Dementia and Geriatric Cognitive Disorders

Research Article

Dement Geriatr Cogn Disord 2023;52:249–257
DOI: 10.1159/000531819

Received: April 18, 2023
Accepted: June 30, 2023
Published online: July 21, 2023

Predicting Alzheimer's Disease with Interpretable Machine Learning

Maoni Jia[a] Yafei Wu[a] Chaoyi Xiang[a] Ya Fang[a, b]

[a]Center for Aging and Health Research, School of Public Health, Xiamen University, Xiamen, China; [b]National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China

Keywords
Alzheimer's disease · Prediction model · Machine learning · Interpretability analysis

Abstract
Introduction: This study aimed to develop novel machine learning models for predicting Alzheimer's disease (AD) and identify key factors for targeted prevention. ***Methods:*** We included 1,219, 863, and 482 participants aged 60+ years with only sociodemographic, both sociodemographic and self-reported health, both the former two and blood biomarkers information from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Machine learning models were constructed for predicting the risk of AD for the above three populations. Model performance was evaluated by discrimination, calibration, and clinical usefulness. SHapley Additive exPlanation (SHAP) was applied to identify key predictors of optimal models. ***Results:*** The mean age was 73.49, 74.52, and 74.29 years for the three populations, respectively. Models with sociodemographic information and models with both sociodemographic and self-reported health information showed modest performance. For models with sociodemographic, self-reported health, and blood biomarker information, their overall performance improved substantially, specifically, logistic regression performed best, with an AUC value of 0.818. Blood biomarkers of ptau protein and plasma neurofilament light, age, blood tau protein, and education level were top five significant predictors. In addition, taurine, inosine, xanthine, marital status, and L.Glutamine also showed importance to AD prediction. ***Conclusion:*** Interpretable machine learning showed promise in screening high-risk AD individual and could further identify key predictors for targeted prevention.

Introduction

Dementia has become a major threat to human health and quality of life. According to the World Health Organization, more than 55 million people were suffering from dementia, and this number is expected to reach 78 and 139 million by 2030 and 2050 [1]. Notably, Alzheimer's disease (AD) is the most common cause of dementia [2]. The onset of AD is latent and irreversible, and it would cause adverse outcomes such as disability and death [3]. Therefore, development of prediction models can help early identification of AD, and further provide implications for targeted prevention and medical decisions.

With the generation of huge volume of medical and health data, it has significantly stimulated the rapid development and application of data mining techniques in the field of healthcare [4]. Studies on data-driven

Maoni Jia and Yafei Wu contributed equally to this work.

karger@karger.com
www.karger.com/dem

Correspondence to:
Ya Fang, fangya@xmu.edu.cn

Karger

diagnostic or prognostic models for diseases are on the rise, evolving from classical statistical methods to machine learning models [5]. It is worth noting that machine learning algorithms often show superior performance to statistical models when encountering high-dimensional and complex healthcare data, and played an increasingly critical role in medical study [6]. Nevertheless, machine learning methods were considered to be less interpretable compared to statistical models [7]. Hence, interpretable techniques are emerging in recent years such as SHapley Additive exPlanation (SHAP), which can be leveraged to shed light on the inner decision-making process and identify significant variables [8].

Previous studies have examined a range of risk factors of AD in the elderly population including sociodemographic factors such as age, gender, and education, disease history (comorbidities) such as diabetes, cardiovascular disease, unhealthy lifestyles such as smoking and alcohol consumption, and blood biological markers [9]. Presently, cerebrospinal fluid amyloid-beta42 (Aβ42) and tau protein are recognized fluid markers [10], but the invasive nature of lumbar puncture limits their widespread use. Moreover, magnetic resonance imaging and positron emission tomography perform more reliable in cognitive impairment assessment, but the high cost and limited availability also constrain their broader applications [11].

In this context, we aimed to examine the following questions: First, was the state of art machine learning models utilizing multimodal data (sociodemographic, self-reported health, and blood biomarker information) a suitable tool for AD prediction among older adults? Second, what key predictors could be identified for AD prediction when using the SHAP analysis? Third, how does blood biomarker information contribute to AD prediction?

Methods

Participants

Data were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu), which is open-access and has high quality for cognitive research. For more information, please refer to the official website at www.adni-info.org. ADNI research began in 2004 (ADNI-1), including study extension in 2009 (ADNI-GO) and renewals in 2011 (ADNI-2) and 2016 (ADNI-3). The ADNI study protocol was reviewed and approved by Ethics Committees at each of the participating site. This full list of participating site and Ethics Committees can be found at previous publication [12]. All subjects provided written informed consent through the participating ADNI sites. We conducted sample selection in three scenarios: (1) participants who had only sociodemographic information (scenario 1); (2) participants who had both sociodemographic and self-reported health information (scenario 2); (3) participants who had sociodemographic, self-reported health, and routine blood biomarker information (scenario 3). Finally, there were 1,219, 863, and 482 subjects included, respectively. The detailed sample information was listed in online supplementary Figure S1 (for all online suppl. material, see https://doi.org/10.1159/000531819). The variables selected for each scenario were presented in online supplementary Methods S1 and supplementary Table S1.

The ADNI study is designed to diagnose and treat patients with cognitive impairment early, therefore, more AD patients were included in the database. In this study, out of the total samples in each scenario, there were 400, 341, and 183 old adults had AD, respectively. The mean age was 73.49, 74.52, and 74.29 years for each scenario. The proportion of male participants was 48.1%, 51.2%, and 50.0%, respectively. More detailed descriptive information of participants in 3 scenarios and the comparisons between AD and normal populations were shown in online supplementary Tables S2–S4.

Outcome

In this study, we aimed to predict AD (normal control vs. AD) for older adults aged 60+ years old. In ADNI study, normal control was mainly defined by following criteria: MMSE scores between 24 and 30, Clinical Dementia Rating score and Memory Box score of 0, an absence of significant impairment in cognitive functions or activities of daily living. AD was mainly defined by following criteria: MMSE scores between 20 and 26, CDR score of 0.5 or 1.0, and meets the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association criteria for probable AD [13]. A more detailed description of normal control and AD could be found in online supplementary Table S5 and at http://adni.loni.usc.edu/methods/documents.

Candidate Variables and Preprocessing

In the current study, we collected 3 kinds of variables including sociodemographic, self-reported health, and blood biomarker information from ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. Variables with missing rate of more than 20% were removed. For variables with missing rate of less than 20%, missForest was employed to fill the missing data, which can deal with both categorical and continuous variables based on random forest (RF) [14]. One-hot encoding was adopted to encode categorical variables [15]. All continuous variables were standardized for better model convergence.

Model Construction and Evaluation

Based on the biomedical research guideline recommendations [16], we selected six commonly used models including logistic regression with penalty (LR) [17], support vector machine (SVM) [18], decision trees [19], random forest (RF) [20], extreme gradient boosting (XGB) [21], and artificial neural network (ANN) [22], to test their ability in predicting AD.

We used the six machine learning algorithms to construct AD prediction models in each scenario. Specifically, in scenario 1, only sociodemographic information was included. Then the self-reported health information were added in scenario 2. Further, blood biomarkers were integrated in scenario 3. In addition, we further explored the value of blood biomarkers in AD prediction and hoped to find some key blood biomarkers.

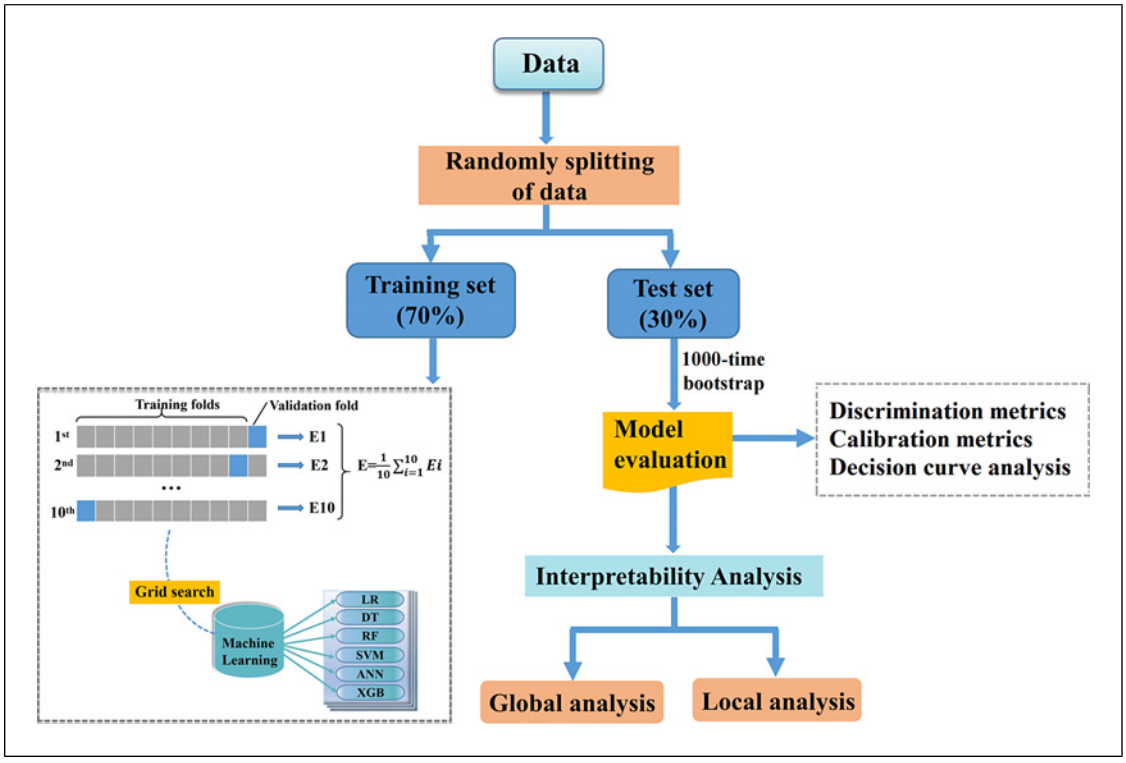

Fig. 1. Flowchart of the workflow. LR, logistic regression; DT, decision tree; RF, random forest; SVM, support vector machine; ANN, artificial neural network; XGB, extreme gradient boosting; AUC, area under curve; DCA, decision curve analysis.

We followed a standard procedure of model construction, derived from the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [23]. The whole data were randomly split into training (70%) and testing datasets (30%) considering the distribution of AD versus non-AD patients. Then we conducted 10-fold cross-validation in training dataset for hyperparameters tuning. The main hyperparameters used for our prediction models were shown in online supplementary Table S6. Discrimination, calibration, and clinical usefulness metrics were used to evaluate the model performance in the test set. Balanced accuracy, sensitivity, specificity, precision, F1 score, area under the receiver operating characteristic curve (AUROC) were considered for model discrimination. Brier score was used to evaluate the calibration results of models. The utility of models in clinical practice was evaluated by decision curve analysis (DCA) based on a range of threshold probabilities. For the above metrics, the 95% confidence interval was also calculated in the testing data with a 1000-time repeated bootstrap method [24]. Python 3.7.6 was adopted for machine learning analyses. The detailed workflow including model construction and evaluation is shown in Figure 1.

Interpretability Analysis

Although machine learning has made great breakthroughs in medicine, their complicated structure make them difficult for decision makers to understand [25]. SHAP, an interpretable technique, was able to uncover the black box of machine learning. SHAP can be used to interpret the model from both global and local perspectives. The global analysis of SHAP was presented by summary plot. The horizontal axis of summary plot denotes the average absolute SHAP value of each feature, a longer horizontal bar means a larger absolute value of SHAP, which implies a more significant feature. Local interpretability provides the details of the predictions by force plot, focusing on explaining how individual predictions are derived. Each feature has its own contribution, and the accumulation of all features ultimately drives the prediction from the base value to the final model output. Features that push the prediction higher are shown in red and that push the prediction lower are shown in blue.

Statistical Analysis

For descriptive analysis, continuous variables were displayed as means (standard deviation) for normal distribution or presented by median (interquartile) for skewed distribution. Categorical variables were presented by numbers (percentage). For comparisons of characteristics, t test or Wilcoxon test and χ^2 test or Fisher's exact test were used for continuous and categorical variables, respectively. The two-tailed value of $p < 0.05$ was regarded as statistically significant. All the above analyses were performed by SPSS.26.0.

Dement Geriatr Cogn Disord 2023;52:249–257
DOI: 10.1159/000531819

Table 1. Performance of machine learning algorithms in AD prediction

	LR	DT	RF	SVM	XGB	ANN
Scenario 1						
Balanced-accuracy	0.738 (0.714, 0.762)	0.722 (0.698, 0.746)	0.749 (0.726, 0.773)	0.717 (0.692, 0.741)	0.746 (0.723, 0.770)	0.644 (0.618, 0.670)
Sensitivity	0.304 (0.256, 0.352)	0.385 (0.336, 0.435)	0.385 (0.337, 0.433)	0.171 (0.132, 0.209)	0.339 (0.291, 0.388)	0.491 (0.440, 0.542)
Specificity	0.882 (0.852, 0.912)	0.832 (0.804, 0.850)	0.869 (0.833, 0.905)	0.823 (0.796, 0.838	0.854 (0.820, 0.879)	0.776 (0.745, 0.786)
Precision	0.652 (0.582, 0.723)	0.565 (0.506, 0.625)	0.651 (0.591, 0.712)	0.633 (0.539, 0.727)	0.666 (0.598, 0.733)	0.428 (0.382, 0.473)
F1	0.414 (0.361, 0.468)	0.458 (0.410, 0.506)	0.484 (0.436, 0.531)	0.268 (0.215, 0.321)	0.449 (0.398, 0.500)	0.457 (0.414, 0.499)
AUC	0.720 (0.691, 0.750)	0.713 (0.683, 0.743)	0.764 (0.736, 0.791)	0.767 (0.738, 0.795)	0.754 (0.726, 0.782)	0.648 (0.617, 0.679)
Brier score	0.184 (0.174, 0.195)	0.209 (0.194, 0.224)	0.175 (0.165, 0.184)	0.178 (0.168, 0.188)	0.179 (0.170, 0.188)	0.356 (0.330, 0.382)
Scenario 2						
Balanced-accuracy	0.621 (0.594, 0.648)	0.626 (0.600, 0.651)	0.614 (0.587, 0.641)	0.668 (0.642, 0.694)	0.660 (0.634, 0.686)	0.575 (0.548, 0.601)
Sensitivity	0.470 (0.427, 0.513)	0.412 (0.369, 0.455)	0.451 (0.407, 0.496)	0.353 (0.312, 0.395)	0.480 (0.436, 0.525)	0.529 (0.485, 0.573)
Specificity	0.776 (0.745, 0.786)	0.837 (0.802, 0.851)	0.824 (0.798, 0.844)	0.806 (0.785, 0.815)	0.806 (0.785, 0.815)	0.786 (0.735, 0.802)
Precision	0.521 (0.474, 0.568)	0.532 (0.481, 0.582)	0.511 (0.463, 0.559)	0.643 (0.586, 0.699)	0.582 (0.534, 0.631)	0.465 (0.423, 0.507)
F1	0.494 (0.454, 0.533)	0.464 (0.423, 0.505)	0.479 (0.439, 0.520)	0.456 (0.413, 0.499)	0.526 (0.486, 0.567)	0.495 (0.457, 0.532)
AUC	0.682 (0.653, 0.711)	0.636 (0.605, 0.667)	0.672 (0.642, 0.702)	0.698 (0.670, 0.727)	0.694 (0.665, 0.723)	0.621 (0.592, 0.650)
Brier score	0.217 (0.207, 0.226)	0.284 (0.265, 0.303)	0.223 (0.212, 0.233)	0.213 (0.204, 0.223)	0.212 (0.203, 0.221)	0.421 (0.394, 0.447)
Scenario 3						
Balanced-accuracy	0.744 (0.673, 0.815)	0.684 (0.603, 0.765)	0.701 (0.625, 0.777)	0.556 (0.490, 0.622)	0.717 (0.640, 0.794)	0.648 (0.570, 0.725)
Sensitivity	0.801 (0.737, 0.864)	0.724 (0.649, 0.798)	0.745 (0.674, 0.816)	0.655 (0.576, 0.734)	0.766 (0.695, 0.836)	0.704 (0.630, 0.779)
Specificity	0.936 (0.900, 0.985)	0.819 (0.783, 0.899)	0.829 (0.793, 0.908)	0.893 (0.857, 0.957)	0.882 (0.846, 0.949)	0.841 (0.805, 0.917)
Precision	0.807 (0.742, 0.871)	0.721 (0.643, 0.799)	0.741 (0.667, 0.815)	0.628 (0.530, 0.726)	0.763 (0.690, 0.836)	0.696 (0.616, 0.776)
F1	0.789 (0.718, 0.859)	0.719 (0.642, 0.796)	0.739 (0.665, 0.813)	0.607 (0.513, 0.701)	0.757 (0.682, 0.832)	0.692 (0.612, 0.772)
AUC	0.818 (0.745, 0.890)	0.702 (0.616, 0.788)	0.802 (0.729, 0.875)	0.382 (0.283, 0.482)	0.788 (0.710, 0.867)	0.730 (0.646, 0.813)
Brier score	0.161 (0.126, 0.195)	0.206 (0.157, 0.255)	0.172 (0.139, 0.205)	0.236 (0.216, 0.255)	0.174 (0.141, 0.207)	0.199 (0.163, 0.235)

All the metrics were calculated in the 30% test data with 1,000-time bootstrap. Scenario 1, only sociodemographic variables; Scenario 2, scenario 1 plus self-reported health information; Scenario 3, scenario 2 plus blood biomarkers. LR; logistic regression; DT, decision tree; RF, random forest; SVM, support vector machine; XGB, extreme gradient boosting; ANN, artificial neural network; AUC, area under the curve.

Results

Performance of Prediction Models

Performance of prediction models in each scenario is shown in Table 1. In scenario 1, all models showed relatively moderate performance in balanced accuracy (ranging from 0.644 to 0.749) and areas under the curve (AUCs) (ranging from 0.648 to 0.767), and these models were inclined to fail to predict AD with lower sensitivity (less than 0.50). Compared to scenario 1, there was a slight improvement in sensitivity (maximum: 0.529) in scenario 2, but the balanced accuracy, precision, and AUC declined slightly, and the Brier score was relatively higher. In scenario 3, the sensitivity of all models improved significantly compared with other two scenarios, the maximum sensitivity, F1, and AUC were 0.801, 0.789, and 0.818 for LR, respectively. Moreover, when only blood biomarkers were incorporated, the overall performance of all models was superior to that in scenario 1 and scenario 2 but slightly lower than that in scenario 3 (online suppl. Table S7), and the sensitivity, F1 and AUCs of LR were 0.745, 0.728, and 0.820, respectively.

For DCA analysis (Fig. 2), when the threshold was set between 0.2 and 0.8 in scenario 1 and scenario 2, certain clinical benefit could be expected from implementing specific interventions on high-risk individuals. While in scenario 3, the threshold was better to be set between 0.2 and 0.6 and LR showed much higher and stable net benefits compared with other models.

Interpretability Analysis of Models

Given that the overall performance of prediction models in scenario 3 was much better, especially LR model for its high sensitivity, F1 score, AUC, and low Brier score, therefore, we selected LR model in scenario 3 for model interpretability analysis to explore the most important predictors, which can be regarded as significant indicators to further tailor the intervention among the high-risk population. Figure 3 presented the top 10 key factors based on the results of the global interpretability analysis, it can be found that the blood biomarkers of ptau protein and plasma neurofilament light (NfL), age, blood tau protein, and education were top five significant predictors (Fig. 3). In addition, taurine, inosine, xanthine, marital status, and L.Glutamine also showed importance for AD prediction. When only utilizing blood biomarkers variables for AD prediction, NfL was recognized as the most significant factor with maximum SHAP value (online suppl. Fig. S2).

Figure 4a, b displayed the contribution of each predictor for the individual decision. For example, from Figure 4a, we can know that male, taurine, age, and ptau protein pushed the model from the base value (−0.8644) to larger direction. Tau protein in blood and education conversely pushed the model down from the base value. Ultimately, with all these factors, the model was pushed from the base value (−0.8644) to 0.1, predicting that the individual was a positive individual (high-risk AD individual). For Figure 4b, apart from tau protein, ptau protein, plasma NfL, age, and taurine pushed the model down from the base value (−0.8644) to a lower value (−2.24); finally, this individual was predicted as normal. Figure 4c, d demonstrated the decision mechanisms for false-negative and false-positive subjects separately.

Discussion

In this study, we proposed machine learning prediction models for AD identification, and interpretable method was utilized to explore the key predictors, and we further discussed the value of blood biomarkers on the performance of AD identification. These models can be employed to optimize risk stratification, and the identified predictors could be served as important clues for targeted prevention.

All models in scenario 1 showed moderate discriminative ability. In scenario 2, models showed lower balanced accuracy, precision, and AUC but relatively higher sensitivity and F1 score compared to models in scenario 1. Presumably because these additional variables may also be substantially associated with various other geriatric diseases such as cancer, cardiovascular diseases, and so on, which undermined their ability to discriminate the onset of AD [26]. Consistent with the findings of previous studies [27], when blood biomarkers were further incorporated in scenario 3, the model performance improved significantly, especially for sensitivity. When only blood biomarkers were involved, all models outperformed than models with only epidemiological data (sociodemographic characteristics and self-reported health information). Comparing with the performance of models in scenario 1 and scenario 2, our results suggested that the value of blood biological markers was somewhat superior to epidemiological information, but there is still a long way to go to explore low invasive but more sensitive indicators for AD identification. In the current study, blood biomarkers of ptau protein and plasma NfL, age, blood tau protein, and education were top five significant variables and could be the priority indicators for AD prediction.

Based on the results of this study, when the models incorporated only epidemiological information, the sensitivity of the models was low and was unable to identify more high-risk AD populations. When blood biological

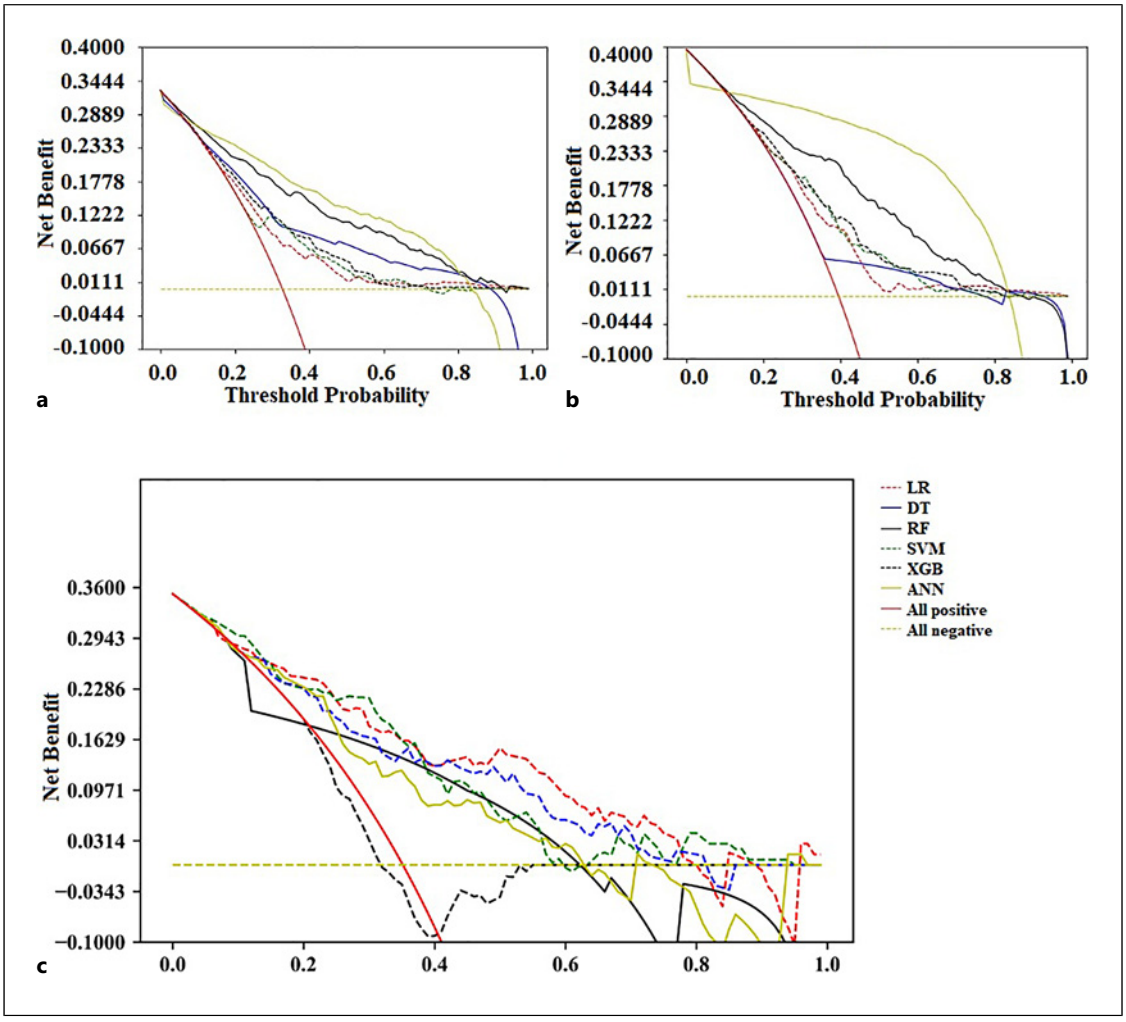

Fig. 2. Decision curve and net benefit of each model. Net benefit: the sum of the gain value of the intervention for the corresponding true-positive population at each threshold and the loss value of the intervention for the false-positive population. LR, logistic regression; DT, decision tree; RF, random forest; SVM, support vector machine; XGB, extreme gradient boosting; ANN, artificial neural network; all positive: the net benefit of providing intervention for all subjects (all subjects would be positive); all negative: the net benefit of providing no intervention for all subjects (all subjects would be negative). **a** Scenario 1 (only sociodemographic variables). **b** Scenario 2 (scenario 1 plus self-reported health information). **c** Scenario 3 (scenario 2 plus blood biomarker).

markers were further incorporated, the sensitivity of the models was considerably improved, which is able to assist in identifying people at high risk of AD. Additionally, in contrast to cerebrospinal fluid and imaging biomarkers, blood biomarkers, detected from peripheral blood samples, contained a variety of metabolites that reflected the physiological activity of many organs, including the brain, which made it available as population prediction tools [28]. Further, it has been confirmed that the biochemical changes in metabolism, followed by energy metabolism, associated with N-acetylaspartate (NAA), inositol (MI), glucose (Glu), glutamate (Gln), and aspartate (Asp) preceded structural changes in AD [29]. Therefore, we constructed models with only blood biomarkers to examine possible low-invasive biological indicators of changes in cognition decline to help early identification of AD. The results identified that NfL chain was a significant predictor based on interpretability analysis. NFL is a crucial component subunit of neurofilament protein (NF). As the major cytoskeletal protein of neuronal axons, NF is highly expressed in neuronal axon sites and maintains the stability of axon morphology and ensuring nerve signaling. The NfL in tissue fluid can diffuse to blood through cerebrospinal fluid, and the concentration of NfL in blood is lower than cerebrospinal fluid. Therefore, NfL is considered to be an promising reflection of neurodegeneration and can be

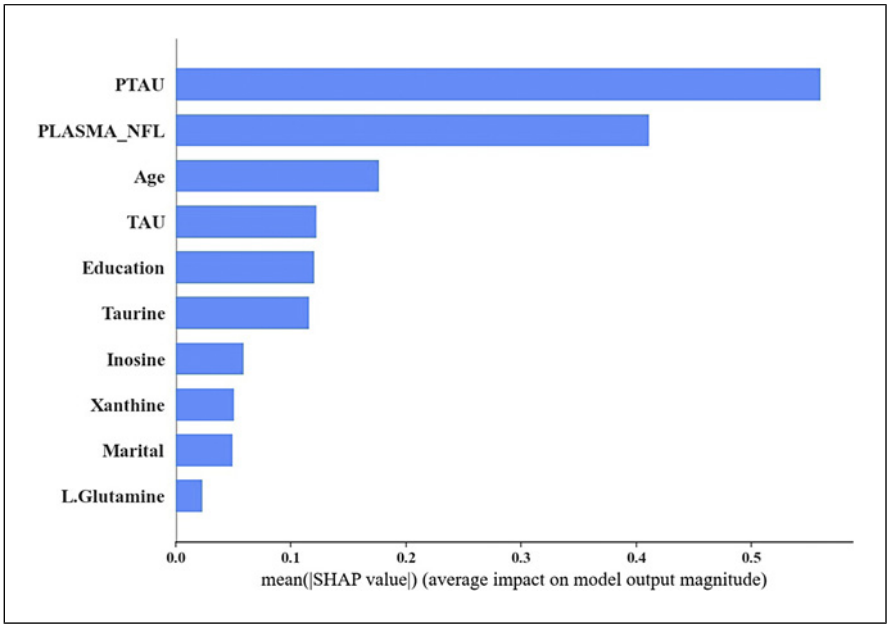

Fig. 3. Interpretability analysis with SHapley Additive exPlanation for logistic regression in scenario 3 (top 10 predictors).

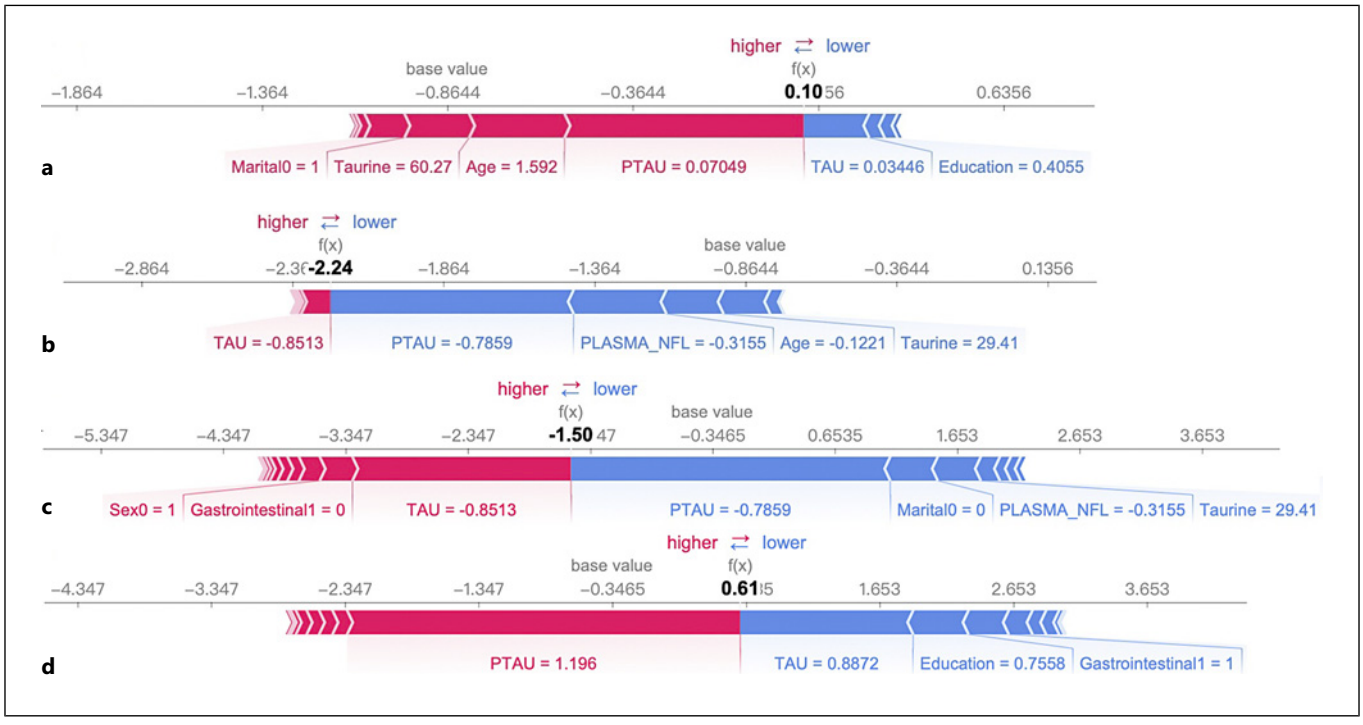

Fig. 4. Individual aspect of interpretability analysis with SHapley Additive exPlanation in scenario 3 (Logistic regression). **a** True positive. **b** True negative. **c** False negative. **d** False positive.

measured in blood [30]. However, the prediction power of plasma metabolites was still limited in the current study. The possible explanation for this was that the method of blood biomarkers assay was taken into consideration, and the ability of the blood markers to predict AD may be improved by more precise tests of blood biomarkers in the future, so the clinical application of blood biomarkers needs to be further examined.

Our prediction models have several clinical implications, the machine learning algorithms not only showed potential for predicting AD with abundant data but also showed stable and high clinical value from the DCA curve. The net benefits in scenario 3 were near 0.2 at the threshold of 0.3 indicating that the designed models could be used to identify 20 positive subjects among 100 individuals [31]. Additionally, the interpretability analyses made it more transparent to give individual decision with presenting the contribution of each predictor.

There are still several limitations. First, this study was a cross-sectional study; therefore, it is necessary to validate our results through a prospective study in future. Second, the sample size was relatively small; we performed bootstrap in order to give objective model evaluation. A large sample is warranted to externally validate our results. Meanwhile, machine learning algorithms didn't outperform LR in our study, which may be also due to the small sample size, and more advanced methods such as convolutional neural network are deserved to be studied in a large dataset. Third, majority of the people recruited in ADNI were from North America, the study population was relatively constrained, which may limit model application and the results need to be generalized among other populations. Fourth, this study only predicts the risk of AD without considering its subtypes because there was no more detailed information about other dementia types, which is deserved to be studied in future. Finally, the positive predicted value was not high enough, indicating some misdiagnosed cases, which needs to be addressed in further investigations.

Conclusion

This study utilized advanced machine learning methods and abundant clinical data to accurately predict AD, which is crucial for optimizing personalized prevention trials and individualized risk management. In addition, interpretable analysis highlighted that blood biomarkers of ptau protein and plasma NfL, and blood tau protein, sociodemographic variables of age and education were the top five key factors for the early identification of AD, and taurine, inosine, xanthine, marital status, and L.Glutamine also showed importance for AD prediction, which may be used for targeted intervention.

Acknowledgments

We thank the staff and the participants of the ADNI study.

Statement of Ethics

The ADNI study is an open-access database, and publicly available for global researchers. The ADNI study was approved by each of the participating sites' Institutional Review Boards (IRBs) and complied with the Declaration of Helsinki. Written informed consent was obtained from all participants after they had received a complete description of the study. A full list of participating site and Ethics Committees can be found in Methods section.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This study was supported by the National Natural Science Foundation of China (No. 81973144).

Author Contributions

M.J., Y.W., C.X., and Y.F. worked together on this article. Specifically, Y.W. conceived and designed the study. Y.W., and M.J. contributed to the data analysis. M.J., and Y.W. drafted the manuscript. C.X. revised the article. Y.F. supervised and revised the article. All authors have read and approved the final manuscript.

Data Availability Statement

The ADNI study is an open-access database, and publicly available at Alzheimer's Disease Neuroimaging Initiative (https://adni.loni.usc.edu/). Further inquiries can be directed to the corresponding author.

References

1 Wang CY, Song PP, Niu YH. The management of dementia worldwide: a review on policy practices, clinical guidelines, end-of-life care, and challenge along with aging population. Biosci Trends. 2022 Apr;16(2):119–29.

2 Zetterberg H, Schott JM. Blood biomarkers for Alzheimer's disease and related disorders. Acta Neurol Scand. 2022;146(1):51–5.

3 2022 Alzheimer's disease facts and figures. Alzheimers Dement. 2022;18(4):700–89.

4 Sarwar T, Seifollahi S, Chan J, Zhang X, Aksakalli V, Hudson I, et al. The secondary use of electronic health records for data mining: data characteristics and challenges. ACM Comput Surv. 2022;55(2):1–40.

5 Shehab M, Abualigah L, Shambour Q, Abu-Hashem MA, Shambour MKY, Alsalibi AI, et al. Machine learning in medical applications: a review of state-of-the-art methods. Comput Biol Med. 2022 Jun;145:105458.
6 Chen SX, Xu C. Handling high-dimensional data with missing values by modern machine learning techniques. J Appl Stat. 2022;50(3):786–804.
7 Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. Can J Cardiol. 2022 Feb;38(2):204–13.
8 Kwon Y, Rivas MA, Zou JYJA. Efficient computation and analysis of distributional shapley values. 2021. abs/2007.01357.
9 Serrano-Pozo A, Das S, Hyman BT. APOE and Alzheimer's disease: advances in genetics, pathophysiology, and therapeutic approaches. Lancet Neurol. 2021;20(1):68–80.
10 Wu C, Wu L, Wang J, Lin L, Li Y, Lu Q, et al. Systematic identification of risk factors and drug repurposing options for Alzheimer's disease. Alzheimers Dement. 2021;7(1):e12148.
11 Blennow KA-O, Zetterberg H. Biomarkers for Alzheimer's disease: current status and prospects for the future. J Intern Med. 2018;284(6):643–63.
12 Podhorna J, Krahnke T, Shear M, Harrison JE; Alzheimer's Disease Neuroimaging Initiative. Alzheimer's disease assessment scale-cognitive subscale variants in mild cognitive impairment and mild Alzheimer's disease: change over time and the effect of enrichment strategies. Alzheimers Res Ther. 2016 Feb 12;8:8.
13 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. Neurology. 1984;34(7):939–44.
14 Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28(1):112–8.
15 Tokuyama Y, Miki R, Fukushima Y, Tarutani Y, Yokohira T. Performance evaluation of feature encoding methods in network traffic prediction using recurrent neural networks. Proceedings of the 2020 8th international conference on information and education technology. Okayama, Japan: Association for Computing Machinery; 2020. p. 279–83.
16 Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res. 2016;18(12):e323.
17 Guan XC, Zhang JH, Chen SY. Logistic regression based on statistical learning model with linearized kernel for classification. Cai. 2021;40(2):298–317.
18 Chauhan VK, Dahiya K, Sharma A. Problem formulations and solvers in linear SVM: a review. Artif Intell Rev. 2019;52(2):803–55.
19 Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. J Chemom. 2004;18(6):275–85.
20 Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci. 2017;9:329.
21 Schapire RE. The strength of weak learnability. Mach Learn. 1990;5(2):197–227.
22 Han SH, Kim KW, Kim S, Youn YC. Artificial neural network: understanding the basic concepts without mathematics. Dement Neurocogn Disord. 2018;17(3):83–9.
23 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(10):735–6.
24 Ötleş E, Seymour J, Wang H, Denton BT. Dynamic prediction of work status for workers with occupational injuries: assessing the value of longitudinal observations. J Am Med Inf Assoc. 2022;29(11):1931–40.
25 Zhou Q, Liao F, Mou C, Wang P. Measuring interpretability for different types of machine learning models. In: Ganji M, Rashidi L, Fung BCM, Wang C, editors. Trends and applications in knowledge discovery and data mining. Cham: Springer International Publishing; 2018. p. 295–308.
26 Licher S, Leening MJG, Yilmaz P, Wolters FJ, Heeringa J, Bindels PJE, et al. Development and validation of a dementia risk prediction model in the general population: an analysis of three longitudinal studies. Am J Psychiatry. 2019 Jul 1;176(7):543–51.
27 Battista P, Salvatore C, Castiglioni I. Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: a machine learning study. Behav Neurol. 2017;2017:1850909.
28 O'Bryant SE, Xiao G, Barber R, Huebinger R, Wilhelmsen K, Edwards M, et al. A blood-based screening tool for Alzheimer's disease that spans serum and plasma: findings from TARC and ADNI. PLoS One. 2011;6(12):e28092.
29 Lin AP, Shic F, Enriquez C, Ross BD. Reduced glutamate neurotransmission in patients with Alzheimer's disease -- an in vivo (13)C magnetic resonance spectroscopy study. Magma. 2003;16(1):29–42.
30 Leuzy A, Cullen NC, Mattsson-Carlgren N, Hansson O. Current advances in plasma and cerebrospinal fluid biomarkers in Alzheimer's disease. Curr Opin Neurol. 2021 Apr 1;34(2):266–74.
31 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3:18.