Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 17 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i17.44173

PAPER

Predicting Amyloid-β Positivity in Alzheimer's Disease: Comprehensive Analysis of Feature Selection and Machine Learning Models for Accurate Identification

Xing Wei[1,2], Na Gao[1,3], Mengru Xu[1,3], Mostafa Rezaeitalesh[2], Nan Mu[2,4], Zonghan Lyu[2], Michael Ngala[5], Jingfeng Jiang[2], Xuelian Chang[6](✉)

[1]Department of Computer Science, Bengbu Medical College, Bengbu, Anhui, China

[2]Department of Biomedical Engineering, Michigan Technical University, Houghton, Michigan, USA

[3]School of Nursing, Bengbu Medical College, Bengbu, Anhui, China

[4]College of Computer Science, Sichuan Normal University, Chengdu, China

[5]Department of Computer Science, Michigan Technical University, Houghton, Michigan, USA

[6]School of Basic Medicine, Bengbu Medical College, Bengbu, Anhui, China

xuelianchang@126.com

ABSTRACT

To accurately identify individuals at risk of Alzheimer's disease (AD), it is crucial to develop precise tools for predicting amyloid-β (Aβ) positivity in the brain. We used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to analyze 1,377 human subjects. These participants were divided into five groups: cognitive normal (CN), subjective memory complaints (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and confirmed AD. Each group was further divided into ten subgroups based on sex, resulting in a comprehensive analysis. The dataset was used to create and evaluate the performance of 15 machine learning (ML) models. A set of 17 potential predictors was generated by combining variables from different categories, including six demographic factors (such as age), ten measurements (such as ADAS13), and APOE4 status. Through ML-based predictive modeling, several cognitive assessment measures, including ADAS13, demonstrate significant importance in multiple ML models. The highest accuracies in the 10 subgroups were 0.875, 0.892, 0.778, 0.850, 0.771, 0.739, 0.781, 0.791, 0.879, and 0.903, respectively. The collection of ML models consists of practical and valuable risk feature scores that can significantly enhance the identification of individuals who are likely to test positive for Aβ.

KEYWORDS

Alzheimer's disease (AD), machine learning (ML), feature selections, accurate identification, amyloid-β positivity prediction

Wei, X., Gao, N., Xu, M., Rezaeitalesh, M., Mu, N., Lyu, Z., Ngala, M., Jiang, J., Chang, X. (2023). Predicting Amyloid-β Positivity in Alzheimer's Disease: Comprehensive Analysis of Feature Selection and Machine Learning Models for Accurate Identification. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(17), pp. 98–114. https://doi.org/10.3991/ijoe.v19i17.44173

Article submitted 2023-08-19. Revision uploaded 2023-09-30. Final acceptance 2023-09-30.

1 INTRODUCTION

Alzheimer's disease (AD) is a chronic neurodegenerative condition that has affected millions of senior citizens worldwide. Furthermore, in the United States alone, an estimated 6.7 million seniors aged 65 and older will be living with AD in 2023 [1], and perhaps many more are showing symptoms of early-onset AD. Typically, patients with AD have an extended asymptomatic period during which neuropathological changes accumulate [2]. One crucial strategy to combat AD is the early identification of individuals before the onset of dementia [3]. Early diagnosis of AD offers various benefits, including timely treatment and intervention to slow the progression of AD and improve overall quality of life [4].

Amyloid-β plays a crucial role as a pathological marker in diagnosing AD and is an early event in the development of AD pathology [5]. Elevated levels of Aβ levels or an increased Aβ42/Aβ40 ratio in cerebrospinal fluid (CSF) have been associated with a higher risk of developing AD, and disease-modifying therapies (DMTs) that target Aβ deposits have been the focus of numerous clinical trials [6].

Positron emission tomography (PET) imaging using Aβ-specific tracers enables the visualization and measurement of amyloid plaques in the brain, facilitating the diagnosis and monitoring of disease progression [7]. However, CSF analysis is an invasive procedure that can potentially cause pain for patients. Additionally, some individuals may not be able to undergo a lumbar puncture due to factors such as back deformity, infection, or the risk of brain herniation [8]. Although amyloid PET imaging is preferred in some instances, its utilization in clinical and trial settings is limited due to patient concerns regarding radiation exposure and high out-of-pocket costs [9].

Given the scale of the affected population, screening using CSF and PET would impose significant pressure on the healthcare system. Access to affordable cognitive assessments can provide significant advantages to individuals worldwide, particularly those in underdeveloped countries. In these regions, low-cost diagnostic methods can be utilized for screening purposes, while more expensive tests can be reserved for high-risk patients only. It is therefore critical to develop an effective, affordable, and non-invasive screening method to identify individuals who are at risk of developing Alzheimer's disease.

Recent studies have demonstrated the efficacy of neuropsychological tests in diagnosing AD and identifying individuals at risk of AD progression [10]. Two notable tests, namely the Montreal Cognitive Assessment (MoCA) [11] and the clinical dementia rating scale sum of boxes (CDRSB) [12], have emerged as promising low-cost screening tools for patients showing signs of early-stage AD. They are standardized tools that can be easily administered and scored, making them suitable for large-scale studies and various clinical settings. However, these tools have limitations, including sensitivity and specificity issues, limited scope, and concerns regarding inter-rater reliability.

Specifically, research has indicated that there is variability in terms of sex and APOE ε4 status within populations, which influences how MCI among patients

with AD may be assessed [13]. Notably, women diagnosed AD tend to experience a more rapid decline in cognitive function compared to their male counterparts. Similarly, females with MCI exhibit greater cognitive decline than males [14]. The presence of APOE ε4 gene carriers also significantly influences cognitive changes. Women generally demonstrate superior verbal memory abilities throughout their lives compared to men, even in factors related to AD, such as beta-amyloid accumulation in the brain [15]. Interestingly, even when measurable pathological changes are present, women with normal cognitive function often display better verbal memory skills. This could potentially affect the accuracy of memory-based assessments when screening women for early AD-related changes [16]. Consequently, clinicians are advised to use them in conjunction with other assessments and clinical judgments to gain a comprehensive understanding of an individual's condition [17].

Predicting Aβ status accurately and cost-effectively is crucial for overcoming diagnostic limitations and reducing healthcare costs [9]. ML methods have emerged as a promising avenue in AD research, offering unique opportunities to predict essential biomarkers such as Aβ and improve diagnosis, prognosis, and risk assessment [18]. However, such studies are still in their infancy because limited validation of personalized predictions of AD status has been reported in the literature [19].

To this end, our study aims to develop a precise ML algorithm for predicting Aβ status by combining diverse scoring results. The objectives are as follows: 1) **Feature importance ranking:** In the ADNI dataset, the random selection algorithm is used to calculate and rank the importance of all features in each model. This step serves as the foundation for our predictive modeling. 2) **Aβ positivity prediction:** Utilizing the significant scores obtained from step 1, we employ 15 ML models to forecast positive results for protein Aβ. These predictions are crucial in evaluating the probability of Aβ positivity among individuals at risk. 3) **Optimal feature set and models:** The evaluation process involves combining the results from steps 1 and 2 to determine the optimal feature set and models. This step is critical for refining our ML algorithm to accurately predict Aβ status. Our comprehensive scoring system includes the Montreal cognitive assessment (MoCA), the functional activities questionnaire (FAQ) [20], the everyday cognition-self-reported (ECog-SP) [21], the longitudinal dementia evaluation-total (LDELTotal) [22], the AD assessment scale-cognitive (ADAS-Cog) [23], and the mini-mental status examination (MMSE) [24]. Our approach utilizes various scoring components to enhance performance for different subpopulations, thereby improving diagnostic accuracy at all stages of Alzheimer's disease.

2 METHODOLOGY

2.1 Study designs and participants

The data used in this project was obtained from a public database called ADNI (http://adni.loni.usc.edu) [25]. Although a continuous study, there have been several phases of ADNI: ADNI-1, ADNI-GO, ADNI-2, and ADNI-3 (current). In each phase, participants with varying cognitive statuses (i.e., CN, SMC, EMCI, LMCI, and confirmed AD) were recruited. Figure 1 illustrates our selection process for individuals from different ADNI studies. This inclusion allows us to incorporate participants at various stages of the disease.

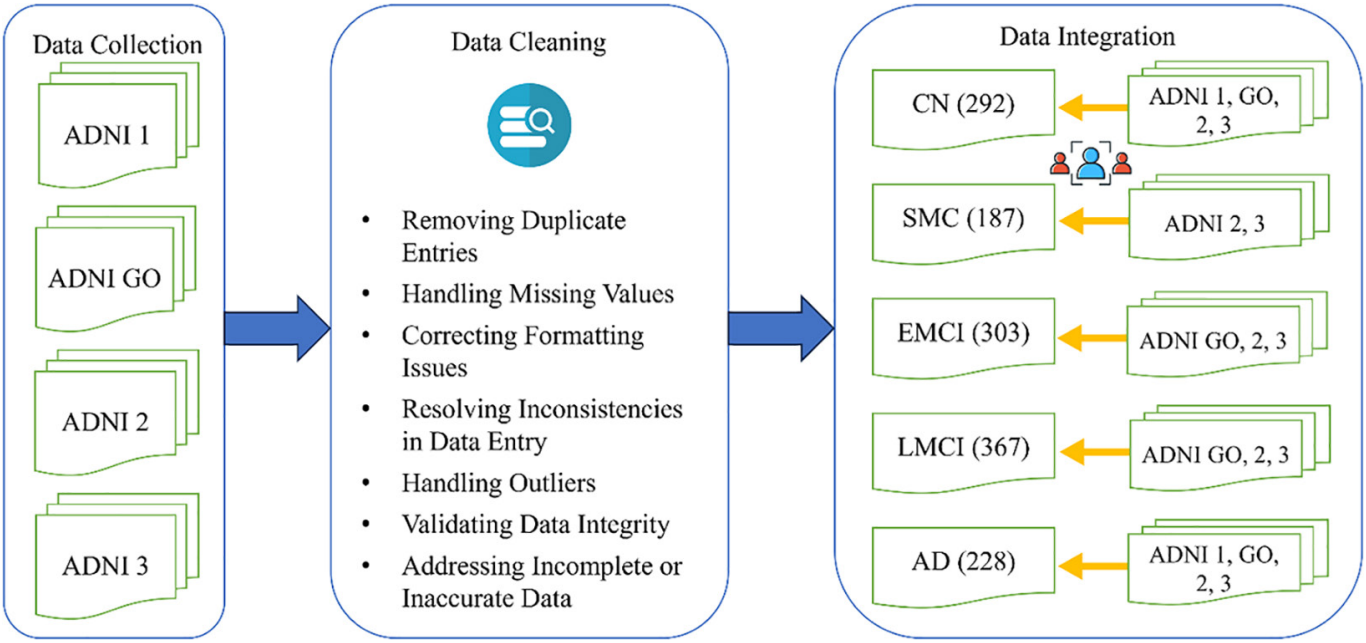

Fig. 1. A flowchart showing the proposed data processing pipeline

The standardized uptake value ratio standardized uptake value ratio (SUVR) [26] was used to quantitatively assess amyloid positivity, allowing for the measurement of amyloid plaque accumulation. Higher SUVR values indicate a greater presence of amyloid. The SUVR was calculated by averaging the weighted mean cortical retention value from the frontal, cingulate, parietal, and temporal regions and dividing it by the SUVR of the cerebellum [27]. In this study, the binary classification of amyloid positivity in the reference model was determined by using a cut-off value of SUVR ≥ 1.11 [28].

In the ADNI study, participants undergo comprehensive assessments, including clinical evaluations, neuropsychological testing, neuroimaging scans, and biomarker measurements [29]. For our analysis, we focused on data collected at baseline, as this is when amyloid PET takes place. The details regarding the sample size and characteristics of the individuals included in our analysis can be found in Table 1.

Table 1. Descriptive statistics of the 10 subgroups from the ADNI

	CN			SMC			EMCI			LMCI			AD			P-Value
	Female	Male	P-Value	Female	Male	P-Value	Female	Male	P-Value	Female	Male	P-Value	Female	Male	P-Value	
N	148	144		112	75		133	170		160	207		99	129		
Amyloid(%)	27(22.3)	25(21.0)	0.084	46(41.1)	23(30.7)	0.045	53(39.8)	74(43.5)	0.052	95(59.4)	130(62.8)	0.051	77(77.8)	109(84.5)	0.020	< 0.0001
AGE	72.9(6.1)	74.0(6.6)	0.005	70.3(6.9)	72.1(6.2)	< 0.001	70.4(8.2)	71.9(6.8)	0.006	72.8(7.7)	74.2(7.7)	0.002	74(8.3)	76(7.9)	0.045	< 0.0001
Edu	15.9(2.6)	17.0(2.7)	< 0.001	16.4(2.4)	17.0(2.2)	0.001	15.6(2.7)	16.5(2.6)	< 0.001	15.7(2.8)	16.4(2.8)	0.003	14.4(2.7)	16.1(2.9)	< 0.001	< 0.0001
Hisp(%)	7(4.7)	2(1.7)	0.015	8(7.9)	6(8.0)	0.345	4(3.6)	4(2.8)	0.544	4(2.6)	3(1.5)	0.489	3.4(3.4)	4(3.1)	0.761	0.867
Race(%)			0.058			0.134			0.416			0.008			0.310	0.0503
White	134(90.5)	126(87.5)		88(78.6)	64(85.3)		117(88.0)	159(93.5)		143(89.4)	195(94.2)		85(85.9)	119(92.2)		
African American	11(7.4)	9(6.3)		16(14.3)	5(6.7)		7(5.3)	3(1.8)		13(8.1)	5(2.4)		8(8.1)	6(4.7)		
Other	3(2.0)	9(6.3)		8(7.1)	6(8.0)		9(6.8)	8(4.7)		4(2.5)	7(3.4)		6(6.1)	4(3.1)		
Marry status(%)			0.167			0.096			< 0.001			< 0.001			0.354	0.0487
Married	88(59.5)	117(81.3)		73(65.2)	62(82.7)		80(60.2)	145(85.3)		99(61.9)	179(86.5)		75(75.8)	120(93.0)		
Never married	9(6.1)	8(5.6)		6(5.4)	2(2.7)		10(7.5)	5(2.9)		4(2.5)	3(1.4)		2(2.0)	0(0.0)		
Divorced	24(16.2)	9(6.3)		14(12.5)	7(9.3)		26(19.5)	11(6.5)		26(16.3)	13(6.3)		5(5.1)	2(1.6)		
Widowed	27(18.2)	10(6.9)		19(17.0)	4(5.3)		17(12.8)	5(2.9)		31(19.4)	10(4.8)		17(17.2)	7(5.4)		
APOE ε4(%)			0.856			0.174			0.496			0.003			0.272	< 0.0001
0 copy	101(68.2)	107(74.3)		66(58.9)	51(68.0)		80(60.2)	89(52.4)		61(38.1)	111(53.6)		33(33.3)	39(30.2)		
1 copy	42(28.4)	29(20.1)		37(33.0)	21(28.0)		40(30.1)	64(37.6)		69(43.1)	71(34.3)		47(47.5)	56(43.4)		
2 copy	5(3.4)	8(5.6)		9(8.0)	3(4.0)		13(9.8)	17(10.0)		30(18.8)	25(12.1)		19(19.2)	34(26.4)		
Family Hist(%)	89(60.1)	81(56.3)	0.274	74(66.1)	43(57.3)	0.317	88(66.2)	103(60.6)	0.309	95(59.4)	116(56.1)	0.407	61(61.6)	84(65.1)	0.135	0.0178
ADAS13	8.3(4.1)	10.5(4.6)	< 0.001	7.9(3.9)	10.1(4.0)	< 0.001	12.3(5.6)	13.5(5.3)	0.064	19.3(7.3)	19.1(6.4)	0.674	30.0(8.4)	30.1(7.8)	0.679	< 0.0001
MoCA	26.5(3.0)	25.7(3.7)	0.107	25.3(2.8)	24.5(2.8)	0.187	23.5(3.2)	23.3(3.5)	0.981	15.9(7.3)	15.5(7.3)	0.085	7.0(7.1)	8.2(7.7)	0.538	< 0.0001
CDRSB	0.0(0.1)	0.0(0.1)	0.283	0.1(0.2)	0.0(0.1)	0.031	1.2(0.8)	1.3(0.8)	0.051	1.6(0.9)	1.7(1.0)	0.692	4.5(1.9)	4.4(1.6)	0.301	< 0.0001
MMSE	29.2(1.1)	29.0(1.3)	0.027	22.8(3.5)	22.9(3.8)	0.504	19.6(6.5)	20.5(6.7)	0.453	24.6(6.0)	24.7(5.9)	0.898	23.2(2.2)	22.9(2.0)	0.656	< 0.0001
LDELTOTAL	13.4(3.2)	12.7(3.1)	0.002	12.7(3.3)	12.5(3.3)	0.623	8.8(2.1)	9.3(1.7)	0.007	3.3(2.6)	4.0(2.6)	< 0.001	1.0(1.6)	1.6(2.0)	0.002	< 0.0001
FAQ	0.1(0.3)	0.2(0.7)	0.214	0.3(1.0)	0.5(1.8)	0.937	1.7(3.4)	2.8(3.8)	0.017	4.0(4.6)	4.0(4.3)	0.549	13.3(7.1)	13.3(7.2)	0.326	< 0.0001
EcogSPTotal	1.8(0.7)	1.8(0.7)	0.351	1.7(0.7)	1.8(0.8)	0.932	2.1(0.8)	2.1(0.9)	0.829	1.9(0.8)	2.0(0.8)	0.211	2.3(0.8)	2.4(0.9)	0.160	< 0.0001
ADAS11	5.3(3.0)	6.6(3.0)	< 0.001	10.3(4.2)	10.4(3.9)	0.527	16.9(6.7)	16.4(6.7)	0.137	16.3(5.7)	15.2(6.1)	0.017	35.8(7.6)	36.7(7.1)	0.250	< 0.0001
ADASQ4	2.3(1.4)	2.3(1.7)	0.319	2.7(1.6)	3.5(1.6)	< 0.001	4.0(2.4)	4.4(2.0)	0.082	6.6(1.6)	6.2(1.8)	0.085	8.7(1.4)	8.6(1.5)	0.069	< 0.0001
GDS	0.8(1.1)	0.7(1.0)	0.333	1.0(1.1)	0.9(1.0)	0.529	2.0(1.5)	1.6(1.5)	0.071	1.8(1.6)	1.6(1.5)	0.373	1.8(1.6)	1.8(1.5)	0.094	< 0.0001

Note: The P-values in the last column are for the difference between the five groups (CN, SMC, EMCI, LMCI, and AD). The data means mean (sd.) or number (percent).

The five diagnostic groups (CN, SMC, EMCI, LMCI, and AD) were determined based on participants' baseline diagnostic results. The amyloid status of participants was obtained either from the baseline visit or from the nearest visit with an available amyloid status that corresponded to the baseline diagnosis. The visit date associated with the amyloid status was used to merge with other data files, such as the memory impairment score. Recognizing the significance of sex as a moderating factor in AD research, we further stratified the data for each diagnosis group into two subgroups: females and males.

2.2 Model creation

Demographics. To construct our model, we aimed to incorporate established risk factors associated with amyloid positivity. Therefore, we utilized six demographic variables extracted from the ADNI dataset: age, years of education, Hispanic ethnicity, race (classified as White, African American, or other), marital status (classified as married, never married, divorced, or widowed), and family history of dementia. Due to the limited number of participants from racial groups other than White or African American, we combined them into a single group. Notably, APOE ε4, along with age and ADAS-cog, were identified as the three significant risk factors for predicting amyloid status [30].

Sex. Women have demonstrated an advantage in verbal memory, suggesting that memory-based measures may be less effective in screening women for early changes associated with AD. Furthermore, the strong memory performance exhibited by women may mask early signs of the disease, making it less effective to predict the presence of brain amyloid using memory test scores, especially in women without detectable memory deficits [31]. Therefore, it is crucial to develop separate statistical prediction models for each subpopulation stratified by sex and APOE ε4 status. This is important because there is heterogeneity observed in individuals with SMC, LMCI, and AD. Based on these differences, we have constructed separate models for females and males at each stage of the disease.

Cognitive tests. This study utilized three subsets of the ADAS-cog (ADAS13, ADAS11, and ADASQ4), which cover a range of 0 to 70, 0 to 70, and 0 to 10, respectively. These subsets were used to comprehensively evaluate cognitive functioning and specifically assess memory function. In addition to the ADAS-cog scores, the ML models incorporated various variables, including the CDRSB scores ranging from 0 to 10 [32], MMSE scores ranging from 0 to 30, LDELTOTAL scores ranging from 0 to 25 [33], FAQ scores ranging from 0 to 30 [20], and Ecog-SP scores ranging from 1 to 5 [34]. LDELTOTAL is a component of the Wechsler Memory Scale-Revised (WMS-R) and serves as a specific measure for assessing memory function.

Performance metrics. The ML model with the highest average accuracy is considered the optimal model. Accuracy is commonly defined in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These terms are used to calculate accuracy in the following manner:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

The Matthews Correlation Coefficient (MCC) is a measure of the quality of binary classification models. It TP, TN, FP, and FN to provide a balanced evaluation of the model's performance. The MCC ranges from –1 to +1, with +1 indicating a perfect

classifier, 0 indicating a random classifier, and –1 indicating a classifier that performs worse than random. The formula to calculate MCC is as follows:

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FN) \times (TP + FP) \times (TN + FP) \times (TN + FN)}} \quad (2)$$

By incorporating these neuropsychological scores into the ML models, our aim was to capture the participants cognitive performance and functional abilities. This would provide a comprehensive assessment of their cognitive status and enhance the predictive capabilities of the model.

2.3 Machine learning models

To improve the accuracy of amyloid pathology prediction models and accommodate diverse data patterns, various ML models were implemented. These models offer flexibility in capturing different types of information. Table 2 presents the ML methods that were used, along with their relevant R packages and corresponding categories.

Table 2. ML models in the R package

ID	ML Models	R Package
1	Boosted classification trees	ada
2	Tree models or rule-based models	C5.0
3	Stochastic gradient boosting	gbm
4	Generalized linear model	glm
5	k-nearest neighbors	knn
6	Linear discriminant analysis	lda
7	Boosted logistic regression	LogitBoost
8	Random forest	ranger
9	Random forest	rf
10	Factor-based linear discriminant analysis	Rflda
11	Recursive partitioning and regression trees	rpart
12	Support vector machines with linear kernel	svmLinear
13	Support vector machines with polynomial kernel	svmPoly
14	Support vector machines with radial basis function kernel	svmRadial
15	Bagged CART	treebag

In the context of our study, Monte Carlo simulations were utilized to assess the performance of our ML models for predicting amyloid positivity. During each Monte Carlo simulation, the complete dataset was randomly divided into a training dataset (80%) and a testing dataset (20%). The caret package was used to train the models by utilizing the train function. Furthermore, the expand-grid function was used to perform a grid search and determine the optimal hyperparameters for each model [35]. The train control function in the R package was also used to

specify the cross-validation settings, such as the number of folds and the type of resampling.

Once the optimal hyperparameters for a given ML model was determined, the model was evaluated on the testing dataset. To ensure the statistical stability of the results, the analysis was repeated 1000 times for each ML model. The training and performance assessment procedures are summarized in Figure 2.

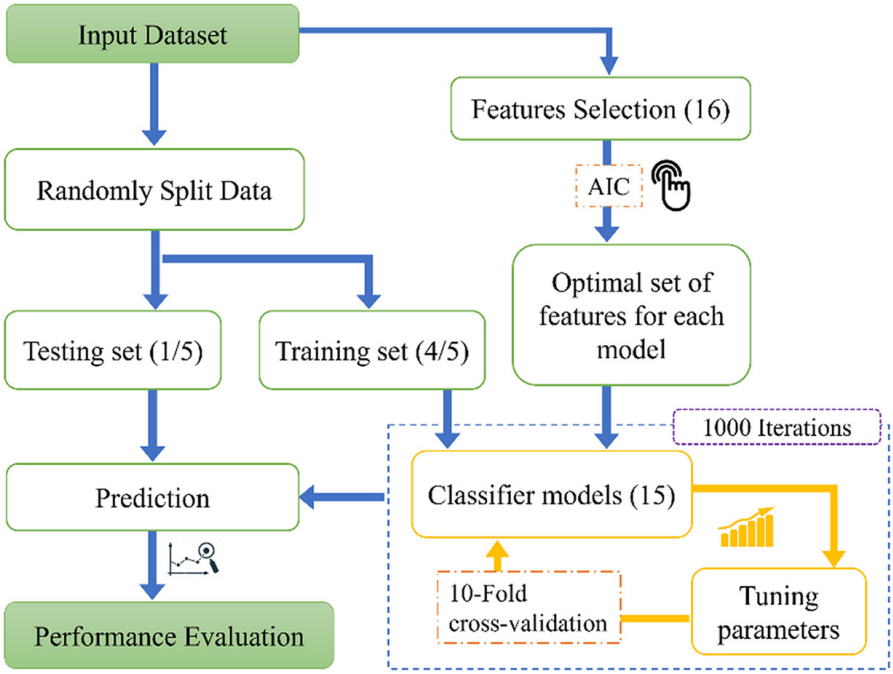

Fig. 2. The evaluation and training process for 15 commonly used ML models

Figure 2, illustrates the evaluation and training process for 15 commonly used ML models. The process includes using 10-fold cross-validation, repeated 1000 times, for each classification model. The outcome of each model is determined by considering the average value obtained through this process.

Feature selections. In this study, we faced the challenge of handling a relatively large number of features, which can lead to computational inefficiencies and increase the risk of overfitting. To address these challenges, we employ the forward model selection method called the Akaike Information Criterion (AIC) [36]. This approach offers a streamlined path to search for the optimal set of features, minimizing computational complexity and facilitating the identification of the most relevant predictors for amyloid positivity.

The process involved fitting F models, with each model incorporating one of the F features. The model with the lowest AIC was chosen, and its corresponding feature was designated as the first feature, denoted as $X_{(1)}$. In the subsequent step, F-1 models were fitted using one of the remaining features, along with $X_{(1)}$ from the previous step. The feature selected for the second position was the one associated with the model that had the lowest AIC among these F-1 models. Let's assume this second feature is $X_{(2)}$. By following this procedure, the ordering of the subsequent F-2 features, $X_{(3)}$, ..., $X_{(F)}$ was determined. For this study, a multiple logistic regression model was used as the statistical model to determine the feature ordering.

To address the computational challenges posed by the large number of feature combinations (2^F), we have implemented a different approach. We focused on F combinations represented by $Z_i = \{X_{(1)}, ..., X_{(I)}\}$, where $i = (1, 2, ..., F)$ This reduced set of combinations provides an efficient means to assess the importance of these features in predicting amyloid positivity. By utilizing this new stochastic ordering approach, we were able to determine the relative significance of each feature more efficiently. Moreover, this approach offers a streamlined path to search for the optimal set of features, minimizing computational complexity and facilitating the identification of the most relevant predictors for amyloid positivity.

3 RESULT AND DISCUSSION

Table 3 presents information on various variables, including gender, amyloid levels, age, education, and Hispanic ethnicity, in the context of different clinical stages of AD and healthy controls. In each group, males were generally older and had a higher level of education compared to females. As AD progressed, the cognitive assessment scores (ADAS13, ADAS11, ADASQ4, CDRSB, and EcogSP total) increased, indicating a greater level of cognitive impairment. Conversely, the scores for MoCA, MMSE, and LDELTOTAL decreased, reflecting a decline in cognitive function. Generally, females exhibited better performance compared to males across these measures.

P-values were computed to compare the five groups (CN, SMC, EMCI, LMCI, and AD) for each characteristic, there were no significant differences. Among these characteristics is Hispanic ethnicity across the five groups. However, the remaining demographic and clinical characteristics exhibited statistically significant variations among the groups. Notably, there were significant variations between genders in terms of amyloid levels, age, education, race, marital status, APOE ε4, family history, ADAS13, MoCA, CDRSB, MMSE, LDELTOTAL, FAQ, EcogSPTotal, ADAS11, ADASQ4, and GDS across different clinical stages of AD and control groups. Consequently, these features were chosen for their role in predicting levels of amyloid. They were included as predictors in the ML models to evaluate their importance and effectiveness in predicting the desired outcomes.

The stochastic ordering method was used to assess the significance of features in each model. However, it is important to note that the rankings of features may differ among models because the predictive relevance of features for amyloid positivity can vary within each specific model. Figure 3 visually presents the results of the analysis, illustrating the number of features investigated for each modelity. It includes a count of features ranging from 1 to 16 and highlights significant values for features from 1 to 5. This range encompasses a diverse range of model complexities.

This range encompasses diverse levels of model complexities. Upon examination of the data, several observations can be inferred:

The assigned scores for each feature vary across different ML models, indicating that the importance of features in predicting the target variable can differ depending on the algorithm used. Among the models, the RF, GBM, and ADA models consistently assign high importance values to most features, indicating their effectiveness in capturing complex relationships.

Features such as ADAS13, CDRSB, LDELTOTAL, and MOCA receive relatively high importance values across multiple models. This suggests that cognitive assessment measures play a crucial role in predicting the target variable, potentially indicating

their significance in evaluating cognitive decline. Additionally, the presence of ADAS13 in several models underscores its significance in assessing cognitive impairment and its potential for identifying individuals at risk of cognitive decline.

Age and the presence of the *APOE4* gene (APOE ε4) are consistently identified as important features in various models [37]. This finding aligns with existing scientific knowledge that emphasizes age and genetic factors as significant contributors to neurodegenerative diseases such as AD. Incorporating these genetic markers enhances the models' ability to predict cognitive decline.

The ECOGSP total feature demonstrates varying importance values across different machine learning models, indicating that it may have a more nuanced relationship with the target variable. Further investigation is needed to understand its precise impact.

Unexpectedly, the performance on the delayed memory task of the GDS, FAQ, and EcogSP total did not significantly contribute to the model's ability to predict underlying amyloid status. This finding suggests that the limited sensitivity and specificity of these measures may not be sufficient for accurately predicting amyloid accumulation during the early stages of AD.

Socio-demographic factors such as PTEDUCAT, PTMARRY, and PTETHCAT generally receive lower importance values, suggesting that they may have a relatively weaker association with the target variable compared to the cognitive and genetic factors.

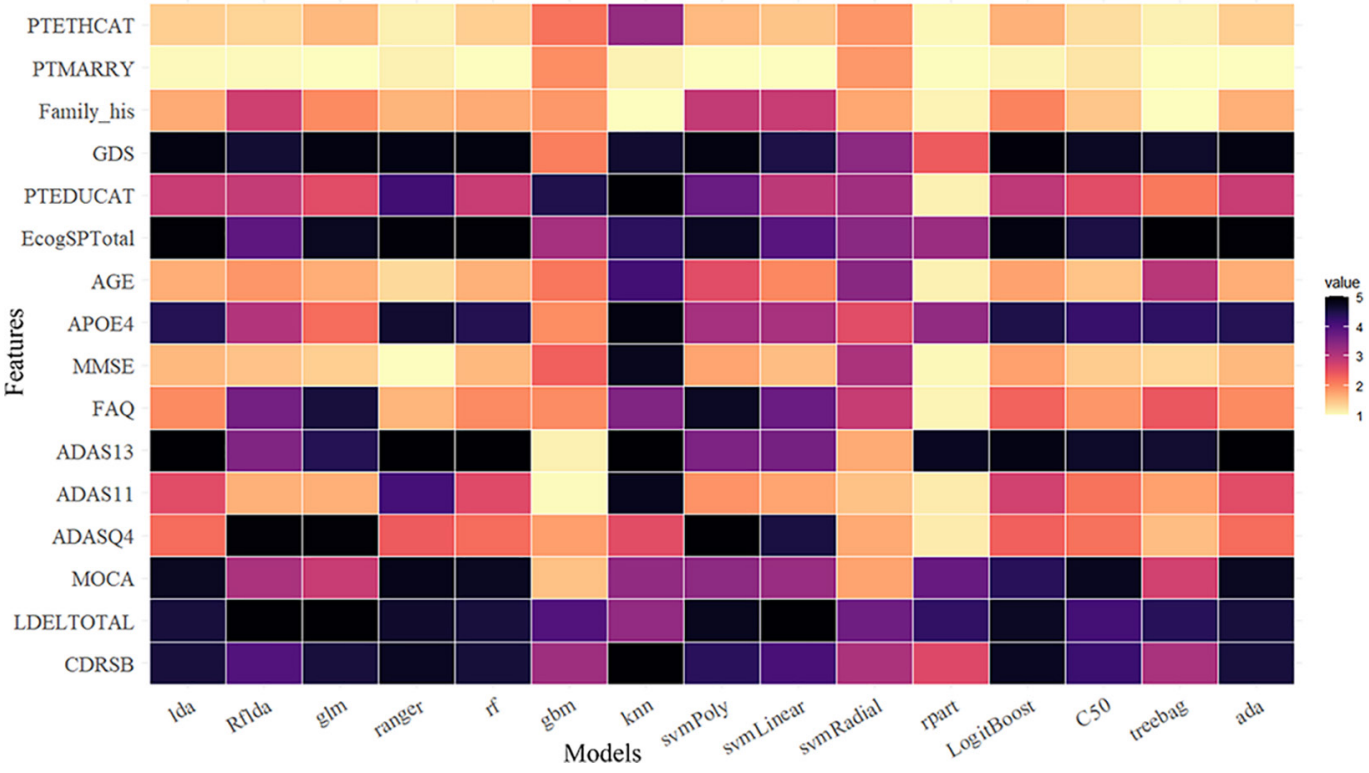

Fig. 3. Heatmap showcasing the importance of features in each model

In Figure 3, each feature is represented by a row, while each column represents a specific ML model. The values within the table indicate the assigned importance scores for each feature by the corresponding ML model. These scores reflect the relative significance of each feature in predicting the target variable. The scale ranges from 1 to 5, with 5 indicating the highest importance and 1 indicating the lowest importance.

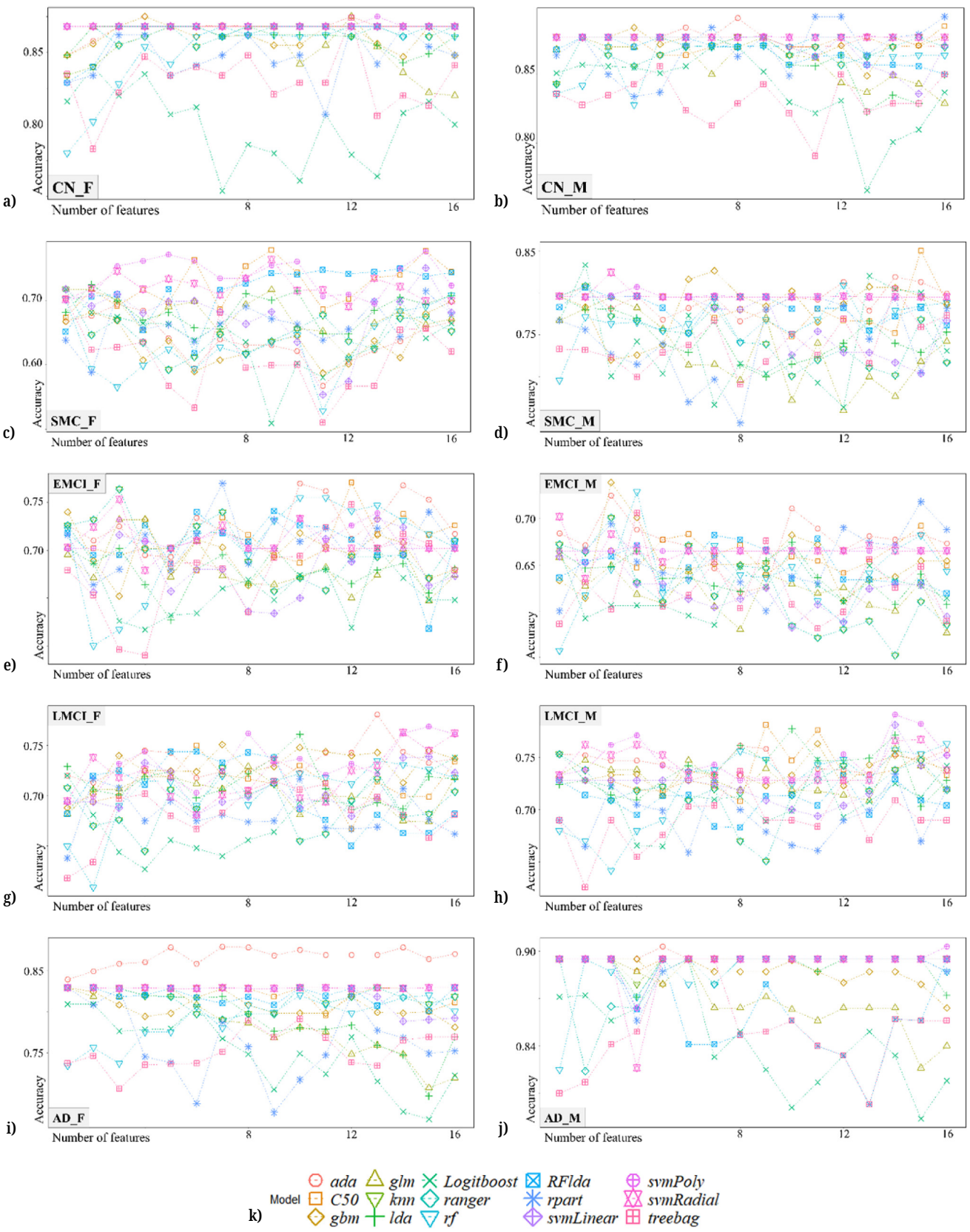

Fig. 4. The relationship between the number of features (from 1 to 16) and the accuracy of 15 ML models in each subgroup. (a) Accuracy of 15 ML models in CN_F; (b) Accuracy of 15 ML models in CN_M; (c) Accuracy of 15 ML models in SMC_F; (d) Accuracy of 15 ML models in SMC_M; (e) Accuracy of 15 ML models in EMCI_F; (f) Accuracy of 15 ML models in EMCI_M; (g) Accuracy of 15 ML models in LMCI_F; (h) Accuracy of 15 ML models in LMCI_M; (i) Accuracy of 15 ML models in AD_F; (j) Accuracy of 15 ML models in AD_M; (k) legend of 15 ML models

Upon examining Figure 4, which depicts the accuracy of 10 subgroups with varying numbers of features, several observations can be made:

The CN group (CN_F: 0.754 to 0.875; CN_M: 0.766 to 0.892) is shown in Figure 4a and b. The ADA and rpart algorithms emerge as the top performers for CN_F and CN_M, respectively. Conversely, models such as Logitboost and svmLinear demonstrate relatively lower scores in comparison. Interestingly, svmPoly and svmRadial consistently achieve similar scores for both subgroups, indicating comparable prediction accuracy for both outcomes.

SMC group (SMC_F: 0.509 to 0.778; SMC_M: 0.645 to 0.850) in Figure 4c and d. The ADA and GBM consistently demonstrate strong predictive capability for both SMC_F and SMC_M, while the Ranger consistently performs well. Models like Logitboost and svmLinear consistently show relatively lower scores. Notably, svmPoly and svmRadial performed similarly well in both subgroups.

EMCI group (EMCI_F: 0.590 to 0.771; EMCI_F: 0.552 to 0.739) in Figure 4e and f. Overall, the GBM, Ranger, and RFlda also show promising performance with above-average scores for both EMCI_F and EMCI_M. Conversely, models such as rpart, svmLinear, and treebag consistently exhibit lower scores.

LMCI group (LMCI_F: 0.609 to 0.781; LMCI_M: 0.626 to 0.791) in Figure 4g and h. The models, such as KNN and rpart, demonstrated lower accuracies. Additionally, decision tree-based models, including treebag, generally yielded lower accuracies compared to other models.

AD group (AD_F: 0.669 to 0.879; AD_M: 0.794 to 0.903) in Figure 4i and j. The AdaBoost, C50, and GBM consistently achieved high accuracy, surpassing other models. Additionally, KNN LDA, and Ranger showed strong performance. Notably, the ADA consistently distinguished between AD_F and AD_M instances, while decision tree-based algorithms C50 and ranger exhibited robust performance in capturing underlying patterns.

Table 3. The optimal ML model and features for each subgroup

Dia.	Sex	Opt. ML	Num	Features
CN	F	ADA	4	CDRSB, LDELTOTAL, MOCA, etc.
CN	F	GBM	4	ADAS13, AGE, MOCA, etc.
CN	F	svmPoly	13	CDRSB, LDELTOTAL, MOCA, etc.
CN	M	rpart	16(All)	LDELTOTAL, MOCA, CDRSB, etc.
SMC	F	C50	15	MMSE, ADAS11, CDRSB, etc.
SMC	M	C50	15	MMSE, ADAS11, CDRSB, etc.
EMCI	F	C50	12	MMSE, ADAS11, CDRSB, etc.
EMCI	M	GBM	3	ADAS13, AGE, MOCA
LMCI	F	ADA	13	CDRSB, LDELTOTAL, MOCA, etc.
LMCI	M	svmPoly	14	CDRSB, LDELTOTAL, MOCA, etc.
AD	F	ADA	7	CDRSB, LDELTOTAL, MOCA, etc.
AD	M	ADA	5	CDRSB, LDELTOTAL, MOCA, etc.
AD	M	LDA	11	MMSE, ADAS13, ADAS11, etc.
AD	M	svmPoly	16(All)	CDRSB, LDELTOTAL, MOCA, etc.

Notes: Dia: Diagnosis; Opt: Optimal.

The ADA model consistently demonstrated strong performance across multiple groups and feature dimensions when compared to other models. This indicates its effectiveness in handling varying numbers of features. Similarly, the GBM, ranger and svmPoly models exhibited robust and consistent performance across different subgroups. On the other hand, simpler models such as decision trees or linear models may offer interpretability, providing insights into the relationships between features and predictions. Notably, ADA or LDA models tended to incorporate a higher number of features while maintaining good generalization and consistent performance.

Table 3 presents the findings of a dataset that includes diverse diagnoses, sex, ML models, the number of features, and their corresponding optimal features. It provides valuable insights into selecting the optimal ML algorithms and the specific features used for each subgroup.

4 CONCLUSION

Our approach introduces a novel feature selection method that utilizes stochastic rankings of features to identify the optimal features within each subgroup, resulting in an optimal ML model. This approach aims to reduce fragility and enhance generalizability, leading to improved interpretability of ML predictions. By selecting the most relevant features, we can improve the model's robustness and its applicability to different scenarios. This claim is supported by the evidence presented in Figures 3 and 4 from our research data and findings.

The unique characteristics and biomarkers associated with cognitive decline and neurodegenerative diseases are highlighted by the distinct combination of features specific to each diagnostic group. This correlation is strongly supported by the data presented in Table 2. The select features encompass a wide range of cognitive, functional, demographic, and genetic factors that greatly contribute to the predictive accuracy of the ML models, as demonstrated through analysis.

Furthermore, the results underscore the significance of specific features and provide insights into the underlying factors that contribute to cognitive decline. This conclusion is supported by the data presented in both Figure 4 and Table 3. The observed variations in optimal ML algorithms and the number of features among different diagnoses and genders suggest that the underlying mechanisms and predictive factors for cognitive impairment may differ across patient groups. This comprehensive set enables early identification and intervention for individuals at risk of Aβ positivity, ultimately leading to improved outcomes.

While the findings offer valuable insights for future research and model selection in predicting disease progression, it is important to acknowledge the study's limitations. Potential confounding factors and the generalizability of the models to diverse populations should be taken into consideration [38]. To enhance the reliability and applicability of the models, future research should prioritize the validation and replication of these findings using larger and more diverse cohorts.

5 ACKNOWLEDGMENT

The authors express their sincere gratitude to the Editor, Associate Editor, and reviewers for their invaluable and insightful comments, which have greatly enhanced the quality of the manuscript. The research is partially supported by the National Natural Science Foundation of China (NSFC) 62006165, a Postdoctoral

Fellowship Award from the American Heart Association (AHA) 23POST1022454, the National Institutes of Health (NIH) R01EB029570A1, and the Education Department of Anhui Province: the Key Research Projects of Overseas Outstanding Young Talents in Colleges and Universities GXGWFX2019031.

6 REFERENCES

[1] Y. Zeng and H. Chen, "Analysis of trends of future home-based care needs and costs for the elderly in China," *Trends and Determinants of Healthy Aging in China*, Springer, pp. 95–120, 2022. https://doi.org/10.1007/978-981-19-4154-2_6

[2] S. T. Ahmed and S. M. Kadhem, "Early alzheimer's disease detection using different techniques based on microarray data: A review," *International Journal of Online & Biomedical Engineering*, vol. 18, no. 4, pp. 106–126, 2022. https://doi.org/10.3991/ijoe.v18i04.27133

[3] L. Guzman-Martinez, C. Calfío, G. A. Farias, C. Vilches, R. Prieto, and R. B. Maccioni, "New frontiers in the prevention, diagnosis, and treatment of Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 82, no. s1, pp. S51–S63, 2021. https://doi.org/10.3233/JAD-201059

[4] J. Rasmussen and H. Langerman, "Alzheimer's disease–why we need early diagnosis," *Degenerative Neurological and Neuromuscular Disease*, pp. 123–130, 2019. https://doi.org/10.2147/DNND.S228939

[5] A. Nakamura, N. Kaneko, V. L. Villemagne, T. Kato, J. Doecke, V. Doré, C. Fowler, Q.-X. Li, R. Martins, and C. Rowe, "High performance plasma amyloid-β biomarkers for Alzheimer's disease," *Nature*, vol. 554, no. 7691, pp. 249–254, 2018. https://doi.org/10.1038/nature25456

[6] A. Leuzy, N. Mattsson-Carlgren, S. Palmqvist, S. Janelidze, J. L. Dage, and O. Hansson, "Blood-based biomarkers for Alzheimer's disease," *EMBO Molecular Medicine*, vol. 14, no. 1, p. e14408, 2022. https://doi.org/10.15252/emmm.202114408

[7] P. N. Young, M. Estarellas, E. Coomans, M. Srikrishna, H. Beaumont, A. Maass, A. V. Venkataraman, R. Lissaman, D. Jiménez, and M. J. Betts, "Imaging biomarkers in neurodegeneration: Current and future practices," *Alzheimer's Research & Therapy*, vol. 12, no. 1, pp. 1–17, 2020. https://doi.org/10.1186/s13195-020-00612-7

[8] W. Bonadio, "Pediatric lumbar puncture and cerebrospinal fluid analysis," *The Journal of Emergency Medicine*, vol. 46, no. 1, pp. 141–150, 2014. https://doi.org/10.1016/j.jemermed.2013.08.056

[9] S. E. Schindler and A. Atri, "The role of cerebrospinal fluid and other biomarker modalities in the Alzheimer's disease diagnostic revolution," *Nature Aging*, vol. 3, no. 5, pp. 460–462, 2023. https://doi.org/10.1038/s43587-023-00400-6

[10] D. Mukherji, M. Mukherji, and N. Mukherji, "Early detection of Alzheimer's disease using neuropsychological tests: A predict–diagnose approach using neural networks," *Brain Informatics*, vol. 9, no. 1, pp. 1–26, 2022. https://doi.org/10.1186/s40708-022-00169-1

[11] C. O'Driscoll and M. Shaikh, "Cross-cultural applicability of the Montreal Cognitive Assessment (MoCA): A systematic review," *Journal of Alzheimer's Disease*, vol. 58, no. 3, pp. 789–801, 2017. https://doi.org/10.3233/JAD-161042

[12] A. P. Lima, R. Castilhos, and M. L. Chaves, "The use of the clinical dementia rating scale sum of boxes scores in detecting and staging cognitive impairment/dementia in Brazilian patients with low educational attainment," *Alzheimer Disease & Associated Disorders*, vol. 31, no. 4, pp. 322–327, 2017. https://doi.org/10.1097/WAD.0000000000000205

[13] F. Panza, V. Frisardi, D. Seripa, G. Logroscino, A. Santamato, B. P. Imbimbo, E. Scafato, A. Pilotto, and V. Solfrizzi, "Alcohol consumption in mild cognitive impairment and dementia: Harmful or neuroprotective?" *International Journal of Geriatric Psychiatry*, vol. 27, no. 12, pp. 1218–1238, 2012. https://doi.org/10.1002/gps.3772

[14] R. Li, and M. Singh, "Sex differences in cognitive impairment and Alzheimer's disease," *Frontiers in Neuroendocrinology*, vol. 35, no. 3, pp. 385–403, 2014. https://doi.org/10.1016/j.yfrne.2014.01.002

[15] S. Subramaniapillai, A. Almey, M. N. Rajah, and G. Einstein, "Sex and gender differences in cognitive and brain reserve: Implications for Alzheimer's disease in women," *Frontiers in Neuroendocrinology*, vol. 60, p. 100879, 2021. https://doi.org/10.1016/j.yfrne.2020.100879

[16] A. J. Karstens, T. R. Maynard, and G. Tremont, "Sex-specific differences in neuropsychological profiles of mild cognitive impairment in a hospital-based clinical sample," *Journal of the International Neuropsychological Society*, pp. 1–10, 2023. https://doi.org/10.1017/S1355617723000085

[17] C. Bastin and E. Salmon, "Early neuropsychological detection of Alzheimer's disease," *European Journal of Clinical Nutrition*, vol. 68, no. 11, pp. 1192–1199, 2014. https://doi.org/10.1038/ejcn.2014.176

[18] C.-Y. Wu, C.-Y. Ho, and Y.-H. Yang, "Developing biomarkers for the skin: Biomarkers for the diagnosis and prediction of treatment outcomes of Alzheimer's disease," *International Journal of Molecular Sciences*, vol. 24, no. 10, p. 8478, 2023. https://doi.org/10.3390/ijms24108478

[19] J. Sui, R. Jiang, J. Bustillo, and V. Calhoun, "Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: Methods and promises," *Biological Psychiatry*, vol. 88, no. 11, pp. 818–828, 2020. https://doi.org/10.1016/j.biopsych.2020.02.016

[20] D. A. González, M. M. Gonzales, Z. J. Resch, A. C. Sullivan, and J. R. Soble, "Comprehensive evaluation of the Functional Activities Questionnaire (FAQ) and its reliability and validity," *Assessment*, vol. 29, no. 4, pp. 748–763, 2022. https://doi.org/10.1177/1073191121991215

[21] R. L. Nosheny, M. R. Camacho, P. S. Insel, D. Flenniken, J. Fockler, D. Truran, S. Finley, A. Ulbricht, P. Maruff, and K. Yaffe, "Online study partner-reported cognitive decline in the Brain Health Registry," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 4, pp. 565–574, 2018. https://doi.org/10.1016/j.trci.2018.09.008

[22] A. L. Gross, P. Y. Khobragade, E. Meijer, and J. A. Saxton, "Measurement and structure of cognition in the longitudinal aging study in India–diagnostic assessment of dementia," *Journal of the American Geriatrics Society*, vol. 68, pp. S11–S19, 2020. https://doi.org/10.1111/jgs.16738

[23] J. K. Kueper, M. Speechley, and M. Montero-Odasso, "The Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog): Modifications and responsiveness in pre-dementia populations. A narrative review," *Journal of Alzheimer's Disease*, vol. 63, no. 2, pp. 423–444, 2018. https://doi.org/10.3233/JAD-170991

[24] S. T. Creavin, S. Wisniewski, A. H. Noel-Storr, C. M. Trevelyan, T. Hampton, D. Rayment, V. M. Thom, K. J. Nash, H. Elhamoui, and R. Milligan, "Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations," *Cochrane Database of Systematic Reviews*, vol. 2016, no. 4, 2016. https://doi.org/10.1002/14651858.CD011145.pub2

[25] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, J. Cedarbaum, M. C. Donohue, R. C. Green, D. Harvey, and C. R. Jack Jr, "Impact of the Alzheimer's disease neuroimaging initiative, 2004 to 2014," *Alzheimer's & Dementia*, vol. 11, no. 7, pp. 865–884, 2015. https://doi.org/10.1016/j.jalz.2015.04.005

[26] S. M. Landau, A. Fero, S. L. Baker, R. Koeppe, M. Mintun, K. Chen, E. M. Reiman, and W. J. Jagust, "Measurement of longitudinal β-amyloid change with 18F-florbetapir PET and standardized uptake value ratios," *Journal of Nuclear Medicine*, vol. 56, no. 4, pp. 567–574, 2015. https://doi.org/10.2967/jnumed.114.148981

[27] P. J. Lao, T. J. Betthauser, A. T. Hillmer, J. C. Price, W. E. Klunk, I. Mihaila, A. T. Higgins, P. D. Bulova, S. L. Hartley, and R. Hardison, "The effects of normal aging on amyloid-β deposition in nondemented adults with Down syndrome as imaged by carbon 11–labeled Pittsburgh compound B," *Alzheimer's & Dementia*, vol. 12, no. 4, pp. 380–390, 2016. https://doi.org/10.1016/j.jalz.2015.05.013

[28] C. Rabe, T. Bittner, A. Jethwa, I. Suridjan, E. Manuilova, M. Friesenhahn, E. Stomrud, H. Zetterberg, K. Blennow, and O. Hansson, "Clinical performance and robustness evaluation of plasma amyloid-β42/40 prescreening," *Alzheimer's & Dementia*, vol. 19, no. 4, pp. 1393–1402, 2023. https://doi.org/10.1002/alz.12801

[29] M. C. Carrillo, L. J. Bain, G. B. Frisoni, and M. W. Weiner, "Worldwide Alzheimer's disease neuroimaging initiative," *Alzheimer's & Dementia*, vol. 8, no. 4, pp. 337–342, 2012. https://doi.org/10.1016/j.jalz.2012.04.007

[30] M. Ba, K. Ng, X. Gao, M. Kong, L. Guan, L. Yu, and Alzheimer's Disease Neuroimaging Initiative, "The combination of apolipoprotein E4, age and Alzheimer's Disease Assessment Scale–Cognitive Subscale improves the prediction of amyloid positron emission tomography status in clinically diagnosed mild cognitive impairment," *European Journal of Neurology*, vol. 26, no. 5, p. 733, 2019. https://doi.org/10.1111/ene.13881

[31] G. Shan, C. Bernick, J. Z. Caldwell, and A. Ritter, "Machine learning methods to predict amyloid positivity using domain scores from cognitive tests," *Scientific Reports*, vol. 11, no. 1, p. 4822, 2021. https://doi.org/10.1038/s41598-021-83911-9

[32] S. E. O'Bryant, L. H. Lacritz, J. Hall, S. C. Waring, W. Chan, Z. G. Khodr, P. J. Massman, V. Hobson, and C. M. Cullum, "Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database," *Archives of Neurology*, vol. 67, no. 6, pp. 746–749, 2010. https://doi.org/10.1001/archneurol.2010.115

[33] E. Fokuoh, D. Xiao, W. Fang, Y. Liu, Y. Lu, and K. Wang, "Longitudinal analysis of APOE-ε4 genotype with the logical memory delayed recall score in Alzheimer's disease," *Journal of Genetics*, vol. 100, pp. 1–9, 2021. https://doi.org/10.1007/s12041-021-01309-y

[34] R. L. Nosheny, C. Jin, J. Neuhaus, P. S. Insel, R. S. Mackin, M. W. Weiner, and Alzheimer's Disease Neuroimaging Initiative investigators, "Study partner-reported decline identifies cognitive decline and dementia risk," *Annals of Clinical and Translational Neurology*, vol. 6, no. 12, pp. 2448–2459, 2019. https://doi.org/10.1002/acn3.50938

[35] C. L. Rodríguez Velasco, E. García Villena, J. Brito Ballester, F. Á. Durántez Prados, E. R. Silva Alvarado, and J. Crespo Álvarez, "Forecasting of post-graduate students' late dropout based on the optimal probability threshold adjustment technique for imbalanced data," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 4, pp. 120–155, 2023. https://doi.org/10.3991/ijet.v18i04.34825

[36] S. Portet, "A primer on model selection using the akaike information criterion," *Infectious Disease Modelling*, vol. 5, pp. 111–128, 2020. https://doi.org/10.1016/j.idm.2019.12.010

[37] M. E. Belloy, V. Napolioni, S. S. Han, Y. Le Guen, M. D. Greicius, and Alzheimer's Disease Neuroimaging Initiative, "Association of Klotho-VS heterozygosity with risk of Alzheimer disease in individuals who carry APOE4," *JAMA Neurology*, vol. 77, no. 7, pp. 849–862, 2020. https://doi.org/10.1001/jamaneurol.2020.0414

[38] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine*, vol. 17, pp. 1–9, 2019. https://doi.org/10.1186/s12916-019-1426-2

7 AUTHORS

Xing Wei is an Associate Professor at Bengbu Medical College, China, and concurrently serves as a visiting scholar at c, USA. Doctor's degree graduated from Xiangya

School of Medicine, Central South University, majoring in medical information management (E-mail: weixing@bbmc.edu.cn; xwei2@mtu.edu).

Na Gao is a postgraduate student at Bengbu Medical College, training in Database Management, Analysis, and Information Management in medical information (E-mail: gaon0713@163.com).

Mengru Xu is a postgraduate student at Bengbu Medical College, specializing Python, C++ and R programming languages (E-mail: xymengru1031@163.com).

Mostafa Rezaeitalesh is a Ph.D. in the Department of Biomedical Engineering at Michigan Technical University (E-mail: srezaeit@mtu.edu).

Nan Mu is a postdoctoral scholar in the Department of Biomedical Engineering at Michigan Technical University (E-mail: nmu2@mtu.edu).

Zonghan Lyu is a Ph.D. in the Department of Biomedical Engineering at Michigan Technical University (E-mail: zonghanl@mtu.edu).

Michael Ngala is a postgraduate student in the Department of Computer Science at Michigan Technical University, specializing in Python, C++ and R programming languages (E-mail: mkngala@mtu.edu).

Jingfeng Jiang is a Professor at the Department of Biomedical Engineering at Michigan Technical University. Research Interests: Biomechanics (Bio-solids and Biofluids), Medical Image/Signal Processing, Machine Learning and Computer Vision with Applications in Medical Imaging (E-mail: jjiang1@mtu.edu).

Xuelian Chang is an Associate Professor at Bengbu Medical College, China. Doctor's degree graduated from Nanjing Medical University, majoring in basic medicine (E-mail: xuelianchang@126.com).