

Identification and cognitive function prediction of Alzheimer's disease based on multivariate pattern analysis of hippocampal volumes

Journal of Alzheimer's Disease

1–10

© The Author(s) 2024

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1387287241296130

journals.sagepub.com/home/alz



Ziwen Gao^{1,2,#}, Wanqiu Zhu^{1,2,#}, Yuqing Li^{1,2}, Wei Ye^{1,2}, Xiao Chen^{1,2}, Shanshan Zhou³, Xiaohu Li^{1,2}, Xiaoshu Li^{1,2}, Yongqiang Yu^{1,2} and the Alzheimer's Disease Neuroimaging Initiative^{*}

Abstract

Background: Alzheimer's disease (AD) is strongly associated with slowly progressive hippocampal atrophy. Elucidating the relationships between local morphometric changes and disease status for early diagnosis could be aided by machine learning algorithms trained on neuroimaging datasets.

Objective: This study intended to propose machine learning models for the accurate identification and cognitive function prediction across the AD severity spectrum based on structural magnetic resonance imaging (sMRI) of the bilateral hippocampi.

Methods: The high-resolution sMRI data of 120 AD dementia patients, 232 amnesic mild cognitive impairment (aMCI) patients, and 206 healthy controls (HCs) were included from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The classification capacity and cognitive predict ability of hippocampal volume was evaluated by multiple pattern analysis using the support vector machine (SVM) and relevance vector regression (RVR) application of the Pattern Recognition for Neuroimaging Toolbox, separately. For validation, the analyses were performed using a biomarker-based regrouping method and another independent local dataset.

Results: The SVM application produced a total accuracy of 94.17%, 80.85%, and 70.74% and area under receiver operating characteristic curves of 0.97, 0.87, and 0.72 between HC versus AD dementia, HC versus aMCI, and aMCI versus AD dementia classification, respectively. The RVR application significantly predicted the baseline and mean cognitive function at three years of follow-up. Qualitatively consistent results were obtained using different regrouping method and the local dataset.

Conclusions: The machine learning methods based on the bilateral hippocampi distinguished across the AD severity spectrum and predicted the baseline and the longitudinal cognitive function with greater accuracy.

Keywords

Alzheimer's disease, amnesic mild cognitive impairment, hippocampus, structural magnetic resonance image, relevance vector regression, support vector machine

Received: 27 April 2024; accepted: 10 September 2024

¹Department of Radiology, The First Affiliated Hospital of Anhui Medical University, Hefei, China

²Anhui Provincial Key Laboratory for Brain Bank Construction and Resource Utilization, Hefei, China

³Department of Neurology, The First Affiliated Hospital of Anhui Medical University, Hefei, China

[#]These authors contributed equally to this work.

^{*}Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://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

Corresponding authors:

Xiaoshu Li, Department of Radiology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022, China.

Email: lixiaoshu2016@163.com

Yongqiang Yu, Department of Radiology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022, China.

Email: cjr.yuyongqiang@vip.163.com

Introduction

Alzheimer's disease (AD) is a chronic and irreversible neurodegenerative disease predominantly afflicting the elderly.¹ It is characterized by the progressive degeneration of neurons and ensuing tissue atrophy, ultimately leading to cognitive impairments and dementia. Around 46.8 million people globally are currently living with AD-associated dementia and the prevalence is expected to triple by the year 2050 due to population aging.² Amnesic mild cognitive impairment (aMCI) is another common age-related condition characterized by cognitive dysfunction, but with intact daily functioning,³ with a high risk of progression to AD dementia.⁴ Until now, the early diagnosis remains a challenge due to the non-specificity of symptoms and the difficulty detecting the underlying pathology using non-invasive approaches.⁵ Despite the lack of curative treatments, the available anti-amyloid- β (A β) treatments (such as lecanemab, donanemab) have been achieved a real breakthrough,⁶ thus early diagnosis is of great significance for timely intervention, which may help preserve cognition, improve patients' living ability and reduce the burden of caregivers.

However, the routine applications of AD detection methods such as positron emission tomography (PET) scans and cerebrospinal fluid (CSF) tests are limited for early diagnosis and large-scale screening due to the complex operation, high cost and invasive characteristics.⁷ In contrast, structural magnetic resonance imaging (sMRI) makes up for the above deficiencies and is the preferred neuroimaging modality for monitoring AD-associated atrophy of brain structure due to the high tissue contrast achievable, permitting accurate measurement of neural surface area, thickness, and volume.⁸ Further, recent morphometric studies have attempted to identify structural biomarkers indicative of disease state for diagnosis and monitoring of disease progression.⁹ In addition, the application of automated machine learning techniques enables to learn more information from images, further improving the accuracy. Previous studies have shown that machine learning classifiers based on sMRI were able to accurately diagnose AD, and the diagnostic accuracy ranging from 80% to 100%, with the hippocampus being the strongest contributor to classification decisions.^{10–12}

The hippocampus is one of the first brain structures exhibiting detectable neurodegenerative changes and atrophy during the progression of AD before onset of cognitive symptoms. Hippocampal volume is currently the best validated sMRI biomarker for AD.¹³ The exceptional spatial resolution and high contrast provided by sMRI allows for the visualization of subtle morphometric changes in the hippocampus that could aid in AD diagnosis,¹⁴ what's more, the hippocampal volume has demonstrated promise for disease monitoring and cognitive function forecasting in clinical trials.^{15,16} Thus, it is necessary to construct a

generalizable machine learning models based on sMRI of bilateral hippocampi for the early diagnosis and cognitive function prediction of AD.

The primary goal of current study was to propose and test a machine learning application for distinguishing across the AD severity spectrum and predicting both the baseline and the longitudinal cognitive function based on objective morphometric values of the bilateral hippocampi derived using sMRI.

Methods

Participants

The data analyzed in the current study were obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI, <http://adni.loni.usc.edu>), an ongoing, multisite and longitudinal dataset. The ADNI was launched in 2003 as a public-private partnership, led by principal investigator Michael W. Weiner, MD. The primary goal of ADNI was to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The study was approved by the Institutional Review Boards (IRBs) of all participating institutions. Informed written consent was obtained from all participants at each site. The detailed criteria of inclusion and exclusion are described on the website (<http://www.adni-info.org>).

In the current study, participants included were recruited from ADNI-2 at baseline, all of them underwent CSF A β_{42} measurement, sMRI scans, a battery of cognitive test at baseline, and had at least one cognitive test during the three years of follow-up. Finally, 558 participants were included (120 AD dementia patients, 232 aMCI patients, and 206 healthy controls (HCs)).

This study was approved by the Medical Research Ethics Committee of the First Affiliated Hospital of Anhui Medical University, China, according to the Declaration of Helsinki, as well as the institutional review board of all participating sites in ADNI. Informed consent was provided by all subjects.

Cognitive assessments

All participants included in the current study received a set of standardized neuropsychological evaluation including the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), Auditory Verbal Learning Test (AVLT) and so on. For a description of the cognitive scales, see the Supplemental Material.

CSF biomarkers

CSF was collected by lumbar puncture, aliquoted to 500 μ L in polypropylene tubes, and stored at -80°C . The

concentrations of CSF biomarkers ($A\beta_{42}$, t-tau, and p-tau) were measured using Elecsys electrochemiluminescence immunoassays according to the methods on the website (<http://adni.loni.usc.edu>).

MRI data acquisition and image pre-processing

The sMRI data of these participants were acquired using MPAGE or equivalent protocols of different resolutions with a slice thickness of 1.2 mm and have been subjected to several preprocessing steps by the research groups belonging to the ADNI. The geometric distortions caused by gradient models were corrected, as were image intensity B1 nonuniformities, an N3 histogram peak sharpening algorithm was applied to reduce image intensity nonuniformities.

The anatomic images were pre-processed using the voxel-based morphometry (VBM) 8 toolbox (<http://dbm.neuro.uni-jena.de/vbm.html>), implemented in Statistical Parametric Mapping (SPM) 8 (www.fil.ion.ucl.ac.uk/spm/software/spm8/), running on Matlab 2018 (Math Works, USA). The detailed steps for image pre-processing can be found in the Supplemental Material.

Machine learning models

Machine learning was performed using the Pattern Recognition for Neuroimaging Toolbox (PRoNT) running in Matlab 2018,¹⁷ and the covariates age, sex, education, and total intracranial volume (TIV) were regressed out. To evaluate the statistical significance of model performance, permutation tests were conducted in which the results were compared to the classification of models retrained by 1000 repetitions of random labels.

Performance of the classification model

Class labels were first assigned to the smoothed images of each subject (AD dementia, aMCI, or HC). A DARTEL gray matter mask was applied as the first-level mask and a mask of the bilateral hippocampi extracted from the Brainnetome atlas¹⁸ was selected as the second-level mask to build the kernel. The hippocampus is further divided into rostral and caudal hippocampus along the long axis in this atlas. Furthermore, a mask of the bilateral hippocampi extracted from the Automatic Anatomical Labelling (AAL) atlas, a commonly used atlas and the atlas obtained from a preliminary release of the EADC-ADNI harmonized segmentation protocol project¹⁹ were selected as the second-level mask to verify the stability of classification results.

Support vector machine (SVM; Cortes and Vapnik) was adopted for the discrimination analysis. It is a generalized linear classifier based on supervised algorithm for pattern recognition. The basic principle is to find a hyperplane in the feature space, so as to distinguish different classes of samples, and maximize the distance from the sample point closest to the hyperplane. It is a common machine learning application for neuroimaging due to its strong capacity for generalization based on relatively small training samples. During the model training phase, weights are assigned to features yielding maximal separation using a hyperplane. Classification labels are then obtained by the total feature weights and the test sample. In this study, a default soft-margin parameter $C = 1$ was selected.

A leave-one-out cross-validation (LOOCV) method was adopted to quantify the classification power of each model generated by the algorithms.^{20,21} Receiver operating characteristic (ROC) curves were then constructed based on LOOCV results to assess the classifiers' performance for distinguishing among AD dementia, aMCI, and HC individuals. The classification performance is expressed as the area under the ROC curve (AUC), and the larger AUC, the better the classification performance.

Performance of the regression model

Due to the relatively high classification accuracy obtained using the whole hippocampus template based on the Brainnetome atlas, a DARTEL gray matter mask was applied as the first-level mask and a mask of the bilateral hippocampi extracted from the Brainnetome atlas was selected as the second-level mask to build the kernel for the regression model.

Relevance vector regression (RVR) is an SVM-based regression model used to predict continuous values. The basic idea of RVR is to find a hyperplane in the feature space so that all data points are closest to that plane. This hyperplane can be thought of as the best fit line. Compared to traditional support vector regression (SVR), RVR is more flexible because it allows data points to have zero weight on the hyperplane, which reduces model complexity and improves prediction accuracy.²² The performance of the prediction model was evaluated using Pearson correlation coefficient (R), mean squared error (MSE), and the coefficient of determination (R^2).

Statistical analysis

The Statistical Package for the Social Sciences 23.0 was applied for the statistical analyses. The differences in continuous variables were compared by one-way analysis of variance (ANOVA) with *post hoc* Bonferroni correction for pair-wise comparisons, while categorical variables

were compared among groups by χ^2 test. Group differences in bilateral hippocampal volume were assessed by constructing a general linear model with age, sex, education, apolipoprotein E (*APOE*) $\epsilon 4$ allele status and TIV as covariates. $p < 0.05$ was considered statistically significant.

Validation analysis

Validation analysis of regrouped participants of ADNI by CSF $A\beta_{42}$ measurements. To represent AD progression in more detail, the participants were further regrouped into different groups according to the CSF $A\beta_{42}$ level based on the following pre-established cutoffs: $A\beta_{42} < 980$ pg/mL A+. ²³ Participants were categorized into the following subgroups: HC A-/A+, cognitive normal with $A\beta$ negative/positive; aMCI A-/A+, aMCI with $A\beta$ negative/positive; AD dementia A-/A+, AD dementia with $A\beta$ negative/positive. The SVM model was applied for distinguishing participants among HC A-, aMCI A+ and AD dementia A+.

Validation analysis of participants from local dataset. To verify the stability of the machine learning model performance, these analyses were repeated using another independent cohort from our local hospital-the First Affiliated Hospital of Anhui Medical University. In total, 423 participants matched for age, sex, and education were included (161 AD dementia patients, 130 aMCI patients, and 132 HCs). The details of participants' inclusion and exclusion criteria, cognitive assessment, MRI data acquisition and image pre-processing steps can be found in the Supplemental Material.

Results

Cohort characteristics

The demographic data of all subjects are summarized in Table 1. No significant differences in age, education, ethnicity, race, sex ratio, or TIV among HC, aMCI, and AD dementia groups were detected (all $p > 0.05$). However, body mass index was significantly higher in the aMCI and HC groups compared to AD dementia group (both $p < 0.05$). As expected, the cognitive assessment scores (MMSE, MoCA, and AVLT at baseline and during follow-up) and the concentration of $A\beta_{42}$ were significantly higher in the HC group compared to AD dementia and aMCI groups, and higher in the aMCI group than the AD dementia group, while the proportion of *APOE* $\epsilon 4$ carriers and the concentration of t-tau and p-tau had the opposite trend (all $p < 0.05$).

Comparison of hippocampal volume

The volumes of the whole, rostral and caudal hippocampus were largest in HC participants, medium in aMCI patients,

and minimal in AD dementia patients ($p < 0.05$, Table 2 and Figure 1).

Performance of the SVM classifier

The predictive power of the classifier was evaluated by calculating the total accuracy, balanced accuracy, sensitivity and specificity in the cross-validation testing phase (Table 3) from ROC curves (Figure 2). Compared to rostral and caudal hippocampus, the SVM model trained using sMRI of the whole hippocampus achieved the highest accuracy, yielding a total accuracy of 94.17%, 80.85%, 70.74%, balanced accuracy of 93.30%, 80.87%, 68.35% (all $p = 0.001$), sensitivity of 94.31%, 83.00%, 56.59%, specificity of 93.91%, 78.59%, 78.92% for the AD dementia versus HC, aMCI versus HC and AD dementia versus aMCI separately (AUC = 0.97, 0.87, 0.72 separately). Similar results were obtained using other two hippocampal masks (Supplemental Table 1).

Performance of the RVR prediction model

As the SVM model based on whole hippocampus achieved highest accuracy, the RVR prediction model used the mask of the whole hippocampus for further analysis (Table 4). The model not only significantly predicted the baseline cognitive performance (MoCA: $R = 0.62$, $p = 0.001$, $R^2 = 0.38$, $p = 0.010$, MSD = 12.36, $p = 0.001$; AVLT-immediate: $R = 0.63$, $p = 0.001$, $R^2 = 0.40$, $p = 0.003$, MSD = 106.29, $p = 0.001$; AVLT-learning: $R = 0.50$, $p = 0.001$, $R^2 = 0.25$, $p = 0.010$, MSD = 5.96, $p = 0.001$), but also significantly predicted the mean cognitive function at three years of follow-up (MoCA: $R = 0.64$, $p = 0.001$, $R^2 = 0.41$, $p = 0.010$, MSD = 18.07, $p = 0.001$; AVLT-immediate: $R = 0.43$, $p = 0.001$, $R^2 = 0.19$, $p = 0.020$, MSD = 110.45, $p = 0.001$; AVLT-learning: $R = 0.38$, $p = 0.001$, $R^2 = 0.15$, $p = 0.023$, MSD = 4.17, $p = 0.001$).

Validation analysis

The SVM classification model yielded similar total accuracy for distinguishing AD dementia A+ from HC A- (95.80%), AD dementia A+ from aMCI A+ (70.80%), and aMCI A+ from HC A- (82.01%) by regrouping ADNI participants according to the concentration of CSF $A\beta_{42}$ (Table 5).

The SVM classification model yielded qualitatively similar results for the local dataset, with a similar total accuracy for distinguishing AD dementia from HC (95.56%), a slightly lower accuracy for distinguishing aMCI from HC (74.05%) and a slightly higher accuracy for distinguishing AD dementia from aMCI (76.63%) (Supplemental Table 6). The RVR model also performed

Table 1. Characteristics of the AD dementia, aMCI, and HC groups in the ADNI database.

	HC	aMCI	AD dementia	<i>F</i>	<i>p</i>
Number (F/M)	86/120	96/136	49/71	0.026	0.987
Age (y)	72.55 ± 5.96	72.86 ± 7.09	74.21 ± 8.25	2.279	0.103
Education (y)	16.53 ± 2.49	16.47 ± 2.68	15.92 ± 2.41	0.539	0.584
BMI	27.37 ± 4.91	27.47 ± 5.31	25.94 ± 5.11	3.923	0.020
Ethnicity				6.292	0.178
Hispanic or Latino	9	4	5		
Not Hispanic or Latino	193	228	114		
Unknown	3	0	1		
Race				6.906	0.330
Asian	3	3	4		
Black	10	4	4		
White	190	224	111		
More than one	3	1	1		
APOE ε4 (%)	63 (30.58)	120 (51.72)	82 (68.33)	46.188	<0.001
MMSE					
Baseline	29.03 ± 1.17	27.82 ± 1.80	23.22 ± 2.03	494.400	<0.001
Follow up frequency	2.66 ± 0.71	3.40 ± 0.88	2.03 ± 0.66	131.741	<0.001
Mean score of follow up	28.90 ± 1.06	26.73 ± 2.48	21.20 ± 3.82	372.424	<0.001
MoCA					
Baseline	25.96 ± 2.40	22.71 ± 3.05	17.41 ± 4.49	260.750	<0.001
Follow up frequency	2.61 ± 0.78	3.41 ± 0.87	1.91 ± 0.86	133.532	<0.001
Mean score of follow up	26.16 ± 2.10	22.31 ± 3.48	15.07 ± 5.85	331.687	<0.001
AVLT-immediate					
Baseline	45.83 ± 10.03	34.31 ± 10.57	22.11 ± 7.10	253.552	<0.001
Follow up frequency	2.72 ± 0.67	3.31 ± 0.84	2.51 ± 0.79	52.660	<0.001
Mean score of follow up	39.60 ± 10.88	32.45 ± 9.98	25.53 ± 10.34	71.900	<0.001
AVLT-learning					
Baseline	6.06 ± 2.32	4.34 ± 2.62	1.75 ± 1.70	129.015	<0.001
Follow up frequency	2.72 ± 0.67	3.31 ± 0.84	2.51 ± 0.79	52.660	<0.001
Mean score of follow up	4.90 ± 1.96	3.78 ± 1.75	2.43 ± 1.71	63.407	<0.001
CSF biomarkers (pg/ml)					
Aβ ₄₂	1250.16 ± 434.91	929.45 ± 399.72	689.77 ± 327.68	80.266	<0.001
t-tau	235.08 ± 92.04	289.43 ± 137.59	379.76 ± 159.56	48.201	<0.001
p-tau	21.28 ± 9.16	28.01 ± 15.16	37.53 ± 16.61	54.003	<0.001
TIV (cm ³)	1375.83 ± 130.63	1401.71 ± 127.01	1393.46 ± 149.98	2.089	0.125

AD: Alzheimer's disease; aMCI: amnesic mild cognitive impairment; HC: healthy control; BMI: body mass index; APOE: apolipoprotein E; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; AVLT: Auditory Verbal Learning Test; CSF: cerebrospinal fluid; Aβ: amyloid-β; TIV: total intracranial volume.

Table 2. Group differences in gray matter volume of bilateral hippocampi between AD dementia, aMCI, and HC groups.

	HC	aMCI	AD dementia	<i>F</i>	<i>p</i>
Whole hippocampal volume	0.61 ± 0.03	0.54 ± 0.05	0.49 ± 0.05	243.115	<0.001
Rostral hippocampal volume	0.69 ± 0.04	0.61 ± 0.06	0.55 ± 0.07	225.135	<0.001
Caudal hippocampal volume	0.54 ± 0.03	0.47 ± 0.05	0.44 ± 0.04	179.312	<0.001

AD: Alzheimer's disease; aMCI: amnesic mild cognitive impairment; HC: healthy control.

well in predicting baseline cognitive function in the local dataset (Supplemental Table 7).

Discussion

We developed and tested generalizable machine learning models using PRoNT application implemented in Matlab

for distinguishing patients on the AD disease spectrum and predict the baseline and longitudinal cognitive performance based on sMRI measures of bilateral hippocampal volume. The main findings were as follows: (1) The SVM model demonstrated high accuracy for distinguishing patients from controls and reasonable accuracy for distinguishing AD dementia from aMCI. (2) The RVR model

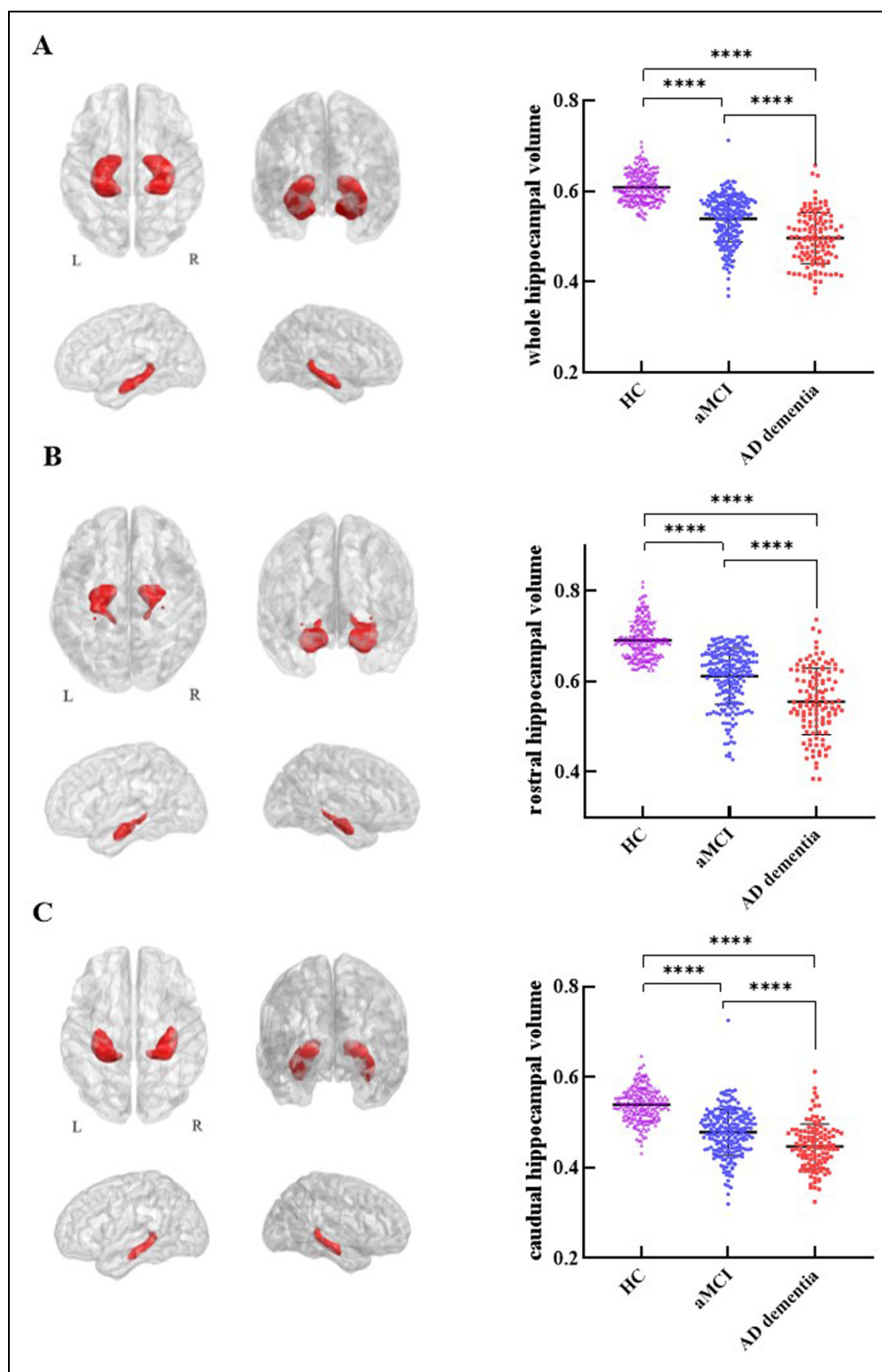
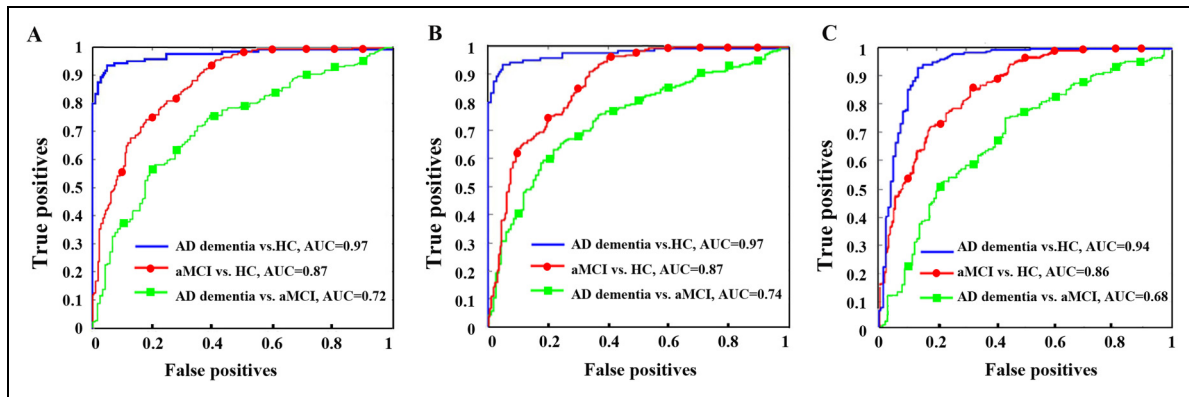


Figure 1. Extraction of bilateral whole, rostral, and caudal hippocampi and intergroup comparisons of hippocampal volume in Alzheimer's disease (AD) dementia, amnesic mild cognitive impairment (aMCI), and healthy control (HC) groups (A-C, Bonferroni correction, **** $p < 0.0001$).

Table 3. Accuracy and area under the ROC curve (AUC) by hippocampal volume in classifying the AD dementia, aMCI, and HC groups.

	Total accuracy	Balanced accuracy	<i>p</i>	Sensitivity	Specificity	AUC
AD dementia versus HC						
Whole hippocampus	94.17%	93.30%	0.001	94.31%	93.91%	0.97
Rostral hippocampus	93.87%	93.23%	0.001	95.63%	92.37%	0.97
Caudal hippocampus	90.80%	89.94%	0.001	88.14%	92.31%	0.94
aMCI versus HC						
Whole hippocampus	80.85%	80.87%	0.001	83.00%	78.59%	0.87
Rostral hippocampus	79.26%	79.74%	0.001	83.48%	75.37%	0.87
Caudal hippocampus	79.26%	79.26%	0.001	81.36%	76.18%	0.86
AD dementia versus aMCI						
Whole hippocampus	70.74%	68.35%	0.001	56.59%	78.92%	0.72
Rostral hippocampus	69.89%	68.71%	0.001	54.93%	80.00%	0.74
Caudal hippocampus	67.61%	60.95%	0.001	53.33%	72.52%	0.68

ROC: receiver operating characteristic; AD: Alzheimer's disease; aMCI: amnesic mild cognitive impairment; HC: healthy control.

**Figure 2.** ROC curves from the support vector machine classifier and the general ROC model for classification of the AD spectrum based on the whole, rostral and caudal hippocampus (A-C). ROC: receiver operating characteristic; AUC: area under the curve; AD: Alzheimer's disease; aMCI: amnesic mild cognitive impairment; HC: healthy control.**Table 4.** The performance of the prediction model.

	R	<i>p</i>	R ²	<i>p</i>	MSD	<i>p</i>
MoCA						
Baseline score	0.62	0.001	0.38	0.010	12.36	0.001
Mean score of follow up	0.64	0.001	0.41	0.010	18.07	0.001
AVLT-immediate						
Baseline score	0.63	0.001	0.40	0.003	106.29	0.001
Mean score of follow up	0.43	0.001	0.19	0.020	110.45	0.001
AVLT-learning						
Baseline score	0.50	0.001	0.25	0.016	5.96	0.001
Mean score of follow up	0.38	0.001	0.15	0.023	4.17	0.001

MoCA: Montreal Cognitive Assessment; AVLT: Auditory Verbal Learning Test; R: Pearson correlation coefficient, MSE: mean squared error; R²: coefficient of determination.

significantly predicted the baseline and mean cognitive function at three years of follow-up. (3) Qualitatively consistent results were obtained by using different mask of

hippocampus, a regrouping method based on the biomarker of CSF A β and a different local dataset.

The SVM model demonstrated high accuracy for distinguishing patients from controls and reasonable accuracy for distinguishing AD dementia from aMCI patients. The model enabled us to distinguish AD dementia patients from controls with up to 93.30% balanced accuracy (AUC=0.97), and the balanced accuracy was 80.87% when distinguishing aMCI patients from controls (AUC=0.87), superior to many established models trained using databases^{24–27} or data from patients recruited by the study site.^{28,29} The high total and balanced accuracy suggested that the model was not affected by the effect of imbalanced test sets and the application of SVM method also makes the model have high efficiency, interpretability and robustness.³⁰ However, the accuracy of this classification model was just acceptable when applied to the identification of aMCI patients and AD dementia patients (balanced accuracy=68.35, AUC=0.72),^{26,31} underscoring the need for training using much large numbers of aMCI patient images. In the current study,

Table 5. Accuracy and area under the ROC curve (AUC) by hippocampal volume in classifying the AD dementia **A+**, aMCI **A+**, and HC **A-** groups.

	Total accuracy	Balanced accuracy	<i>p</i>	Sensitivity	Specificity	AUC
AD dementia A+ versus HC A-						
Whole hippocampus	95.80%	94.90%	0.001	94.96%	95.24%	0.98
Rostral hippocampus	93.85%	93.71%	0.001	94.20%	93.40%	0.98
Caudal hippocampus	88.93%	88.61%	0.001	89.24%	88.46%	0.95
aMCI A+ versus HC A-						
Whole hippocampus	82.01%	81.99%	0.001	80.58%	83.33%	0.91
Rostral hippocampus	79.24%	79.43%	0.001	75.50%	83.33%	0.90
Caudal hippocampus	79.58%	79.76%	0.001	76.00%	83.45%	0.89
AD dementia A+ versus aMCI A+						
Whole hippocampus	70.80%	70.18%	0.001	63.29%	76.60%	0.76
Rostral hippocampus	69.64%	69.61%	0.001	61.38%	76.99%	0.75
Caudal hippocampus	63.85%	61.77%	0.001	55.58%	68.63%	0.72

ROC: receiver operating characteristic; AD: Alzheimer's disease; aMCI: amnesic mild cognitive impairment; HC: healthy control.

we selected SVM for model construction, which is more suitable for small sample size models and has better generalization and explainability, however, this model is limited in scalability, practicability and computational complexity, etc.^{32,33} Continued model training using large numbers of images and the application of other models, such as random forests, may further improve this classification accuracy, thereby facilitating earlier diagnosis and identifying structural features predictive of the aMCI to AD dementia transition.

The hippocampus is one of the first brain regions exhibiting AD pathology and ultimately one of the most severely damaged as focal lesions of apoptotic neurons accumulate, leading to detectable hippocampal shrinkage.³⁴ Consistent with numerous previous reports, we found a decrease in bilateral hippocampal volume of both AD and aMCI patients.^{35,36} The analyses based on the subregions of the hippocampus revealed that the classification accuracy was lower than that of whole hippocampus. The outperformance of the whole hippocampal volume suggested that there are significant inter-group differences in the whole hippocampus, which can provide more information in the classifier than subregions and play a slightly greater role in the diagnosis of AD. Based on findings of hippocampal degeneration in AD models and patient samples, numerous studies over the past decade have used sMRI to non-invasively capture the structural changes to the whole hippocampus induced by the disease.^{37,38} It may be possible to establish models predictive of AD onset and progression based on regional structural changes or combinations of changes in other brain regions.

Globally, the MMSE is the preferred screening tool for AD patients, while the MoCA is applied regularly for the clinical diagnosis of aMCI patients³⁹ as total scores strongly correlate with disease progression (i.e., with the extent of cortical and hippocampal damage). In the current study, the AVLT scale was also included to provide a more targeted assessment of short- and long-term memory deficits.

According to previous literatures, there were strong relationship between AVLT scores and hippocampal volume associated with AD, highlighting the significance of the AVLT for the detection and diagnosis of AD in the early stages.⁴⁰ The results of the current study suggested that the RVR model based on bilateral hippocampal volume can predict baseline and follow-up scores of the above scales, further suggesting that bilateral hippocampi contribute to cognitive dysfunction in AD.

In terms of generalization of the model, similar classification accuracies were obtained by regrouping the participants according to the levels of CSF A β on the basis of clinical diagnosis, which proved that the models were stable. On the other hand, this model had similar predictive performance in both the local dataset and the ADNI database, representing different races, educational levels, and socioeconomic status, indicating that the model can be generalized to independent datasets and thus play a greater role.

This study has several limitations: (1) the participants of the local dataset were mainly Han Chinese, lacking diversity and under-represented, at the same time, the lack of biomarker, genes, socioeconomic status limited the representative power of the local dataset; (2) model training using only sMRI of the hippocampus, and the discriminative ability was relatively weak for aMCI discriminations. Training using image acquired by other modalities, including functional MRI, PET and so on, may improve the accuracy of classification. (3) the segmentation of hippocampus was based on the application of SPM, and the results were not validated using segmentation methods based on other applications, such as Freesurfer; (4) the inclusion/exclusion criteria reduced the likelihood of mixed pathologies and limit the generalization of this dataset.

Conclusion

The SVM and RVR machine learning models based on hippocampal volume can accurately distinguish AD from

matched controls with satisfying generalization ability and significantly predicted the baseline and longitudinal cognitive function. The findings suggest the key role of hippocampus in the early detection and diagnosis of AD, highlighting the value of promoting and applying machine learning classification model based on hippocampus structure in clinical practice. Additional training with larger image samples may further improve classification performance, and enhance the utility of this tool for aiding in AD diagnosis and prognosis.

Acknowledgments

We thank all the participants for their cooperation during this study.

Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH) including 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.

ORCID iD

Yongqiang Yu  <https://orcid.org/0000-0001-8977-2215>

Statements and declarations

Author contributions

Gao Ziwen (Writing – original draft); Wanqiu Zhu (Writing – original draft); Yuqing Li (Data curation); Wei Ye (Data curation); Xiao Chen (Data curation); Shanshan Zhou (Data curation); Xiaohu Li (Data curation); Xiaoshu Li (Writing – review & editing); Yongqiang Yu (Writing – review & editing).

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was financially supported by the National Natural Science Foundation of China (No. 82071905).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

The data of the local dataset used in the current study are available from the corresponding author on reasonable request. The dataset of ADNI used in this analysis can be found at: <https://adni.loni.usc.edu>. The machine learning application used can be downloaded from: MLNL :: PRoNTTo (ucl.ac.uk).

Supplemental material

Supplemental material for this article is available online.

References

- 2022 Alzheimer's disease facts and figures. *Alzheimers Dement* 2022; 18: 700–789.
- Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. *Lancet* 2021; 397: 1577–1590.
- Petersen RC, Lopez O, Armstrong MJ, et al. Practice guideline update summary: mild cognitive impairment: report of the guideline development, dissemination, and implementation subcommittee of the American academy of neurology. *Neurology* 2018; 90: 126–135.
- Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004; 256: 183–194.
- Bature F, Guinn BA, Pang D, et al. Signs and symptoms preceding the diagnosis of Alzheimer's disease: a systematic scoping review of literature from 1937 to 2016. *BMJ Open* 2017; 7: e015746.
- Perneczky R, Jessen F, Grimmer T, et al. Anti-amyloid antibody therapies in Alzheimer's disease. *Brain* 2023; 146: 842–849.
- Gao F, Lv X, Dai L, et al. A combination model of AD biomarkers revealed by machine learning precisely predicts Alzheimer's dementia: China aging and neurodegenerative initiative (CANDI) study. *Alzheimers Dement* 2023; 19: 749–760.
- Dai WZ, Liu L, Zhu MZ, et al. Morphological and structural network analysis of sporadic Alzheimer's disease brains based on the APOE4 gene. *J Alzheimers Dis* 2023; 91: 1035–1048.
- Falahati F, Westman E and Simmons A. Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. *J Alzheimers Dis* 2014; 41: 685–708.
- 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: 519–535.
- Diogo VS, Ferreira HA, Prata D, et al. Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. *Alzheimers Res Ther* 2022; 14: 107.

12. Feng X, Provenzano FA, Small SA, et al. A deep learning MRI approach outperforms other biomarkers of prodromal Alzheimer's disease. *Alzheimers Res Ther* 2022; 14: 45.
13. Jack CR Jr, Barkhof F, Bernstein MA, et al. Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimers Dement* 2011; 7: 474–485. e474.
14. Kalin AM, Park MT, Chakravarty MM, et al. Subcortical shape changes, hippocampal atrophy and cortical thinning in future Alzheimer's disease patients. *Front Aging Neurosci* 2017; 9: 38.
15. Hill DLG, Schwarz AJ, Isaac M, et al. Coalition against Major diseases/European medicines agency biomarker qualification of hippocampal volume for enrichment of clinical trials in predementia stages of Alzheimer's disease. *Alzheimers Dement* 2014; 10: 421–429. e423.
16. Li H, Habes M, Wolk DA, et al. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimers Dement* 2019; 15: 1059–1070.
17. Schrouff J, Rosa MJ, Rondina JM, et al. PRoNT: pattern recognition for neuroimaging toolbox. *Neuroinformatics* 2013; 11: 319–337.
18. Fan L, Li H, Zhuo J, et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cereb Cortex* 2016; 26: 3508–3526.
19. Wolf D, Bocchetta M, Preboske GM, et al. Reference standard space hippocampus labels according to the European Alzheimer's disease consortium-Alzheimer's disease neuroimaging initiative harmonized protocol: utility in automated volumetry. *Alzheimers Dement* 2017; 13: 893–902.
20. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; 7: 263–269.
21. Wee CY, Yap PT, Li W, et al. Enriched white matter connectivity networks for accurate identification of MCI patients. *Neuroimage* 2011; 54: 1812–1822.
22. Xie Y, Sun J, Man W, et al. Personalized estimates of brain cortical structural variability in individuals with autism spectrum disorder: the predictor of brain age and neurobiology relevance. *Mol Autism* 2023; 14: 27.
23. Rauchmann BS, Schneider-Axmann T, Perneczky R, et al. Associations of longitudinal plasma p-tau181 and NfL with tau-PET, Abeta-PET and cognition. *J Neurol Neurosurg Psychiatry* 2021; 92: 1289–1295.
24. Cui R and Liu M. Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis. *IEEE J Biomed Health Inform* 2019; 23: 2099–2107.
25. Sorensen L, Igel C, Pai A, et al. Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *Neuroimage Clin* 2017; 13: 470–482.
26. Lotjonen J, Wolz R, Koikkalainen J, et al. Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease. *Neuroimage* 2011; 56: 185–196.
27. Gerardin E, Chetelat G, Chupin M, et al. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage* 2009; 47: 1476–1486.
28. Feng F, Huang W, Meng Q, et al. Altered volume and structural connectivity of the hippocampus in Alzheimer's disease and amnesic mild cognitive impairment. *Front Aging Neurosci* 2021; 13: 705030.
29. Moller C, Pijnenburg YA, van der Flier WM, et al. Alzheimer disease and behavioral variant frontotemporal dementia: automatic classification based on cortical atrophy for single-subject diagnosis. *Radiology* 2016; 279: 838–848.
30. Guesne SJJ, Hanser T, Werner S, et al. Mind your prevalence!. *J Cheminform* 2024; 16: 43.
31. Poloni KM and Ferrari RJ. Automated detection, selection and classification of hippocampal landmark points for the diagnosis of Alzheimer's disease. *Comput Methods Programs Biomed* 2022; 214: 106581. 2.
32. Noble WS. What is a support vector machine? *Nat Biotechnol* 2006; 24: 1565–1567.
33. Frizzell TO, Glashutter M, Liu CC, et al. Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: a systematic review. *Ageing Res Rev* 2022; 77: 101614.
34. 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: 106–152.
35. Jacquemont T, De Vico Fallani F, Bertrand A, et al. Amyloidosis and neurodegeneration result in distinct structural connectivity patterns in mild cognitive impairment. *Neurobiol Aging* 2017; 55: 177–189.
36. Filippi M, Basaia S, Canu E, et al. Changes in functional and structural brain connectome along the Alzheimer's disease continuum. *Mol Psychiatry* 2020; 25: 230–239.
37. Arvidsson I, Strandberg O, Palmqvist S, et al. Comparing a pre-defined versus deep learning approach for extracting brain atrophy patterns to predict cognitive decline due to Alzheimer's disease in patients with mild cognitive symptoms. *Alzheimers Res Ther* 2024; 16: 61.
38. Leandrou S, Lamnisis D, Bougias H, et al. A cross-sectional study of explainable machine learning in Alzheimer's disease: diagnostic classification using MR radiomic features. *Front Aging Neurosci* 2023; 15: 1149871.
39. Han G, Maruta M, Ikeda Y, et al. Relationship between performance on the Mini-mental state examination sub-items and activities of daily living in patients with Alzheimer's disease. *J Clin Med* 2020; 9: 1537.
40. Stricker NH, Lundt ES, Albertson SM, et al. Diagnostic and prognostic accuracy of the cogstate brief battery and auditory verbal learning test in preclinical Alzheimer's disease and incident mild cognitive impairment: implications for defining subtle objective cognitive impairment. *J Alzheimers Dis* 2020; 76: 261–274.