Am J of Geriatric Psychiatry 33:5 (2025) 487–499

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ajgponline.org

Regular Research Article

Predicting Progression to Dementia Using Auditory Verbal Learning Test in Community-Dwelling Older Adults Based On Machine Learning

Xin-Yan Xie, Ph.D., Lin-Ya Huang, M.S., Dan Liu, M.D., Ph.D., Gui-Rong Cheng, M.D., Fei-Fei Hu, Ph.D., Juan Zhou, M.S., Jing-Jing Zhang, M.S., Gang-Bin Han, M.S., Jing-Wen Geng, M.S., Xiao-Chang Liu, M.S., Jun-Yi Wang, M.S., De-Yang Zeng, B.S., Jing Liu, M.S., Qian-Qian Nie, M.S., Dan Song, M.S., Shi-Yue Li, M.S., Cheng Cai, M.S., Yu-Yang Cui, M.S., Lang Xu, M.D., Ph.D., Yang-Ming Ou, M.D., Ph.D., Xing-Xing Chen, M.D., Ph.D., Yan-Ling Zhou, M.D., Ph.D., Yu-Shan Chen, M.D., Ph.D., Jin-Quan Li, M.D., Ph.D., Zhen Wei, M.D., Ph.D., Qiong Wu, M.D., Ph.D., Yu-Fei Mei, Ph.D., Shao-Jun Song, M.D., Ph.D., Wei Tan, M.D., Ph.D., Qian-Hua Zhao, M.D., Ph.D., Ding Ding, M.D., Ph.D., Yan Zeng, M.D., Ph.D.

ARTICLE INFO

Article history:
Received May, 29 2024
Revised October, 22 2024
Accepted October, 30 2024

ABSTRACT

Background: *Primary healthcare institutions find identifying individuals with dementia particularly challenging. This study aimed to develop machine learning models for identifying predictive features of older adults with normal cognition to develop dementia.* ***Methods:*** *We developed four machine learning*

From the Hubei Provincial Clinical Research Center for Alzheimer's Disease (XYX, LYH, DL, GRC, FFH, JZ, JJZ, GBH, JWG, XCL, JYW, DYZ, JL, QQN, DS, SYL, CC, YYC, LX, YMO, XXC, YLZ, YSC, JQL, ZW, QW, YFM, YZ), Tian You Hospital Affiliated to Wuhan University of Science and Technology, Wuhan; Geriatric Hospital Affiliated to Wuhan University of Science and Technology (XYX, DL, GRC, FFH, LX, YMO, XXC, YLZ, JQL, QW, YFM, WT, YZ), Wuhan; School of Public Health (XYX, DL, LX, YMO, YSC, JQL, ZW, YZ), Wuhan University of Science and Technology, Wuhan; Reproductive Medicine Center (SJS), Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan; Department of Neurology (QHZ, DD), Huashan Hospital, Fudan University, Shanghai; National Center for Neurological Disorders (QHZ, DD), Huashan Hospital, Fudan University, Shanghai; and the National Clinical Research Center for Aging and Medicine (QHZ, DD), Huashan Hospital, Fudan University, Shanghai. Send correspondence and reprint requests to Yan Zeng, Ph.D., Wuhan University of Science and Technology, West Huangjiahu Road, Hongshan District, Wuhan, 430065, China. Send correspondence and reprint requests to Ding Ding, Ph.D., Department of Neurology, Huashan Hospital, No. 12 Middle Wulumuqi Road, Shanghai, 200040, China. e-mails: dingding@huashan.org.cn zengyan68@wust.edu.cn

Joint first authors: Xin-Yan Xie and Lin-Ya Huang.

https://doi.org/10.1016/j.jagp.2024.10.016

Key Words:
Community-dwelling older adults
Dementia
Auditory verbal learning test
Prospective cohort studies
Machine learning
Predictive model

models: logistic regression, decision tree, random forest, and gradient-boosted trees, predicting dementia of 1,162 older adults with normal cognition at baseline from the Hubei Memory and Aging Cohort Study. All relevant variables collected were included in the models. The Shanghai Aging Study was selected as a replication cohort (n = 1,370) to validate the performance of models including the key features after a wrapper feature selection technique. Both cohorts adopted comparable diagnostic criteria for dementia to most previous cohort studies. ***Results:*** *The random forest model exhibited slightly better predictive power using a series of auditory verbal learning test, education, and follow-up time, as measured by overall accuracy (93%) and an area under the curve (AUC) (mean [standard error]: 088 [0.07]). When assessed in the external validation cohort, its performance was deemed acceptable with an AUC of 0.81 (0.15). Conversely, the logistic regression model showed better results in the external validation set, attaining an AUC of 0.88 (0.20).* ***Conclusion:*** *Our machine learning framework offers a viable strategy for predicting dementia using only memory tests in primary healthcare settings. This model can track cognitive changes and provide valuable insights for early intervention.* (Am J Geriatr Psychiatry 2025; 33:487−499)

Highlights

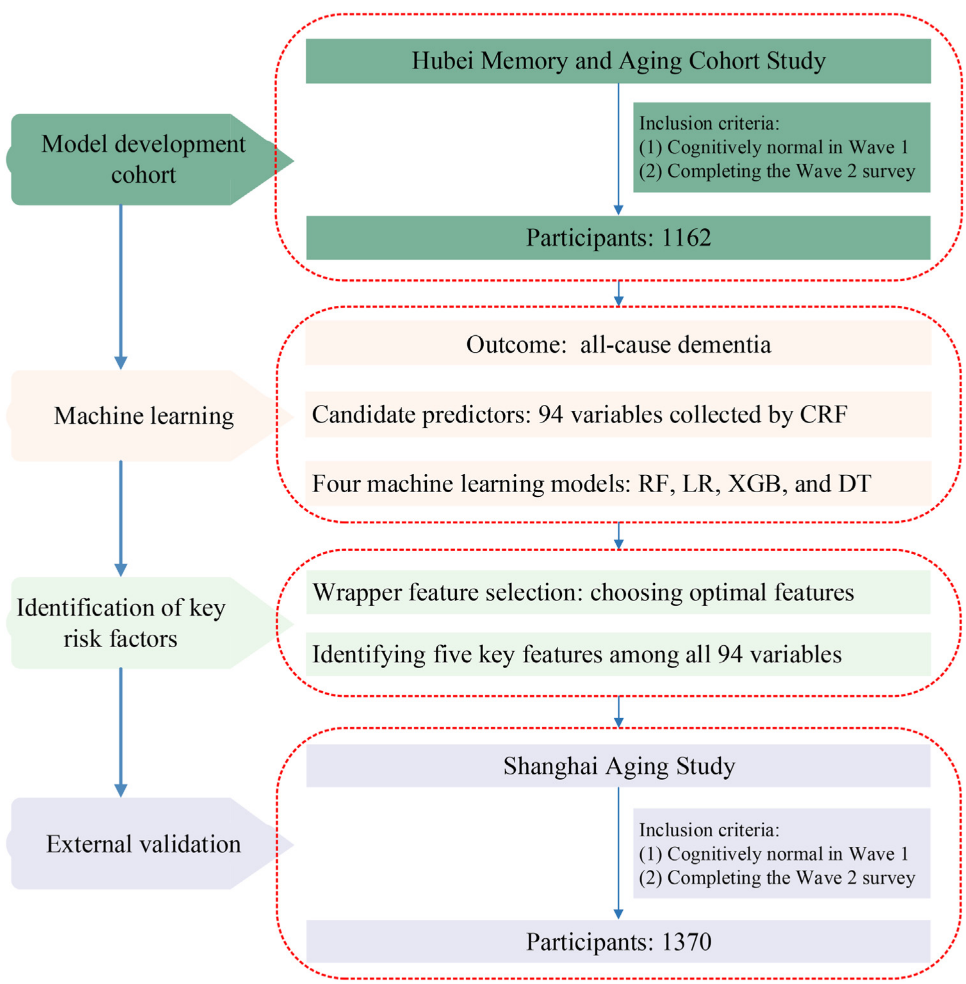
- **What is the primary question addressed by this study?**
  This study explores the predictive features of dementia among older adults.
- **What is the main finding of this study?**
  The random forest model, incorporating five key features, including age, follow years, and Auditory verbal learning test, exhibited better performance.
- **What is the meaning of the finding?**
  These findings offer a viable strategy for predicting dementia using only memory tests in primary healthcare settings.

INTRODUCTION

The increasing prevalence of dementia and its precursor, mild cognitive impairment (MCI), poses a significant public health challenge worldwide. Having the largest older population worldwide, China has more than 38.77 million individuals with MCI,[1] accounting for a quarter of the global population of patients with dementia.[2] Owing to the complex nature of its etiology and clinical heterogeneity[3] and especially to the limited diagnostic resources and a shortage of specialists,[4] the diagnostic rate of dementia remains extremely low despite the first-line position of primary healthcare institutions in early identification and diagnosis of dementia.[5] Primary healthcare institutions across China find it particularly challenging to identify individuals with dementia, compounded by the lack of evidence-based medicine and applicable guidelines specific to China.[6]

Collaborative efforts among researchers have resulted in the development of machine learning algorithms designed to predict and identify dementia. These algorithms exploit the rich information contained in dense, high-dimensional data.[7−13] In a study involving 15,307 participants, the analysis of extensive datasets containing demographic details, lifestyle patterns, psychological health, and chronic ailments indicated that machine learning algorithms could effectively forecast the onset of dementia within 2 years among patients receiving care at memory clinics, using only six variables.[13] A systematic review of 92 studies demonstrated that machine learning, particularly when leveraging neuroimaging techniques and neural networks, can equal or even surpass clinical predictions in various dementia-related tasks.[14] Although multiple studies have demonstrated the potential of machine learning in identifying those at risk earlier in the disease trajectory, several significant limitations exist. Firstly, some studies were limited by their cross-sectional design. Information collected at a single time-point may not be sufficient to establish causality between predictive factors and disease development.[15] Secondly, most machine learning models utilize brief cognitive assessments in their cognitive modules, such as the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA),[16,17] which tend to overlook specific features presented by certain cognitive subdomains. Although a longitudinal cohort study that used multidimensional cognitive assessments for machine learning model prediction revealed the significant role of vocabulary and episodic memory domains in predicting cognitive impairment,[18] this study had a small sample size (n = 253), which poses the risk of overfitting the model.[18] Thirdly, most studies relied solely on a single dataset for internal validation and lacked external validation, which may result in overfitting of the model and difficulties in generalizing the results to different populations.[19] Finally, the complexity can be a challenge when implementing machine learning models in real-world settings. It is important to ensure that these models are not only accurate but also interpretable and user-friendly for healthcare professionals, particularly in community settings.[19]

FIGURE 1. Flowchart of developing machine learning algorithms based on two prospective cohort studies in China. Abbreviations: LR, logistic regression; RF, random forest; DT, decision tree; XGB, gradient-boosted trees; CRF, case report form.



To address the limitations of existing machine learning models in predicting dementia and increase their applicability in community settings, this study aimed to develop machine learning models for identifying predictive features of older adults with normal cognition to develop dementia. Internal validation was conducted using the Hubei Memory and Aging Cohort Study (HMACS), and external validation using the Shanghai Aging Study (SAS) (Fig. 1). Both cohorts have comparable study designs, procedures, and diagnostic criteria dementia to most previous cohort studies.[20,21] Four widely used and explanatory machine learning models in predicting cognitive decline were employed in this study.[19,22,23] By identifying individuals at a high risk of dementia, interventions can be targeted to those who need mostly, potentially preventing or delaying the onset of cognitive decline.

METHODS

Study Sample of Model Development

We utilized data previously collected from the HMACS database. The HMACS was designed as a community-based cohort study that included urban (31 neighborhoods) and rural (48 villages) settings. Based on the electronic health records maintained in communities and health centers, all eligible geriatric residents living within the sampled neighborhoods and villages were invited to complete cognitive tests and clinical assessments in the first wave of the survey (Wave 1, 2018−2021). The HMACS collects data from all participants every 2−3 years (Wave 2, 2022−2023) and investigates how social, family, behavioral, economic, and environmental factors affect aging and cognition. Based on the results of the study visit, an expert consensus panel identified either normal cognition, MCI, or dementia as the clinical diagnosis. In the current study, participants from the HMACS were included if they met the following criteria: (1) provided informed consent or witnessed oral consent; (2) demonstrated cognitive normality in Wave 1 and completed the Wave 2 survey; (3) neither had severe hearing loss nor difficulties understanding spoken or written Chinese; (4) did not have physician-diagnosed dementia, mental disorders, or intellectual disability; and (5) were able to communicate and participate in the cognitive examinations. A total of 1,351 individuals with an average follow-up duration of approximately 2 years were included. Details of the protocol and associated guidelines have been previously published.[24]

This study was approved by the Medical Ethics Committee of the Wuhan University of Science and Technology granted the study ethical approval (protocol code: 201845). Before the commencement of the study, all participants provided written or oral consent. No information identifying any individual was disclosed. All procedures involving human participants complied with the 1964 Declaration of Helsinki and its subsequent guidelines.

Outcome

The outcome of interest was the incident all-cause dementia within approximately 2 years of baseline assessment in the HMACS cohort. There was insufficient information to categorize the population into different types of dementia based on etiology. Diagnosis of dementia established according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria,[25] which includes prior normal cognitive function, decline in cognitive abilities or abnormal mental behavior, decline in cognitive abilities affecting work or daily life, and the absence of any other psychiatric disorder or delirium that could explain cognitive decline.

Candidate Predictors of Model Development

We included all relevant variables obtained during the initial visit using a standardized case report form in the HMACS cohort (Table S1). Variables containing free text values, such as medication names, were excluded, as were variables that remained constant across all participants, such as the visit number. Consequently, we identified a total of 94 variables, encompassing 9 variables related to demographic characteristics, 30 variables related to behavior and lifestyle, 5 variables related to physical examination, 27 variables related to medical history, 6 variables related to peripheral organ function assessment, 2 variables related to psychological status, and 15 variables derived from the neuropsychological test battery.

Participants enrolled in the program underwent comprehensive clinical interviews with neurologists who used a neuropsychological battery. This battery was used to evaluate various cognitive domains, including global cognition, executive functioning/attention, language, visuospatial functioning, and memory. The cognitive assessment tools employed included the MMSE, Montreal Cognitive Assessment-basic (MoCA-B), Verbal Fluency Test (VFT), Auditory Verbal Learning Test- Huashan Version (AVLT), Clock Drawing Test, Digit Span Test (DST), Trail Making Test-A/B, Boston Naming Test (BNT), and Subjective Cognitive Decline Scale. The AVLT measures short-term memory (trials 1−3), delayed recall (trials 4 and 5) at 5 and 20 min, category-cued recall (trial 6), and recognition (trial 7). Further details are provided in Table S1.

Model Development in the HMACS Cohort

We implemented four machine learning algorithms: logistic regression (LR),[26] decision tree (DT),[27] random forest (RF),[28] and gradient-boosted trees (XGB)[29] (eMethods in the Supplement). These models were selected based on their distinct attributes: LR for its interpretability, DT for its visual and rule-based decision-making, RF for its robustness and accuracy through ensemble methods, and XGB for its high performance and ability to model complex interactions. This diverse selection aims to leverage both interpretability and predictive power to enhance our model's efficacy and reliability. Moreover, these methods have been widely used and demonstrated good performance in previous studies, further supporting their suitability for our research.[19,22,23] These algorithms performed a classification task: determining whether a participant belonged to class 0 (predicted to retain normal cognition 2 years from Wave 1) or class 1 (predicted to experience incident dementia during follow-up on the variables recorded at the first visit). To implement the machine learning algorithms, we used the Python scikit-learn library (Python Software Foundation) with 10-fold cross-validation.

Identification of Key Features

The utilization of a machine learning approach may present a potential limitation due to the high number of variables required. With an increasing number of variables, the implementation of the model in a clinical setting may become less feasible, and the interpretability of the model may be compromised. Thus, a wrapper feature selection technique was used to identify the key features among all the variables from the HMACS cohort, which could improve model performance and reduce complexity and computational costs.[30] It can also improve the accuracy, reduce the overfitting, speed up training, improve data visualization, and increase the explainability of the model. Further details regarding the wrapper approach are provided in the eMethods section of the Supplement.

External Validation

To validate our results, we selected the SAS as the replication cohort. In the external validation set, we validated the machine learning models containing the key features after the wrapper feature selection technique. The SAS cohort was established to prospectively investigate cognitive impairment among individuals aged ≥ 60 years residing in downtown Shanghai, China. All participants underwent extensive epidemiological, neurological, and neuropsychological assessments, with consensus diagnostic criteria applied uniformly across all participants.[20] The inclusion criteria for the study population included in the current study were aligned with the standards set for HMACS. Finally, 1,370 individuals who were cognitively normal in Wave 1 and completed the Wave 2 survey were included in the external validation dataset.

Statistical Analysis

Continuous variables are presented as medians and interquartile ranges (IQR) due to their skewed distribution, as assessed by histograms. Categorical variables are presented as frequencies and percentages. The distribution of variables between the two groups was compared using the Mann-Whitney *U* test for continuous variables and the chi-square test for categorical variables. The missing data rate for all variables was less than 20%, with the exception of the DST, Trail Making Test, BNT, CDT, and AVLT tests in the HMACS cohort, which exhibited a missing rate of approximately 50%. Before imputing the missing data, we considered the pattern of missingness in the data. Firstly, we generated a correlation matrix of the variables with missing data and found that many variables were significantly correlated, which ruled out the possibility of missing completely at random. We then fitted a logistic regression model for each field's missing data indicator against each of the other fields in turn, examining the p-values. P-values were adjusted using a Šidák correction, replacing the threshold α with 1-(1-α)1/#tests (8.307×10^{-6}). Our analysis indicated that the missing values of the variables did not exhibit any correlation with other variables, thereby confirming that the missing data in our study were indeed missing at random (MAR). Machine-learning methods capable of managing multivariate inputs have proven effective across various data processing domains,[31–34] especially in situations marked by extreme missingness (greater than 90%),[35] such as Inverse Distance Weighting, Support Vector Regressor, and Random Forest Regressor (RFR). In our research, we utilized the RFR from the Python library sklearn, ensemble to address missing values. This selection was based on its superior performance relative to other methods, assessed using the Nash-Sutcliffe efficiency metric. The static imputation process involved several steps, including data pre-processing, profiling, analyzing variable correlations, and imputation. Categorical variables are represented as numerical values in the machine learning models using dummy coding. Continuous variables were standardized using z-scores. The performance of the machine learning models was evaluated based on their overall accuracy, sensitivity, and specificity, with equal weighting given to false-positive and false-negative errors. The positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUC) were utilized to summarize the model performance, and the mean performance measures and standard errors (SE) were also obtained. All data processing and analyses were conducted using R, version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria), and Python, version 3.9 (Python Software Foundation, Wilmington, DE, USA).

RESULTS

Baseline Characteristic in the HMACS Cohort

The final analytic sample comprised 1,162 participants (median [IQR] age, 70.00 [67.00, 74.00] years; 610 [52.5%] women and 552 [47.5%] men) in the HMACS cohort. The sample characteristics are listed in Table 1. Within a span of approximately 2 years from baseline, dementia was observed in 110 participants, accounting for 9.5% of the total sample. Compared to individuals who maintained normal cognitive function, those who developed dementia in Wave 2 were characterized by older age, a higher proportion of females, lower educational attainment, a higher representation of physically demanding occupations, a higher proportion of unmarried individuals, lower monthly income with unstable earnings, and lower levels of physical and intellectual activities (Table 1) in Wave 1. In addition, this group of individuals also exhibited poorer performance on neuropsychological tests, including MMSE (Mann-Whitney *U* test, Z = -8.152, p <0.001) and MoCA-B (Mann-Whitney *U* test, Z = -5.998, p <0.001) for assessing global cognition and VFT (Mann-Whitney *U* test, Z = -3.120, p = 0.002) for language domain.

Performance of Machine Learning Models

Table 2 shows the model discrimination results for the predicting of dementia, including overall accuracy, sensitivity, specificity, PPV, NPV, and AUC including a total of 94 variables from the HMACS cohort. RF model demonstrated an overall accuracy of 92% and an AUC of 0.83 (mean [SE]: 0.07). In comparison, the performance of DT and LR models was slightly inferior, with LR achieving an overall accuracy of 88% and an AUC of 0.80 [0.07], while DT recorded an accuracy of 83% and an AUC of 0.82 [0.04]. XGB model exhibited the weakest performance, with an AUC value of 0.86 (infinity). The receiver operating characteristic curve for each machine learning model demonstrated similarities (Figure 2a).

Identification of Key Features

The wrapper feature selection technique was used to identify key features among all 94 variables[30] (Table S1). Figure 3 shows the number of optimal feature sequences obtained after the sequential feature selection. Given the generally higher AUC values of the RF model compared to the others and the highest value observed in the RF model in Figure 3, the feature variables that

TABLE 1. Demographic Characteristics of Study Participants from HMACS in Wave 1.

Variables	Cognitively Normal (n = 1,052)	All-Cause Dementia (n = 110)	χ^2/Z	p-Value
Demographic characteristics				
Age, median (IQR)	70.00 (67.00, 73.00)	73.00 (69.00, 78.00)	−5.884	**< 0.001**
Sex, male, n (%)	527 (50.1)	25 (22.7)	29.911	**< 0.001**
Years of education, median (IQR)	7.00 (0.00, 12.00)	0.00 (0.00, 0.00)	−9.498	**< 0.001**
BMI, median (IQR)	23.88 (21.88, 25.96)	22.98 (20.89, 26.04)	−1.504	0.133
Job before retirement, n (%)			49.549	**< 0.001**
White collar worker	259 (27.1)	3 (3.0)		
Blue collar worker	593 (62.0)	98 (97.0)		
Others	104 (10.9)	0 (0.0)		
Marriage, married, n (%)	750 (71.8)	67 (61.5)	5.141	**0.023**
Stable income, yes, n (%)	875 (90.5)	78 (77.2)	16.730	**< 0.001**
Monthly income, ≤ 3000 CNY, n (%)	538 (58.5)	80 (90.9)	35.612	**< 0.001**
Disease history				
Hypertension, yes, n (%)	544 (52.0)	65 (59.1)	2.003	0.157
Diabetes, yes, n (%)	200 (19.2)	18 (16.4)	0.519	0.471
Behavior and lifestyle				
Alcohol consumption, yes, n (%)	350 (34.6)	25 (23.8)	4.982	**0.026**
Smoking, yes, n (%)	372 (35.6)	18 (16.7)	15.625	**< 0.001**
Tea consumption, yes, n (%)	621 (64.2)	70 (68.6)	0.784	0.376
Physical activity, yes, n (%)	804 (76.6)	62 (56.4)	21.679	**< 0.001**
Intellectual activity, yes, n (%)	604 (60.0)	20 (19.2)	63.578	**< 0.001**
Living alone, yes, n (%)	143 (14.9)	22 (20.4)	2.257	0.133
Neuropsychological tests, median (IQR)				
MMSE	29.00 (26.00, 30.00)	25.00 (22.00, 27.00)	−8.152	**< 0.001**
MoCA-B	25.00 (21.00, 27.00)	20.50 (18.00, 25.00)	−5.998	**< 0.001**
VFT	15.00 (12.00, 18.00)	13.00 (9.00, 16.00)	−3.120	**0.002**
AVLT trail 1	3.00 (2.00, 5.00)	3.00 (2.00, 3.50)	−1.503	0.133
AVLT trail 2	4.50 (3.00, 6.00)	5.00 (3.50, 6.00)	−0.143	0.886
AVLT trail 3	5.00 (4.00, 7.00)	5.00 (4.00, 7.00)	−0.176	0.860
AVLT trail 4	5.00 (3.00, 7.00)	5.00 (3.25, 6.75)	−0.363	0.717
AVLT trail 5	4.00 (3.00, 7.00)	5.00 (3.00, 6.00)	−0.308	0.758
AVLT trail 6	4.00 (2.00, 7.00)	4.00 (0.00, 6.00)	−0.872	0.383
AVLT trail 7	21.00 (18.00, 22.00)	21.00 (19.00, 22.25)	−0.561	0.575
CDT	23.00 (11.00, 26.00)	23.00 (19.00, 26.00)	−0.839	0.402
DST	14.00 (11.00, 17.00)	14.00 (10.00, 17.00)	−0.030	0.976
BNT	22.00 (18.00, 26.00)	21.00 (17.00, 25.00)	−1.123	0.261
TMT-A, seconds	84.00 (62.00, 110.00)	83.00 (60.00, 133.00)	−0.590	0.555
TMT-B, seconds	190.00 (148.00, 240.00)	193.50 (139.50, 252.00)	−0.033	0.973
Depression symptoms[a], yes, n (%)	39 (4.2)	5 (5.0)	0.145	0.703
ADL, median (IQR)	20.00 (20.00, 21.00)	21.00 (20.00, 23.00)	−5.763	**< 0.001**

[a] Depression symptoms were evaluated by Geriatric depression scale-15.

The distribution of variables between the two groups was compared using the Mann-Whitney *U* test (Z) for continuous variables and the chi-square test (χ^2) for categorical variables.

The chi-square degree of freedom (df) was 1 for all variables except for Job before retirement (df = 2).

HMACS: Hubei memory and aging cohort study; BMI: body mass index; MMSE: mini-mental state examination; MoCA-B: Montreal cognitive assessment basic; VFT: verbal fluency test; AVLT: auditory verbal learning test; CDT: clock drawing test; DST: digit span test; BNT: Boston naming test; TMT-A: trail making test-A; TMT-B: trail making test-B; ADL: activities of daily living; IQR: interquartile range.

yielded the highest AUC value for the RF model were selected. Subsets included education, follow-up time, AVLT trial 1, AVLT trial 3, and AVLT trial 6. The selected features yielding the classification performance are listed in Table 2, and a visualization of the receiver operating characteristic curves is shown in Figure 2b. All machine learning models incorporating the five key features demonstrated slightly improvement of AUC values compared with the basic machine learning models. The AUC for the RF, LR, DT, and XGB models were 0.88 (0.07), 0.81 (0.11), 0.86 (0.04), and 0.89 (0.21), respectively.

External Validation

The validation cohort comprised 1,370 participants (median [IQR] age, 69.85 [64.27, 75.72] years; 736

TABLE 2. Performance Measures for the Prediction of Incident Cognitive Impairment.

	Performance Measures, Mean (SE)					
	Overall Accuracy	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Area Under the Curve
HMACS, including all predictors						
LR	0.88 (0.02)	0.18 (0.10)	0.95 (0.02)	0.27 (0.03)	0.92 (0.02)	0.80 (0.07)
DT	0.83 (0.03)	0.55 (0.08)	0.86 (0.03)	0.29 (0.03)	0.95 (0.02)	0.82 (0.04)
RF	0.92 (0.01)	0.36 (0.13)	0.98 (0.01)	0.62 (0.03)	0.94 (0.02)	0.83 (0.07)
XGB	0.91 (0.01)	0.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.91 (0.02)	0.86 (Inf)
HMACS, five key features after wrapper selection						
LR	0.91(0.01)	0.13 (0.14)	0.99 (0.01)	0.50 (0.03)	0.92 (0.02)	0.81 (0.11)
DT	0.82(0.03)	0.68 (0.07)	0.84 (0.03)	0.30 (0.03)	0.96 (0.01)	0.86 (0.04)
RF	0.93(0.01)	0.36 (0.15)	0.99 (0.01)	0.73 (0.03)	0.94 (0.02)	0.88 (0.07)
XGB	0.91 (0.01)	0.05 (0.20)	1.00 (0.00)	1.00 (0.00)	0.90 (0.02)	0.89 (0.21)
SAS, external validation						
LR	0.97 (0.01)	0.10 (0.30)	1.00 (0.00)	1.00 (0.00)	0.97 (0.01)	0.88 (0.20)
DT	0.89 (0.03)	0.50 (0.09)	0.91 (0.02)	0.17 (0.02)	0.98 (0.01)	0.80 (0.04)
RF	0.97 (0.01)	0.20 (0.23)	0.99 (0.01)	0.67 (0.03)	0.97 (0.01)	0.81 (0.15)
XGB	0.96 (0.01)	0.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.96 (0.01)	0.80 (Inf)

HMACS: Hubei memory and aging cohort study; SAS: Shanghai aging study; LR: logistic regression; RF: random forest; DT: decision tree; XGB: gradient-boosted trees; SE: standard error; Inf: infinity.

[53.7%] women and 634 [46.3%] men) in the SAS cohort. The characteristics of the samples are listed in Table S2. Within a span of 5.3 years from baseline, dementia was observed in 52 participants, accounting for 3.8% of the total sample. Compared to individuals with normal cognitive function, those who experienced dementia in Wave 2 were older (Mann-Whitney U test, $Z = -7.285$, $p < 0.001$) and had lower educational attainment (Mann-Whitney U test, $Z = -4.748$, $p < 0.001$) in Wave 1. In addition, this group of individuals also exhibited poorer performance on neuropsychological tests, including the MMSE for assessing global cognition and AVLT trials 1−7 for the memory domain (Table S2).

Table 2 presents the performance of the model evaluated in the SAS cohort including the five key features after the wrapper feature selection technique. Among the models evaluated during internal validation, the RF model, which showed slightly better performance, also slightly outperformed the DT and

FIGURE 2. Receiver operating characteristic curves. [A] Four machine learning models include all candidate predictors in HMACS; [B] four machine learning models incorporating five key features identified by wrapper feature selection in HMACS. Abbreviations: LR, logistic regression; RF, random forest; DT, decision tree; XGB, gradient-boosted trees. HMACS: Hubei Memory and Aging Cohort Study.

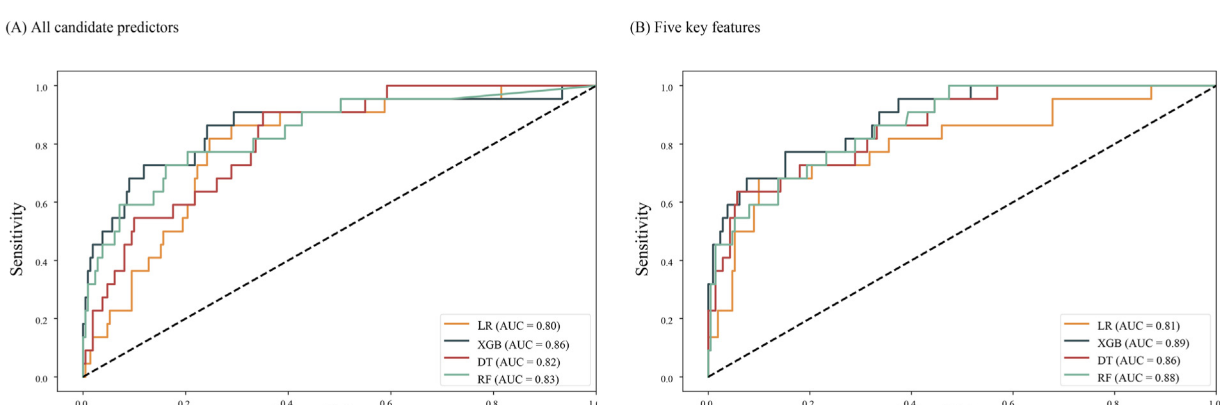
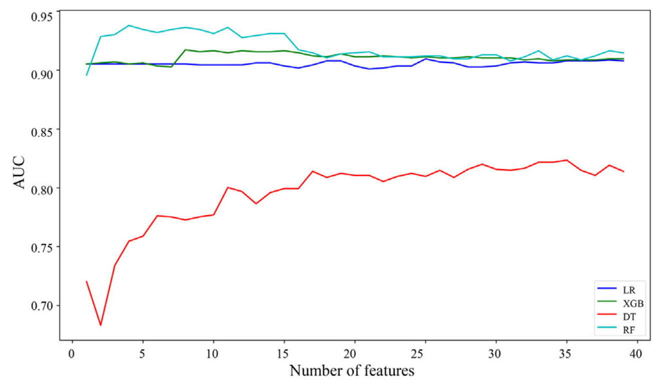


FIGURE 3. Area under the curve (AUC) versus number of features used for training the four machine learning models. Abbreviations: LR, logistic regression; RF, random forest; DT, decision tree; XGB, gradient-boosted trees.



XGB models in the external validation cohort. In addition, the LR model performed better in the external validation, with an AUC value of 0.88 (0.20).

DISCUSSION

Based on two prospective cohort studies conducted in China, we devised machine learning algorithms to forecast the advancement of dementia in older adults living in communities with normal cognition. Internal and external validations were performed. Our results indicate that the RF model, incorporating five key features, including education, follow-up time, and AVLT (trials 1, 3, and 6), exhibited better performance of accuracy and AUC values. The performances of the XGB, DT, and LR models were slightly inferior. Our model can be integrated into community primary health systems, utilizing health examination data for individuals aged 65 and above to screen high-risk populations for cognitive impairment and implement preventive interventions.[36] By predicting with an effective model, timely interventions can be applied before significant cognitive decline occurs. Such interventions, including enhancing health education and promoting healthy lifestyle changes, could have a substantial impact, as delaying the onset by just one year could prevent over 9 million dementia cases by 2050.[37]

Prior studies on machine learning to predict the risk of cognitive impairment focused on high-dimensional data, including positron emission tomography scans, cerebrospinal fluid biomarkers neuroimaging, neurophysiology, genetics, and proteomics,[38–40] which are not commonly available in community settings. Our machine learning models complement these analyses and have the advantage of incorporating only five key features to predict the outcome of dementia. Most of these key features are derived from the AVLT, a widely used tool in standard memory tests known for its high sensitivity and specificity in identifying MCI and dementia.[41] Neuroimaging studies have linked the immediate, short-term, and delayed recall components of the AVLT to the structural integrity of the hippocampus.[42] Compared with the MMSE and MoCA, the AVLT is more sensitive to cognitive decline,[43] consistent with our findings. Among memory dimensions, delayed recall is considered the most sensitive predictive factor for cognitive impairment,[43,44] and both short- and long-term delayed recall are equally valuable.[45] Immediate recall acts as a determinant of delayed recall,[46] while cue recall is impaired in the process of cognitive impairment.[47] Our study streamlined the conventional neuropsychological battery by selecting the most representative items (AVLT trials), significantly reducing the time and manpower costs. Thus, our model may facilitate large-scale cognitive screening in communities, providing a convenient method for community healthcare workers to identify high-risk individuals at an early stage.

Previous machine learning algorithms have already been used in population surveys to identify persons with cognitive impairment but often relied solely on internal validation.[19,48] Our model was derived from follow-up data from the HMACS and externally validated using longitudinal data from an independent SAS cohort. The two cohorts, originating from southeast and midwest China, were rigorously designed with consistent and reliable diagnostic criteria.[20,21]

Compared to most reported machine learning algorithms for predicting cognitive impairment (Table S3), our models demonstrated slightly better performance. Two large-scale studies, one using Taiwan's National Health Insurance Research Database and the other using general practice patient records from the UK, utilized machine learning models to predict dementia and achieved AUC values of 0.74 and 0.63, respectively.[8,49] In addition, owing to limitations in electronic medical records, these studies did not incorporate comprehensive cognitive assessments as predictive variables. Based on the 4-year follow-up data from the China Health and Retirement Longitudinal Study cohort, a cognitive impairment prediction model exhibited an AUC of 0.79.[50] The model included only MMSE scores and did not incorporate specific cognitive domains as predictor variables. A study by James et al.[13] overcame this difficulty by utilizing the US National Alzheimer Coordinating Center to construct predictive models with excellent performance (AUC: 0.91−0.92) using machine learning algorithms. Although this study used a large sample from the United States, no external validation was conducted, making generalizing the model and expanding its applicability challenging. Our model was externally validated using an independent domestic cohort, demonstrating its robust applicability within similar cultural and linguistic contexts.

Limitations

Firstly, the machine learning models used in our study and their external validation were derived from a population of the same ethnicity. Caution should be exercised when extrapolating the model to other ethnic groups. However, at each sampling site (neighborhoods and villages), general population sampling was conducted to obtain a representative cohort. Moreover, we ensured high data quality by imposing strict inclusion criteria based on sample characteristics. All field interviewers received comprehensive training, and stringent quality control measures were implemented throughout the study to improve data the reliability and validity. Secondly, our study did not incorporate imaging and biomarker data for identifying dementia. Although these data could enhance the precision of the model, they are not easily accessible in primary healthcare institutions. Our objective is to establish predictive models more suitable for large-scale community screening. Thirdly, this study focused solely on all-cause dementia, without differentiating between various dementia types based on their etiology. This lack of classification restricts the model's ability to account for the heterogeneity among dementia subtypes, potentially affecting the generalizability and specificity of the findings. Future research should consider incorporating etiological distinctions to enhance the model's accuracy and applicability. Finally, the presence of missing values poses a limitation to the study, potentially affecting the robustness of the findings. While we employed advanced imputation techniques such as the RFR, the high missing rate in certain variables may still introduce bias and limit the generalizability of our results. Additionally, RFR is considered a static imputation method. While it can be a quick and easy solution for handling MAR data, it is essential to recognize its disadvantages. Future research should aim to minimize missing data through improved data collection methods and consider more sophisticated imputation techniques to enhance the validity of conclusions drawn from the analysis.

An important issue to consider was that although the SAS cohort had a longer follow-up period than the HMACS cohort, it exhibited a lower prevalence of dementia. This discrepancy may be attributed to the differences in cohort composition: the HMACS cohort included both urban and rural populations, while the SAS cohort consisted solely of urban residents. Variations in lifestyle, access to healthcare, and environmental factors between urban and rural settings could significantly influence cognitive health and its progression. Rural populations might experience distinct risk and protective factors compared to urban populations, potentially affecting the rate of cognitive decline and the performance of predictive models.[51] Additionally, the urban residents in the SAS cohort might benefit from improved access to healthcare and cognitive interventions, which could contribute to a slower progression of cognitive decline in this cohort.[52] Furthermore, although the diagnostic methods for dementia were consistent across both cohorts, differences in the execution or reporting of cognitive assessments could still impact observed rates of dementia. These differences highlight the importance of considering sample characteristics when interpreting study results. Therefore, caution is needed when generalizing predictive models across diverse populations. Future research should focus on accounting for these variations to enhance the reliability of cognitive decline predictions.

CONCLUSION

Our machine learning framework achieved high accuracy in predicting dementia through community screenings and was validated using an independent cohort. This model could provide valuable insights for early intervention through accessible information in community settings. To explore more effective models for reducing the incidence of dementia, it is essential to include populations with extended follow-up periods to investigate machine learning models with greater accuracy. This, in turn, can help advance the timing of preventive interventions.

AUTHOR CONTRIBUTIONS

Xin-Yan Xie: Writing—original draft, Methodology, Formal analysis, Writing—review and editing, Funding acquisition. Lin-Ya Huang: Writing—original draft, Investigation, Writing—review and editing. Dan Liu: Methodology, Writing—review and editing, Project administration. Gui-Rong Cheng: Data curation, Project administration, Writing—review and editing. Fei-Fei Hu: Methodology, Visualization, Writing—review and editing. Juan Zhou: Data curation, Investigation, Writing—review and editing. Jing-Jing Zhang: Investigation, Writing—review and editing. Gang-Bin Han: Investigation, Writing—review and editing. Jing-Wen Geng: Investigation, Writing—review and editing. Xiao-Chang Liu: Investigation, Writing—review and editing. Jun-Yi Wang: Investigation, Writing—review and editing. De-Yang Zeng: Investigation, Writing—review and editing. Jing Liu: Investigation, Writing—review and editing. Qian-Qian Nie: Investigation, Writing—review and editing. Dan Song: Investigation, Writing—review and editing. Shi-Yue Li: Investigation, Writing—review and editing. Cheng Cai: Investigation, Writing—review and editing. Yu-Yang Cui: Investigation, Writing—review and editing. Lang Xu: Supervision, Funding acquisition, Writing—review and editing. Yang-Ming Ou: Supervision, Writing—review and editing. Xing-Xing Chen: Supervision, Writing—review and editing. Yan-Ling Zhou: Supervision, Writing—review and editing. Yu-Shan Chen: Supervision, Writing—review and editing. Jin-Quan Li: Supervision, Writing—review and editing. Zhen Wei: Supervision, Writing—review and editing. Qiong Wu: Supervision, Writing—review and editing. Yu-Fei Mei: Project administration, Writing—review and editing. Shao-Jun Song: Project administration, Formal analysis, Writing—review and editing. Wei Tan: Project administration, Writing—review and editing. Qian-Hua Zhao: Project administration, Funding acquisition, Writing—review and editing. Ding Ding: Conceptualization, Writing—review and editing, Data curation. Yan Zeng: Conceptualization, Writing—review and editing, Funding acquisition, Data curation.

DISCLOSURES

The authors declare no conflicts of interest related to this paper. This work was supported by grants from Science and Technology Innovation 2030 Major Projects (2022ZD0211600), National Natural Science Foundation of China (82371444, 82304172, 82473701, 82071200, 82071272, 82171491 and 72174159), and National College Students Innovation and Entrepreneurship Training Program (202310488042). The financial contributors had no role in the design, analysis, or writing of this article.

DATA STATEMENT

The data has not been previously presented orally or by poster at scientific meetings.

ACKNOWLEDGMENTS

We thank the HMACS research, SAS research, and the study's respondents for their contributions to this work.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.jagp.2024.10.016.

References

1. Jia L, Du Y, Chu L, et al: Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. Lancet Public Health 2020; 5:e661–e671
2. GBD 2016 Neurology Collaborators: Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol 2019; 18:459–480
3. Jack CR, Bennett DA, Blennow K, et al: NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia 2018; 14:535–562
4. Liu L, Liu S, Zhang L, et al: Cascaded multi-modal mixing transformers for Alzheimer's disease classification with incomplete data. NeuroImage 2023; 277:120267
5. Wang H, Xie H, Qu Q, et al: The continuum of care for dementia: needs, resources and practice in. China J Glob Health 2019; 9:020321
6. Ren R, Qi J, Lin S, et al: The China Alzheimer report 2022. Gen Psych 2022; 35:e100751
7. Hu M, Shu X, Yu G, et al: A risk prediction model based on machine learning for cognitive impairment among Chinese community-dwelling elderly people with normal cognition: development and validation study. J Med Internet Res 2021; 23:e20298
8. Ford E, Rooney P, Oliver S, et al: Identifying undetected dementia in UK primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. BMC Med Inform Decis Mak 2019; 19:248
9. Wang W, Peng J, Hou J, et al: Predicting mild cognitive impairment progression to Alzheimer's disease based on machine learning analysis of cortical morphological features. Aging Clin Exp Res 2023; 35:1721–1730
10. Noh B, Yoon H, Youm C, et al: Prediction of decline in global cognitive function using machine learning with feature ranking of gait and physical fitness outcomes in older adults. IJERPH 2021; 18:11347
11. Pereiro AX, Valladares-Rodríguez S, Felpete A, et al: Relevance of complaint severity in predicting the progression of subjective cognitive decline and mild cognitive impairment: a machine learning approach. JAD 2021; 82:1229–1242
12. Liu H, Zhang X, Liu H, et al: Using machine learning to predict cognitive impairment among middle-aged and older Chinese: a longitudinal study. Int J Public Health 2023; 68:1605322
13. James C, Ranson JM, Everson R, et al: Performance of machine learning algorithms for predicting progression to dementia in memory clinic patients. JAMA Netw Open 2021; 4:e2136553
14. Martin SA, Townend FJ, Barkhof F, et al: Interpretable machine learning for dementia: a systematic review. Alzheimer's & Dementia 2023; 19:2135–2149
15. Aschwanden D, Aichele S, Ghisletta P, et al: Predicting cognitive impairment and dementia: a machine learning approach. J Alzheimers Dis 2020; 75:717–728
16. Noh B, Yoon H, Youm C, et al: Prediction of decline in global cognitive function using machine learning with feature ranking of gait and physical fitness outcomes in older adults. IJERPH 2021; 18:11347
17. Wang M, Greenberg M, Forkert ND, et al: Dementia risk prediction in individuals with mild cognitive impairment: a comparison of Cox regression and machine learning models. BMC Med Res Methodol 2022; 22:284
18. Pereiro AX, Valladares-Rodríguez S, Felpete A, et al: Relevance of complaint severity in predicting the progression of subjective cognitive decline and mild cognitive impairment: a machine learning approach. JAD 2021; 82:1229–1242
19. Hu M, Shu X, Yu G, et al: A risk prediction model based on machine learning for cognitive impairment among Chinese community-dwelling elderly people with normal cognition: development and validation study. J Med Internet Res 2021; 23:e20298
20. Ding D, Zhao Q, Guo Q, et al: The Shanghai Aging Study: study design, baseline characteristics, and prevalence of dementia. Neuroepidemiology 2014; 43:114–122
21. Li L, Cheng G-R, Liu D, et al: The Hubei memory and Aging Cohort Study: study design, baseline characteristics, and prevalence of cognitive impairments. J Alzheimers Dis 2022; 85:561–571
22. Zhou Y, Han W, Yao X, et al: Developing a machine learning model for detecting depression, anxiety, and apathy in older adults with mild cognitive impairment using speech and facial expressions: a cross-sectional observational study. Int J Nurs Stud 2023; 146:104562
23. Hsiu H, Lin S-K, Weng W-L, et al: Discrimination of the cognitive function of community subjects using the arterial pulse spectrum and machine-learning analysis. Sensors (Basel) 2022; 22:806
24. Cheng G, Liu D, Huang L, et al: Prevalence and risk factors for subjective cognitive decline and the correlation with objective cognition among community-dwelling older adults in China: results from the Hubei memory and aging cohort study. Alzheimer's & Dementia 2023; 19:5074–5085
25. Roehr B: American psychiatric association explains DSM-5. BMJ 2013; 346:f3591
26. Jr DWH, Lemeshow S, RX S: Applied Logistic Regression. Third Edition Hoboken, New Jersey: John Wiley & Sons, Inc., 2013
27. Suthaharan S: Decision tree learning. Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning. Boston, MA: Springer US, 2016:237–269
28. Breiman L: Random Forests. Mach Learn 2001; 45:5–32
29. Friedman JH: Stochastic gradient boosting. Computa Stat Data Analy 2002; 38:367–378
30. Tekin Erguzel T, Tas C, Cebi M: A wrapper-based approach for feature selection and classification of major depressive disorder-bipolar disorders. Comput Biol Med 2015; 64:127–137
31. Zhang X, Yan C, Gao C, et al: Predicting missing values in medical data via XGBoost regression. J Healthc Inform Res 2020; 4:383–394
32. Babar B, Luppino LT, Boström T, et al: Random forest regression for improved mapping of solar irradiance at high latitudes. Solar Energy 2020; 198:81–92
33. Rodríguez R, Pastorini M, Etcheverry L, et al: Water-quality data imputation with a high percentage of missing values: a machine learning approach. Sustainability 2021; 13:6318
34. Petrazzini BO, Naya H, Lopez-Bello F, et al: Evaluation of different approaches for missing data imputation on features associated to genomic data. BioData Mining 2021; 14:44
35. Aguilera H, Guardiola-Albert C, Serrano-Hidalgo C: Estimating extremely large amounts of missing precipitation data. J Hydroinformatics 2020; 22:578–592
36. Tang D, Wang J: Basic public health service utilization by internal older adult migrants in China. Int J Environ Res Public Health 2021; 18:270
37. Frankish H, Horton R: Prevention and management of dementia: a priority for public health. Lancet (London, England) 2017; 390:2614–2615

38. Rutkowski TM, Abe MS, Sugimoto H, et al: Mild cognitive impairment detection with machine learning and topological data analysis applied to EEG time-series in facial emotion oddball paradigm. Annu Int Conf IEEE Eng Med Biol Soc 2023; 2023:1–4
39. Martí-Juan G, Sanroma-Guell G, Piella G: A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease. Comput Meth Prog Biomed 2020; 189:105348
40. Haran JP, Bhattarai SK, Foley SE, et al: Alzheimer's Disease microbiome is associated with dysregulation of the anti-inflammatory P-glycoprotein pathway. mBio 2019; 10, e00632-19
41. Callahan CD, Johnstone B: The clinical utility of the rey auditory-verbal learning test in medical rehabilitation. J Clin Psychol Med Settings 1994; 1:261–268
42. Wolk DA, Dickerson BC: Alzheimer's Disease Neuroimaging Initiative: fractionating verbal episodic memory in Alzheimer's disease. Neuroimage 2011; 54:1530–1539
43. Bao X, Li W, Liu Y, et al: Impairment of delayed recall as a predictor of amnestic mild cognitive impairment development in normal older adults: a 7-year of longitudinal cohort study in Shanghai. BMC Psychiatry 2023; 23:892
44. Vyhnalek M, Jester DJ, Andel R, et al: Contribution of memory tests to early identification of conversion from amnestic mild cognitive impairment to. Dementia J Alzheimers Dis 2022; 88:1397–1409
45. Xu Y, Chen K, Zhao Q, et al: Short-term delayed recall of auditory verbal learning test provides equivalent value to long-term delayed recall in predicting MCI clinical outcomes: a longitudinal follow-up study. Appl Neuropsychol Adult 2020; 27:73–81
46. Venkatesan UM, Margolis SA, Tremont G, et al: Forward to the past: revisiting the role of immediate recognition in the assessment of episodic memory. J Clin Exp Neuropsychol 2020; 42:160–170
47. Browning CA, Thompson CL, Kochan NA, et al: Prospective memory function predicts future cognitive decline and incident dementia. J Gerontol B Psychol Sci Soc Sci 2023; 78:819–829
48. Yim D, Yeo TY, Park MH: Mild cognitive impairment, dementia, and cognitive dysfunction screening using machine learning. J Int Med Res 2020; 48:300060520936881
49. Huang S-T, Hsiao F-Y, Tsai T-H, et al: Using hypothesis-led machine learning and hierarchical cluster analysis to identify disease pathways prior to dementia: longitudinal cohort Study. J Med Internet Res 2023; 25:e41858
50. Liu H, Zhang X, Liu H, et al: Using machine learning to predict cognitive impairment among middle-aged and older Chinese: a longitudinal study. Int J Public Health 2023; 68:1605322
51. Livingston G, Huntley J, Liu KY, et al: Dementia prevention, intervention, and care: 2024 report of the lancet standing commission. The Lancet 2024; 404:572–628
52. Yang Q, Lyu X, Lin Q, et al: Effects of a multicomponent intervention to slow mild cognitive impairment progression: a randomized controlled trial. Int J Nurs Stud 2022; 125:104110