Exploring the Distinctions Between Mild Cognitive Impairment and Alzheimer's Disease

Julia Eng[1] Yvette Vargas[1] Olivia Willard[1] Kent Tran[2] Ted Yoon[2] Yen Phuong Lam[3] Wesley Zeng[4]

[1]University of California, Irvine [2]California State University, Fullerton [3]Cypress College [4]Cornell University

Abstract

Alzheimer's disease is a progressive brain disorder that gradually impairs cognitive functions, including memory, thinking skills, and eventually the ability to perform simple, everyday tasks. As the US population ages, the prevalence of Alzheimer's disease is expected to rise dramatically, with projections of up to 13 million affected individuals by 2050.[1] Given this looming public health challenge, our study aims to deepen our understanding of the distinctions that can be found between the diagnosis of Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). The core difference ultimately relates to difficulties doing simple everyday tasks like paying bills, remembering dates, preparing meals, and doing taxes.

Background Information and Scientific Objective

With data on patients across different cognition levels, our research aims to distinguish between Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) patients through predictive modeling. By leveraging easily accessible data from neuropsychological tests, depression and behavioral surveys, and demographic information—compared to more complex and costly methods such as MRI scans and biomarkers—we aim to develop more effective diagnosis methods for this debilitating condition.

Data Structure

Our data is provided by the National Alzheimer's Coordinating Center (NACC) and includes three main datasets from the patients' first visits:

1. **UDS**: demographics (DEMO), neuropsychological tests (NPT), depression scales (GDS), and behavioral surveys (FAS) for each subject.
2. **MRI**: image derived features from structural MRI scans.
3. **CSF**: biomarker data.

Data Wrangling

To achieve our goal, we focused mainly on the Uniform Data Set (UDS) as we had significantly more data with about 45,000 subjects compared to only 3,000 and 2,000 subjects in the MRI and CSF datasets, respectively. We divided the UDS dataset into two subsets: one distinguishing between healthy and unhealthy patients, and another distinguishing the unhealthy patients as MCI or AD. Furthermore, we conducted exploratory data analysis to investigate the associations between cognitive status and the variables in each category of the UDS dataset.

Significant UDS Variables

DEMO	*F-value*	NPT	*F-value*	GDS	χ^2*-value*	FAS	χ^2*-value*
EDUC	466	VEG	9700	MEMPROB	5157	REMDATES	27650
SEX	285	ANIMALS	9102	DROPACT	1845	TAXES	26800
AGE	208	TRAILB	4798	WRTHLESS	1199	BILLS	26785
		TRAILA	4798	BETTER	1182	TRAVEL	25780
		CRAFTDRE	3335	BORED	1147	MEALPREP	21109
		MINTTOTS	1063	HELPLESS	1083	GAMES	20423

Table 1. DEMO (Demographic Information), NPT (Neuropsychological Tests), GDS (Depression Scale), FAS (Behavioral Survey). Table of significant variables determined by ANOVA and χ^2 tests.

ANOVA H_0: There is no significant mean difference in cognitive status given one of the UDS variables.
χ^2 *Test* H_0: There is no relationship between the two categorical variables.

Exploratory Data Analysis

Our response variable, cognitive status, has four categories including healthy control (HC), impaired (Impaired-not-MCI), mild cognitive impairment (MCI), and Alzheimer's Disease (AD). Due to the small sample size of 2,000 impaired subjects out of the 45,000 total, we decided to exclude this category from further modeling and only focus on HC, MCI, and AD.

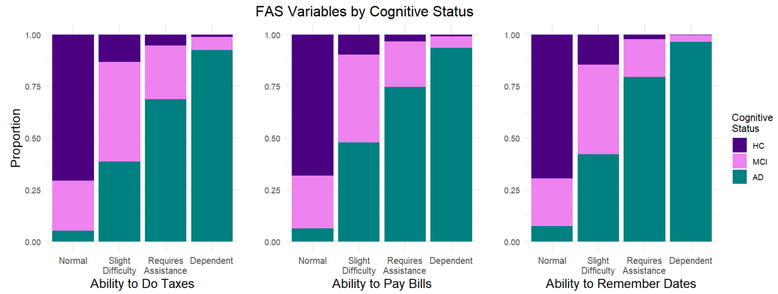

Figure 1. As dependency increases in a person's ability to do taxes, bills, or remembering dates; the proportion of HC (dark purple) patients decreases, while MCI (light purple) and AD (teal) subjects increase.

Modeling

To better understand the distinctions between MCI and AD, we trained the following models:

- Random Forest Classifier
- Logistic Regression
- Single Layered Neural Network
- Support Vector Machine

Starting with the table of variables, we trained a random forest model to classify MCI vs. AD and generated a variable importance plot to further narrow down the variables to build our logistic regression model.

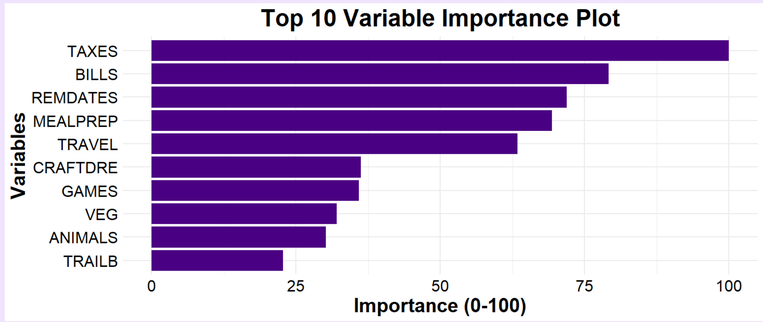

Figure 2. Random Forest MCI vs. AD classification variable importance plot.

Predicted Odds of AD

Ability	Taxes	Pay Bills	Remember Dates
Has difficulty but self sufficient	1.08x	1.71x	1.44x
Requires assistance	2.36x	1.45x	2.49x
Dependent	2.37x	2.98x	4.90x

Table 2. The logistic regression model produces predicted odds times the reference level of normal ability to perform each of the following variables listed. ***Note: There are additional covariates not listed.***

Results

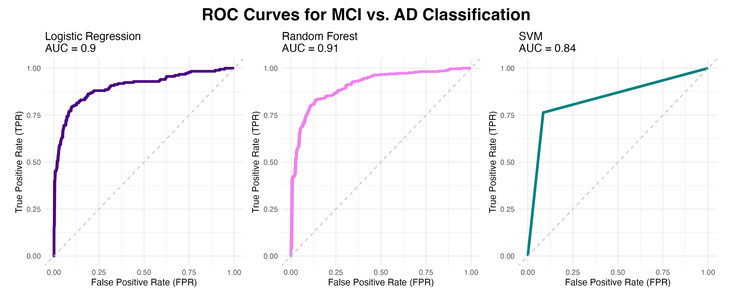

Figure 3. Receiver operating characteristic (ROC) curves show that the random forest and logistic regression model perform similarly in accuracy, as measured by the area under the curve (AUC).

- False Positive Rate (1-Specificity): Proportion of AD cases incorrectly classified as MCI.
- True Positive Rate (Sensitivity): Proportion of MCI cases correctly identified.

Conclusion: The factors that best classify between MCI and AD are simple everyday tasks and NPTs, which can be evaluated using quick and simple surveys and tests. This accessibility allows for easier and earlier detection of AD, which is critical since patients can begin treatment sooner, improving health outcomes on a larger scale.

Future Work: We would like to expand our exploration of additional variables in the MRI and CSF datasets as they remain unused in our current study.

Acknowledgements

This project was funded by the National Science Foundation HDR DSC with grants #2123366, #2123380 and #2123384. We want to thank our PI, Yueqi Ren, and NACC for providing the dataset. We also thank our faculty: Dr. Babak Shahbaba, Dr. Jessica Jaynes, Dr. Roberto Pelayo, Dr. Sam Behseta, Dr. Mine Dogucu, Professor Alma Castro, and our TAs Aubree Krager and Cadee Pinkerton for guiding us throughout the program.

References

[1] Alzheimer's Association. "2023 Alzheimer's Disease Facts and Figures" Alzheimers Dement. 2023; 19: 1598-1695. doi:10.1002/alz.13016. Accessed 19 July 2024.