


Predicting and Evaluating Cognitive Status in Aging Populations Using Decision Tree Models

American Journal of Alzheimer's
Disease & Other Dementias®
Volume 40: 1–13
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15333175251339730
journals.sagepub.com/home/aja



Zhidi Luo, MS^{1,2} , Stella (Ping) Wang, MS, MPH^{1,3},
Emily H. Ho, PhD³, Lihua Yao, PhD³, and Richard C. Gershon, PhD^{3,4}

Abstract

Objective: To improve the identification of cognitive impairment by distinguishing normal cognition (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD). **Methods:** A recursive partitioning tree model was developed using AR-MADA data and the NIH Toolbox, a multidimensional health assessment tool. It incorporated demographic and clinical assessment variables to predict NC, MCI, and AD. Model performance was evaluated using AUC, precision, recall, and F1 score. Robustness was tested through 5-fold cross-validation, sensitivity, scenario, and subgroup analyses. **Results:** The model achieved macro-AUC and micro-AUC scores of 0.92 and 0.91 (training) and 0.89 and 0.86 (testing). Key predictors included the Picture Sequence Memory Test and List Sorting Working Memory Test. Cross-validation yielded 70.22% accuracy and a Kappa of 0.52. **Conclusion:** Machine learning effectively uses a small set of assessments to distinguish NC, MCI, and AD, offering a valuable tool to support clinical decision-making. Future research should validate this model across diverse populations.

Keywords

cognitive impairment, Alzheimer's disease, predictive modeling, machine learning, ARMADA, NIH Toolbox

Introduction

Alzheimer's disease (AD) is the most prevalent type of dementia, affecting an estimated 32 million people globally.¹ In the United States, AD impacts about 10.7% of those aged 65 and older, with its occurrence rising with age.² This condition significantly diminishes individuals' quality of life, impairing their daily functioning, interpersonal relationships, and social engagement. It is the fifth-leading cause of death for Americans over 65, with 121 499 deaths recorded in 2019.² Moreover, caregivers experience substantial challenges and burdens in providing care to individuals with dementia and AD. In 2022, the estimated costs for healthcare, long-term care, and hospice for older adults with dementia totaled \$321 billion.²

Mild cognitive impairment (MCI) is a critical precursor to AD. Individuals with MCI exhibit significant cognitive decline that exceeds normal age-related changes yet does not meet the clinical criteria for dementia.^{3,4} Approximately 22%

of those aged 71 and older are affected by MCI.⁵ The risk of progressing from MCI to dementia is estimated to be 3 to 5 times greater than for those with normal cognitive functions,^{6–8} with an annual rate of progression estimated to be

¹Health Sciences Integrated Program, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

²Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

³Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

⁴Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Corresponding Author:

Zhidi Luo, Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, 710 N Lake Shore Dr, Floor 9, Chicago, IL 60611, USA.

Email: zhidi.luo@northwestern.edu

Data Availability Statement included at the end of the article



Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

12% in the general population and 17% to 20% in those with subtypes of prodromal AD and those with a history of stroke.⁵

Accurately distinguishing between normal cognitive function, MCI, and AD is essential. Early and precise diagnosis of MCI and AD allows for interventions that can slow disease progression, thereby enhancing the quality of life for those affected.⁹ Early detection of cognitive decline helps tailor specific care strategies that meet individual needs effectively, enabling patients and families to better manage the disease's impacts. From a societal point of view, an accurate diagnosis informs more strategic healthcare planning and resource allocation, ensuring that patients receive appropriate treatments at the optimal time, and healthcare systems can better manage the substantial care demands associated with these conditions. This accuracy not only improves patient outcomes but also contributes to the overall efficiency of healthcare services, reducing unnecessary costs and maximizing the use of available medical and supportive resources.^{10,11}

The study “Advancing Reliable Measurement in Alzheimer’s Disease and Cognitive Aging” (ARMADA), a longitudinal multi-site study,^{12,13} used the NIH Toolbox® (NIHTB), a “common currency” measuring multidimensional aspects of health across cognition, emotion, motor, and sensory function for individuals aged 65 to 85 across the cognitive spectrum from normal aging to Alzheimer’s dementia, aiming to predict cognitive decline and its association with Alzheimer’s disease biomarkers.¹² NIHTB is an iPad-based standardized assessment platform that can be easily administered in various settings with minimal training. It offers psychometrically robust, adaptable measures that ensure measurement consistency, support data sharing, and facilitate the integration of findings within research environments. This study aims to improve the identification of cognitive impairment by analyzing the ARMADA dataset to pinpoint key measures across domains that effectively differentiate individuals with normal cognition (NC), MCI, and AD.

Methods

Data Source and Population

This study used the ARMADA dataset,¹² which includes a diverse sample of adults aged 65 and above. The dataset includes assessment data from the baseline visit (year 1), with additional follow-up assessments at 12-month (year 2) and 24-month (year 3) intervals. The sample includes adults with normal cognitive functions (NC), those clinically diagnosed with mild cognitive impairment (MCI), and those diagnosed with AD.

Participants aged over 85 were excluded due to differing recruitment criteria.¹² While participants aged 65 to 85 were selected to represent a range of cognitive health statuses and demographic groups proportionate to the general U.S. population, those over 85 were recruited solely on cognitively

normal individuals and primarily from one site.^{12,14} This group was excluded to maintain consistency in cognitive health representation.

Year 2 and year 3 data were also excluded due to loss of follow-up. In Year 2, 174 participants remained from the original 319 in Year 1, representing a 54.5% retention rate. In Year 3, 107 participants remained from the original 319 in Year 1, representing a 33.5% retention rate. This is largely due to the COVID-19 pandemic, which rendered in-person assessment difficult.

Variables

The following variables are used in this paper.

- (1) Dependent variable: diagnosis outcome, which includes normal cognitive function, mild cognitive impairment, and Alzheimer’s disease.
- (2) Demographic variables: race, age, gender, and education level.
- (3) Clinical assessment uncorrected score variables are (see Table 1 for details): Dimensional Change Card Sort (DCCS), Eriksen Flanker task (Flanker), List Sorting Working Memory (LSWM), Oral Reading Recognition (ORR), Pattern Comparison Processing Speed (PCPS), Picture Sequence Memory (PSM), Picture Vocabulary (TPVT), Anger – Affect (AngerAff), Anger – Hostility (AngerHost), Anger – Physical Aggression (AngerPhysAg), Emotional Support (EmoSup), Fear – Affect (FearAff), Fear – Somatic Arousal (FearSom), Friendship (Friend), General Life Satisfaction (GenLS), Instrumental Support (InstSup), Loneliness (Lone), Meaning and Purpose (MeanP), Perceived Hostility (PHost), Positive Affect (PosAff), Perceived Rejection (PReject), Perceived Stress (PStress), Sadness (Sad), Self-Efficacy (SelfEff), Grip Strength Test – Dominant hand (Gripd), Grip Strength Test – Non-dominant hand (Gripnd), 9-Hole Pegboard Dexterity Test – Dominant hand (Peg9hd), 9-Hole Pegboard Dexterity Test – Non-dominant hand (Peg9hnd), 2-Minute Walk Endurance Test (Walk2Min), 4-Meter Walk Gait Speed Test (Walk4M), Odor Identification Test (Odor), Pain Interference (PainInt), Visual Acuity Test (VisualAc), Words-In-Noise Test – left (WINL), Words-In-Noise Test – right (WINR)

Data Imputation

In this study, all records have less than 10% missing assessment scores. Missing clinical assessment scores were imputed using Multiple Imputation by Chained Equations (MICE) with random forest.¹⁵ This method imputed missing values for each assessment score based on all available

Table 1. Clinical Assessment Score Variable Details.

	Name	Description
DCCS	Dimensional Change Card Sort	A cognitive flexibility task that assesses a person's ability to switch between different rules or sorting criteria
Flanker	Eriksen Flanker task	A test of attention and response inhibition, where individuals must focus on a target stimulus while ignoring distracting stimuli
LSWM	List Sorting Working Memory	Measures working memory by asking individuals to sort items based on size or category
ORR	Oral Reading Recognition	Assesses reading ability and word recognition through a reading-aloud task
PCPS	Pattern Comparison Processing Speed	Evaluates how quickly a person can compare and identify matching visual patterns
PSM	Picture Sequence Memory	Tests episodic memory by requiring individuals to recall the order of pictures presented in a sequence
TPVT	Picture Vocabulary Test	A measure of receptive vocabulary that assesses understanding of word meanings
AngerAff	Anger – Affect	Measures the emotional experience of anger, specifically the intensity of anger feelings
AngerHost	Anger – Hostility	Evaluates cognitive and affective aspects of hostility
AngerPhysAg	Anger – Physical Aggression	Assesses tendencies towards physical expressions of anger, such as aggression
EmoSup	Emotional Support	A measure of perceived emotional support from others
FearAff	Fear – Affect	Assesses the emotional experience of fear, such as feelings of anxiety or apprehension
FearSom	Fear – Somatic Arousal	Measures physical symptoms associated with fear, such as increased heart rate or sweating
Friend	Friendship	Evaluates the quality and depth of friendships, including feelings of companionship
GenLS	General Life Satisfaction	Assesses overall satisfaction with one's life
InstSup	Instrumental Support	Measures perceived availability of help with practical tasks or problem-solving from others
Lone	Loneliness	Evaluates feelings of social isolation or loneliness
MeanP	Meaning and Purpose	Assesses an individual's sense of meaning and purpose in life
PHost	Perceived Hostility	Measures perceptions of hostility or antagonism from others
PosAff	Positive Affect	Evaluates the frequency and intensity of positive emotions, such as happiness and joy
PReject	Perceived Rejection	Assesses feelings of being rejected or excluded by others
PStress	Perceived Stress	Measures the level of stress an individual feels in response to life circumstances
Sad	Sadness	Assesses the emotional experience of sadness or depressive feelings
SelfEff	Self-Efficacy	Measures belief in one's ability to accomplish tasks or goals
Gripd	Grip Strength Test – Dominant hand	Measures the grip strength of the dominant hand as an indicator of physical strength
Gripnd	Grip Strength Test – Non-dominant hand	Measures grip strength in the non-dominant hand
Peg9hd	9-Hole Pegboard Dexterity Test – Dominant hand	Assesses fine motor skills by timing how long it takes to place and remove pegs using the dominant hand
Peg9hnd	9-Hole Pegboard Dexterity Test – Non-dominant hand	Measures fine motor dexterity in the non-dominant hand
Walk2Min	2-Minute Walk Endurance Test	Assesses cardiovascular fitness and endurance by measuring how far a person can walk in 2 min
Walk4M	4-Meter Walk Gait Speed Test	Evaluates gait speed and mobility by timing a person walking four meters
Odor	Odor Identification Test	Tests the ability to identify various odors as a measure of olfactory function
PainInt	Pain Interference	Assesses the impact of pain on daily functioning and activities
VisualAc	Visual Acuity Test	Measures the clarity or sharpness of vision
WINL	Words-In-Noise Test – left	Evaluates the ability to hear and understand words in noisy environments using the left ear
WINR	Words-In-Noise Test – right	Assesses the ability to hear and understand words in noisy environments using the right ear

assessment scores in the dataset, as well as the demographic characteristics of the individual.

Descriptive Analysis

Descriptive analysis was conducted on demographic and clinical measures of all subjects. The analysis included counts, percentages, means, standard deviations, and statistical tests to compare differences among individuals with NC, MCI, and AD. Categorical variables (race, gender, and education level) were assessed using chi-square tests, while continuous variables (age and clinical assessments) were analyzed using t-tests.

Recursive Partitioning Tree Model

The study sample was randomly divided into a training set, which comprised 80% of the sample, and a test set, which comprised 20%. With this split, a recursive partitioning tree model¹⁶ was used to classify individuals into three categories: NC, MCI, and AD. This model recursively splits the dataset based on the most significant variables at each decision point. It selects the variables that best separate the categories. It creates a decision tree that classifies individuals based on demographic information (race, age, gender, and education level) and clinical assessment scores from cognitive, motor, emotional, and sensory measures.

Model Evaluation and Interpretation

To evaluate the model's performance, both multi-class macro-AUC and micro-AUC metrics were used.^{17,18} Macro-AUC provides insight into how well the model performs for each class, regardless of class size, while micro-AUC gives more weight to larger classes and reflects the model's overall performance across the entire dataset. For both training and test sets, macro- and micro-AUC were calculated. The model is considered to perform well with higher AUC values, where 1 indicates perfect classification, and 0.5 suggests random chance. An AUC of 0.7 to 0.8 indicates acceptable discrimination, while values above 0.8 indicate excellent discrimination.¹⁹

Additionally, we assessed the One-Vs-All ROC curve to gain insights into how well the model can identify individual categories within a multi-class classification problem.²⁰ Furthermore, we assessed the model's performance using precision, recall, and F1 score. This helped us understand how accurately the model identified each cognitive state and balanced the trade-off between false positives and false negatives.

Interpreting the decision tree model involved visualizing its structure. Each split highlighted the most informative variables influencing classification. Higher-level splits indicated more influential factors in distinguishing cognitive states. This

visualization can provide clear insights into the key drivers of cognitive function and impairment.

Model Validation

To evaluate the performance and generalizability of the classification model, we employed a five-fold cross-validation,²¹⁻²³ which provides a reliable performance estimate while mitigating overfitting. The dataset was randomly divided into five subsets. Each subset maintained a representative distribution of the cognitive states: NC, MCI, and AD. In each iteration, one fold served as the test set, while the remaining four were used for training. This process was repeated until each fold had been tested. Performance was evaluated using accuracy, calculated as the proportion of correctly classified instances, and Cohen's Kappa, which accounts for chance agreement and offers a more robust measure of effectiveness. Kappa above 0.4 indicates moderate agreement, while a value above 0.6 indicates substantial agreement.²⁴

Model Comparison and Benchmarking

The performance of the recursive partitioning tree model was compared with three benchmark models: Support Vector Machine (SVM),²⁵ Random Forest,²⁶ and Neural Network.²⁷ The SVM model was implemented with a radial basis function kernel. The Random Forest model was specified with 500 trees, and the maximum depth of each tree was left unrestricted. The Neural Network model consisted of a feedforward architecture with one hidden layer containing 10 nodes, a maximum of 200 iterations, and a weight decay of 0.01 to help prevent overfitting.

These models were chosen to evaluate whether they provide superior classification performance in identifying individuals with NC, MCI, and AD. All models were trained and evaluated on the same dataset, and their performance was assessed using macro- and micro-AUC metrics to compare classification results.

Sensitivity Analysis

To further evaluate the robustness of the model and understand the influence of individual predictors on its performance, we conducted a sensitivity analysis using two approaches: Global Sensitivity Analysis (GSA)²⁸⁻³⁰ and Scenario Sensitivity Analysis.^{31,32}

GSA used Sobol's indices to quantify the contribution of each predictor variable to the output variance of the model.²⁸⁻³⁰ This method provides a detailed understanding of how uncertainty in model inputs propagates to uncertainty in model predictions. By calculating the main effect (first-order) and total effect (total Sobol index) for each predictor, we gained a better understanding of their individual and combined impacts. The main effect

represents the contribution of a single predictor to the output variance, assuming all other variables are held constant. In contrast, the total effect captures both the individual contribution of a predictor and its interactions with other variables. This analysis highlighted the most influential predictors, allowing us to prioritize them for further exploration.

Additionally, we performed a Scenario Sensitivity Analysis^{31,32} by systematically removing the most important predictors identified in GSA and observing how the model's performance was affected. This approach enabled us to evaluate the model's stability and adaptability by assessing its ability to maintain performance in the absence of key variables. Furthermore, we monitored which alternative variables were selected by the model in these scenarios, thus revealing potential compensatory relationships among predictors. Collectively, these sensitivity analyses not only highlighted the critical predictors that significantly influence the model's outcomes, but also enhanced our understanding of the model's resilience and flexibility in varying conditions.

Subgroup Analysis

To explore potential variations in the model's performance across different age groups, we conducted a subgroup analysis stratifying the dataset into two distinct categories based on the median age: individuals aged 72 years or younger and those aged 73 years or older. This division allowed us to investigate how age might influence the model's predictive capabilities. Evaluation metrics such as macro- and micro-AUC were used to assess the model's performance within each age group.

Results

Study Cohort and Descriptive Analysis

To establish the cohort for this study (Figure 1), 850 records for 462 unique individuals in the ARMADA database were extracted. The final study population included 319 unique individuals aged between 65 and 85 with baseline assessment data.

Based on the descriptive analysis (Tables 2 and 3), out of the 319 individuals in the study cohort, 161 had normal cognitive function, 90 had MCI, and 68 had AD. The population has an average age of 73.4 (SD = 5.32), with 84.6% white, 55.2% female, and 45.8% with postgraduate degrees, while 34.2% with undergraduate degrees. There is a statistically significant difference among NC, MCI, and AD regarding gender (female percentage: NC: 67.1%; MCI: 40%; AD: 47.1%; p -value < .001). In addition, the NC group has a higher level of education in general, while the AD group has a lower level of education (NC: high school or less: 3.7%; postgraduate: 55.9%; AD: high school or less: 11.8%, postgraduate: 27.9%; p -value < .001).

With respect to clinical assessments, there are statistically significant differences for all cognition tests, including DCCS, Flanker, LSWM, ORR, PCPS, PSM, and TPVT, with NC showing the highest average scores and AD the lowest. Additionally, there is a significant difference for a few of the emotion tests, including AngerHost Test (NC: 42.7; MCI: 45.2; AD: 42.6; p -value = .038), GenLS Test (NC: 56.4; MCI: 53.4; AD: 54.1; p -value = .033), and SelfEff Test (NC: 52.0; MCI: 50.6; AD: 48.6; p -value = .023). Finally, many motor tests and sensation tests also show significant differences. For motor tests, Peg9hd and

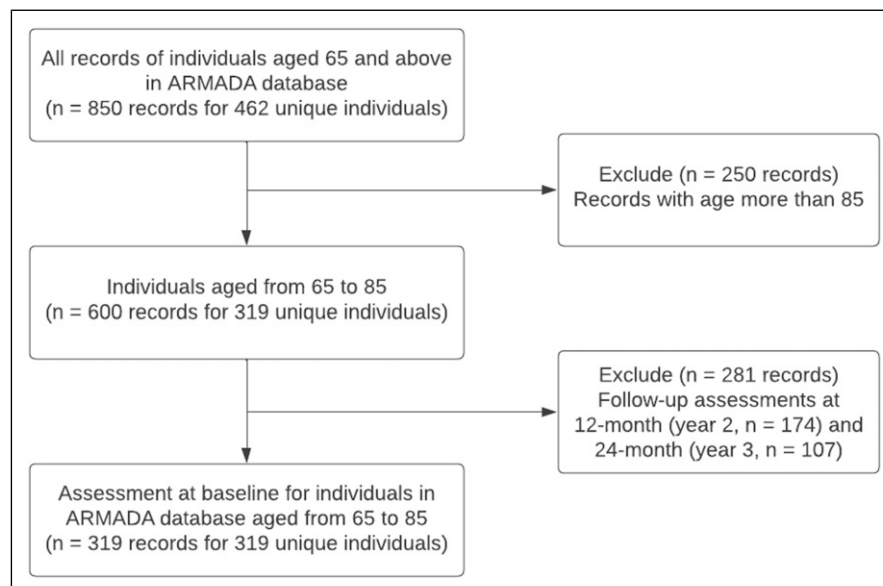


Figure 1. Flow Diagram of the Study Population.

Table 2. Demographic Characteristics by Cognitive States.

	Overall (N = 319)	NC (N = 161)	MCI (N = 90)	AD (N = 68)	p-value
Age	73.4 (5.32)	72.7 (5.07)	74.3 (5.07)	74.0 (6.01)	.05
Race					.239
White	270 (84.6%)	133 (82.6%)	75 (83.3%)	62 (91.2%)	
Non-White	49 (15.4%)	28 (17.4%)	15 (16.7%)	6 (8.8%)	
Gender					<.001
Female	176 (55.2%)	108 (67.1%)	36 (40.0%)	32 (47.1%)	
Male	143 (44.8%)	53 (32.9%)	54 (60.0%)	36 (52.9%)	
Education					<.001
High School or less	20 (6.3%)	6 (3.7%)	6 (6.7%)	8 (11.8%)	
Postgraduate Degree	146 (45.8%)	90 (55.9%)	37 (41.1%)	19 (27.9%)	
Some College, no degree	44 (13.8%)	13 (8.1%)	21 (23.3%)	10 (14.7%)	
Undergraduate Degree	109 (34.2%)	52 (32.3%)	26 (28.9%)	31 (45.6%)	

Notes for abbreviations: NC: normal cognition; MCI: mild cognitive impairment; AD: Alzheimer's disease.

Peg9hnd show significant differences (NC: 95.4 and 96.7; MCI: 91.5 and 93.5; and AD: 84.3 and 86.6; p -value <.001), and Walk2Min Test (NC: 89.3; MCI: 85.8; AD: 76.5; p -value <.001). For sensation tests, there is a significant difference for Odor Identification test (NC: 91.3; MCI: 75.1; AD: 60.0; p -value <.001), and Words-In-Noise Test, both left (NC: 17.3; MCI: 14.6; AD: 13.9; p -value <.001) and right (NC: 18.6; MCI: 16.3; AD: 14.9; p -value <.001).

Recursive Partitioning Tree Model

The study cohort was randomly split into 255 (80%) for training set and 64 (20%) for test set. The macro-AUC and micro-AUC were 0.92 and 0.91 for the training set, suggesting that the model achieved high accuracy in predicting cognitive states. When applied to the independent test set, the macro-AUC and micro-AUC scores were slightly lower, at 0.89 and 0.86, respectively. Despite this minor reduction in performance, the results suggest that the model retained its predictive power when faced with unseen data.

In addition to the AUC metrics, the model's confusion matrix (Figure 2) and One-Vs-All ROC curves (Figure 3) were presented, and precision and recall scores were evaluated for each cognitive category (Table 4). The precision for AD was at 0.78, indicating that 78% of the time, the model's predictions for AD were correct. The recall for AD was notably high at 0.84, demonstrating the model's proficiency in identifying most actual cases of AD, which is essential for timely clinical interventions. Similarly, for normal cognitive function, the precision and recall scores of 0.83 and 0.88, respectively, suggested strong reliability in the model's predictions. Conversely, the model exhibited some challenges with MCI classification, as indicated by a precision score of 0.73 and a recall score of 0.61. This discrepancy may suggest difficulties in accurately distinguishing MCI from other cognitive states, possibly due

to shared characteristics between MCI and AD, as well as between MCI and NC.

For interpretation (Figure 4), the decision tree selected used only nine assessments for prediction, starting with the PSM for episodic memory, and were further assessed using the LSWM. The majority of individuals with NC were predicted by high PSM scores, indicating intact episodic memory function. Conversely, most individuals with MCI were characterized by low PSM scores and high LSWM scores, suggesting compromised episodic memory alongside preserved working memory capacity. In contrast, most individuals with AD demonstrated low scores on both PSM and LSWM, indicating impairment in both episodic and working memory domains. This could be attributed to the progressive nature of AD, where both episodic and working memory types are affected as the disease advances. To enhance prediction accuracy, the tree model branches further based on specific PSM and LSWM score ranges. Those with high PSM (PSM \geq 90) and high LSWM (LSWM \geq 88) scores were further evaluated using PHost score to discern between NC and MCI. Those with low PSM (PSM < 90) and high LSWM (LSWM \geq 72) scores were subsequently evaluated using the PCPS test, followed by either GenLS Test and AngerHost Test to distinguish between NC and MCI, or the Peg9hnd Test and Walk4M Test to distinguish between MCI and AD. Finally, those with low scores on both PSM (PSM < 90) and LSWM (LSWM < 72) were assessed using the Flanker task to discern between MCI and AD.

5-Fold Cross-Validation

The results from the five-fold cross-validation demonstrate an overall accuracy of 70.22%, which suggests a solid level of performance in distinguishing among the cognitive states, and a Kappa statistic of 0.52, which indicates moderate agreement. These findings suggest that the model

Table 3. Clinical Assessment by Cognitive States.

	Overall (N = 319)	NC (N = 161)	MCI (N = 90)	AD (N = 68)	p-value
DCCS	94.5 (13.2)	100 (7.79)	93.7 (10.9)	82.1 (17.1)	<.001
Flanker	88.0 (11.9)	93.2 (7.45)	88.1 (8.81)	75.7 (14.7)	<.001
LSWM	88.7 (17.0)	98.4 (10.5)	84.7 (14.5)	71.3 (16.4)	<.001
ORR	110 (6.88)	112 (5.43)	108 (7.88)	106 (6.92)	<.001
PCPS	80.0 (16.5)	86.8 (14.0)	78.2 (13.6)	66.3 (16.7)	<.001
PSM	90.9 (13.4)	98.8 (13.1)	85.5 (9.00)	79.6 (4.31)	<.001
TPVT	116 (10.8)	120 (9.18)	112 (9.45)	109 (10.9)	<.001
AngerAff	47.4 (9.38)	47.2 (8.30)	47.3 (9.83)	47.9 (11.2)	.851
AngerHost	43.4 (8.06)	42.7 (7.49)	45.2 (8.56)	42.6 (8.42)	.038
AngerPhysAg	45.3 (6.34)	44.6 (5.16)	45.7 (6.99)	46.3 (7.73)	.146
EmoSup	47.4 (8.32)	47.8 (8.51)	47.0 (8.35)	46.9 (7.88)	.651
FearAff	50.2 (9.54)	49.3 (8.76)	51.6 (9.60)	50.3 (11.0)	.168
FearSom	47.6 (8.80)	47.5 (7.91)	47.4 (8.89)	48.1 (10.6)	.856
Friend	50.3 (9.00)	51.0 (9.69)	50.0 (8.16)	49.0 (8.29)	.285
GenLS	55.0 (9.35)	56.4 (8.86)	53.4 (9.10)	54.1 (10.4)	.033
InstSup	50.9 (9.54)	49.8 (10.4)	51.4 (9.30)	52.8 (7.32)	.081
Lone	50.2 (9.38)	49.8 (9.18)	49.9 (8.98)	51.4 (10.4)	.467
MeanP	49.5 (8.69)	50.3 (8.76)	48.7 (8.28)	48.6 (9.02)	.267
PHost	49.0 (8.18)	48.9 (7.79)	49.7 (8.56)	48.5 (8.66)	.623
PosAff	48.0 (7.33)	48.7 (7.44)	47.5 (7.48)	47.2 (6.80)	.272
PReject	48.2 (9.03)	47.9 (8.59)	48.9 (9.16)	48.1 (9.93)	.686
PStress	46.1 (9.20)	45.2 (8.10)	47.2 (9.71)	46.8 (10.8)	.197
Sad	48.2 (10.1)	47.3 (10.1)	49.0 (9.33)	49.4 (10.9)	.245
SelfEff	50.9 (8.77)	52.0 (8.34)	50.6 (8.79)	48.6 (9.34)	.023
Gripd	96.6 (10.1)	96.4 (9.64)	97.6 (10.2)	95.7 (11.1)	.496
Gripnd	96.6 (10.1)	96.6 (9.91)	97.3 (9.74)	95.6 (11.2)	.591
Peg9hd	91.9 (13.3)	95.4 (10.3)	91.5 (11.1)	84.3 (18.3)	<.001
Peg9hnd	93.7 (10.1)	96.7 (8.32)	93.5 (7.65)	86.6 (12.9)	<.001
Walk2Min	85.6 (14.1)	89.3 (12.4)	85.8 (13.1)	76.5 (15.2)	<.001
Walk4M	3.85 (1.77)	3.83 (2.32)	3.65 (0.685)	4.16 (1.09)	.202
Odor	80.1 (24.9)	91.3 (19.5)	75.1 (24.4)	60.0 (22.6)	<.001
PainInt	49.9 (8.22)	50.0 (8.18)	50.2 (8.80)	49.2 (7.57)	.723
VisualAc	92.5 (13.0)	93.4 (11.9)	91.0 (14.8)	92.6 (13.1)	.394
WINL	15.8 (7.00)	17.3 (6.21)	14.6 (7.40)	13.9 (7.56)	<.001
WINR	17.2 (6.99)	18.6 (6.03)	16.3 (6.87)	14.9 (8.43)	<.001

Notes for abbreviations: NC: normal cognition; MCI: mild cognitive impairment; AD: Alzheimer's disease; DCCS: Dimensional Change Card Sort; Flanker: Eriksen Flanker task; LSWM: List Sorting Working Memory; ORR: Oral Reading Recognition; PCPS: Pattern Comparison Processing Speed; PSM: Picture Sequence Memory; TPVT: Picture Vocabulary; AngerAff: Anger – Affect; AngerHost: Anger – Hostility; AngerPhysAg: Anger – Physical Aggression; EmoSup: Emotional Support; FearAff: Fear – Affect; FearSom: Fear – Somatic Arousal; Friend: Friendship; GenLS: General Life Satisfaction; InstSup: Instrumental Support; Lone: Loneliness; MeanP: Meaning and Purpose; PHost: Perceived Hostility; PosAff: Positive Affect; PReject: Perceived Rejection; PStress: Perceived Stress; Sad: Sadness; SelfEff: Self-Efficacy; Gripd: Grip Strength Test – Dominant hand; Gripnd: Grip Strength Test - Non-dominant hand; Peg9hd: 9-Hole Pegboard Dexterity Test - Dominant hand; Peg9hnd: 9-Hole Pegboard Dexterity Test - Non-dominant hand; Walk2Min: 2-Minute Walk Endurance Test; Walk4M: 4-Meter Walk Gait Speed Test; Odor: Odor Identification Test; PainInt: Pain Interference; VisualAc: Visual Acuity Test; WINL: Words-In-Noise Test – left; WINR: Words-In-Noise Test – right.

makes informed decisions based on the input features and effectively captures the underlying patterns in the data beyond random guessing.

Model Comparison and Benchmarking

The Recursive Partitioning Tree model was compared to three benchmark models: SVM, Random Forest, and Neural Network (Table 5). The Recursive Partitioning Tree model

achieved macro-AUC of 0.92 and micro-AUC of 0.91 on the training set, and 0.89 and 0.86 on the test set, respectively. The SVM model demonstrated high performance with a macro-AUC of 0.98 and micro-AUC of 0.97 on the training set, but performance dropped to 0.87 and 0.83 on the test set, suggesting some overfitting. The Random Forest model showed perfect results on the training set with a macro-AUC and micro-AUC of 1.00, but decreased on the test set, with a macro-AUC of 0.88 and a micro-AUC of 0.84. The Neural

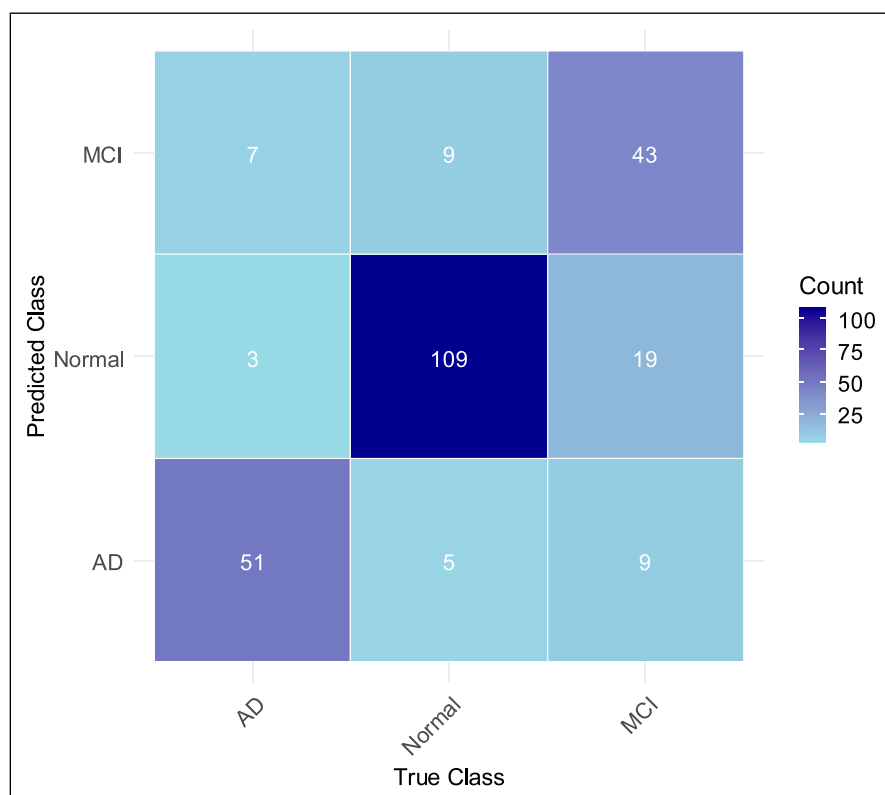


Figure 2. Confusion Matrix.

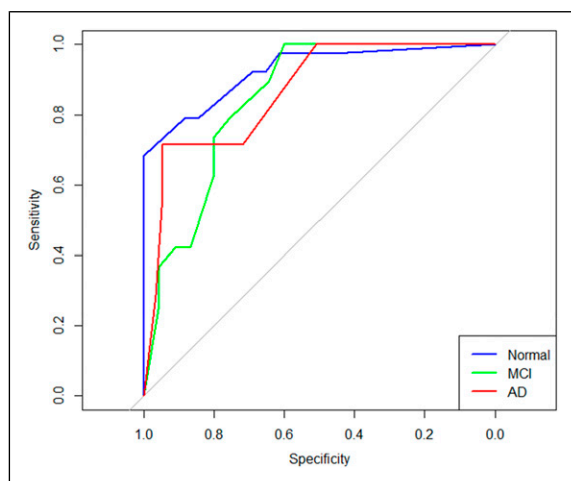


Figure 3. One-Vs-All ROC Curves.

Network model had the weakest performance, achieving a macro-AUC of 0.66 and micro-AUC of 0.81 on the training set, and 0.68 and 0.72 on the test set.

In summary, while SVM, Random Forest, and Recursive Partitioning Tree performed well, Neural Network struggled. SVM exhibited the highest performance on the training set but was overfitted, while Random Forest and Recursive Partitioning Tree offered more balanced results, with Recursive

Table 4. Precision, Recall, and F1 Score.

	Precision	Recall	F1
AD	0.78	0.84	0.81
Normal	0.83	0.89	0.86
MCI	0.73	0.61	0.66

Partitioning Tree showing the most consistent performance across both training and test sets.

Global Sensitivity Analysis

The Global Sensitivity Analysis (Figure 5) indicates that PSM emerges as the most significant predictor, with a main effect of 0.43, which underscores its strong influence on model outcomes. LSWM follows as the second most important predictor, with an index of 0.24, suggesting a meaningful contribution to the model as well. Conversely, PHost, PCPS, and GenLS have notably lower indices of 0.11, 0.01, and 0.10, respectively, indicating their lesser impact on the model's predictive power. Additionally, the total effect reveals interaction among parameters, particularly evident in the shifts of total indices (total effect) when key predictors are modified. PSM's total effect of 0.74 confirms that it is consistently a key predictor, both on

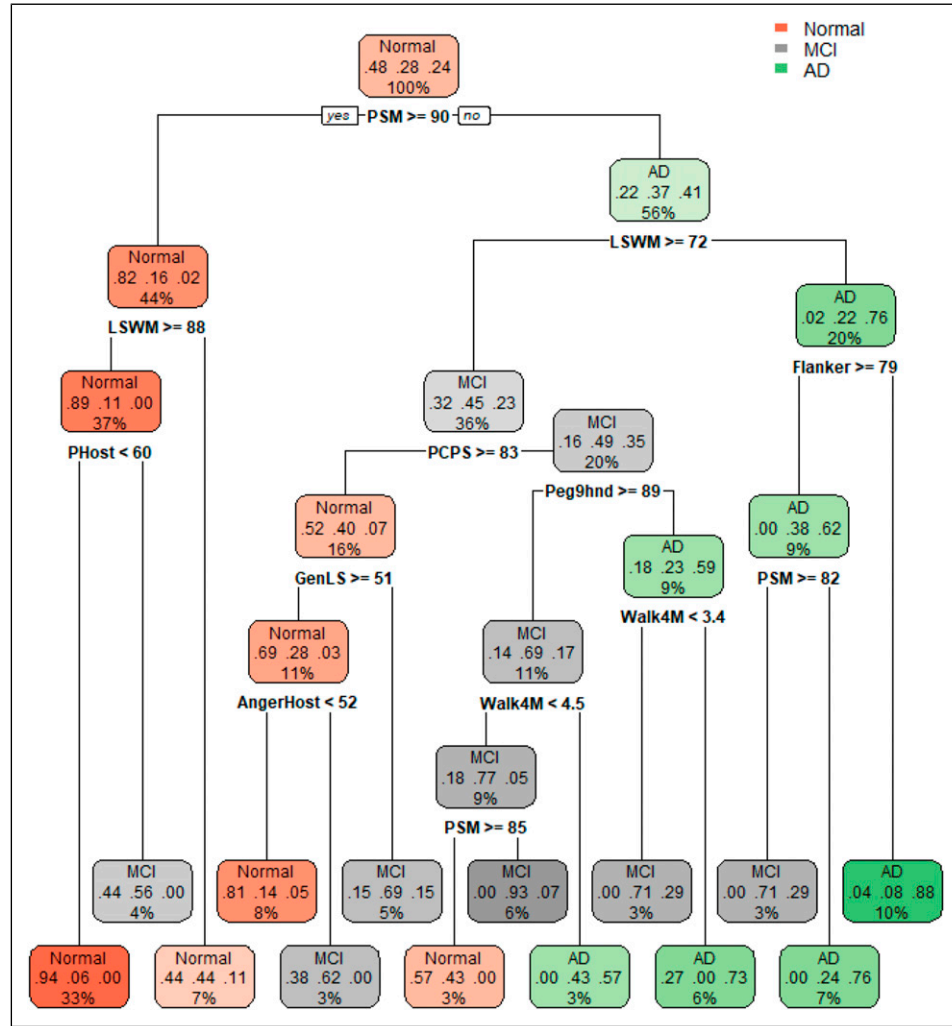


Figure 4. Decision Tree for Cognitive Assessment. Notes for Interpretation: The Percentage Displayed at the Bottom of Each Box Represents the Proportion of Data Contained Within that Box. The Three Numbers Within Each Box Indicate the Percentage of Data Corresponding to the NC, MCI, and AD Groups. For Example, the Bottom Left Box Shows a Percentage of 33%, Indicating that it Includes 33% of the Sample. The Three Numbers - 0.94, 0.06, and 0.00 - Indicate that 94% of the Samples in this Box Belong to the NC Group, While 6% are in the MCI Group. Notes for Abbreviations: MCI: Mild Cognitive Impairment; AD: Alzheimer's Disease; PSM: Picture Sequence Memory; LSWM: List Sorting Working Memory; PHost: Perceived Hostility; PCPS: Pattern Comparison Processing Speed; GenLS: General Life Satisfaction; AngerHost: Anger – Hostility; Peg9hnd: 9-Hole Pegboard Dexterity Test - Non-dominant Hand; Walk4M: 4-Meter Walk Gait Speed Test; Flanker: Eriksen Flanker Task

Table 5. AUC Comparison for Recursive Partitioning Tree, SVM, Random Forest, and Neural Network

	Training set AUC		Test set AUC	
	Macro-AUC	Micro-AUC	Macro-AUC	Micro-AUC
Recursive Partitioning Tree	0.92	0.91	0.89	0.86
SVM	0.98	0.97	0.87	0.83
Random Forest	1	1	0.88	0.84
Neural Network	0.66	0.81	0.68	0.72

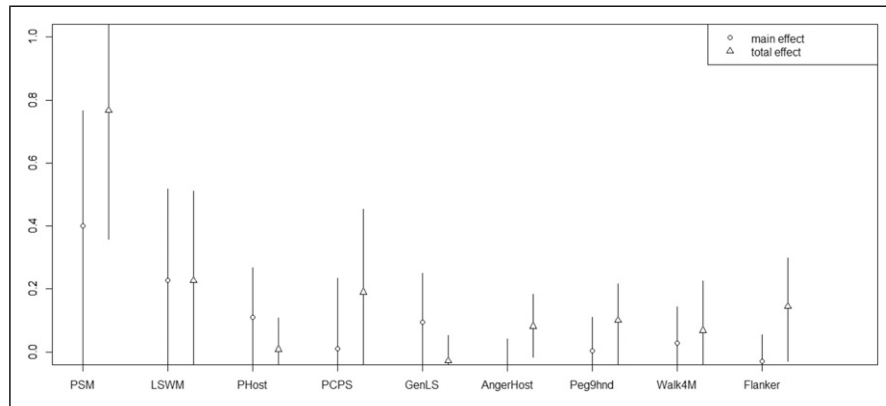


Figure 5. Sobol's Indices From Global Sensitivity Analysis

its own and when interacting with other variables. The total effect's broader confidence intervals for other predictors, such as AngerHost and Flanker, suggest that these variables might need closer examination.

Scenario Analysis

The scenario sensitivity analysis revealed valuable insights into the robustness of the model's performance when key predictors were removed. In the first scenario, we removed PSM, the most important predictor, which resulted in a decrease in both in-sample macro- and micro-AUC from 0.92 to 0.86 and from 0.91 to 0.87, respectively, while out-of-sample AUC fell from 0.89 to 0.76 and from 0.86 to 0.75. Despite this reduction, the model maintained an acceptable predictive power, with LSWM becoming the most significant predictor. However, the AUC values in this scenario were lower than when PSM was included, indicating that LSWM is not fully interchangeable with PSM but serves as an acceptable substitute in PSM's absence. Secondary predictors PCPS and Flanker also gained prominence, underscoring the model's adaptability by leveraging alternative cognitive features for classification.

In the second scenario, LSWM was removed, and in-sample AUC values remained similar to the original model (0.90 macro-AUC and 0.91 micro-AUC), but out-of-sample performance dropped from 0.89 to 0.76 (macro-AUC) and from 0.86 to 0.75 (micro-AUC). This finding further suggests that the model can still effectively capture relevant cognitive patterns even when a key predictor is absent. This pattern suggests that while the model can achieve strong predictive performance without LSWM on the training data, the predictor plays a key role in generalizing to an independent dataset. LSWM, therefore, appears crucial for maintaining the model's stability and predictive accuracy across broader contexts. In the absence of LSWM, Flanker and PHost gained prominence as significant predictors. Flanker, linked to cognitive control, indicates the model's reliance on cognitive processes, while PHost adds a social-emotional dimension, implying that perceived hostility may interact with or influence cognitive performance.

Subgroup Analysis

In the subgroup analysis, the model's performance was evaluated for two distinct age groups: individuals aged 62 to 72 and those aged 73 to 84. The younger group (62 to 72 years) exhibited training AUC values of 0.94 and out-of-sample values of 0.77 (for both macro- and micro-AUC), while the older group (73 to 84 years) showed training AUC values of 0.90, with out-of-sample values of 0.76 (macro) and 0.78 (micro), demonstrating the model's consistent effectiveness across both demographics. The minimal performance differences across age groups and the alignment of predictive factors with those of the overall model highlight its robustness, suggesting that the model effectively captures critical variables related to cognitive impairments regardless of age. Additionally, both age-specific models exhibit features that closely align with those of the overall model, suggesting that the model effectively captures critical variables relevant to cognitive impairments regardless of age.

Discussion

Summary of Findings and Key Implications

Using the ARMADA dataset, this study used a decision tree model to identify key clinical assessments across cognition, emotion, motor, and sensation domains to distinguish individuals with normal cognition, MCI, and AD. PSM and LSWM were the most influential features, while PCPS, Peg9hnd, Walk4M, and Flanker also contributed to classification. Additionally, self-reported measures such as PHost, GenLS, and AngerHost played a role in distinguishing cognitive states. The predictive power of PSM and LSWM aligns with prior research emphasizing the centrality of cognitive assessments in delineating cognitive states.^{33,34} Furthermore, the inclusion of additional predictive variables is consistent with the emerging evidence that non-cognitive factors, such as fine motor skills³⁵⁻³⁷ and emotional regulation,³⁸ may offer valuable insights for differentiating cognitive states.

The robustness and generalizability of the model were evaluated using a range of analytical approaches. Five-fold cross-validation was employed to reduce the risk of overfitting and to provide a more reliable estimate of model accuracy. Global Sensitivity Analysis was conducted to identify the most influential predictors, notably emphasizing the PSM as a critical variable in the model. Furthermore, scenario analysis was performed to examine how the model operates, ensuring its adaptability and robustness. Subgroup analysis was also utilized to confirm that the model's effectiveness remained consistent across different age groups, thereby validating its applicability in diverse clinical settings. These methodologies enriched the interpretation of the model's performance, reinforcing its robustness and generalizability.

Importance of Non-Cognitive Factors

This study not only incorporated a variety of cognitive assessments, such as PSM and LSWM, but also highlighted the significance of non-cognitive measures, including fine motor skills (Peg9hnd), gait speed (Walk4M), emotional regulation (PHost, AngerHost), and general life satisfaction (GenLS). These non-cognitive variables provided additional layers of insight that enhanced the model's predictive power, as they are likely to reflect broader neurological and psychological processes that influence cognitive function. Previous research has shown that fine motor skills and emotional regulation are not only markers of neurological integrity, but also influence cognitive performance in individuals with varying levels of cognitive decline.³⁸⁻⁴¹ Additionally, integrating these non-cognitive measures into longitudinal studies could offer a deeper understanding of their role in cognitive decline progression.⁴⁰

Comparison with Prior Studies

Early diagnosis enables prompt access to support services, therapeutic options, and lifestyle adjustments that may positively influence cognitive outcomes. While traditional diagnostic procedures for cognitive impairment often rely on a combination of biomarker testing and neuropsychological assessments, which often require specialized equipment,⁴² this study explores the potential of a data-driven approach using a machine-learning predictive model. Our model incorporates four cognitive tests, two motor function tests, and three self-reported emotion measures to differentiate between NC, MCI, and AD. This method adds to the growing evidence of the utility of machine-learning models in identifying people with MCI and AD.

This study achieved macro-AUC and micro-AUC scores of 0.92 and 0.91 on the training set, and 0.89 and 0.86 on the test set, respectively. By comparison, Park et al⁴³ reported AUCs of 69.5 with linear regression, 70.6 with support vector machine (SVM), and 76.1 for random forest. Moore et al reported an AUC of 0.82 using random forest,⁴⁴ while Revathi et al⁴⁵

reported AUCs of 0.90 with SVM and 0.74 with random forest, though it is unclear if these values are in-sample or out-of-sample predictions. In conclusion, this study's model demonstrated competitive and, in most cases, superior predictive performance compared to previous approaches.

Limitations

This study has several limitations that may affect the generalizability of its findings. The sample is predominantly white and highly educated, which may limit the applicability of our results to more diverse populations. Cognitive risk factors, healthcare access, and socio-economic influences vary across racial and educational groups, suggesting that the model may require further validation and refinement in underrepresented populations. The lack of racial diversity also prevents subgroup analyses on cognitive outcomes across different demographic groups. Additionally, the exclusion of individuals over 85 years, a group at elevated risk for Alzheimer's disease, limits our ability to examine cognitive risk factors specific to the oldest adults, reducing the comprehensiveness of our findings on late-life cognitive impairment. Furthermore, the omission of follow-up data from years 2 and 3 due to participant attrition constrains the broader applicability of the results, and it restricts our evaluation of the model's predictive performance over time. Moreover, while our decision tree model achieved high accuracy in predicting cognitive states, it may not capture the full complexity of cognitive impairment, and further refinement and validation are warranted. Lastly, the cross-sectional nature of our study precludes causal inference, and longitudinal research is needed to validate our findings and assess the predictive utility of our model over time.

Future Research

Future research should aim to address the limitations identified in this study and further explore the potential of non-cognitive measures in cognitive health assessment. Longitudinal studies that follow participants over time would provide valuable insights into the trajectory of cognitive decline and the predictive validity of our model. In particular, Random Survival Forests could be a promising method for analyzing survival data and assessing the model's ability to predict cognitive decline over time.^{46,47} Additionally, efforts to refine and validate our decision tree approach in diverse populations and clinical settings would enhance its utility and applicability. Furthermore, incorporating other machine learning methods, such as random forests, support vector machines, or neural networks, could provide a more robust comparison and potentially improve predictive accuracy. Finally, investigating the underlying mechanisms linking non-cognitive factors to cognitive impairment could shed light on novel targets for intervention and prevention strategies. Overall, continued research in this area holds promise for advancing our

understanding of cognitive health and improving diagnostic and therapeutic approaches for individuals at risk of cognitive decline.

ORCID iD

Zhidi Luo  <https://orcid.org/0009-0006-7475-179X>

Ethical Statement

Ethical Considerations

This study is a secondary analysis using data collected from the original study, which received ethical approval from Northwestern University. The data used for this analysis is de-identified and adheres to privacy and confidentiality guidelines. Given the secondary nature of the analysis and the use of de-identified data, no additional ethical approval was required.

Author Contributions

All authors contributed to the study's conception and design. Material preparation and data collection were performed by Dr. Emily H. Ho, Dr. Lihua Yao, and Dr. Richard C. Gershon. Study design and analysis were performed by Zhidi Luo, Stella (Ping) Wang, supervised by Emily H. Ho. The first draft of the manuscript was written by Zhidi Luo, and all authors critically reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by Federal funds from the National Institute on Aging, National Institutes of Health, under grant No. U2CAG057441 (Gershon, Weintraub). The analysis presented in this manuscript did not receive any additional funding.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data is planned to be made publicly available within the next 12 months. Until then, it can be accessed upon request.

References

1. Gustavsson A, Norton N, Fast T, et al. Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's Dement.* 2023;19(2):658-670. doi:[10.1002/alz.12694](https://doi.org/10.1002/alz.12694)
2. 2022 Alzheimer's disease facts and figures. *Alzheimer's & Dementia.* 2022;18(4):700-789. doi:[10.1002/alz.12638](https://doi.org/10.1002/alz.12638)
3. Jack CR, Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.* 2018;14(4):535-562. doi:[10.1016/j.jalz.2018.02.018](https://doi.org/10.1016/j.jalz.2018.02.018)
4. Kasper S, Bancher C, Eckert A, et al. Management of Mild Cognitive Impairment (MCI): the need for national and international guidelines. *World J Biol Psychiatr.* 2020;21(8):579-594. doi:[10.1080/15622975.2019.1696473](https://doi.org/10.1080/15622975.2019.1696473)
5. Plassman BL, Langa KM, Fisher GG, et al. Prevalence of cognitive impairment without dementia in the United States. *Ann Intern Med.* 2008;148(6):427-434. doi:[10.7326/0003-4819-148-6-200803180-00005](https://doi.org/10.7326/0003-4819-148-6-200803180-00005)
6. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol.* 1999;56(3):303-308. doi:[10.1001/archneur.56.3.303](https://doi.org/10.1001/archneur.56.3.303)
7. Petersen RC, Roberts RO, Knopman DS, et al. Mild cognitive impairment: ten years later. *Arch Neurol.* 2009;66(12):1447-1455. doi:[10.1001/archneurol.2009.266](https://doi.org/10.1001/archneurol.2009.266)
8. Yaffe K, Petersen RC, Lindquist K, Kramer J, Miller B. Subtype of mild cognitive impairment and progression to dementia and death. *Dement Geriatr Cogn Disord.* 2006;22(4):312-319. doi:[10.1159/000095427](https://doi.org/10.1159/000095427)
9. Sabbagh MN, Boada M, Borson S, et al. Rationale for early diagnosis of mild cognitive impairment (MCI) supported by emerging digital technologies. *J Prev Alzheimers Dis.* 2020;7(3):158-164. doi:[10.14283/jpad.2020.19](https://doi.org/10.14283/jpad.2020.19)
10. Alderwick H, Hutchings A, Briggs A, Mays N. The impacts of collaboration between local health care and non-health care organizations and factors shaping how they work: a systematic review of reviews. *BMC Public Health.* 2021;21(1):753. doi:[10.1186/s12889-021-10630-1](https://doi.org/10.1186/s12889-021-10630-1)
11. Lyketsos CG. Prevention of unnecessary hospitalization for patients with dementia: the role of ambulatory care. *JAMA.* 2012;307(2):197-198. doi:[10.1001/jama.2011.2005](https://doi.org/10.1001/jama.2011.2005)
12. Weintraub S, Karpouzian-Rogers T, Peipert JD, et al. ARMADA: assessing reliable measurement in Alzheimer's disease and cognitive aging project methods. *Alzheimer's Dement.* 2022;18(8):1449-1460. doi:[10.1002/alz.12497](https://doi.org/10.1002/alz.12497)
13. Karpouzian-Rogers T, Ho E, Novack M, et al. Baseline characterization of the ARMADA (assessing reliable measurement in Alzheimer's disease) study cohorts. *Alzheimer's Dement.* 2023;19(5):1974-1982. doi:[10.1002/alz.12816](https://doi.org/10.1002/alz.12816)
14. Mather MA, Ho EH, Bedjeti K, et al. Measuring multidimensional aspects of health in the oldest old using the NIH Toolbox: results from the ARMADA study. *Arch Clin Neuropsychol.* 2024;39(5):535-546. doi:[10.1093/arclin/acad105](https://doi.org/10.1093/arclin/acad105)
15. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res.* 2011;20(1):40-49. doi:[10.1002/mp.329](https://doi.org/10.1002/mp.329)
16. Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees. *Wadsworth Int Group.* 1984;37(15):237-251.
17. Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for Multiple class classification problems. *Mach Learn.* 2001;45(2):171-186. doi:[10.1023/A:1010920819831](https://doi.org/10.1023/A:1010920819831)
18. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett.* 2006;27(8):861-874. doi:[10.1016/j.patrec.2005.10.010](https://doi.org/10.1016/j.patrec.2005.10.010)
19. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5(9):1315-1316. doi:[10.1097/JTO.0b013e3181ec173d](https://doi.org/10.1097/JTO.0b013e3181ec173d)

20. Rifkin R, Klautau A. In defense of one-vs-all classification. *Journal of Machine Learning Research*. Published online 2004.
21. Kuhn M. Futility analysis in the cross-validation of machine learning models. Published online 2014. doi:[10.48550/ARXIV.1405.6974](https://doi.org/10.48550/ARXIV.1405.6974)
22. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Stat Comput*. 2011;21(2):137-146. doi:[10.1007/s11222-009-9153-8](https://doi.org/10.1007/s11222-009-9153-8)
23. Wong TT, Yeh PY. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans Knowl Data Eng*. 2020;32(8):1586-1594. doi:[10.1109/TKDE.2019.2912815](https://doi.org/10.1109/TKDE.2019.2912815)
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
25. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297. doi:[10.1007/BF00994018](https://doi.org/10.1007/BF00994018)
26. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
27. Ripley BD. *Pattern Recognition and Neural Networks*. Cambridge University Press; 1996. doi:[10.1017/CBO9780511812651](https://doi.org/10.1017/CBO9780511812651)
28. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput Phys Commun*. 2010;181(2):259-270. doi:[10.1016/j.cpc.2009.09.018](https://doi.org/10.1016/j.cpc.2009.09.018)
29. Iooss B, Lemaître P. A review on global sensitivity analysis methods. In: Dellino G, Meloni C, eds. *Uncertainty management in simulation-optimization of complex systems*. Operations Research/Computer Science Interfaces Series. Springer US. doi:[10.1007/978-1-4899-7547-8_5](https://doi.org/10.1007/978-1-4899-7547-8_5); 2015;Vol 59:101-122.
30. Sobol' IM, Tarantola S, Gatelli D, Kucherenko SS, Mauntz W. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliab Eng Syst Saf*. 2007;92(7):957-960. doi:[10.1016/j.ress.2006.07.001](https://doi.org/10.1016/j.ress.2006.07.001)
31. Gao L, Bryan BA, Nolan M, Connor JD, Song X, Zhao G. Robust global sensitivity analysis under deep uncertainty via scenario analysis. *Environ Model Software*. 2016;76:154-166. doi:[10.1016/j.envsoft.2015.11.001](https://doi.org/10.1016/j.envsoft.2015.11.001)
32. Borgonovo E, Plischke E. Sensitivity analysis: a review of recent advances. *Eur J Oper Res*. 2016;248(3):869-887. doi:[10.1016/j.ejor.2015.06.032](https://doi.org/10.1016/j.ejor.2015.06.032)
33. Ding H, Wang B, Hamel AP, et al. Prediction of progression from mild cognitive impairment to Alzheimer's disease with longitudinal and multimodal data. *Front Dement*. 2023;2:1271680. doi:[10.3389/frdem.2023.1271680](https://doi.org/10.3389/frdem.2023.1271680)
34. Thabtah F, Spencer R, Ye Y. The correlation of everyday cognition test scores and the progression of Alzheimer's disease: a data analytics study. *Health Inf Sci Syst*. 2020;8(1):24. doi:[10.1007/s13755-020-00114-8](https://doi.org/10.1007/s13755-020-00114-8)
35. Hesseberg K, Tangen GG, Pripp AH, Bergland A. Associations between cognition and hand function in older people diagnosed with mild cognitive impairment or dementia. *Dement Geriatr Cogn Dis Extra*. 2020;10(3):195-204. doi:[10.1159/000510382](https://doi.org/10.1159/000510382)
36. Liu X, Abudukeremu A, Jiang Y, et al. Fine or gross motor index as a simple tool for predicting cognitive impairment in elderly people: findings from the Irish longitudinal study on ageing (TILDA). *J Alzheimers Dis*. 2021;83(2):889-896. doi:[10.3233/JAD-210704](https://doi.org/10.3233/JAD-210704)
37. Ntracha A, Iakovakis D, Hadjidimitriou S, Charisis VS, Tsolaki M, Hadjileontiadis LJ. Detection of mild cognitive impairment through natural language and Touchscreen typing processing. *Front Digit Health*. 2020;2:567158. doi:[10.3389/fdgth.2020.567158](https://doi.org/10.3389/fdgth.2020.567158)
38. Liu M, Ma J, Fu CY, et al. Dysfunction of emotion regulation in mild cognitive impairment individuals combined with depressive disorder: a neural mechanism study. *Front Aging Neurosci*. 2022;14:884741. doi:[10.3389/fnagi.2022.884741](https://doi.org/10.3389/fnagi.2022.884741)
39. de Paula JJ, Albuquerque MR, Lage GM, Bicalho MA, Romano-Silva MA, Malloy-Diniz LF. Impairment of fine motor dexterity in mild cognitive impairment and Alzheimer's disease dementia: association with activities of daily living. *Br J Psychiatry*. 2016;38(3):235-238. doi:[10.1590/1516-4446-2015-1874](https://doi.org/10.1590/1516-4446-2015-1874)
40. Sayyid ZN, Wang H, Cai Y, et al. Sensory and motor deficits as contributors to early cognitive impairment. *Alzheimer's Dement*. 2024;20(4):2653-2661. doi:[10.1002/alz.13715](https://doi.org/10.1002/alz.13715)
41. Zhang M, Ho E, Nowinski CJ, et al. The paradox in positive and negative aspects of emotional functioning among older adults with early stages of cognitive impairment. *J Aging Health*. 2024;36(7-8):471-483. doi:[10.1177/08982643231199806](https://doi.org/10.1177/08982643231199806)
42. Chen YX, Liang N, Li XL, Yang SH, Wang YP, Shi NN. Diagnosis and treatment for mild cognitive impairment: a systematic review of clinical practice guidelines and consensus statements. *Front Neurol*. 2021;12:719849. doi:[10.3389/fneur.2021.719849](https://doi.org/10.3389/fneur.2021.719849)
43. Park JH, Cho HE, Kim JH, et al. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *npj Digit Med*. 2020;3(1):46. doi:[10.1038/s41746-020-0256-0](https://doi.org/10.1038/s41746-020-0256-0)
44. Moore PJ, Lyons TJ, Gallacher J, Alzheimer's Disease Neuroimaging Initiative. Random forest prediction of Alzheimer's disease using pairwise selection from time series data. *PLoS One*. 2019;14(2):e0211558. doi:[10.1371/journal.pone.0211558](https://doi.org/10.1371/journal.pone.0211558)
45. Revathi A, Kaladevi R, Ramana K, Jhaveri RH, Rudra Kumar M, Sankara Prasanna Kumar M. Early detection of cognitive decline using machine learning algorithm and cognitive ability test. *Secur Commun Network*. 2022;2022:1-13. doi:[10.1155/2022/4190023](https://doi.org/10.1155/2022/4190023)
46. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Published online 2008.
47. Dey AKNS, Teja TS, Juneja A. Some variations on ensemble random survival forest with application to cancer research. Published online 2017. doi:[10.48550/ARXIV.1709.05515](https://doi.org/10.48550/ARXIV.1709.05515)