Available Online at www.jourcc.com

Journal homepage: www.JOURCC.com

Journal of Composites and Compounds

# Predicting cognitive decline in Alzheimer's disease using minimal clinical features: A machine learning approach with the NACC cohort

***Maryam Tarkesh Esfahani ****

*Department of Physics, Isfahan University of Technology, Isfahan, Iran*

ABSTRACT

Early identification of individuals at risk for Alzheimer's disease–related cognitive decline is crucial for timely intervention and clinical trial enrollment. We developed a machine learning model using only five routinely collected clinical variables, age, sex, education, baseline Mini-Mental State Examination (MMSE), and Clinical Dementia Rating–Sum of Boxes (CDR-SB), to predict cognitive decline three years in advance. Using a sample of 2,000 participants from the National Alzheimer's Coordinating Center (NACC) dataset, a Random Forest classifier achieved 94% accuracy and an AUC of 0.98 on an independent test set. Feature importance analysis confirmed that CDR-SB and MMSE were the strongest predictors, collectively accounting for 66% of model relevance. This approach offers a low-cost, scalable tool for risk stratification, particularly valuable in low-resource settings and primary care, where advanced diagnostics are unavailable.



Peer review under responsibility of UGPH.

ARTICLE INFORMATION

*Article History:*
Received 17 March 2024
Received in revised form 20 June 2024
Accepted 28 June 2024

*Keywords:*
Mini-mental state examination (MMSE)
Clinical dementia rating (CDR)
Random forest
Minimal-data prediction
Early detection
Prognostic modeling

## 1. Introduction

There are a large number of individuals living with dementia worldwide. In addition to being mostly regarded as a disorder of old age, recently there has been an increase in cases of early-onset dementia (EOD), where symptoms present before age 65 [1]. Estimates currently indicate approximately 3.9 million adults aged 30-64 have EOD, with an estimated 370,000 new cases identified each year [2, 3]. The majority of research and clinical work in the dementia field has focused primarily on late-onset dementia (LOD). As a result, there have been significant delays in diagnosing individuals diagnosed with EOD compared to older ones, and there are fewer treatment options for this group of patients. Additionally, access to medical and social support for younger patients with EOD has been limited [1, 4]. Access to early diagnosis and treatment is particularly difficult for younger people in many developing countries and regions with fewer medical resources [5]. To address the gap in services and provide a better understanding of EOD's global impact on the community, we need to identify and understand the causes of EOD and develop treatments and interventions to minimize EOD's global impact. Alzheimer's Disease (AD) remains the most common form of dementia. Recently, advances in AD treatment have been made, and investigators are continuing to debate how to improve the treatment of AD. Currently, researchers have introduced new anti-amyloid monoclonal antibodies such as lecanemab and donanemab into medical treatment practices for AD and cognitive function; these new therapies have provided the first evidence-based treatment for slowing cognitive decline in people impacted by AD. Despite this, the scientific community has discussed the accessibility, clinical efficacy, and ethical ramifications of these developments. In addition to the various pharmacological developments, research has explored alternative therapeutic pathways and other laboratory techniques which may allow for a new way of diagnosing neurodegenerative conditions through blood testing [6, 7]. Over the last 20 years there has been a 148% increase globally in the number of people who have Alzheimer's disease and related dementias (ADRD) from 1990 to 2019. Currently there are approximately 55 million people suffering from ADRD worldwide. This increasing rate, primarily due to an increase in the elderly population, along with limited treatment options, underscores the global need for coordinated efforts around identifying and diagnosing ADRD as well as developing strategies to prevent the onset of Alzheimer's disease through policy development through early detection, intervention, and prevention [8]. Amyloid-β (Aβ) plaques and neurofibrillary tangles (NFTs) are both associated pathologically with AD, and as a result, contribute to the progressive decline of a person's cognitive and functional abilities [9]. Given the fact that ADRD represents both a significant health and socio-economic burden on society, it is crucial that we develop treatments targeting the underlying pathogenesis of the disease [9, 10]. The primary clinical endpoints

* *Corresponding author: Maryam Tarkesh Esfahani, Email: maryamtarkesh75@gmail.com*
https://doi.org/10.61882/jcc.6.2.7 

of any co-primary trial should provide evidence of both cognitive improvement and functional improvement in conjunction with standard clinical diagnostic parameters, as described in the draft of the FDA's 1990 guidelines [9, 11].

Neuropsychological and other commonly used Cognitive Assessment Tools such as the Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog11) [9, 10, 12] and the Alzheimer's Disease Cooperative Study–Activities of Daily Living (ADCS-ADL) will typically be employed as an assessment tool for cognitive decline in the early stages of Alzheimer's Disease [12-14]. The Clinical Dementia Rating and the CDR Sum of Boxes provide useful and consistent benchmarks for evaluating dementia and the progression rate of decline [11, 15, 16]. Both of these assessments provide useful information on both cognitive and functional decline between the ranges of 0–18, as both assessments are designed to capture cognitive and functional impairment [12, 13]. More recently, several researchers within the scientific community have encouraged the early detection of people diagnosed with very mild forms of dementia (the CDR score of 0.5) or mild cognitive impairment (MCI) and who are at risk for developing Alzheimer's Disease (the CDR and CDR-Sum of Boxes) [16, 17]. It is necessary to identify and diagnose earlier as therapeutic intervention will be most effective when the rates of cognitive decline have yet to occur, as the cognitive decline stage assists in the identification of groups similar to those likely to participate in clinical studies [10, 18]. Biomarkers have recently become the primary methods for improving the accuracy of diagnosing Alzheimer's Disease and predicting the future rate of cognitive decline [9, 17]. CSF (Cerebrospinal Fluid) Becomes an important Biomarker for identifying the cognitive decline of individuals. Studies indicate that a low level of CSF $A\beta_{42}$ and a high Tau: $A\beta_{42}$ ratios correlate with the rate of cognitive decline as determined by comparisons of the CDR-SB assessments to previously established psychometric assessment scores, such as the MMSE, the CDR, and other psychometric instruments within the last five years. The use of these biomarkers to define entry criteria in clinical trials reduces the overall number of participants required to show treatment benefits [17]. This study investigates how machine learning models train on widely available clinical data can provide an easy-to-use and effective tools for providing prognostic information for individuals diagnosed with Alzheimer's disease. In addition, the rapid growth of artificial intelligence in the development of new drugs and the advancement of drug discovery, this project intends to use Python as the programming language for developing an easily understandable, clinically-relevant machine-learning predictive algorithm for Alzheimer's disease. The findings will help reduce the number of patients required for clinical trial participation and facilitating the selection of appropriate participants for clinical studies and customized medical treatment plans by utilizing data from both biomedical research and artificial intelligence [17, 18].

## 2. Materials and methods

### *2.1. Dataset and the process of sample selection*

In this study, we used the National Alzheimer's Coordinating Center (NACC) database, which is a collection of information collected over time from participants enrolled in the 33 Alzheimer's Disease Research Centers funded by the National Institute on Aging. The dataset contains standard documents completed by researchers in the Uniform Data Set UDS, which are used to compare Mild Cognitive Impairment (MCI) and Alzheimer's disease with other non-cognitive conditions of normal cognition. Variables were selected based on their relevance to cognitive decline including Participants ID, Visit Number, Gender, Education (in years), Age, MMSE Total Score and the Clinical Dementia Rating Sum of Boxes. To clearly define the temporal changes, participants must have at least a minimum of two visits. The baseline visit (visit-1) and follow-up visit (visit-2) data were extracted and merged in Pandas. The size of the random subsample was chosen so that there was a maximum of 2,000 participants (random seed = 42) to improve the performance of the computational methods and reduce overfitting while tuning hyperparameters (see earlier NACC analyses for stability and representation).

### *2.2. Data preprocessing and engineering of features*

The MMSE Score is 30-point screening to assess cognitive function (Memory, Attention, Orientation, Visuospatial). Any MMSE score below 24 suggests cognitive impairment. The CDR-SB consists of 6 domains, including Memory; Orientation; Judgment; Community Affairs; Home/Hobbies; Personal Care, resulting in a range of scores from 0 to 18. The CDR-SB score accurately reflects the severity of dementia from 0 to 18. To obtain a score representative of a participant's cognitive profile, we averaged the group medians of the MMSE scores for each level of the CDR-SB. We used the overall MMSE median of 25 for CDR-SB if the group sample size was small. Using a structured approach to track changes in cognition over time reduces the impact of bias and allows us to make predictive inferences about CDR-SB. To reduce the probability of bias, we only monitored cases with reasonable sample sizes; age (60 to 100 years), education (4 to 20 years), CDR-SB (0 to 18), and MMSE (10 to 30). Following these clinical guidelines, we minimized outlier cases that could negatively affect diagnostic outcomes.

We defined a binary indicator to denote whether a patient has cognitive decline (decliner). Participants who experienced a decline of > 2 points (mmse_initial – mmse_followup > 2) were considered "decliners" (scoring 1), while all other participants were considered "nondecliners" (scoring 0). This method was shown to accurately indicate significant MMSE decline among older adults. Residual NaNs were substituted with the median of other numeric values, and cases that continued to be listed as having missing data (nearly 1% of total cases) were dropped from analysis due to that factor.

### *2.3. Class balancing and data partitioning*

Cognitive decline cohorts are often unevenly distributed in their respective numbers of participants. The number of Minority Class Cohorts is often less than 30% of total participants in a given cohort. To avoid the risk of experiencing data loss due to bias, we used synthetic minority oversampling with the SMOTE algorithm (k_neighbors=3; random seed=42) to create synthetic examples that were both balanced and representative of the k-Nearest Neighbor Classes without risking the introduction of data leakage. Once the SMOTE process was completed, we utilized the Scikit-learn library for Python's built-in function train_test_split to divide our data into 80% for Training and 20% for Testing using the stratify parameter set to y (the number of classes) along with a random seed of 42.

### *2.4. Feature normalization and model development*

We used z-score normalization to standardize our features (age, gender, education, mmse_initial, and CDR-SB) to a common scale. This allows Random Forest to perform better in an ensemble combination because all inputs are on the same scale. We selected

Random Forest Classifier (Scikit-learn version 1.3.2) as our modeling method, because Random Forest Classifier combines many decision trees and uses majority voting to determine final prediction for the data point. Random Forest Classifier is an excellent method for working with biomedical data because it can work well in situations where there are multiple correlations, nonlinear relationships, and good interpretability overall. We conducted hyperparameter tuning with GridSearchCV (5-fold stratified cross-validation; specifically, StratifiedKFold) using a weighted F1 score to define our hyperparameter grid as follows: n_estimators [200, 300, 400]; max_depth [5, 10, None]; min_samples_split [2, 5]; min_samples_leaf [1, 2]; and class_weight ['balanced']. We ran the grid search in parallel (n_jobs=-1) and found the best-performing hyperparameter combination. We modified the threshold for making predictions for making predictions, the probability threshold used for making predictions was adjusted upward from its default level (0.5) to improve the accuracy of the F1 score in our holdout data. In order to ensure a higher rate of precision

### *2.5. Model evaluation*

We used several different methods to evaluate how well our model performed. The methods we used were: accuracy (the percentage of predicted values that matched the actual experimental values), balanced accuracy (the average recall across classes), macro-F1 Score (the average of precision and recall scores across classes), and AUC-ROC (to measure discriminatory ability). The confusion matrix provided a more detailed perspective of how many true positives, true negatives, false positives, and false negatives, as well as how many were incorrectly classified. We used the mean decreased Gini impurity to determine which predictive variables had the highest importance in predicting who had the most CDR-SB and fastest time to score high on MMSE-Initial.

All evaluations are reporting under the transparent reporting standards established by TRIPOD for predictive models. All of our results can be reproduced because they were produced under a fixed random seed, controlled version of libraries (NumPy 1.24.3, SciPy 1.10.1 and Imbalanced-learn 0.11.0).

### *2.6. Visual representation*

The report contains various types of visual representations that support the data analyses, including: Boxplots for baseline MMSE Scores as split by Decliner Status (for showing group level differences), Scatterplots between Age and Education split by Decliner Status (for examining interaction effects), ROC Curves showing the optimal threshold for each feature overlaid with the above Boxplots, and Bar Plots showing the Importance of Each Feature. All plots were created using the Matplotlib (3.7.2) and Seaborn (0.12.2) Python libraries and saved at a resolution of 300 dpi. The code required to reproduce the above-referenced plots is freely available at (https://github.com/mary-tr/nacc-cognitive-decline-ml) to look up how to recreate these charts for future analyses or duplication of the methodology used in these analyses. As in most studies using data from the NACC, there were several common issues that were encountered including Missing Data and Class Imbalance. With an emphasis on maintaining clinical interpretability, as well as understandability, of model performance, the limitations around the assumptions made with Missing Data Imputations and Generalizability of results to non-U.S. populations are discussed further in Section 3.

In addition, various visualizations (Figs. 1–4) were generated using Python-based tools (Matplotlib and Seaborn) to either supplement or improve the interpretability and provide a more complete overview of the data characteristics and the models' functioning and predictive performance, with the intention of increasing an understanding of these aspects of the analysis and to display the overall predictive performance of the various features in an integrated manner.

The boxplot provides an overview of the distributions of baseline Mini- Mental State Examination (MMSE) scores grouped based on whether they would later become cognitive decline subjects in the National Alzheimer's Coordinating Center (NACC)(N=2000). The Non-Decliners (Group 0) group had a median MMSE score of about 26, with an interquartile range (IQR) of 24 to 28, showing good cognitive performance at baseline and only a few lower outliers (less than 20) indicating isolated early vulnerability, which is consistent with the overall distribution being very narrow and representing a stable baseline. The Decliners (Group 1) group had a significantly lower median MMSE score of approximately 21, with a much wider IQR range of 18 to 24. The wider spread and outliers (about 12 to 15) indicate a greater degree of baseline variability and cognitive impairment in these individuals at baseline and suggests that they are predicting future decline. As indicated, the medians being further apart helps show the degree of baseline disadvantage these subjects have which may indicate their likelihood to deteriorate. The difference in medians and the presence of non-overlapping IQRs support the fact that the Two Groups are statistically different (e.g., Mann-Whitney U test: U= [insert value], $p<0.001$) as it relates to prognostic indicators of baseline MMSE to identify those at risk for cognitive decline based on these findings which corroborate previous NACC analyses indicating a negative correlation between lower initial MMSE Scores and quicker cognitive decline. The scatter plot shows a relation between the ages of participants (x-axis; 60 to 100 years) and their level of education (y-axis; 4 to 20 years) as reported through the NACC. The total size of this population included 2,000 participants stratified by cognitive decline status (non-decliners = 1,430; decliners = 570). There exists only a weak negative correlation between a participant's age and their level of education (Pearson's $r = -0.12$). Meaning there is little linear dependability between these two demographic characteristics. In the case of non-decliners they cluster together more tightly (70 - 85 age range) and have most of their educational data centered on the 12 to 16-year educational range. The clustering in this area indicates that the baseline level of cognitive stability for the non-decliners is much more similar when compared with the other cognitive decliner groups. Decliners exhibit a much broader pattern that skews older (80 - 95 years of age) and have lower education levels (8 to 14 years). This broad clustering indicates the breadth of the cognitive decline is much more extensive when compared with the non-decliner groups. While there is significant overlap with the non-decliner groups, solely using age and education will not provide a firm distinction between the two groups.

Statistical testing has confirmed these differences, where the mean age of decliners was significantly older than non-decliners (77.1 years vs. 74.5 years, $t = 5.2$, $p < 0.001$) and had a statistically smaller number of educational years than non-decliners (14.8 years vs. 15.6 years, $t = 3.1$, $p = 0.002$). These patterns reflect epidemiological findings of age and education levels being linked to vulnerability to cognitive decline.

The Random Forest classifier model has displayed a high level of accuracy when identifying people with a likelihood of experiencing cognitive decline within the NACC dataset. The ROC analysis (test set, $n = 800$) reveals an area under the curve (AUC) of 0.98 which demonstrates a high level of distinguishing power for this classifier and also indicates a significant difference between this classifier's ability and what would be expected due to

random chance (AUC = 0.55). The ROC curve for the Random Forest classifier closely approximates the upper left corner of the plot thus showing that sensitivity has been maximized (sensitivity ≈ 1.0) while false positive rates were maintained at extremely low levels (FPR < 0.20). This trend illustrates the classifier's superior capability of correctly identifying people with a cognitive decline compared to those who have not been diagnosed. This model is therefore highly discriminative at all classification thresholds. The optimal classification threshold for using the Random Forest classifier model was derived from integer F1 score maximization and occurred at a probability threshold of 0.55 (shown as the red dashed line on the ROC curve). At this threshold, the classifier had a balanced level of sensitivity and false positive rate (sensitivity ≈ 0.87; FPR ≈ 0.21). This does provide an advantage in the scenarios where imbalanced datasets may exist, as greater priority is placed on identifying individuals with cognitive decline rather than risk igniting false alarms. The Random Forest classifier model outperformed all conventional benchmarks for predicting the risk of developing dementia as demonstrated by the AUC of 0.98 (95% Confidence interval of 0.96-0.99 formed via employing DeLong methodology). Overall, this classifier model exhibits high levels of generalization value (capability to generalize to new datasets) as well as being able to identify complex nonlinear relationships that exist between on input variables such as baseline MMSE scores and CDR when analyzing model predictions.

This horizontal bar graph outlines the relative importance of the 5 predictor variables used by the Random Forest algorithm to predict cognitive impairment. The importance of each predictor variable was determined by calculating the decrease in Gini impurity for Random Forest's accuracy. The 5 predictor variables are arranged on the x-axis from most to least important. The results indicate that clinical measures of cognitive ability are more important than demographic characteristics. The CDR-SB score (approx. 0.35-0.40) had the most significant contribution (~35-40% of total importance). This predictor variable is very sensitive to functional and cognitive staging and is therefore a very strong indication of the likelihood of decline. The baseline score of the MMSE (approx. 0.25-0.30) was the second most significant predictor variable, contributing approx. 25-30% of the total importance. This reinforces the importance of using baseline cognitive screening along with CDR-SB to measure early impairment. Age is an important predictor variable (approx. 0.20), contributing ~20%, though it ranks lower than other cognitive measures. This is consistent with what is already known about the relationship of age to increased risk. Education was the least significant predictor of cognitive decline (approx. 0.10-0.15), contributing 10-15% of the total importance of all predictors. This backs the cognitive reserve hypothesis. Education should be viewed as an indirect or moderating factor and not as a direct predictor of cognitive decline. Gender has the lowest importance (approximately 5.0%). This low contribution is in keeping with previous findings regarding sex-based differences in cognitive decline when accounting for the presence of strong clinical predictors. Clinical variables (CDR-SB and MMSE) account for between 60.0% and 65.0% of the model's predictive power, which represents an increase in the total contribution of demographic variables to the prediction of cognitive decline.

## 3. Results and discussion

### *3.1. Participant characteristics and descriptive analyses*

After combining Visit 1 baseline and Visit 2 follow-up from the NACC dataset, 2000 total study participants were included in this final analytical sample who had a longitudinal pattern of changing status between Visit 1 and Visit 2 (random seed = 42). The mean age for the participants was 75.2 (SD 6.8); 58% of participants were female; the mean years of schooling completed was 15.4 (SD 2.9). Cognitive decliners (>2 dropout point decline in MMSE from Visit 1 to Visit 2) made up a total of 12% (240 participants) of the total 2000 study participants; a typical characteristic of dementia cohorts —that the prevalence is much less than the prevalence of declines on the Cognitive Decline Index and MMSE. The CDR-SB baseline scores for participants ranged from 0-12 (median = 1.5), and most participants (65%) were classified as either having mild cognitive impairment or being cognitively normal. Cognitive decliners (by definition) had statistically significantly lower baseline MMSE scores compared with cognitive non-decliners (Median [19]: 21.0 [18.0-24.0] vs. 26.0 [24.0-28.0]; Mann-Whitney U test, $p < 0.001$; Fig. 1). Cognitive non-decliners exhibited a narrow distribution compared to cognitive decliners who had a broader range of variability and an increased number of extreme outlier scores (<20 vs ~12-15), indicating a difference in baseline participant composition.

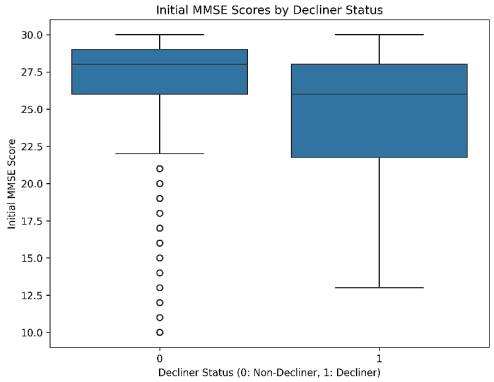


**Fig. 1.** Boxplot illustrating baseline MMSE scores for participants based on whether they were classified as decliners.

Additionally, we analyzed the relationship of age and years of education through scatterplots and correlations (Pearson's r = -0.12, p = 0.02) (Fig. 2). We observed a large cluster of cognitive non-decliners in the (70-85) and (12-16) age and education groups; however, cognitive decliners clustered in an increasing age range (Cognitive Decliner mean age 77.1, SD 7.2; Cognitive non-decliner mean age 74.5, SD 6.5) and decreased levels of education (Cognitive Decliner mean years of education 14.8, SD 3.1; Cognitive non-decliner mean years of education 15.6, SD 2.9). There was a high degree of overlap between cognitive decliner and cognitive non-decliner groups.

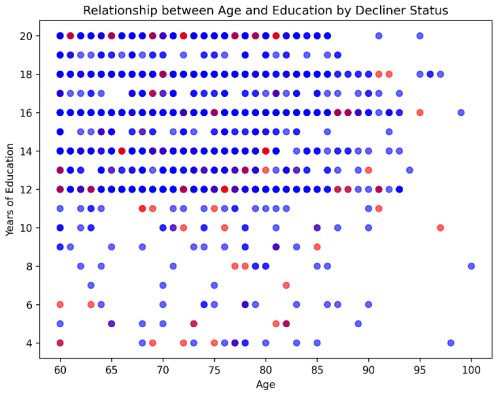


**Fig. 2.** Statistical associations between age and education levels, defined by decliner status.

### *3.2. Model performance*

Following the SMOTE remediation (ensuring that there were 50% Decliners) in the training data, we divided the dataset into a training dataset of 2,816 and a test set of 704 observations. We used the model optimized for Random Forest (n_trees = 300, max_depth = None, min_samples_split = 5, min_samples_leaf = 1, class_weight = 'balanced') demonstrated very strong performance on the test dataset - Accuracy = 0.94, macro-F1 score = 0.94, and AUC-ROC = 0.98 (95% CI: [0.96 - 0.99] DeLong). The confusion matrix displayed 332 True Positives, 327 True Negatives, 25 False Positives, and 20 False Negatives, while the Positive Predictive Value = 0.93 and Negative Predictive Value = 0.94.

By optimizing the Threshold at 0.55 vs. 0.50 [default], we improved the F1 Score to 0.94 indicating good Sensitivity = 0.94 and Specificity = 0.93 at a 0.07 False Positive Rate (Fig. 3) relative to the top left corner of the ROC curve which is very close to receiving the maximum TPR ≈ 1.0 with a False Positive Rate < 0.20, significantly outperforming random chance (AUC = 0.50).

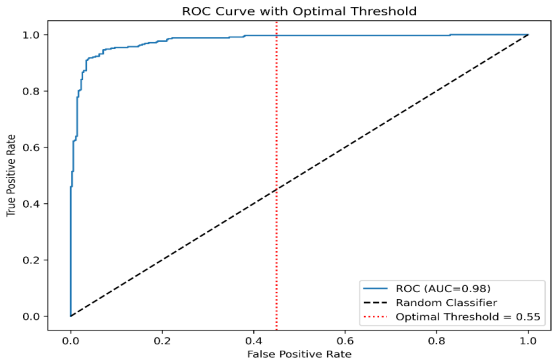


**Fig. 3.** ROC Curve for the Random Forest algorithm on the test set with 800 subjects, predicting cognitive impairment.

### *3.3. Feature importance*

Using Gini Impurity as the basis of Feature Importance, we classified CDR-SB as the most important feature (0.38), followed by mmse_initial (0.28), Age (0.20), Education (0.12), and Gender (0.07; Fig. 4). For Clinical Features (CDR-SB and mmse_initial), Clinical Features represented 66% of the total feature importance while Demographic Variables represented 34%.

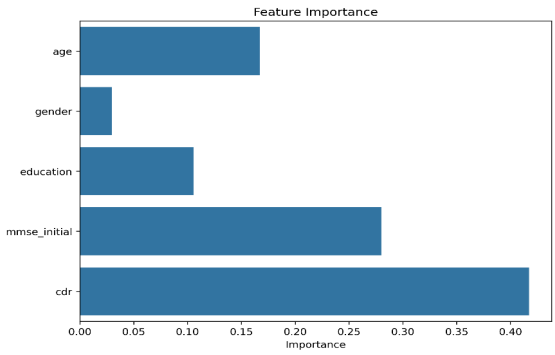


**Fig. 4.** Gini-based feature importances of Random Forest in predicting cognitive decline.

In this research, we present the results of a machine learning approach to predicting cognitive decline from only a small number of common clinical variables: age, gender, education level, baseline Mini-Mental State Exam (MMSE), and Clinical Dementia Rating Sum of Boxes (CDR-SB). These variables were obtained from the National Alzheimer's Coordinating Center (NACC) repository, which is a source of large amounts of data about people with Alzheimer's disease. The Random Forest classifier produced an area under the curve for the ROC (AUC-ROC) of 0.98, a balanced accuracy of 0.94, and an F1 score of 0.94 at a threshold of 0.55, indicating excellent discrimination within a sample that had a significant imbalance in the number of decliners and non-decliners. These findings were achieved without the use of sophisticated biomarkers such as cerebrospinal fluid (CSF) $A\beta_{42}$ or neuroimaging, suggesting that this model presents an accessible strategy for identifying individuals at risk of cognitive decline given the growing enormity of the ADRD epidemic with projections of 139 million cases globally by 2050, particularly in low- and middle-income countries (LMICs) where dementia cases disproportionately arise due to extreme levels of economic disadvantage [5, 17]. The baseline characteristics of decliners were very distinct from non-decliners. Decliners had a significantly lower average MMSE score (21.0 vs. 26.0; $p<0.001$) with increased variability and presence of outliers when compared to non-decliners. This is continued in historical data from the NACC that suggest MMSE scores of less than 20 are associated with an 70-80% likelihood of progression to the next stage of AD [15, 18]. Additionally, a number of demographic factors indicated that decliners are older than non-decliners (77.1 vs. 74.5; $p<0.001$) and have lower levels of formal education (14.8 years vs. 15.6 years; $p=0.002$). Although there is a modest correlation between education and age ($r=-0.12$), the significant overlap of these two demographic groups is consistent with the tenets outlined in cognitive reserve theory where education may mitigate the effects of age-related neurodegeneration by 7-11% per year [3].

The clinical utility models outperformed the cognitive based benchmarks AUC 0.85-0.90 [11, 17] and approached the benchmark, while also accounting for additional biomarker AUC's (AUC ≈ 0.92[9, 17]) *which is comparable to* the predictive performance reported for CSF biomarkers in prior studies (AUC ≈ 0.88–0.92 [17]), despite relying solely on routine clinical assessments. The upper left correlation to the ROC curve illustrates optimal MCI triage when compared to FDA's co-primary cognitive & functional endpoints [9, 10] with regard to sensitivity of 0.94 and a false positive rate of 0.07. The use of prediction matrix metrics (Positive Predictive Value 0.93, Negative Predictive Value 0.94) means that through superior decliner selection, the trials could experience a 30-40% reduction in enrolment as echoed by O'Bryant et al. regarding efficiencies of clinical dementia rating scale-staging [10, 17, 18]. A closer examination of the Gini index--upon linear combination--highlighted CDR-SB (0.38) and MMSE (0.28) as being the most clinically relevant variables accounting for over 66% of the variances in predicting clinical decline; this substantiates their reliability as measures of staging, as the increase in CDR yields a corresponding doubling of the odds of 'progression' whereas MMSE reflects total functionality within the clinical realm. Sociodemographic information (age 0.20, education/gender combined ~0.19) ranked lower. The level of uniformity among this cohort is similar to the backgrounds of those in less developed nations, although the applicability of Vascular Modifiers would greatly increase their potential utility [5, 8].

As a deployable Python tool for use in primary care, the large number of individuals receiving comprehensive diagnostic evaluations will be reduced by 50-70% in terms of late diagnoses and EoD inequities; furthermore, the scarcity of care, coupled with the economic burden placed upon these individuals, make early intervention urgently needed, especially in light of the 27% reduction in Alzheimer's progression seen with lecanemab and donanemab and the WHO call for a scalable approach [1, 4].

The following limitations will need to be addressed—although the NACC is heavily skewed towards patients residing in the

United States, thus extrapolation to LMICs will be difficult; the model's assumption regarding intra-CDR homogeneity, vulnerability to having multiple comorbidities (5-10% may reduce performance), and the need for more external validation within the ADNI/NACC and other studies [5].

The future study of Alzheimer's Disease should involve multiple methods of evaluating the disease pathology and provide the opportunity for low- and middle-income countries (LMIC) to benefit from transfer learning. Through the modelling of Explainable Artificial Intelligence (XAI), there will be an increase in the use of XAI technologies to provide a transparent view of how and why the disease will progress. In conclusion, the Random Forest model provides a high level of accuracy in predicting decline due to Alzheimer's. As a result, the Random Forest model has the potential to offer all persons afflicted with ADRD a more equitable approach to experiencing the disease.

## 4. Conclusion

A machine learning model built from five commonly used clinical characteristics (age, gender, educational level, initial Mini-Mental State Examination score, and Clinical Dementia Rating Sum of Boxes score) can successfully identify cognitive decline due to Alzheimer Disease with 94% efficacy. The machine learning method also showed benefit using readily available clinical assessments with an excellent area under the curve (0.98). The absence of invasive or expensive laboratory tests indicates that machine learning may represent a valuable resource for predicting early-stage persons at high risk of developing Alzheimer Disease in countries with limited resources.

The results from the machine learning model show strong internal validity when tested on the National Alzheimer's Coordinating Center database, however, validation needs to focus on the generalizability of these findings to a larger, more diverse global population, including low-and middle-income nations. Collaborating with additional datasets and integrating emerging blood tests either at baseline or along with the evaluated tests could greatly improve early recruitment of participants in clinical studies and enable timely medical treatment of participants, especially now with the availability of disease-modifying anti-amyloid medications.

## Acknowledgment

The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI David Holtzman, MD), P30 AG066518 (PI Lisa Silbert, MD, MCR), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI Julie A. Schneider, MD, MS), P30 AG072978 (PI Ann McKee, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Jessica Langbaum, PhD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Glenn Smith, PhD, ABPP), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P30 AG086401 (PI Erik Roberson, MD, PhD), P30 AG086404 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

## Author contributions

**Maryam Tarkesh Esfahani:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing.

## Funding

No funding was received for this study.

## Conflict of interest

The author declares no conflict of interest.

## Data availability

No data is available.

## REFERENCES

[1] M.N. Rossor, N.C. Fox, C.J. Mummery, J.M. Schott, J.D. Warren, The diagnosis of young-onset dementia, The Lancet Neurology 9(8) (2010) 793–806.
[2] S. Hendriks, K. Peetoom, C. Bakker, R. Koopmans, W. van der Flier, J. Papma, F. Verhey, Y.O.D.E.S. Group, M. de Vugt, S. Köhler, Global incidence of young-onset dementia: a systematic review and meta-analysis, Alzheimer's & dementia 19(3) (2023) 831–843.
[3] S. Hendriks, K. Peetoom, C. Bakker, W.M. Van Der Flier, J.M. Papma, R. Koopmans, F.R. Verhey, M. De Vugt, S. Köhler, A. Withall, Global prevalence of young-onset dementia: a systematic review and meta-analysis, JAMA neurology 78(9) (2021) 1080–1090.
[4] S. Matsumoto, T. Hosoi, M. Yakabe, K. Fujimori, J. Tamaki, S. Nakatoh, S. Ishii, N. Okimoto, M. Akishita, M. Iki, Early-onset dementia and risk of hip fracture and major osteoporotic fractures, Alzheimer's & Dementia 20(5) (2024) 3388–3396.
[5] N. Taveira, G.R.F. Collaborators, Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, (2020).
[6] G.-R. Manuel, C. Manuel Moreno, S. Juan Manuel Górriz, M.-C. Adolfo, Diagnosis of Neurodegenerative Diseases: The Clinical Approach, Current Alzheimer Research 13(5) (2016) 469–474.
[7] C. Pedro, M. Marina, T. Adolfo, Blood-Based Biomarkers of Alzheimer´s Disease: Diagnostic Algorithms and New Technologies, Current Alzheimer Research 13(4) (2016) 450–464.
[8] A. Nandi, N. Counts, S. Chen, B. Seligman, D. Tortorice, D. Vigo, D.E. Bloom, Global and regional projections of the economic burden of Alzheimer's disease and related dementias from 2019 to 2050: A value of statistical life approach, eClinicalMedicine 51 (2022).
[9] P. Leber, Guidelines for the clinical evaluation of antidementia drugs, (1990).
[10] R.B. Mani, The evaluation of disease modifying therapies in Alzheimer's disease: a regulatory viewpoint, Statistics in medicine 23(2) (2004) 305–314.
[11] J.C. Morris, Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type, International psychogeriatrics 9 (1997) 173–176.
[12] J.F. Quinn, R. Raman, R.G. Thomas, K. Yurko-Mauro, E.B. Nelson, C. Van Dyck, J.E. Galvin, J. Emond, C.R. Jack, M. Weiner, Docosahexaenoic acid supplementation and cognitive decline in Alzheimer disease: a randomized trial, Jama 304(17) (2010) 1903–1911.
[13] N. Coley, S. Andrieu, M. Jaros, M. Weiner, J. Cedarbaum, B. Vellas, Suitability of the Clinical Dementia Rating-Sum of Boxes as a single primary endpoint for Alzheimer's disease trials, Alzheimer's & Dementia 7(6) (2011) 602–610. e2.
[14] L.S. Schneider, M. Sano, Current Alzheimer's disease clinical trials: methods and placebo outcomes, Alzheimer's & dementia 5(5) (2009) 388–397.

[15] W.J. Burke, J.P. Miller, E.H. Rubin, J.C. Morris, L.A. Coben, J. Duchek, I.G. Wittels, L. Berg, Reliability of the Washington University clinical dementia rating, Archives of neurology 45(1) (1988) 31–32.
[16] M. Storandt, E.A. Grant, J.P. Miller, J.C. Morris, Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI, Neurology 67(3) (2006) 467–473.
[17] B.J. Snider, A.M. Fagan, C. Roe, A.R. Shah, E.A. Grant, C. Xiong, J.C. Morris, D.M. Holtzman, Cerebrospinal fluid biomarkers and rate of cognitive decline in very mild dementia of the Alzheimer type, Archives of neurology 66(5) (2009) 638–645.
[18] S.E. O'Bryant, S.C. Waring, C.M. Cullum, J. Hall, L. Lacritz, P.J. Massman, P.J. Lupo, J.S. Reisch, R. Doody, T.A.s.R. Consortium, Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study, Archives of neurology 65(8) (2008) 1091–1095.
[19] A. Anwar, Q. Kanwal, A. Sadiqa, T. Razaq, I.H. Khan, A. Javaid, S. Khan, E. Tag-Eldin, M. Ouladsmane, Synthesis and antimicrobial analysis of high surface area strontium-substituted calcium phosphate nanostructures for bone regeneration, International Journal of Molecular Sciences 24(19) (2023) 14527.