The Journal of Prevention of Alzheimer's Disease 12 (2025) 100083

Contents lists available at ScienceDirect

The Journal of Prevention of Alzheimer's Disease

journal homepage: www.elsevier.com/locate/tjpad

Original Article

Joint ensemble learning-based risk prediction of Alzheimer's disease among mild cognitive impairment patients

Tianyuan Guan[a,b,#], Lei Shang[a,b,#], Peng Yang[a,b], Zhijun Tan[a,b], Yue Liu[a,b], Chunling Dong[a,b], Xueying Li[a,b], Zuxuan Hu[a,b], Haixia Su[a,b,*], Yuhai Zhang[a,b,*]

[a] Department of Health Statistics, School of Public Health, Airforce Medical University, Xian, Shaanxi, China
[b] Ministry of Education Key Lab of Hazard Assessment and Control in Special Operational Environment, Xian, Shaanxi, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Prediction model
Mild cognitive impairment
Joint ensemble learning
Extreme gradient boosting
NACC

ABSTRACT

Objective: Due to the recognition for the importance of early intervention in Alzheimer's disease (AD), it is important to focus on prevention and treatment strategies for mild cognitive impairment (MCI). This study aimed to establish a risk prediction model for AD among MCI patients to provide clinical guidance for primary medical institutions.

Methods: Data from MCI subjects were obtained from the NACC. Importance ranking and the SHapley Additive exPlanations (SHAP) method for the Random Survival Forest (RSF) and Extreme Gradient Boosting (XGBoost) algorithms in ensemble learning were adopted to select the predictors, and hierarchical clustering analysis was used to mitigate multicollinearity. The RSF, XGBoost and Cox proportional hazard regression (Cox) models were established to predict the risk of AD among MCI patients. Additionally, the effects of the three models were evaluated.

Results: A total of 3674 subjects with MCI were included. Thirteen predictors were ultimately identified. In the validation set, the concordance indices were 0.781 (RSF), 0.781 (XGBoost), and 0.798 (Cox), and the Integrated Brier Score was 0.087 (Cox). The prediction effects of the XGBoost and RSF models were not better than those of the Cox model.

Conclusion: The ensemble learning method can effectively select predictors of AD risk among MCI subjects. The Cox proportional hazards regression model could be used in primary medical institutions to rapidly screen for the risk of AD among MCI patients once the model is fully clinically validated. The predictors were easy to explain and obtain, and the prediction of AD was accurate.

1. Introduction

Dementia is a syndrome characterized by progressive cognitive impairment, and it leads to significant declines in patients' daily living abilities, learning abilities, working abilities, and social interaction abilities. Dementia can be categorized into degenerative and nondegenerative types according to its different causes and diseases. Alzheimer's disease (AD) is a degenerative type of dementia, and the National Institute on Aging and Alzheimer's Association (NIA-AA) defines AD as β-amyloid and pathological tau deposition. In the 2023 NIA-AA guidelines [1], the stage 0 was added based on the numeric clinical stages1 through 6 from the 2018 guidelines [2], which means no evidence of clinical change and biomarkers in normal range. In the clinic, the three cognition stages are commonly used by practitioners [3]: cognitively unimpaired (CU), mild cognitive impairment (MCI) [4], and dementia. According to WHO statistics [2], there were more than 50 million dementia patients worldwide in 2023, and this number is predicted to increase to 139 million by 2050, with AD accounting for 60–70 % of the cases. The number of MCI patients is far greater than the number of AD patients. There are approximately 6.38 million AD patients in the United States, and there are approximately 12.82 million MCI patients [5]. The academic community generally believes that MCI is an important transitional stage in the progression of AD and that MCI patients are at a high risk of AD. Studies have shown that 10 % ~ 15 % MCI cases [6] per year [7] are associated with a high risk of progressing to dementia. However, some patients remain at the MCI stage and even improve to normal cognition. Therefore, risk prediction and individualized treatment may be important at the MCI stage. While biomarkers are crucial for staging AD, there are limitations [1]. When detecting mild AD neuropathological changes (ADNPCs), the sensitivity of neu-

* Corresponding authors.
E-mail addresses: suhaixia1225@163.com (H. Su), zhyh@fmmu.edu.cn (Y. Zhang).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.tjpad.2025.100083
Received 19 November 2024; Received in revised form 26 December 2024; Accepted 24 January 2025
Available online 5 February 2025

ropathological examination is greater than that of biomarkers (positron emission tomography [PET], cerebrospinal fluid and blood) [1]. Similarly, the cost of examinations such as biomarkers is relatively high, and their use for early screening of disease is inconvenient. In recent years, demographic characteristics [8–11] and neuropsychological assessment scales [12,13] have been used to predict the progression from MCI to AD [14], thus facilitating the development of individualized treatment plans for MCI patients.

Survival analysis in the medical field [15] is widely used to explore the probability that an individual or group will survive or die within a period and to obtain the relevant predictors of this outcome. The traditional Cox proportional hazard regression model is widely used in survival analysis. However, this model also has several limitations. One is the difficulty in addressing nonlinear relationships between logarithmic proportional risk and combinations of predictor variables. Moreover, if there is a high correlation (collinearity) among multiple predictors in the survival data, the traditional Cox model may be unstable and inaccurate. Machine learning and deep learning are characterized by the ability to address nonlinearity, strongly correlated variables and accurate predictions. In recent years, machine learning and deep learning have been widely used in the risk prediction of cognitive impairment [16]. Ensemble learning is a machine learning method that achieves better overall forecasting performance by combining the forecast results of multiple basic learners (weak classifiers or regressors). The advantages of ensemble learning compared with single models include higher prediction accuracy and good model generalization ability. Ensemble learning can process complex high-dimensional data and has robustness. SHapley additive explanation (SHAP) can help explain the prediction process of the black-box model [15], visualize the detailed relationships between the prediction results and the predictors, improve the interpretability and credibility of the model, help understand the decision process and identify important predictors. Ensemble learning can be used to construct a risk prediction model of the progression from MCI to AD, and the SHAP method can help to identify important predictors. Moreover, the Cox proportional hazards model has the advantage of being easier to interpret and can be used to screen high-risk populations in primary medical institutions.

In this study, we conducted survival analysis [17] via the XGBoost [16,18] and RSF models [16,19] in ensemble learning based on data from the National Alzheimer's Coordinating Center (NACC). The SHAP and permutation importance methods were used to interpret the model and identify important predictors [20]. A Cox proportional hazards regression model was subsequently constructed to predict the risk of AD among MCI patients and might have the potential to help inform interventional plans.

2. Methods

2.1. Data sources

This study used the UDS version 3.0 from the National Alzheimer's Coordinating Center (NACC) (https://www.naccdata.org) [21]. The NACC follow-up data were collected by Alzheimer's Disease Research Centers (ADRCs) located at medical institutions across the United States. Data collection began in September 2005 and was frozen in December 2023. UDS 3.0 aims to address the deficiencies of the previous version and includes updated data on episodic memory, processing speed, executive function, language and constructive ability, etc. [22].

2.2. Study subjects

This study used data from 27,418 subjects in the UDS 3.0. The inclusion criteria were as follows: 1. subjects with MCI (note: the first visit at which MCI was diagnosed was considered the baseline for this analysis). The diagnoses of MCI and AD were determined by the investigator and documented in the database. The variables of survival events are presented in the footnotes of Supplementary eTable 2 in the Appendix. 2. subjects with ≥ 2 follow-up visits. The exclusion criterion was as follows: subjects with more than 30 % missing data for preselected predictor variables. Ultimately, 3674 research subjects were included. The dataset was divided into a training set and a validation set at a 7:3 ratio. The training set is used to train models, whereas the validation set is used to verify the effects of the established models. In the training set, approximately 26 % of the subjects developed AD. In the validation set, approximately 27 % of the subjects developed AD. The flow chart of the subject selection process is shown in Fig. 1.

2.3. Data preprocessing

2.3.1. Definition of survival event outcome

This study defined the survival event outcome as follows: in the most recent follow-up, the subjects were clinically evaluated as having dementia and diagnosed with AD. If AD was still not diagnosed at the last follow-up, the patient was censored. The time was measured in months.

2.3.2. Missing data

The overall rate of missing data was 4.75 %. The missing value statistics of predictors are shown in Supplementary eTable 1 in the Appendix. The missing data were not completely missing at random [23]; therefore, the random forest imputation method was used for data imputation [24].

2.3.3. Variable screening

Based on previous studies, 58 predictors that may be associated with the pathogenesis of AD were prescreened [25], including demographic characteristics, lifestyle characteristics, comorbidities, and neuropsychological measurements, etc. The codes of candidate predictors in the NACC are shown in Supplementary eTable 2 in the Appendix. Ultimately, a total of 40 predictors were preliminarily included. The 40 preliminarily selected predictors are shown in Supplementary eTable 1, and a description of the main cognitive tests is provided in Supplementary eMethods in the Appendix.

The 40 preliminarily selected predictors were further selected according to the predictor importance ranking methods of the RSF (Bagging: Random Survival Forest) and XGBoost (Boosting: Extreme Gradient Boosting) evaluation models. For each method, the SHAP and permutation importance methods were used. A total of four ranking results by predictors were generated.

To avoid collinearity, hierarchical clustering was performed on the 40 preliminarily selected predictors, and a heatmap was drawn according to Spearman's rank correlation [26]. If the predictors were strongly correlated, they were clustered into one category. A threshold of 0.75 was used to transect the tree diagram, and the model output the classification results, which were then combined with the ranking results of the importance of the predictors to screen out meaningful classifications.

The variable selection criteria were as follows. 1. The predictors that satisfy the following three conditions were retained simultaneously: predictors with permutation importance scores greater than 0, predictors with SHAP values for RSF greater than 1, and predictors with SHAP values for XGBoost greater than 0.01. 2. Based on the results of the four predictor importance–ranking methods, predictors with high collinearity were excluded. 3. Variables with clinical significance and easy collection were prioritized for retention.

2.4. Model establishment

Using the final selected predictors, four models were established: the RSF, XGBoost, and Cox proportional hazards regression models. Since there would be some costs involved in testing the ApoEε4 gene, two Cox model were established (including or excluding the variable "number of ApoEε4 alleles") to test whether excluding the ApoEε4 would significantly impact the model's effect.

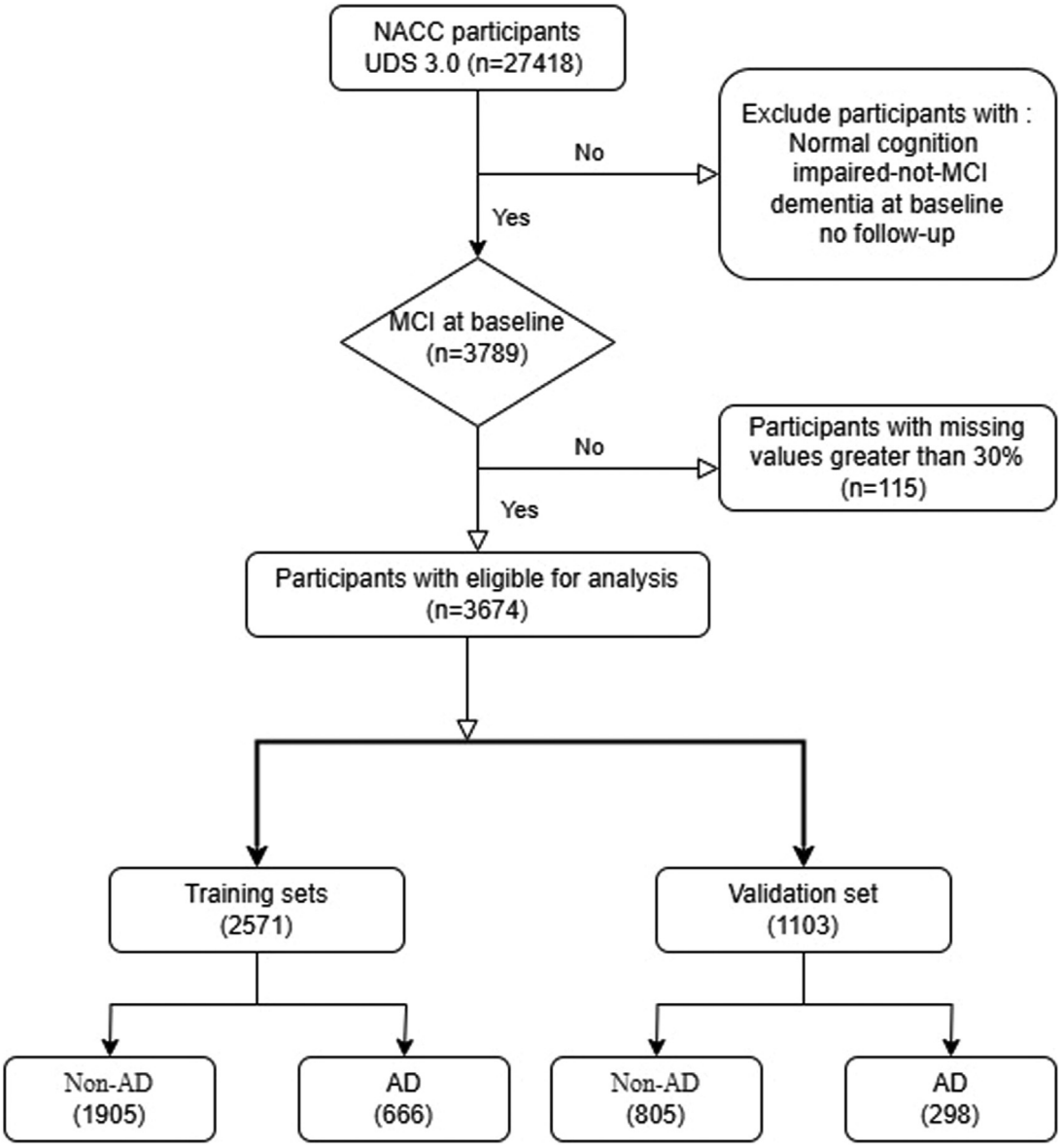

Fig. 1. Flow chart of the subject selection process.

2.4.1. RSF model

The RSF model was established by using the bagging algorithm represented by Random Survival Forest in ensemble learning. The parameters were set to n_estimators=1000, min_samples_split=10, min_samples_leaf=10, max_depth=10, and max_features='sqrt'. The remaining parameters were default in the scikit survival library.

2.4.2. XGBoost model

The XGBoost model was established by using the boosting algorithm represented by XGBoost in ensemble learning. The parameters were set as follows: n_estimators=200, objective= 'survival:cox', eval_metric= 'cox-nloglik', eta= 0.03, tree_method='approx', max_depth=3, subsample =1, colsample_bytree=1, gamma=0.5. The remaining parameters were set to default values.

2.4.3. Cox model

The final selected predictors were used to establish a Cox model. Many studies suggest that the ApoEε4 gene cannot be ignored, but detection of the ApoEε4 gene is uncommon in actual work. Therefore, we also established a Cox model that did not include the ApoEε4 gene in the predictor (excluding ApoEε4) and compared it with the effect of the Cox model including this gene.

2.4.4. Model validation

The fivefold cross-validation method was used to perform internal cross-validation on the training set. The survival analysis model used the concordance index (C-index), Integrated Brier Score (IBS) and AUC at different time points to assess the effect. We also measured the model accuracy, sensitivity, specificity, precision, and F1 score. Moreover, the SHAP method was used to visualize the contribution of each predictor to the target.

2.5. Statistical analysis

In descriptive statistics, continuous data that follow a normal distribution are represented by means and standard deviations, whereas data that do not follow a normal distribution are represented by medians and interquartile ranges. Categorical data are presented as frequencies and percentages. The measurement data were compared between groups via Wilcoxon rank sum tests. The count data were compared between

groups via chi-square tests. Univariate Cox proportional hazards regression was used to calculate the risk of MCI progressing to AD caused by each predictor.

Risk prediction scores were determined for all study subjects via the Cox proportional hazards regression model. Those with risk prediction scores greater than the median were regarded as the high-risk group, and those with scores lower than the median were defined as the low-risk group. The survival of different risk groups was determined via the Kaplan–Meier method. The survival times of the two groups were compared via the log-rank test.

Data processing, analysis, model building, and result visualization were implemented in Python (v3.7 and v3.9) and IBM SPSS Statistics 26.

3. Results

3.1. Basic characteristics of the subjects

A total of 3674 subjects were included in this study, of which 82.2 % (n = 3020) were white and 50.3 % (n = 1849) were women. The mean age of the subjects was 75 years (IQR 69–81), the mean education level was 16 years (IQR 14–18), and the mean BMI was 26.4 (23.7–29.5). The median follow-up time was 24.5 months (IQR 13.4–40.4). During the follow-up, a total of 964 subjects progressed from MCI to AD. Among the subjects who progressed to AD, compared with the subjects who did not progress to AD by the last follow-up, 87.8 % were white (n = 846, p value<0.001), and 50.3 % were women (n = 485, p value=0.991). The subjects who progressed to AD had a mean age of 76 years (IQR 70–82, p value<0.001), a mean education level of 16 years (IQR 14–18, p value=0.115), and a BMI of 25.4 (23–28.18, p value<0.001). A total of 45.3 % of the subjects who developed AD did not have the ApoEε4 gene (n = 437, p value <0.001). The main predictors at baseline are shown in Table 1, and details of the other baseline predictors are shown in Supplementary eTable 3 in the Appendix.

3.2. Selection results of predictors

3.2.1. Importance ranking results

In the RSF model and XGB model, there were 32 predictors with permutation importance scores greater than 0 for the RSF model and 30 predictors with permutation importance scores greater than 0 for the XGB model. The two results were similar, but the rankings were slightly different. There were 27 predictors whose SHAP scores in the RSF model were greater than 1. There were 25 predictors for which the SHAP score of the XGBoost model was greater than 0.01. Nineteen predictors were selected, all of which satisfied the abovementioned conditions. Among these predictors, Craft Story 21 Recall – Delayed and Benson Complex Figure Copy were the most important among the four ranking results. The total score of the CDR scale was not selected after the ranking, but the ranking of each dimension in the CDR scale was higher, which has an essential effect on the prediction. The specific results are shown in Fig. 2.

3.2.2. Hierarchical clustering results

Some of the preliminarily selected predictors were strongly correlated with one another. The Spearman rank–order correlation hierarchical clustering method was used to eliminate variables with a high degree of collinearity. Details about the clustering results are shown in Fig. 3. The result of hierarchical clustering dendrogram of predictors is shown in Supplementary eTable 4 in the Appendix. All the predictors were classified into 24 categories, the predictors in each category were strongly correlated with one another. For example, as judged by physicians, impairment in executive function compared with previous judgments is strongly correlated with the judgment function dimension of the CDR scale. The correlation of different dimensions in the same test was also strong, such as Craft Story 21 Recall – Immediate and Craft Story 21 Recall – Delayed.

3.2.3. Final selection results

Based on the results of variable importance ranking and hierarchical clustering, the top 13 variables were ultimately included as in the Cox proportional hazards regression model. The descriptive statistics and hazard ratios (HRs) of the univariate Cox proportional hazards regression analyses are shown in Table 1.

The 13 predictors were as follows: 3 sociodemographic characteristics (age, BMI, and years of education); the neuropsychological measurement scales, including the Montreal Cognitive Assessment (MoCA) score and the scores of the memory, judgment, and social activity dimensions in the Clinical Dementia Rating (CDR); the Craft Story 21 Recall–Delayed (CRAFTDVR), Benson Complex Figure Copy (UDSBENTD) and Category Fluency (VEG and ANIMALS) scores; companion reports on whether the subjects' memory ability decreased compared to before (DECIN); and the number of ApoEε4 alleles.

3.3. Results and visualization of the RSF model and XGB model

The SHAP bee colony diagram is shown in Fig. 4. Each subject was represented as a point on each predictor. The size of the predictor is represented by color, with red indicating high values and blue indicating low values. The vertical axis represents different predictors. Positive values on the horizontal axis indicate that the predictors have a positive effect on the prediction results (increase the possibility of the occurrence of output AD); negative values indicate that the predictors have a negative impact on the prediction results (reduce the possibility of the occurrence of AD). The average SHAP value of the predictors represents the importance to the model [20,27]. A comprehensive analysis of the graph revealed that the higher the scores of the Benson Complex Figure Copy, Craft Story 21 Recall–Delayed, Category Fluency, and MoCA scores are, the lower the risk of developing AD. The higher the score of each module of the CDR scale was, the higher the risk of developing AD. In terms of sociodemographic characteristics, the risk of developing AD was higher in subjects with older age or lower BMI. From the perspective of the SHAP value, the importance of sociodemographic characteristics in the model was lower than the predictive effect of neuropsychological scales and tests on AD. The SHAP value of the ApoEε4 allele was greater than that of the sociodemographic characteristics and individual neuropsychological scales and tests, but its importance ranking in the model was lower than the observation of the subject's memory ability by companions. The results of the univariate analysis of these predictors via Cox proportional hazards regression revealed that the interpretation directions of the HR value and the SHAP value were the same. The specific HR values are shown in Table 1.

3.4. Cox model results

The results of the multivariate Cox model and the multivariate Cox model (excluding ApoEε4) established in this study were similar (Table 2). In the Cox model, the MoCA scale results revealed that a higher MoCA score was associated with a reduced risk of developing AD. For the CDR scale, the higher the scores of the memory, social activity, and judgment dimensions are, the higher the risk of AD. The risk of developing AD increased with increasing age, increasing education level, and lower BMI. The risk increased by 0.019 times for every one-year increase in the subject's age (0.011–0.027, p value<0.001). The risk of disease increased by 0.049 times for each additional year of education (0.026–0.072, p value<0.001). For every 1-unit increase in BMI, the risk of disease was 0.965 times greater (0.950–0.980, p value<0.001). The risk of a subject's memory loss having a companion's report is 1.919 times greater than that of nonreported memory loss (1.509–2.440, p value<0.001). The risk of the progression of AD in subjects with one

Table 1
The baseline characteristics of the NACC subjects included in the study.

Participants Characteristics	Overall (n = 3674)	Non-AD group[c] (n = 2710)	AD group[d] (n = 964)	p-value	Univariate cox regression	
					Hazrad ratio	p-value
Follow up, month	24.5[13.4–40.4]	24.8[13.4–41.6]	23.7[13.3–37.5]	<0.001		
Age, year	75[69–81]	75[69–80]	76[70–82]	<0.001	1.020[1.013–1.028]	<0.001
Geriatric Depression Scale, score	2[1–3]	2[1–3]	2[1–3]	0.860	1.027[1.001–1.054]	0.043
BMI, kg/ m^2	26.4[23.7–29.5]	26.8[24.1–30]	25.4[23–28.18]	<0.001	0.937[0.923–0.950]	<0.001
Education, year	16[14–18]	16[14–18]	16[14–18]	0.115	1.009[0.988–1.030]	0.415
RACE[a], n (%)						
Black	501(13.6 %)	424(15.6 %)	77(8.0 %)	<0.001	Reference	
Whites	3020(82.2 %)	2174(80.2 %)	846(87.8 %)		2.087[1.652–2.637]	<0.001
SEX, n (%)						
Male	1825(49.7 %)	1346(49.7 %)	479(49.7 %)	0.991	reference	
Female	1849(50.3 %)	1364(50,3 %)	485(50,3 %)		1.004[0.885–1.139]	0.955
DECIN[b], n (%)						
No	820 (22.3 %)	737(27.2 %)	83(8.6 %)	<0.001	reference	
Yes	2854(77.7 %)	1973(72.8 %)	881(91.4 %)		3.564[2.846–4.464]	<0.001
Category Fluency-ANIMALS, score	16[13–19]	17[13–20]	15[12–18]	<0.001	0.926[0.914–0.939]	<0.001
Category Fluency- Vegetables, score	11[8–13]	11[9–14]	10[7–12]	<0.001	0.900[0.885–0.917]	<0.001
Trail Making Test Part A, score	38[29–49]	37[29–48]	39[30–52]	<0.001	1.010[1.008–1.013]	<0.001
Trail Making Test Part B, score	112[80–176]	109[79–169]	120[85–192]	<0.001	1.003[1.002–1.004]	<0.001
Montreal Cognitive Assessment, score	23[21–25]	23[21–25]	22[19–24]	<0.001	0.876[0.861–0.891]	<0.001
Craft Story 21 Recall-Immediate, score	12[9–15]	13[10–15]	11[7–13]	<0.001	0.891[0.878–0.904]	<0.001
Craft Story 21 Recall-Delayed, score	11[7–16]	12[8–17]	7[2–12]	<0.001	0.897[0.888–0.907]	<0.001
Multilingual Naming Test, score	29[27–31]	29[27–31]	29[26–30]	<0.001	0.949[0.934–0.964]	<0.001
Benson Complex Figure Copy, score	8[5–10]	8[6–11]	6[2–8]	<0.001	0.847[0.834–0.861]	<0.001
Clinical Dementia Rating-Global, n (%)						
No impairment	388(10.6 %)	357(13.2 %)	31(3.2 %)	<0.001	reference	
Questionable impairment	3238(88.1 %)	2336(86.2 %)	902(93.6 %)		3.681[2.573–5.266]	<0.001
Mild impairment	48(1.3 %)	17(0.6 %)	31(3.2 %)		18.440[11.195–30.372]	<0.001
Clinical Dementia Rating-JUDGMENT, n (%)						
No impairment	1757(47,8 %)	1415(52.2 %)	342(35.5 %)		reference	
Questionable impairment	1672(45.5 %)	1153(42.5 %)	519(53.8 %)	<0.001	2.017[1.759–2.313]	<0.001
Mild impairment	242(6.6 %)	141(5.2 %)	101(10.5 %)		4.076[3.257–5.101]	<0.001
Moderate impairment	3(0.1 %)	1(0.0 %)	2(0.2 %)		64.543[15.832–263.132]	<0.001
Clinical Dementia Rating-MEMORY, n (%)						
No impairment	476(13.0 %)	440(16.2 %)	36(3.7 %)	<0.001	reference	
Questionable impairment	2605(70.9 %)	1973(72.8 %)	632(65.6 %)		3.167[2.263–4.431]	<0.001
Mild impairment	589(16.0 %)	297(11.0 %)	292(30.3 %)		8.724[6.170–12.337]	<0.001
Moderate impairment	4(0.1 %)	0(0.0 %)	4(0.4 %)		108.122[38.190–306.112]	<0.001
Clinical Dementia Rating-COMMUN, n (%)						
No impairment	2653(72.2 %)	2086(77.0 %)	567(58.8 %)	<0.001	reference	
Questionable impairment	934(25.4 %)	581(21.4 %)	353(36.6 %)		2.377[2.080–2.717]	<0.001
Mild impairment	84(2.3 %)	42(1.5 %)	42(4.4 %)		4.640[3.384–6.362]	<0.001
Moderate impairment	3(0.1 %)	1(0.0 %)	2(0.2 %)		4.762[1.187–19.100]	0.028
Number of APOE e4 alleles, n (%)						
No e4 allele	2225(60.6 %)	1788(66.0 %)	437(45.3 %)	<0.001	reference	
1 copy of e4 allele	1213(33.0 %)	802(29.6 %)	411(42.6 %)		1.742[1.523–1.994]	<0.001
2 copies of e4 allele	236(6.4 %)	120(4.4 %)	116 (12.0 %)		2.404[1.959–2.950]	<0.001

[a] Whites and black or African American individuals made up 95.8 % of the subjects. Other races include American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, Asian and Other (specify). Comparisons were made between blacks and whites across different groups.
[b] Dose the companion report a decline in the subject's memory?.
[c] Non-AD group: the subjects who did not progress to AD at the last follow-up.
[d] AD group: the subjects who progressed from MCI to AD.

ApoEε4 allele only is 1.476 times greater than that in subjects without the ApoEε4 allele (1.285–1.696, p value<0.001), and the risk of the development of AD in subjects with two ApoEε4 alleles is 1.545 times greater than that in subjects without the ApoEε4 allele (1.238–1.929, p value<0.001).

3.5. Model evaluation

Based on the different effects of the four models on the training set and validation set, the performance of the models was evaluated.

For the training set, the C-index of the RSF model in Fig. 5A was 0.861, and the AUC of the model under the total time was 0.845. The AUC values for the first, second, third, and fifth years were 0.885, 0.882, 0.879, and 0.862, respectively. The accuracy and sensitivity of this model were 0.765 and 0.782, respectively. The specificity was 0.759, the precision was 0.531, and the F1 score was 0.633. As shown in Fig. 5B, the C-index of the XGBoost model was 0.840, and the AUC of the model under the total time was 0.826. The AUC values for the first, second, third, and fifth years were 0.872, 0.863, 0.857, and 0.838, respectively. The accuracy and sensitivity of this model were 0.754 and 0.760, respectively. The specificity was 0.752, the precision was 0.517, and the F1 score was 0.616. The C-index of the Cox model in Fig. 5C was 0.793, and the Integrated Brier Score was 0.087. The AUC of the model under the total time was 0.780. The AUC values for the first, second, third, and fifth years were 0.824, 0.817, 0.808, and 0.793, respectively. The accuracy and sensitivity of this model were 0.702 and 0.758, respectively. The specificity was 0.682, the precision was 0.455, and the F1 score was 0.568. The C-index of the Cox proportional hazards regression model without the ApoEε4 gene was 0.788. The AUC of the model under the total time was 0.771. The AUC values for the first, second, third, and fifth years were 0.826, 0.811, 0.801, and 0.784, respectively. The accuracy and sensitivity of this model were 0.697 and 0.749, respectively.

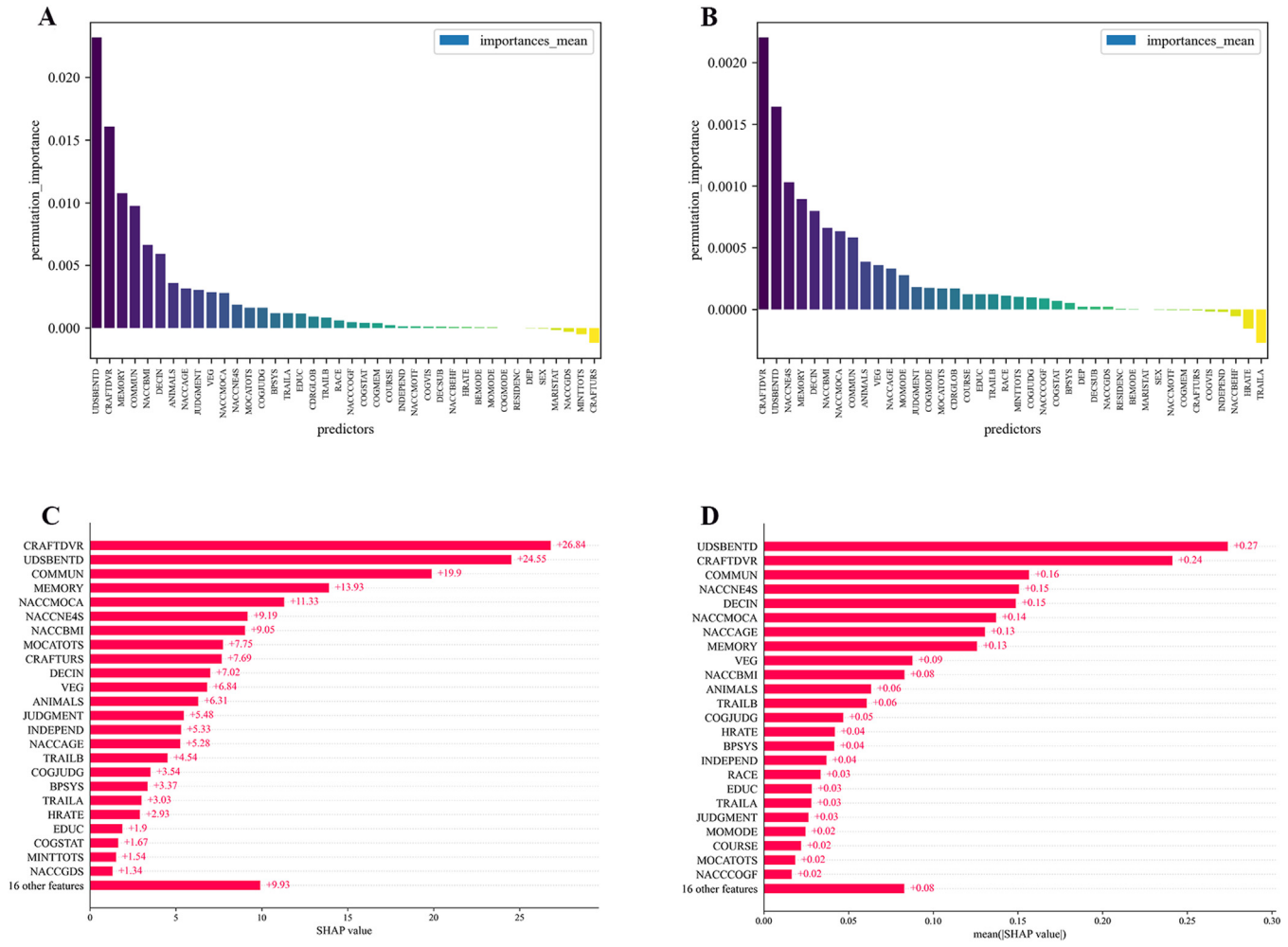

Fig. 2. Ranking of Predictor Importance.
(A) Permutation importance chart in the RSF model. (B) Permutation importance chart of the XGBoost model. (C) SHAP bar chart of the RSF model. (D) SHAP bar chart of the XGBoost model.

The specificity was 0.679, the precision was 0.450, and the F1 score was 0.562. The details for specific indicators are shown in Tables 3.

In the validation set, the C-index of the RSF model in Fig. 5D was 0.781, and the AUC of the model under the total time was 0.778. The AUC values for the first, second, third, and fifth years were 0.846, 0.825, 0.794, and 0.779, respectively. The accuracy and sensitivity of this model were 0.675 and 0.795, respectively. The specificity was 0.630, the precision was 0.443, and the F1 score was 0.569. The C-index of the XGBoost model in Fig. 5E was 0.781, the Integrated Brier Score was 0.145, and the AUC of the model under the total time was 0.782. The AUC values for the first, second, third, and fifth years were 0.858, 0.826, 0.795, and 0.780, respectively. The accuracy and sensitivity of this model were 0.699 and 0.748, respectively. The specificity was 0.681, the precision was 0.465, and the F1 score was 0.573. As shown in Fig. 5F, the C-index of the Cox model was 0.797, and the Integrated Brier Score was 0.087. The AUC of the model under the total time was 0.804. The AUC values for the first, second, third, and fifth years were 0.869, 0.840, 0.814, and 0.802, respectively. The accuracy and sensitivity of this model were 0.735 and 0.721, respectively. The specificity was 0.740, the precision was 0.507, and the F1 score was 0.596. The C-index of the Cox model without the ApoEε4 gene was 0.798. The AUC of the model under the total time was 0.799. The AUC values for the first, second, third, and fifth years were 0.858, 0.841, 0.815, and 0.798, respectively. The accuracy and sensitivity of this model were 0.736 and 0.711, respectively. The specificity was 0.745, the precision was 0.508, and the F1 score was 0.593. The details for specific indicators are shown in Tables 3. AUC values for different years of each neuropsychological measurement scale in Cox model are shown in Supplementary eTable 5 in the Appendix. Overall evaluation criteria and AUC values for different years for each model of 12 predictors excluding *ApoEε4* are shown in Supplementary eTable 6 and eTable 7 in the Appendix.

Survival curves were drawn via the Kaplan–Meier method for different risk groups [17,28]. As shown in Fig. 6, the two curves have no intersection. The confidence interval of the median survival time of the high-risk group was 36.84–40.13 months, and the median survival time of the low-risk group was 85.97 months and above. The survival curve revealed that the survival rate of the low-risk group was significantly greater than that of the high-risk group; the 5-year survival rate of the low-risk group was 78 %, whereas the 5-year survival rate of the high-risk group was only 29 %. There was a significant difference between the two groups ($P < 0.001$).

4. Discussion

In this study, we used the NACC database and combined the predictor selection method with ensemble learning. A total of 13 predictors were included to establish four prediction models. The Cox model is a simple and interpretable model that avoids the disadvantages of the black box in machine learning, and its purpose is to predict the risk of MCI patients developing into AD patients. Because the number of ApoEε4

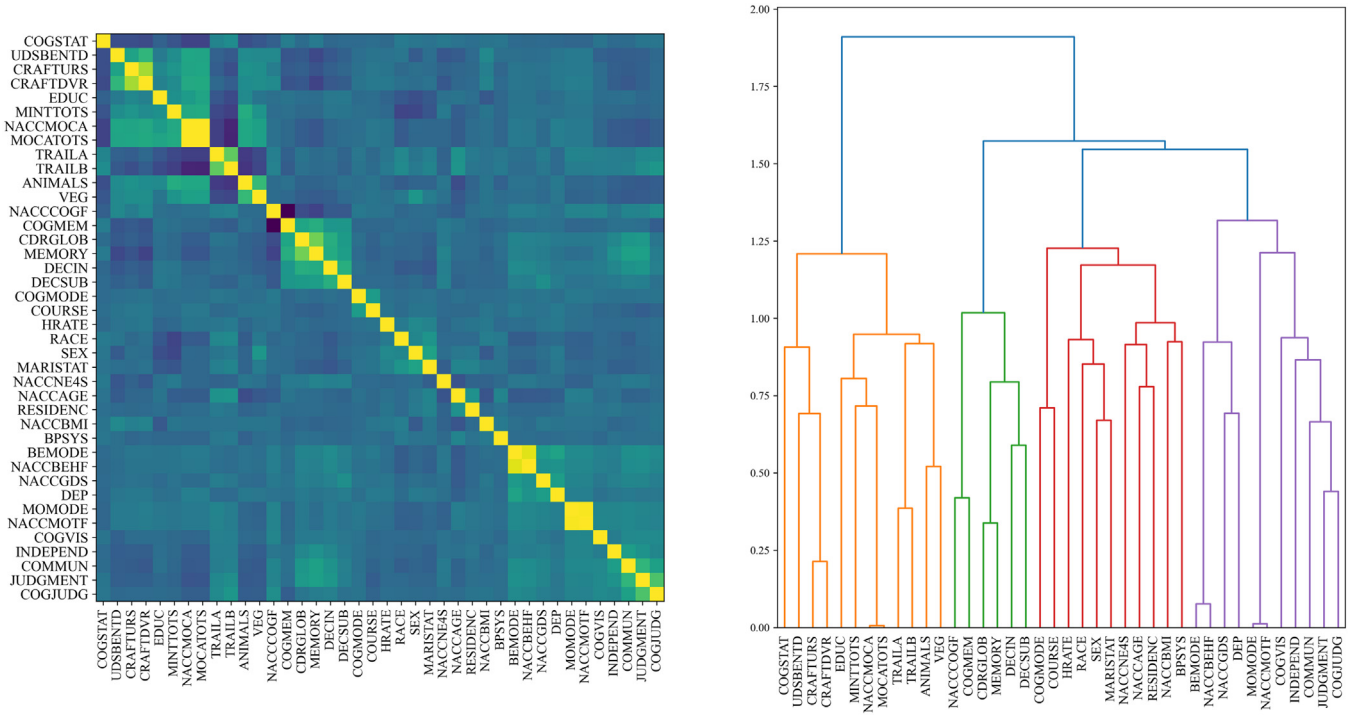

Fig. 3. Correlation heatmap and hierarchical clustering dendrogram of predictors.

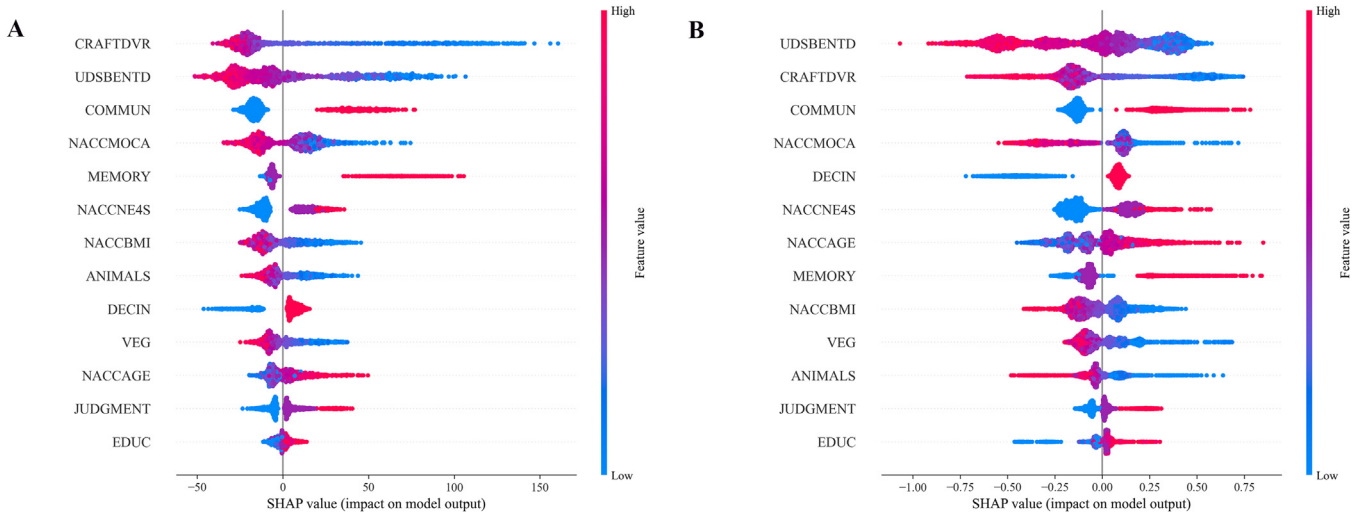

Fig. 4. SHAP visualization of the modeling of selected predictors.
(A) SHAP bar chart of the RSF model. (B) SHAP bar chart of the XGBoost model.

alleles is difficult or expensive to obtain in clinical practice, the prediction effect of the Cox model (excluding ApoEε4) created after excluding this predictor did not significantly differ. The Cox model is relatively easy to interpret compared to other black-box models in machine learning. The C-index of the model is approximately 0.80. The ROC curve of the model revealed that the AUC of the model under the total time was 0.80, and the AUC of prediction within two years reached 0.84, indicating good predictive ability in the short term. Once the model undergoes full clinical validation, the corresponding calculator could be developed to facilitate applications.

In our study, the C-index of the validation set of the model with the addition of the *ApoEε4* gene was 0.797, whereas that without the *ApoEε4* gene was 0.798, indicating that there was no significant improvement. Liu et al. [29] considered that among MCI and AD patients, the prevalence of APOE, a genetic risk predictor for AD, is lower in the NACC database than in the ADNI database. This difference may be attributed to the strong clinical heterogeneity in NACC, and *ApoEε4* might be associated with other types of dementia, such as vascular dementia [29]. Sun et al. [30] reported that *ApoEε4* promoted the development of AD by affecting the clearance of β-amyloid (Aβ). Lautner et al. [31] reported that the CSF Aβ level was related to the degree of AD development but was not related to the ApoE genotype. Therefore, Aβ test results may be more important than the ApoE genotype, and the addition of *ApoEε4* in the model had little effect on risk prediction. Owing to excessive missing cerebrospinal fluid examination results and radiomics examination results, this model was not compared with the results adding these two types of predictors [32], but these predictors may have a greater effect on the progression of AD [29]. Many scholars have shown that the effects

Table 2
Multivariate Cox proportional hazard regression model.

	multivariate cox model			multivariate cox model (exclude ApoEε4)		
	B	Sig.	Hazrad ratio	B	Sig.	Hazrad ratio
Age, y	0.018	<0.001	1.019[1.011–1.027]	0.013	<0.001	1.013[1.006–1.021]
BMI	−0.036	<0.001	0.965[0.950–0.980]	−0.039	<0.001	0.962[0.947–0.976]
Category Fluency- Vegetables	−0.034	0.002	0.967[0.947–0.987]	−0.031	0.004	0.970[0.950–0.990]
Category Fluency-Animals	−0.028	0.001	0.972[0.957–0.988]	−0.026	0.001	0.974[0.959–0.990]
Montreal Cognitive Assessment	−0.050	<0.001	0.951[0.930–0.972]	−0.051	<0.001	0.950[0.929–0.972]
Education,year	0.048	<0.001	1.049[1.026–1.072]	0.046	<0.001	1.047[1.024–1.071]
Craft Story 21 Recall-Delayed	−0.043	<0.001	0.958[0.946–0.970]	−0.046	<0.001	0.955[0.943–0.967]
Benson Complex Figure Copy	−0.071	<0.001	0.931[0.914–0.949]	−0.078	<0.001	0.925[0.908–0.942]
Rported memory decline	0.652	<0.001	1.919[1.509–2.440]	0.673	<0.001	1.961[1.542–2.492]
Clinical Dementia Rating-MEMORY		<0.001			<0.001	
No impairment						
	reference			reference		
Questionable impairment	0.530	0.003	1.700[1.200–2.407]	0.531	0.003	1.701[1.201–2.408]
Mild impairment	0.834	<0.001	2.302[1.587–3.340]	0.852	<0.001	2.344[1.616–3.400]
Moderate impairment	2.389	<0.001	10.908[3.663–32.478]	2.301	<0.001	9.982 [3.356–29.685]
Clinical Dementia Rating-COMMUN		<0.001			<0.001	
No impairment						
	reference			reference		
Questionable impairment	0.286	<0.001	1.331[1.140–1.554]	0.292	<0.001	1.339[1.147–1.563]
Mild impairment	0.540	0.004	1.717[1.193–2.471]	0.525	0.005	1.691[1.174–2.435]
Moderate impairment	1.167	0.112	3.211[0.763–13.511]	1.120	0.126	3.065[0.729–12.892]
Clinical Dementia Rating-JUDGMENT		<0.001			<0.001	
No impairment						
	reference			reference		
Questionable impairment	0.277	<0.001	1.319[1.131–1.539]	0.269	<0.001	1.309[1.123–1.527]
Mild impairment	0.586	<0.001	1.797[1.370–2.355]	0.579	<0.001	1.785[1.360–2.343]
Moderate impairment	4.062	<0.001	58.080[13.684–246.514]	4.060	<0.001	57.968 [13.600–247.076]
Number of APOE e4 alleles		<0.001				
No e4 allele						
	reference					
1 copy of e4 allele	0.389	<0.001	1.476[1.285–1.696]			
2 copies of e4 allele	0.435	<0.001	1.545[1.238–1.929]			

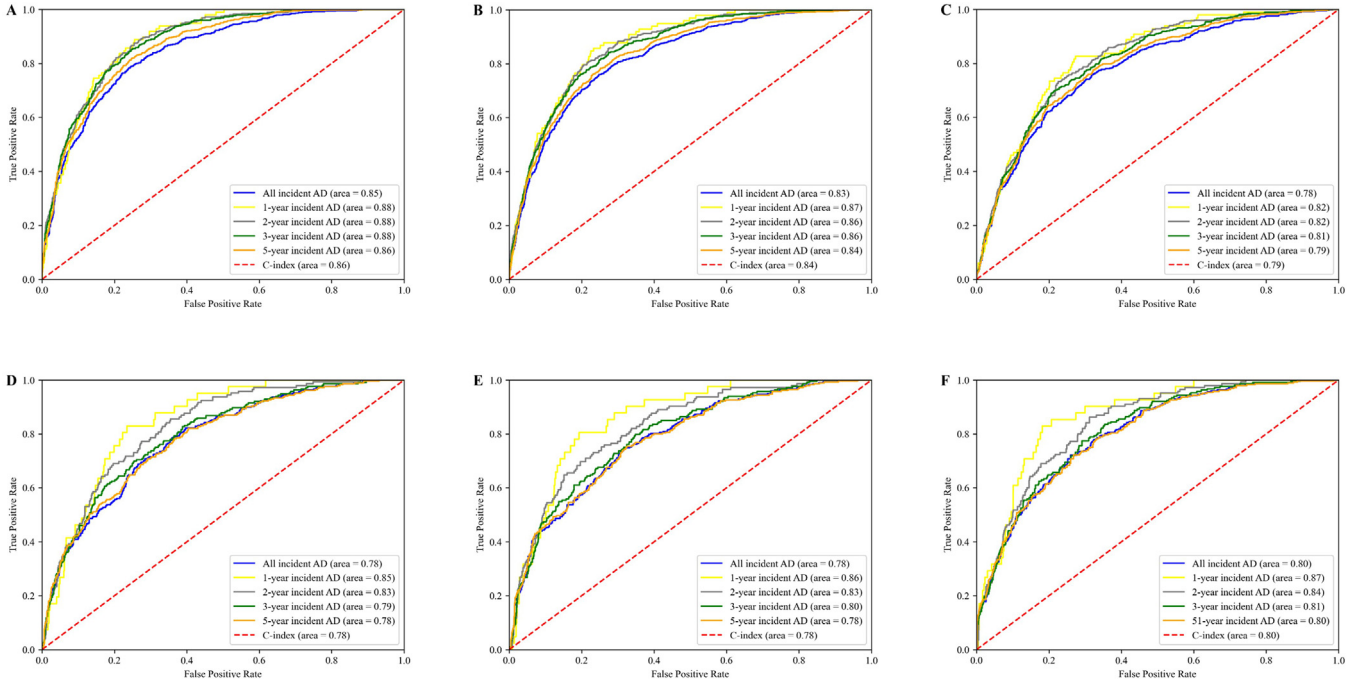

Fig. 5. Train-AUC plots and Test-AUC plots of the NACC on AD at different incident times.
(A) RSF model-training set. (B) XGBoost model-training set. (C) Cox model-training set. (D) RSF model-validation set. (E) XGBoost model-validation set. (F) Cox model-validation set.

Table 3
Overall evaluation for each model.

	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUC	C-index
RSF model							
Training set	0.765	0.782	0.759	0.531	0.633	0.845	0.861
Validation set	0.675	0.795	0.630	0.443	0.569	0.778	0.781
XGB model							
Training set	0.754	0.760	0.752	0.517	0.616	0.826	0.840
Validation set	0.699	0.748	0.681	0.465	0.573	0.782	0.781
Cox model							
Training set	0.702	0.758	0.682	0.455	0.568	0.780	0.793
Validation set	0.735	0.721	0.740	0.507	0.596	0.804	0.797
Cox model (excluding ApoEε4)							
Training set	0.697	0.749	0.679	0.450	0.562	0.771	0.788
Validation set	0.736	0.711	0.745	0.508	0.593	0.799	0.798

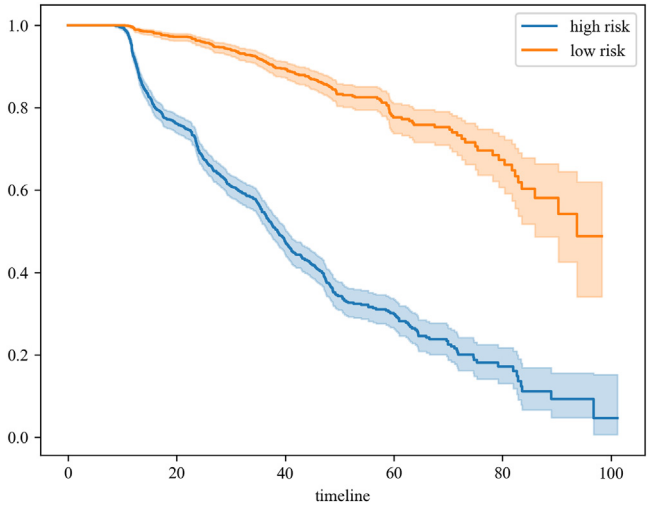

Fig. 6. Survival curves of the two different risk groups.

of the Benson Complex Figure Copy, Craft Story 21 Recall–Delayed, and Category Fluency Test scores are associated with changes in cognitive ability [33–35]. In the model developed herein, the lower the scores of the Benson Complex Figure Copy and the Craft Story 21 Recall–Delayed tests are, the higher the risk of developing AD; these were the strongest predictors in each model. The lower the MoCA scale score and the higher the score of each dimension in the CDR are, the greater the risk of developing AD [36]. In previous studies, Bernier et al. [37] reported that the MoCA had high accuracy in screening patients with MCI and AD. Alatrany et al. [15] revealed the key role of each dimension of the CDR scale in predicting AD [38–40]. Alatrany et al. [15] reported that dimensions of the CDR scale, such as MEMORY, JUDGMENT, COMMUN, and ORIENT, have an impact on the pathogenesis of AD. UDS-related studies have also described the importance of neuropsychological tests and other related predictors [41]. Companion reports of patients' memory loss play an important role in risk prediction and reflect the current severity of patients' condition [16,42]. In terms of sociodemographic characteristics, the risk of AD increases with increasing age, decreasing BMI, and increasing education level. Qizilbash et al. [43] suggested that diabetes and other diseases can affect the development of dementia by affecting lipid metabolism and that obesity does not increase the risk of AD. In 2022, Yuan et al. [44] used the UK Biobank study and reported that high BMI was a protective predictor for the development of AD. These findings are consistent with the results of our study.

In terms of model evaluation [15,18,45], the accuracy, sensitivity, specificity, precision, and F1 score of the training set of the machine learning model were significantly higher than those of the validation set. The AUC values for different years in the training set were also greater than those in the validation set, whereas the Cox model did not significantly differ between the training set and the validation set, indicating that the machine learning model has better learning ability. The AUC of the training set was slightly greater than the AUC of the validation set, which met our expectations, indicating that the three models do not experience severe overfitting. The AUC values of the training set and the validation set of the three models for three years were 0.8 or above, indicating that the overall effect of the models was relatively good. In the Cox model, the model with the *ApoEε4* gene had a greater effect than the model without the ApoEε4 gene. The sensitivity was 1 % higher, but the accuracy was not increased. These results suggest that although this gene does not significantly improve the overall prediction result, it does improve the accuracy of prediction for future diseased subjects. In contrast, the model with this gene has a relatively low specificity. However, the overall prediction performance of the two models did not significantly differ. In the validation set, the Cox model had the highest F1 score, indicating superior performance. A comparison of the AUC values of the three models in different forecast years revealed that the risk prediction of the three models in the first two years was more accurate, with the highest AUC value of 0.87. It is important to analyze the C-index from the perspective of survival analysis indicators. The C-index can avoid the influence of censoring on the results of the curve analysis. Overall, the C-index of the Cox model reached up to 0.797. Therefore, the model established in this study has good accuracy.

Compared with models that include predictors obtained from genome sequencing, cerebrospinal fluid examination of lumbar puncture, or expensive MRI or PET imaging, the predictors in this study could be collected from simple medical history or neuropsychological tests. Consequently, this model can reduce the examination cost for patients. After external validation and clinical validation, the model is expected to be used to identify the risk in patients with self-reported memory concerns or those with cognition-related functional decline.

This study has certain limitations. First, the average follow-up time of all the subjects who met the inclusion criteria was only 24.8 months (IQR 13.4–41.6). Therefore, this model is more accurate in predicting the risk of MCI patients progressing to AD in approximately 2 years, and the prediction ability decreases as the prediction time increases. Second, the models are dominated by white subjects, with a small proportion of Asians. Moreover, the backgrounds of the subjects who can participate in the cohort may not be representative of all people. For example, the educational level of research subjects is generally higher than that of society. The universality of the application of this model still needs to be improved [22]. Third, although this study used methods such as cluster analysis and SHAP values to eliminate collinearity and select predictors, the variables in the model were still weakly correlated (Supplementary eFig. 1 and eFig. 2). Fourth, the present study lacks external validation. Owing to the lack of suitable external validation data, this study did not perform external validation.

Glossary

AUC: In statistics, the Area Under the Curve (AUC) is a metric used to evaluate the performance of a binary classifier. It is calculated as the integral of the Receiver Operating Characteristic (ROC) curve over the false positive rate (FPR) range from 0 to 1.

MCI: Mild Cognitive Impairment (MCI) is defined as an intermediate state between normal cognitive aging and dementia, characterized by a subtle yet noticeable decline in cognitive abilities compared to others of the same age group, which is more pronounced than typical age-related changes but does not significantly disrupt daily activities

Abbreviations

AD: Alzheimer's disease
ADNI: Alzheimer's Disease Neuroimaging Initiative
CSF: Cerebrospinal fluid
NACC: National Alzheimer's Coordinating Center
SHAP: SHapley Additive exPlanations
RSF: Random Survival Forest
XGBoost: Extreme Gradient Boosting
CU: Cognitively unimpaired
MCI: Mild cognitive impairment
PET: Positron emission tomography
ADNPCs: AD neuropathological changes
MoCA: Montreal Cognitive Assessment
CDR: Clinical Dementia Rating
AUC: Area Under the Curve
C-index: Concordance index
IBS: Integrated Brier Score

Funding

This work was supported by the National Natural Science Foundation of China (82073662, 82473726).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, I have not used any AI.

Data availability statement

Availability of data and materials: All of the source data are available in the cited studies and NACC (https://www.naccdata.org). Extracted data and materials are reported in Table 1 to Table 3 in the manuscript and eTable 1 to eTable 7 in the supplementary materials. Additional information is available upon reasonable request.

Declaration of competing interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yuhai Zhang reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Tianyuan Guan: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Lei Shang:** Writing – review & editing, Visualization, Investigation, Formal analysis, Data curation. **Peng Yang:** Writing – review & editing, Visualization, Validation, Software. **Zhijun Tan:** Writing – review & editing, Software, Methodology, Data curation. **Yue Liu:** Software, Data curation. **Chunling Dong:** Project administration, Data curation. **Xueying Li:** Resources, Data curation. **Zuxuan Hu:** Visualization, Validation. **Haixia Su:** Supervision, Methodology, Conceptualization. **Yuhai Zhang:** Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Data availability

Joint ensemble learning-based risk prediction of Alzheimer's disease among mild cognitive impairment patients (Reference data) (Mendeley Data).

Acknowledgements

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).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tjpad.2025.100083.

References

[1] Jack CR Jr, Andrews JS, Beach TG, Buracchio T, Dunn B, Graf A, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: alzheimer's association workgroup. Alzheimers Dement 2024;20:5143–69. doi:10.1002/alz.13859.
[2] Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement 2018;14:535–62. doi:10.1016/j.jalz.2018.02.018.
[3] Battle DE. Diagnostic and statistical manual of mental disorders (DSM). CoDAS 2013;25:191–2.
[4] Anand S, Schoo C. ***Mild cognitive impairment***. Treasure Island, FL: StatPearls Publishing; 2024.
[5] Rajan KB, Weuve J, Barnes LL, McAninch EA, Wilson RS, Evans DA. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060). Alzheimers Dement 2021;17:1966–75. doi:10.1002/alz.12362.
[6] Petersen RC. Mild cognitive impairment. Continuum (Minneapolis, Minn) 2016;22:404–18. doi:10.1212/CON.0000000000000313.
[7] Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, et al. Mild cognitive impairment: ten years later. Arch Neurol 2009;66:1447–55. doi:10.1001/archneurol.2009.266.
[8] Xia X, Jiang Q, McDermott J, Han J-DJ. Aging and Alzheimer's disease: comparison and associations from molecular to system level. Aging Cell 2018;17:e12802. doi:10.1111/acel.12802.
[9] Zhu D, Montagne A, Zhao Z. Alzheimer's pathogenic mechanisms and underlying sex difference. Cell Mol Life Sci 2021;78:4907–20. doi:10.1007/s00018-021-03830-w.

[10] Xu W, Tan L, Wang HF, Tan MS, Tan L, Li JQ, et al. Education and risk of dementia: dose-response meta-analysis of prospective cohort studies. Mol Neurobiol 2016;53:3113–23. doi:10.1007/s12035-015-9211-5.
[11] Gonneaud J, Bedetti C, Pichet Binette A, Benzinger TLS, Morris JC, Bateman RJ, et al. Association of education with Aβ burden in preclinical familial and sporadic Alzheimer disease. Neurology 2020;95:e1554. doi:10.1212/WNL.0000000000010314.
[12] Li H, Habes M, Wolk DA, Fan Y. Alzheimer's Disease neuroimaging initiative, the Australian imaging biomarkers lifestyle study of aging. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. Alzheimers Dement 2019;15:1059–70. doi:10.1016/j.jalz.2019.02.007.
[13] Han X, Zhang Y, Shao Y. Alzheimer's disease neuroimaging initiative. Application of concordance probability estimate to predict conversion from mild cognitive impairment to Alzheimer's disease. Biostat Epidemiol 2017;1:105–18. doi:10.1080/24709360.2017.1342187.
[14] Amini S, Hao B, Yang J, Karjadi C, Kolachalama VB, Au R, et al. Predicting transitions from mild cognitive impairment to dementia within 5-years. Alzheimers Dement 2023;19:e076873. doi:10.1002/alz.076873.
[15] Alatrany AS, Khan W, Hussain A, Kolivand H, Al-Jumeily D. An explainable machine learning approach for Alzheimer's disease classification. Sci Rep 2024;14:2637. doi:10.1038/s41598-024-51985-w.
[16] Fouladvand S, Noshad M, Goldstein MK, Periyakoil VJ, Chen JH. Mild cognitive impairment: data-driven prediction, risk factors, and workup. AMIA Jt Summits Transl Sci Proc 2023;2023:167–75.
[17] Flynn R. Survival analysis. J Clin Nurs 2012;21:2789–97. doi:10.1111/j.1365-2702.2011.04023.x.
[18] Liu P, Fu B, Yang SX, Deng L, Zhong X, Zheng H. Optimizing survival analysis of XGboost for ties to predict disease progression of breast cancer. IEEE Trans Biomed Eng 2021;68:148–60. doi:10.1109/tbme.2020.2993278.
[19] Wang J, Wang Z, Liu N, Liu C, Mao C, Dong L, et al. Random forest model in the diagnosis of dementia patients with normal mini-mental state examination scores. J Pers Med 2022;12:37. doi:10.3390/jpm12010037.
[20] Yi F, Yang H, Chen D, Qin Y, Han H, Cui J, et al. XGBoost-SHAP-based interpretable diagnostic framework for Alzheimer's disease. BMC Med Inform Decis Mak 2023;23:137. doi:10.1186/s12911-023-02238-9.
[21] Beekly DL, Ramos EM, Lee WW, Deitrich WD, Jacka ME, Wu J, et al. The national Alzheimer's coordinating center (NACC) database: the uniform data set. Alzheimer Dis Assoc Disord 2007;21:249–58. doi:10.1097/wad.0b013e318142774e.
[22] Weintraub S, Besser L, Dodge HH, Teylan M, Ferris S, Goldstein FC, et al. Version 3 of the Alzheimer disease centers' neuropsychological test battery in the uniform data set (UDS). Alzheimer Dis Assoc Disord 2018;32:10–17. doi:10.1097/WAD.0000000000000223.
[23] Little RJ. Missing data analysis. Annu Rev Clin Psychol 2024;20:149–73. doi:10.1146/annurev-clinpsy-080822-051727.
[24] Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 2012;28:112–18. doi:10.1093/bioinformatics/btr597.
[25] Etgen T, Sander D, Bickel H, Förstl H. Mild cognitive impairment and dementia: the importance of modifiable risk factors. Dtsch Arztebl Int 2011;108:743–50. doi:10.3238/arztebl.2011.0743.
[26] Cheng C, Elmer S, Batterham R, Hawkins M, Osborne RH. Measuring health literacy to inform actions to address health inequities: a cluster analysis approach based on the Australian national health literacy survey. J Public Health 2024. doi:10.1093/pubmed/fdae165.
[27] Sarica A, Aracri F, Bianco MG, Arcuri F, Quattrone A, Quattrone A, et al. Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to Alzheimer's disease. Brain Inform 2023;10:31. doi:10.1186/s40708-023-00211-w.
[28] Telisnor G, Lim A, Zhang Z, Lou X, Nassour I, Salloum RG, et al. Analysis of pancreatic cancer treatment and survival disparities in Florida throughout the COVID-19 pandemic. J Natl Med Assoc 2024;116:328–37. doi:10.1016/j.jnma.2024.07.004.
[29] Liu S, Masurkar AV, Rusinek H, Chen J, Zhang B, Zhu W, et al. Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. Sci Rep 2022;12:17106. doi:10.1038/s41598-022-20674-x.
[30] Sun YY, Wang Z, Huang HC. Roles of ApoE4 on the pathogenesis in Alzheimer's disease and the potential therapeutic approaches. Cell Mol Neurobiol 2023;43:3115–36. doi:10.1007/s10571-023-01365-1.
[31] Lautner R, Palmqvist S, Mattsson N, Andreasson U, Wallin A, Pålsson E, et al. Apolipoprotein E genotype and the diagnostic accuracy of cerebrospinal fluid biomarkers for Alzheimer disease. JAMA Psychiatry 2014;71:1183–91. doi:10.1001/jamapsychiatry.2014.1060.
[32] Chandra A, Dervenoulas G, Politis M. Alzheimer's disease neuroimaging initiative. Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. J Neurol 2019;266:1293–302. doi:10.1007/s00415-018-9016-3.
[33] Petrillo K, Javed B, Toosizadeh N. Association between dual-task function and neuropsychological testing in older adults with cognitive impairment. Exp Gerontol 2023;178:112223. doi:10.1016/j.exger.2023.112223.
[34] Gómez-Valadés A, Martínez R, Rincón M. Designing an effective semantic fluency test for early MCI diagnosis with machine learning. Comput Biol Med 2024;180:108955. doi:10.1016/j.compbiomed.2024.108955.
[35] Jiskoot LC, Russell LL, Peakman G, Convery RS, Greaves CV, Bocchetta M, et al. The Benson complex figure test detects deficits in visuoconstruction and visual memory in symptomatic familial frontotemporal dementia: a GENFI study. J Neurol Sci 2023;446:120590. doi:10.1016/j.jns.2023.120590.
[36] Kühnel L, Berger AK, Markussen B, Raket LL. Simultaneous modeling of Alzheimer's disease progression via multiple cognitive scales. Stat Med 2021;40:3251–66. doi:10.1002/sim.8932.
[37] Bernier PJ, Gourdeau C, Carmichael PH, Beauchemin JP, Voyer P, Hudon C, et al. It's all about cognitive trajectory: accuracy of the cognitive charts-MoCA in normal aging, MCI, and dementia. J Am Geriatr Soc 2023;71:214–20. doi:10.1111/jgs.18029.
[38] Huang HC, Tseng YM, Chen YC, Chen PY, Chiu HY. Diagnostic accuracy of the clinical dementia rating scale for detecting mild cognitive impairment and dementia: a bivariate meta-analysis. Int J Geriatr Psychiatry 2021;36:239–51. doi:10.1002/gps.5436.
[39] Chaves MLF, Camozzato AL, Godinho C, Kochhann R, Schuh A, de Almeida VL, et al. Validity of the clinical dementia rating scale for the detection and staging of dementia in Brazilian patients. Alzheimer Dis Assoc Disord 2007;21:210–17. doi:10.1097/wad.0b013e31811ff2b4.
[40] Chang YL, Bondi MW, McEvoy LK, Fennema-Notestine C, Salmon DP, Galasko D, et al. Global clinical dementia rating of 0.5 in MCI masks variability related to level of function. Neurology 2011;76:652–9. doi:10.1212/WNL.0b013e31820ce6a5.
[41] Pang Y, Kukull W, Sano M, Albin RL, Shen C, Zhou J, et al. Predicting progression from normal to MCI and from MCI to AD using clinical variables in the National Alzheimer's coordinating center uniform data set version 3: application of machine learning models and a probability calculator. J Prev Alzheimer's Dis 2023;10:301–13. doi:10.14283/jpad.2023.10.
[42] Lista S, Molinuevo JL, Cavedo E, Rami L, Amouyel P, Teipel SJ, et al. Evolving evidence for the value of neuroimaging methods and biological markers in subjects categorized with subjective cognitive decline. J Alzheimer's Dis 2015;48:S171–91. doi:10.3233/jad-150202.
[43] Qizilbash N, Gregson J, Johnson ME, Pearce N, Douglas I, Wing K, et al. BMI and risk of dementia in two million people over two decades: a retrospective cohort study. Lancet Diabetes Endocrinol 2015;3:431–6. doi:10.1016/s2213-8587(15)00033-9.
[44] Yuan S, Wu W, Ma W, Huang X, Huang T, Peng M, et al. Body mass index, genetic susceptibility, and Alzheimer's disease: a longitudinal study based on 475,813 participants from the UK Biobank. J Transl Med 2022;20:417. doi:10.1186/s12967-022-03621-2.
[45] Mishra S. A comparative study for time-to-event analysis and survival prediction for heart failure condition using machine learning techniques. J Electron Electromed Eng Med Inform 2022;4:115–34. doi:10.35882/jeeemi.v4i3.225.