper, we categorized predictors into four groups: demographic, lifestyle, medical, and cognitive, and ran models for each:

Category/ Model Accuracy	Logistic Regression	Elastic Net	Random Forest	XGBoost
Demographic	0.646512	0.646512	0.646512	0.646512
Lifestyle	0.679070	0.679070	0.672093	0.676744
Medical	0.693023	0.693023	0.648837	0.665116
Cognitive	0.795349	0.795349	0.837209	0.837209

- Cognitive features performed best, with Random Forest and XGBoost achieving similar accuracies.

Inference

Our findings underscore the significant impact of accurate Alzheimer's diagnosis for early intervention and improved patient outcomes. Cognitive and functional assessments were the most predictive features, while variables like age and traditional symptoms proved less significant.

Future work includes validating the model in real-world, diverse populations to refine its predictive capabilities and address ethical concerns such as the implications of false positives and negatives in diagnosis.

What We Learned

- Effective communication and task coordination were critical for this collaborative effort.
- Data-driven insights often challenge assumptions, such as age being an insignificant factor for Alzheimer's diagnosis.
- Ethical considerations are paramount in healthcare applications, especially concerning medical diagnostic models.