Age and Ageing 2023; **52:** 1–8
https://doi.org/10.1093/ageing/afad182

RESEARCH PAPER

Investigating predictors of progression from mild cognitive impairment to Alzheimer's disease based on different time intervals

Yafei Wu[1,2,†], Xing Wang[1,2,†], Chenming Gu[1,2], Junmin Zhu[1,2], Ya Fang[1,2,3]

[1]School of Public Health, Xiamen University, Xiamen, Fujian, China
[2]Key Laboratory of Health Technology Assessment of Fujian Province, Xiamen University, Xiamen, Fujian, China
[3]National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, Fujian, China

Address correspondence to: Ya Fang, School of Public Health, Xiamen University, Xiang'an South Road, Xiang'an District, Xiamen, Fujian, China. Email: fangya@xmu.edu.cn
[†]Yafei Wu and Xing Wang contributed equally to this work.

Abstract

Background: Mild cognitive impairment (MCI) is the early stage of AD, and about 10–12% of MCI patients will progress to AD every year. At present, there are no effective markers for the early diagnosis of whether MCI patients will progress to AD. This study aimed to develop machine learning-based models for predicting the progression from MCI to AD within 3 years, to assist in screening and prevention of high-risk populations.
Methods: Data were collected from the Alzheimer's Disease Neuroimaging Initiative, a representative sample of cognitive impairment population. Machine learning models were applied to predict the progression from MCI to AD, using demographic, neuropsychological test and MRI-related biomarkers. Data were divided into training (56%), validation (14%) and test sets (30%). AUC (area under ROC curve) was used as the main evaluation metric. Key predictors were ranked utilising their importance.
Results: The AdaBoost model based on logistic regression achieved the best performance (AUC: 0.98) in 0–6 month prediction. Scores from the Functional Activities Questionnaire, Modified Preclinical Alzheimer Cognitive Composite with Trails test and ADAS11 (Unweighted sum of 11 items from The Alzheimer's Disease Assessment Scale-Cognitive Subscale) were key predictors.
Conclusion: Through machine learning, neuropsychological tests and MRI-related markers could accurately predict the progression from MCI to AD, especially in a short period time. This is of great significance for clinical staff to screen and diagnose AD, and to intervene and treat high-risk MCI patients early.

Keywords: machine learning, Alzheimer's disease, mild cognitive impairment, predictors, older people

Key Points

- Machine learning especially multi-model fusion strategies had great potential in distinguishing mild cognitive impairment (MCI) and Alzheimer's disease (AD) patients.
- The prediction power of models declined over time, with the highest area under ROC curve (AUC) of 0.98 in 0–6 month prediction.
- Scores from Functional Activities Questionnaire, Modified Preclinical Alzheimer Cognitive Composite with Trails test and ADAS11 (Unweighted sum of 11 items from The Alzheimer's Disease Assessment Scale-Cognitive Subscale) were key predictors for AD prediction.

Introduction

Alzheimer's disease (AD) is a common progressive neurodegenerative disease in older people with insidious onset and slow deterioration in cognition, and functional ability. AD accounts for about 60–80% of all dementia patients [1]. So far, there are no effective drugs and treatments that can completely cure AD, but only delay its deterioration [2]. Mild cognitive impairment (MCI) is an early stage of AD, epidemiological studies have shown that about 10–12% of patients with MCI will progress to AD every year [3, 4]. Therefore, it is of great significance for the prevention of AD to accurately predict whether MCI will develop into AD and to formulate intervention plans.

Patients with MCI have noticeable declines in memory, language and other cognitive functions, and some impairment may also be found in clinical examination, but not enough to affect daily living function [5]. However, MCI patients show different heterogeneity over time [6], which provides the possibility for early prediction. Previous studies have identified some biomarkers for early prediction of AD, such as hippocampus volume, tau and β-amyloid protein aggregation, and neuropsychological test scale scores [7–12]. While these studies have some limitations, for example, utilisation of cross-sectional data [13], inclusion of too many variables [14], utilisation of single model [15, 16] and lack of model application [17]. Therefore, a simple, accurate and reliable method is needed, which could assist clinical workers in early detection and timely prevention.

Machine learning (ML) has shown outstanding performance in disease risk prediction [18, 19]. It can learn complex relationships among variables from large amounts of data [20]. However, ML also faces major challenge such as poor interpretability and limited applications. Interpretability refers to the ability of a model to present its predictions in a way that humans can understand [21]. For model applications, nomogram is a simple and easy tool to make diagnosis and assist clinical decision-making [22, 23].

The purpose of this study is to predict the risk of progression from MCI to AD within 3 years using the ADNI database. We also examined the key predictors for AD prediction in different time intervals.

Materials and methods

Participants

Data was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI, https://ida.loni.usc.edu) database, The ADNI project aims to comprehensively assess the progress of MCI and early AD by combining longitudinal magnetic resonance imaging (MRI), positron emission tomography (PET) and biomarker data and clinical neuropsychological scale assessment. The samples are from ADNI-1, ADNI-GO, ADNI-2 and ADNI-3 queues. The diagnostic criteria of MCI and AD can be found in Supplementary Table S1 and at the ADNI website (https://clinicaltrials.gov/ct2/show/NCT02854033). The follow-up interval for ADNI was 6–12 months. The inclusion criteria of the samples in this study were: (i) diagnosed with MCI at baseline; (ii) had a follow-up time of 3 years; (iii) had outcome (cognitive status) information. Participants diagnosed with bidirectional changes during follow-up were excluded [24]. Finally, a total of 627 patients with baseline MCI were screened, of whom 269 patients progressed to AD at a 3-year follow-up and 358 patients remained MCI. The detailed sample selection process was shown in Supplementary Figure S1.

Predictors and data preprocessing

We evaluated multiple types of variables such as demographic, neuropsychological test, MRI-related markers, etc. Some variables that are not helpful to model construction, such as patient id, MRI image id, test time etc., were excluded (Supplementary Table S2), and then variables with a missing rate of more than 20% were also excluded (Supplementary Table S3). For the remaining variables with missing value (Supplementary Table S4), we used the missForest algorithm to fill it. MissForest algorithm is a nonparametric imputation method, and is superior to traditional filling methods, such as hot deck filling, mean filling, and median filling, etc. [25]. Assignment of categorical variables is shown in Supplementary Table S5. Finally, we included 29 variables in four categories, as shown in Supplementary Table S6. In order to reduce model computation and generalise model performance, we map all variables to the same dimensional range (0–1) to ensure that they are in uniform and comparable dimensions. Specifically, the Max-Min normalisation method was applied for continuous variables like age and education. While the neuropsychological test variables were normalised by dividing their scores by the total scores of the corresponding scale. For the intracranial volume, due to the inconsistency of cranial size in each individual, the proportional method [26] was used to normalise it. For the other MRI-related markers, we normalised the brain area volume of different individuals by dividing the intracranial volume of the individual.

Machine learning model

According to the No free lunch theorem, different ML models have different performance in different tasks [27]. We selected as many as possible models related to traditional machine learning, ensemble learning and artificial neural network. Specifically, Logistic Regression, K Nearest Neighbours, Decision Tree, Random Forest, Extra Trees, Support Vector Machine, Multi-layer Perceptron, Gradient Boosting Tree and eXtreme Gradient Boosting were included. The principles of these ML models can be viewed in Supplementary Method 1. In addition to applying the above single model, we also tried to further improve the performance by using multi-model fusion strategies such as Voting, Boosting and Bagging. The detailed description of multi-model fusion can be seen in the Supplementary Method 1.

All patients were firstly assigned to two parts (70% versus 30%). The 30% of data was used as test set, and was only

applied to verify the performance of models. A 5-fold cross validation was further used in the 70% of data, that is, the 70% of data was equally divided into 5-fold (14% of the whole data for each fold). Among the 5-fold, four of them (56% of the whole data) were used as training set, and another fold was used as validation set (14% of the whole data). The training set (56%) and validation set (14%) were used to do feature selection and hyperparameter tunning, and the feature selection and hyperparameter tunning were performed separately in the 70% of data with 5-fold cross validation. After hyperparameter tunning, the prediction models were refitted in the training set with optimal hyperparameters and validated in the validation set for five times, and an average result was obtained to select the optimal model. Finally, the test set (30%) was used to assess the performance of optimal model. Hyperparameters of all models were tuned with Bayesian optimization. In order to reduce the number of features and avoid model overfitting, the Least Absolute Shrinkage And Selection Operator (LASSO) was applied to screen features [28]. More detailed information of LASSO can be accessed at the Supplementary Method 2. All models were constructed using Python (Version 3.6) on the PyCharm professional (Version 2021.3.2) platform. The various packages and versions used are shown in Supplementary Table S7, and detailed machine learning model construction process is shown in Supplementary Figure S2.

Evaluation of model performance and variable importance

Models were evaluated by accuracy, precision, recall, F1-score and AUC. The calculation for these metrics is shown in Supplementary Table S8. Calibration curve, expected calibration error (ECE) and brier score were used to evaluate the calibration results. More information about ECE, Brier score and DCA can be found in the Supplementary Method 3. For model interpretability, we applied the feature importance and permutation importance to output the key predictors of optimal model. Furthermore, we also evaluated model performance at different time points, that is, the global model was assessed by grouping the outcomes of the test data at different time points (0–6, 6–12, 12–18, 18–24 and 24–36 months). Further, we also visualised the logistic regression with nomogram utilising the LASSO selected variables. All the codes of model construction and evaluation can be accessed through the following link https://github.com/jaysonwong-maker/github.

Statistical analysis

SPSS (Version 26) was used for statistical analysis. Continuous variable is expressed as mean±standard deviation ($\bar{x} \pm s$) or median (interquartile range, IQR), categorical variable is expressed as number and percentage (%). The t-test and non-parametric Kruskal–Wallis test were used to compare the groups for continuous variables. The x^2 test was used to compare the frequency of categorical variables. P value of <0.05 was considered statistically significant.

Result

Baseline characteristics of the patients

Supplementary Table S9 summarises the differences of baseline variables between the MCI and AD groups. The AD group was older than MCI group. All neuropsychological tests were statistically significant between two groups. There was statistical significance between APOE genotypes. Except for intracranial volume, other MRI-related features showed statistical significance. Supplementary Table S10 summarises the differences of baseline variables between the 70% of data (training and validation sets) and 30% of data (test set), and there was no statistical difference between the two sets.

Performance evaluation of machine learning models

Table 1 shows the performance of machine learning models in predicting AD within 3 years in validation set. In full-variable set, RF ranked first with its accuracy, precision, recall, F1-score and AUC reaching 0.80, 0.79, 0.79, 0.78 and 0.89, respectively. After using voting, bagging and boosting fusion strategies, the performance had slightly improved compared with single model, especially for Bagging. Eight variables were selected by LASSO, that is, ADAS11, RAVLT_immediate, RAVLT_learning, Functional Activities Questionnaire (FAQ), Fusiform, MidTemp, ICV and mPACCtrailsB. In LASSO set, MLP had the best performance, with its accuracy, precision, recall, F1-score and AUC being 0.81, 0.80, 0.78, 0.78 and 0.90, respectively. After multi-model fusion, boosting fusion based on logistic regression achieved best performance (accuracy: 0.83, precision: 0.80, AUC: 0.91). Supplementary Table S11–S14 showed the detailed results of model fusion. Generally, Adaboost (base estimator: logistic regression) in LASSO set was much better considering its higher performance and less variables. Therefore, the AdaBoost model based on logistic regression is taken as the final optimal model, and the accuracy, precision, recall, F1-score and AUC of the final model on the test set are 0.83, 0.82, 0.71, 0.76 and 0.89, respectively.

Supplementary Figure S3 shows the ROC curve of the nine single models and best fusion model (AdaBoost with logistic regression as base estimator) in the validation set. Supplementary Figure S4 shows the calibration curve of the best fusion model on test set. The expected calibration error and brier score of the best fusion model were 0.07 (standard deviation: 0.02), and 0.13 (standard deviation: 0.04), respectively. The DCA of the best fusion model in the test set was shown in Figure 1, a positive net benefit was observed within the whole threshold, suggesting its value in clinical application. Supplementary Figure S5 and Supplementary Table S15 show the importance rank of variables by the best fusion model. It could be seen that there is little difference in the feature importance ranking of the output of the two methods (feature importance and permutation importance). Generally, FAQ, mPACCtrailsB, RAVLT_immediate and

Table 1. Performance of prediction models on the validation set before and after feature selection

Model	Full-variable set ($\bar{x} \pm s$)					LASSO set ($\bar{x} \pm s$)				
	Accuracy	Precision	Recall	F1-Score	AUC	Accuracy	Precision	Recall	F1-Score	AUC
LR	0.81 ± 0.05	0.81 ± 0.11	0.79 ± 0.04	0.79 ± 0.03	0.89 ± 0.04	0.80 ± 0.02	0.80 ± 0.05	0.77 ± 0.09	0.78 ± 0.03	0.90 ± 0.02
KNN	0.76 ± 0.05	0.73 ± 0.07	0.74 ± 0.09	0.73 ± 0.06	0.83 ± 0.04	0.77 ± 0.01	0.74 ± 0.04	0.76 ± 0.06	0.75 ± 0.03	0.85 ± 0.02
SVM	0.80 ± 0.04	0.79 ± 0.09	0.79 ± 0.05	0.78 ± 0.03	0.88 ± 0.04	0.79 ± 0.01	0.78 ± 0.04	0.76 ± 0.06	0.77 ± 0.02	0.89 ± 0.02
DT	0.74 ± 0.08	0.73 ± 0.13	0.71 ± 0.09	0.71 ± 0.09	0.74 ± 0.08	0.73 ± 0.03	0.71 ± 0.06	0.68 ± 0.03	0.69 ± 0.02	0.72 ± 0.02
MLP	0.77 ± 0.03	0.75 ± 0.05	0.74 ± 0.08	0.74 ± 0.04	0.85 ± 0.04	0.81 ± 0.02	0.80 ± 0.05	0.78 ± 0.08	0.78 ± 0.02	0.90 ± 0.02
RF	0.80 ± 0.03	0.79 ± 0.08	0.79 ± 0.04	0.78 ± 0.03	0.89 ± 0.02	0.79 ± 0.03	0.77 ± 0.07	0.76 ± 0.08	0.76 ± 0.02	0.88 ± 0.02
ET	0.80 ± 0.03	0.79 ± 0.09	0.78 ± 0.07	0.78 ± 0.03	0.89 ± 0.05	0.79 ± 0.04	0.78 ± 0.08	0.77 ± 0.03	0.77 ± 0.02	0.89 ± 0.02
GBT	0.79 ± 0.01	0.78 ± 0.05	0.76 ± 0.06	0.77 ± 0.02	0.88 ± 0.02	0.80 ± 0.03	0.79 ± 0.07	0.77 ± 0.07	0.78 ± 0.02	0.88 ± 0.02
XGB	0.79 ± 0.02	0.78 ± 0.04	0.75 ± 0.06	0.76 ± 0.03	0.88 ± 0.03	0.78 ± 0.02	0.75 ± 0.04	0.76 ± 0.04	0.75 ± 0.01	0.86 ± 0.02
Voting	0.79 ± 0.04	0.77 ± 0.10	0.77 ± 0.04	0.77 ± 0.04	0.89 ± 0.04	0.79 ± 0.04	0.77 ± 0.05	0.78 ± 0.07	0.77 ± 0.03	0.89 ± 0.02
Bagging[a]	0.80 ± 0.03	0.79 ± 0.07	0.77 ± 0.06	0.78 ± 0.04	0.90 ± 0.04	0.81 ± 0.03	0.79 ± 0.07	0.80 ± 0.07	0.79 ± 0.02	0.90 ± 0.02
AdaBoost[a]	0.82 ± 0.02	0.81 ± 0.07	0.80 ± 0.04	0.80 ± 0.02	0.89 ± 0.02	0.83 ± 0.03	0.80 ± 0.06	0.74 ± 0.11	0.76 ± 0.04	0.91 ± 0.03

All metrics were the average results by 5-fold cross validation. LR, logistic regression; KNN, K Nearest Neighbours; SVM, Support Vector Machine; DT, Decision Tree; MLP, Multi-layer Perception; RF, Random Forest; ET, Extra Trees; GBT, Gradient Boosting Tree; XGB, eXtreme Gradient Boosting; LASSO, least absolute shrinkage and selection operator; AUC, the area under the receiver operating characteristic curve. [a]Only results of the optimal fusion models were shown, for detailed fusion results, please see Tables S11–S14.

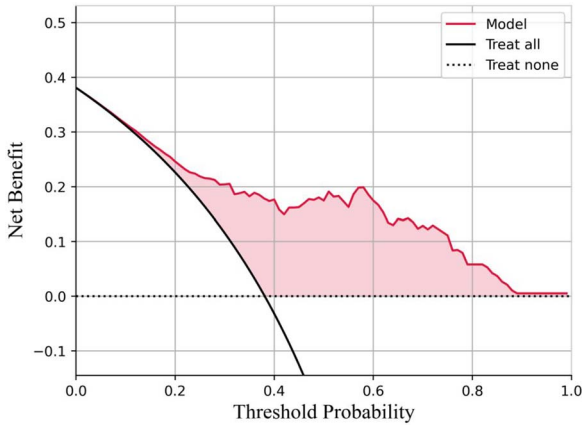

Figure 1. DCA curve for the best fusion model (AdaBoost with logistic regression as the base estimator) on test set. The horizontal coordinate is the threshold probability and the ordinate is the net benefit. As can be seen from the figure, the final model has positive benefits in the whole threshold interval, which means that it has certain clinical value.

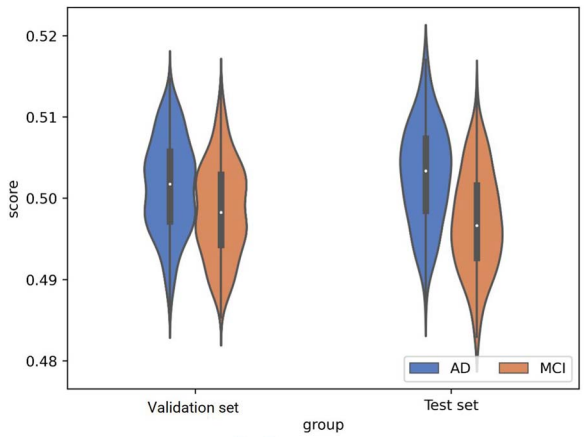

Figure 2. Violin box diagram of prediction score. Score represents the predicted probability of the final model (AdaBoost with logistic regression as the base estimator) for each individual, statistically significant differences of prediction scores were observed between the MCI and AD groups in the validation (left) and test (right) sets (independent samples T-test). It can be seen that the final model has the ability to distinguish between MCI and AD.

ADAS11 were important predictors, demonstrating their stability in early prediction of progression to AD in patients with MCI. Besides, we compared the distribution of probability score between high- and low-risk patients based on the best fusion model. In the validation and test sets, the score of the high-risk group was significantly higher than that of the low-risk group (Figure 2), and the difference was statistically significant ($P < 0.05$).

We further constructed a nomogram using the eight variables selected by LASSO (Supplementary Figure S6), and the calibration curve of the nomogram is shown in Supplementary Figure S7, suggesting a good calibration result.

Model performance and variable importance at different time intervals

We further examined the prediction power of optimal fusion model (AdaBoost with logistic regression as base estimator) in different time intervals. The results showed that model had high AUC during the entire time intervals, among which the prediction power in 0–6 month was the highest (AUC: 0.98, Table 2). The results of importance ranking of predictors at different time intervals are shown in Figure 3. Generally, scores from the FAQ, ADAS11 and Modified Preclinical Alzheimer Cognitive Composite with Trails test (mPACCtrailsB) ranked top 3.

Table 2. Performance evaluation of the best fusion model at different time intervals on test set

Time interval	Sample	Accuracy ($\bar{x} \pm s$)	Precision ($\bar{x} \pm s$)	Recall ($\bar{x} \pm s$)	F1-score ($\bar{x} \pm s$)	AUC ($\bar{x} \pm s$)
0–6 month	130	0.93 ± 0.05	0.59 ± 0.04	0.89 ± 0.08	0.71 ± 0.10	0.98 ± 0.01
6–12 month	135	0.92 ± 0.04	0.64 ± 0.08	0.89 ± 0.06	0.74 ± 0.17	0.94 ± 0.03
12–18 month	122	0.92 ± 0.11	0.31 ± 0.23	0.80 ± 0.04	0.44 ± 0.22	0.89 ± 0.09
18–24 month	140	0.86 ± 0.20	0.57 ± 0.12	0.52 ± 0.24	0.55 ± 0.12	0.84 ± 0.10
24–36 month	130	0.88 ± 0.01	0.44 ± 0.17	0.54 ± 0.27	0.48 ± 0.37	0.80 ± 0.15

Note: The best fusion model refers to AdaBoost (base estimator: logistic regression). AUC, the area under the receiver operating characteristic curve. The global model was assessed by grouping the outcomes of the test data at different time intervals (0–6, 6–12, 12–18, 18–24 and 24–36 months). There are 13, 18, 5, 23, and 13 patients with MCI progression in 0-6, 6-12, 12-18, 18-24, and 24-36 months, respectively.

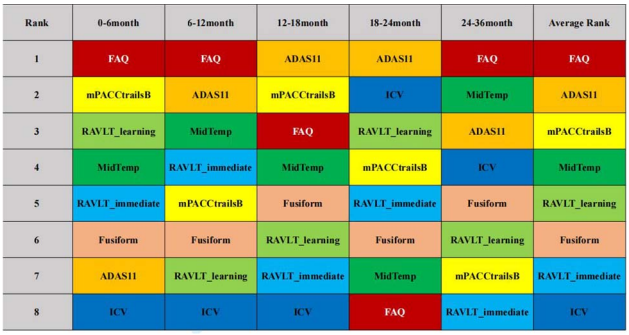

Figure 3. Ranking of feature importance at different time intervals. The same colours represent the same features. It can be seen that the importance of features changes in different time intervals, generally, FAQ, ADAS11 and mPACCtrailsB ranked the top 3.

Discussion

Based on the ADNI database, this study constructed a fusion model for AD prediction within different time intervals. Meanwhile, key predictors were identified and easily used nomogram was validated, which was valuable for risk screening and prevention.

Neuropsychological tests play an important role in early diagnosis and prediction of AD [29]. As a non-invasive and simple evaluation tool, neuropsychological tests are widely used in brain science research [30, 31]. Numerous neuropsychological test variables were included in this study, and were statistically significant between the MCI and AD groups. After LASSO processing, five of the eight selected variables were neuropsychological tests, which confirmed the importance of neuropsychological tests. We found that demographic characteristics were not statistically significant between the MCI and AD groups except for age, which was consistent with the results of Mouchet [31] et al. After feature selection, no demographic variables were included, indicating that age may not be accurate in distinguishing MCI and AD.

A retrospective study of AD prediction pointed out that the average accuracy of most studies using traditional machine learning algorithm was 75.4%, and the average accuracy of neural network was about 78.5% [32]. In the current study, we demonstrated the prediction power of machine learning models, especially for fusion strategies. The AdaBoost fusion model based on logistic regression achieved high AUC, especially in 0–6 month (AUC: 0.98). In addition, we selected eight variables with LASSO method, and a nomogram was further constructed to ensure the convenience of practical application. Our results indicated that the score of the high-risk group was significantly higher than that of the low-risk group, which is similar to previous finding [32]. The DCA curve shows that the optimal fusion model has a positive net benefit within the whole threshold probability interval, which indicates that our model has certain clinical application value. For clinical implications, our models were helpful to the early identification of high-risk populations and the identified key predictors in different time intervals were crucial for targeted prevention.

Feature importance ranking is an important method to help people understand machine learning prediction results. Two feature importance ranking algorithms were selected to eliminate the differences brought by different methods. Whatever feature importance ranking methods were utilised, FAQ, mPACCtrailsB, RAVLT_immediate and ADAS11 are important predictors. The results in different time intervals also prove the importance of FAQ, ADAS11 and mPACCtrailsB. This suggests that the neuropsychological tests were important differentiating factors between MCI and AD patients. Although the differences of these predictors between two groups may be small (feature importance is < 0.1), ML is able to identify such small differences and make accurate predictions, which are almost impossible for medical professionals to tell apart. As one of the widely used neuropsychological scales, FAQ scale is used to assess the physical condition, psychological condition, completion of social role functions and factors affecting daily performance of patients in completing daily activities. The speed and degree of FAQ changes have certain significance for the assessment of clinical dementia function. Some studies have pointed out that the score of FAQ is the most accurate variable in predicting the progression from MCI to AD (AUC > 0.9) [33], which is consistent with the results in this study. mPACC test is one of the important variables in this study. It is pointed out in previous study [34] that mPACC test can reliably measure signs of cognitive decline at an early stage, and our study also supports this view. The score of The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog11) is also one of the important predictors. Previous finding [35] has shown that ADAS-Cog11 can predict the progression from MCI to AD within 18 months. Our best fusion model shows that the accuracy was around

0.89 in 18-month prediction, while it declined over time, which may be related to a decline in the ranking of ADAS-Cog11. Rey's Auditory Verbal Learning is usually used to measure episodic language memory. In this study, the model finally incorporates two scales, instantaneous memory and learning memory. The importance of the score of the instantaneous memory scale is high, suggesting that the instantaneous memory impairment may occur before AD. Although the impaired memory function in patients with MCI does not affect daily life, it may be an important risk factor for the progression of MCI to AD, which may require further study in the future. In addition, cranial volume (ICV), volume of fusiform gyrus (Fusiform) and volume of the middle temporal lobe (MidTemp) are also important variables, which represent more changes in brain region volume. The reason for these changes may be the accumulation of changes in a single intracranial region or multiple intracranial regions, such as changes in average cortical thickness, atrophy of the hippocampus and atrophy of the medial temporal lobe structure, which are often associated with cognitive decline [36–38].

It can be seen from the identified important variables that the neuropsychological scale can reflect the progress of MCI patients in 3 years to a certain extent. Brain region volume and other variables showed relatively low predictive power, indicating that structural changes may not occur in the early process from MCI to AD, but functional changes at the cellular level. This is consistent with the findings of some retrospective studies [39]. On the other hand, the limits of the neuropsychological scale in assisting the diagnosis of MCI to AD are not clear, and it is not clear whether it is still valid at a long stage (e.g. 10 years), which needs further research in the future.

Our study was also flawed. First, the overrepresentation of whites and non-Hispanics/Latinos limited the application of models in other populations, however race was not included as a variable in the final model, and we also didn't introduce more variables such as genes and PET-CT images, which are expensive and invasive [33]. Second, the included population is older people, and the value of models in other age groups is not yet clear. Third, the number of our models and fusion methods are limited, and more advanced algorithms are warrant to be studied. In addition, we only forecast the risk of progression within 3 years, which may need to be further extended to long-term prediction in the future. Finally, it is necessary to be cautious of credibility when applying machine learning models, that is, the performance of our model is subject to change outside of this dataset, which needs to be refined in future studies to explore the generalisability of our models in a wider range of populations and regions.

Conclusion

In this study, we proposed an MCI progression prediction model. Eight variables were selected by LASSO method, the AdaBoost model based on logistic regression achieved the best performance (AUC: 0.98) in 0–6 month prediction. Scores from the FAQ, mPACCtrailsB and ADAS11 (Unweighted sum of 11 items from The Alzheimer's Disease Assessment Scale-Cognitive Subscale) were key predictors. In future studies, long-term prediction with more variables and advanced methods is needed, so as to provide clues for the prevention and treatment of AD.

Supplementary Data: Supplementary data mentioned in the text are available to subscribers in *Age and Aging* online.

Acknowledgements: We thank the Alzheimer's disease Neuroimaging initiative (ADNI) for providing us with the data. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www. fnih.org). The grantee organisation is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.
Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Declaration of Conflict of Interest: None.

Declaration of Sources of Funding: This swork was supported by the National Natural Science Foundation of China (Grant number 81973144) and the National Key R&D Program of China (No. 2022YFC3603000). Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian

Institutes of Health Research is providing funds to support ADNI clinical sites in Canada.

Data Availability: The data supporting the findings of this study are openly available at https://adni.loni.usc.edu/.

References

1. Reitz C, Mayeux R. Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. Biochem Pharmacol 2014; 88: 640–51.
2. Neugroschl J, Wang S. Alzheimer's disease: diagnosis and treatment across the spectrum of disease severity. Mt Sinai J Med 2011; 78: 596–612.
3. Petersen RC, Roberts RO, Knopman DS *et al.* Mild cognitive impairment: ten years later. Arch Neurol 2009; 66: 1447–55.
4. Schneider JA, Arvanitakis Z, Leurgans SE, Bennett DA. The neuropathology of probable Alzheimer disease and mild cognitive impairment. Ann Neurol 2009; 66: 200–8.
5. Vega JN, Newhouse PA. Mild cognitive impairment: diagnosis, longitudinal course, and emerging treatments. Curr Psychiatry Rep 2014; 16: 490. https://doi.org/10.1007/s11920-014-0490-8.
6. Liampas I, Siokas V, Ntanasi E *et al.* Cognitive trajectories preluding the imminent onset of Alzheimer's disease dementia in individuals with normal cognition: results from the HELIAD cohort. Aging Clin Exp Res 2023; 35: 41–51.
7. Madusanka N, Choi HK, So JH, Choi BK, Park HG. One-year follow-up study of hippocampal subfield atrophy in Alzheimer's disease and normal aging. Curr Med Imaging Rev 2019; 15: 699–709.
8. Devanarayan P, Devanarayan V, Llano DA, Initia ADN. Identification of a simple and novel cut-point based cerebrospinal fluid and MRI signature for predicting Alzheimer's disease progression that reinforces the 2018 NIA-AA research framework. J Alzheimers Dis 2019; 68: 537–50.
9. deToledo-Morrell L, Stoub TR, Bulgakova M *et al.* MRI-derived entorhinal volume is a good predictor of conversion from MCI to AD. Neurobiol Aging 2004; 25: 1197–203.
10. Querbes O, Aubry F, Pariente J *et al.* Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain 2009; 132: 2036–47.
11. Hojjati SH, Ebrahimzadeh A, Khazaee A, Babajani-Feremi A, Alzheimer's Disease Neuroimaging Initiative. Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. J Neurosci Methods 2017; 282: 69–80.
12. Naseri NN, Wang H, Guo J, Sharma M, Luo WJ. The complexity of tau in Alzheimer's disease. Neurosci Lett 2019; 705: 183–94.
13. Nagaraj S, Duong TQ. Deep learning and risk score classification of mild cognitive impairment and Alzheimer's disease. J Alzheimers Dis 2021; 80: 1079–90.
14. Zhao X, Sui H, Yan C *et al.* Machine-based learning shifting to prediction model of deteriorative MCI due to Alzheimer's disease - a two-year follow-up investigation. Curr Alzheimer Res 2022; 19: 708–15.
15. Massetti N, Russo M, Franciotti R *et al.* Erratum to: a machine learning-based holistic approach to predict the clinical course of patients within the Alzheimer's disease spectrum. J Alzheimers Dis 2022; 90: 931. https://doi.org/10.3233/JAD-229016.
16. Wang L, Li P, Hou M, Zhang X, Cao X, Li H. Construction of a risk prediction model for Alzheimer's disease in the elderly population. BMC Neurol 2021; 21: 271. https://doi.org/10.1186/s12883-021-02276-8.
17. Velazquez M, Lee Y, Alzheimer's Disease Neuroimaging Initiative. Random forest model for feature-based Alzheimer's disease conversion prediction from early mild cognitive impairment subjects. PloS One 2021; 16: e0244773. https://doi.org/10.1371/journal.pone.0244773.
18. Fleuren LM, Klausch TLT, Zwager CL *et al.* Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020; 46: 383–400.
19. Sajjadian M, Lam RW, Milev R *et al.* Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. Psychol Med 2021; 51: 2742–51.
20. Pellegrini E, Ballerini L, Hernandez M *et al.* Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. Alzheimers Dement (Amst) 2018; 10: 519–35.
21. Gondrie MJA, Janssen KJM, Moons KGM, van der Graaf Y. A simple adaptation method improved the interpretability of prediction models for composite end points. J Clin Epidemiol 2012; 65: 946–53.
22. Raschke RA, Reilly BM, Guidry JR, Fontana JR, Srinivas S. The weight-based heparin dosing nomogram compared with a "standard care" nomogram. A randomized controlled trial. Ann Intern Med 1993; 119: 874–81.
23. Wu S, Zheng J, Li Y *et al.* A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. Clin Cancer Res 2017; 23: 6904–11.
24. Tang L, Wu X, Liu H *et al.* Individualized prediction of early Alzheimer's disease based on magnetic resonance imaging radiomics, clinical, and laboratory examinations: a 60-month follow-up study. J Magn Reson Imaging 2021; 54: 1647–57.
25. Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 2012; 28: 112–8.
26. Jack CR Jr, Twomey CK, Zinsmeister AR, Sharbrough FW, Petersen RC, Cascino GD. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. Radiology 1989; 172: 549–54.
27. Service TC. A no free lunch theorem for multi-objective optimization. Inf Process Lett 2010; 110: 917–23.
28. Shu ZY, Mao DW, Xu YY, Shao Y, Pang PP, Gong XY. Prediction of the progression from mild cognitive impairment to Alzheimer's disease using a radiomics-integrated model. Ther Adv Neurol Disord 2021; 14: 1756286421102955. https://doi.org/10.1177/17562864211029551.
29. Veitch DP, Weiner MW, Aisen PS *et al.* Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's disease neuroimaging initiative. Alzheimers Dement 2019; 15: 106–52.
30. Carcamo J, Kociolek AJ, Fernandez KK *et al.* Neuropsychological predictors of severe functional dependency in a multiethnic community cohort of individuals with Alzheimer's disease. J Alzheimers Dis 2021; 83: 539–44.

31. Kwak S, Oh DJ, Jeon YJ *et al.* Utility of machine learning approach with neuropsychological tests in predicting functional impairment of Alzheimer's disease. J Alzheimers Dis 2022; 85: 1357–72.
32. Grueso S, Viejo-Sobera R. Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. Alzheimers Res Ther 2021; 13: 162. https://doi.org/10.1186/s13195-021-00900-w.
33. Huang K, Lin Y, Yang L *et al.* A multipredictor model to predict the conversion of mild cognitive impairment to Alzheimer's disease by using a predictive nomogram. Neuropsychopharmacology 2020; 45: 358–66.
34. Donohue MC, Sperling RA, Salmon DP *et al.* The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol 2014; 71: 961–70.
35. Bertolini D, Loukianov AD, Smith A *et al.* Forecasting progression of mild cognitive impairment (MCI) and Alzheimer's disease (AD) with digital twins. Alzheimers Dement 2021; 17: 1–3. https://doi.org/10.1002/alz.054414.
36. Kehoe EG, McNulty JP, Mullins PG, Bokde AL. Advances in MRI biomarkers for the diagnosis of Alzheimer's disease. Biomark Med 2014; 8: 1151–69.
37. Illan-Gala I, Pegueroles J, Montal V *et al.* Challenges associated with biomarker-based classification systems for Alzheimer's disease. Alzheimers Dement (Amst) 2018; 10: 346–57.
38. Dubois B, Feldman HH, Jacova C *et al.* Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol 2014; 13: 614–29.
39. Khan S, Barve KH, Kumar MS. Recent advancements in pathogenesis, diagnostics and treatment of Alzheimer's disease. Curr Neuropharmacol 2020; 18: 1106–25.

Received 11 April 2023; editorial decision 4 August 2023