AJNR

Deep Learning-Based Prediction of PET Amyloid Status Using MRI

Donghoon Kim, Jon André Ottesen, Ashwin Kumar, Brandon C. Ho, Elsa Bismuth, Christina B. Young, Elizabeth Mormino, Greg Zaharchuk and for the Alzheimer's Disease Neuroimaging Initiative

AJNR Am J Neuroradiol published online 27 June 2025
http://www.ajnr.org/content/early/2025/06/27/ajnr.A8899

Published June 27, 2025 as 10.3174/ajnr.A8899

ORIGINAL RESEARCH

Deep Learning-Based Prediction of PET Amyloid Status Using MRI

Donghoon Kim, Jon André Ottesen, Ashwin Kumar, Brandon C. Ho, Elsa Bismuth, Christina B. Young, Elizabeth Mormino, and Greg Zaharchuk for the Alzheimer's Disease Neuroimaging Initiative

ABSTRACT

BACKGROUND AND PURPOSE: Identifying amyloid-beta (Aß)–positive patients is essential for Alzheimer's disease (AD) clinical trials and disease-modifying treatments but currently requires PET or cerebrospinal fluid sampling. Previous MRI-based deep learning models, using only T1-weighted (T1w) images, have shown moderate performance.

MATERIALS AND METHODS: Multi-contrast MRI and PET-based quantitative Aß deposition were retrospectively obtained from three public datasets: ADNI, OASIS3, and A4. Aß positivity was defined using each dataset's recommended centiloid threshold. Two EfficientNet models were trained to predict amyloid positivity: one using only T1w images and another incorporating both T1w and T2-FLAIR. Model performance was assessed using an internal held-out test set, evaluating AUC, accuracy, sensitivity, and specificity. External validation was conducted using an independent cohort from Stanford Alzheimer's Disease Research Center. DeLong's and McNemar's tests were used to compare AUC and accuracy, respectively.

RESULTS: A total of 4,056 exams (mean [SD] age: 71.6 [6.3] years; 55% female; 55% amyloid-positive) were used for network development, and 149 exams were used for external testing (mean [SD] age: 72.1 [9.6] years; 58% female; 56% amyloid-positive). The multi-contrast model outperformed the single-modality model in the internal held-out test set (AUC: 0.67, 95% CI: 0.65–0.70, $P < 0.001$; accuracy: 0.63, 95% CI: 0.62–0.65, $P < 0.001$) compared to the T1w-only model (AUC: 0.61; accuracy: 0.59). Among cognitive subgroups, the highest performance (AUC: 0.71) was observed in mild cognitive impairment. The multi-contrast model also demonstrated consistent performance in the external test set (AUC: 0.65, 95% CI: 0.60–0.71, $P = 0.014$; accuracy: 0.62, 95% CI: 0.58–0.65, $P < 0.001$).

CONCLUSIONS: The use of multi-contrast MRI, specifically incorporating T2-FLAIR in addition to T1w images, significantly improved the predictive accuracy of PET-determined amyloid status from MRI scans using a deep learning approach.

ABBREVIATIONS: Aß = amyloid-beta; AD = Alzheimer's disease; AUC = area under the receiver operating characteristic curve; CN = cognitively normal; MCI = mild cognitive impairment; T1w = T1-wegithed; T2-FLAIR = T2-weighted fluid attenuated inversion recovery; FBP = ^{18}F-florbetapir; FBB = ^{18}F-florbetaben; SUVR = standard uptake value ratio

Received April 4, 2025; accepted after revision June 13, 2025.
From the Department of Radiology (D.K., J.A.O., A.K., B.C.H., E.B.,), and the Department of Neurology and Neurological Sciences (C.B.Y., E.M.), Stanford University, Stanford, California, USA, and Computational Radiology and Artificial Intelligence (CRAI) Research Group (J.A.O), Oslo University Hospital, Oslo, Norway.

This study was funded by NIH R56 AG071558, NIH P30 AG066515, NIH U24 AG074855, NIH K99 AG071837, and Alzheimer's Association AARFD-21-849349. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript. Dr. Greg Zaharchuk reported receiving royalties from Cambridge University Press; travel support and honoraria for lectures from Biogen, Bracco; various patents; equity in Subtle Medical. Dr. Elizabeth Mormino has been a paid consultant for Roche, Genentech, Eli Lilly, and Neurotrack. The other authors declare no completing interests.

Please address correspondence to Greg Zaharchuk, MD, PhD, Department of Radiology, Stanford University, 1201 Welch Rd, Stanford, CA, 94305, USA; gregz@stanford.edu

SUMMARY SECTION

PREVIOUS LITERATURE: MRI-based machine-learning studies for predicting PET-determined amyloid status have mostly used hand-crafted features from T1w scans reporting modest AUCs of 0.60–0.74 and often depending on extra clinical or genetic data. Only a limited number of deep learning studies have utilized whole-brain images, and these have been restricted to a single T1w contrast without robust external validation. Consequently, the generalizability of such models remains uncertain, and the potential added value of widely available supplementary sequences, such as T2-FLAIR, has not been systematically assessed.

KEY FINDINGS: Incorporating T2-FLAIR alongside T1w inputs improved the AUC from 0.61 to 0.67 on the internal held-out test set and from 0.54 to 0.65 on the external test set. The multi-contrast model achieved its highest performance within the MCI subgroup, with an AUC of 0.71.

KNOWLEDGE ADVANCEMENT: This work delivers the externally validated, image-only deep-learning model for amyloid status using widely available multi-contrast MRI, empirically confirming the incremental value of T2-FLAIR. It establishes a 3D EfficientNet baseline for opportunistic amyloid screening and provides openly available code and pretrained weights to support further research in MRI-based amyloid prediction.

Copyright 2025 by American Society of Neuroradiology.

INTRODUCTION

Alzheimer's disease (AD) is a progressive disorder that includes an asymptomatic phase, a mild cognitive impairment (MCI) phase, and a dementia phase. Early detection of neuropathological changes in AD is valuable as it potentially enables earlier treatment.[1] AD is commonly defined by its neuropathological hallmarks. Among these, amyloid-beta (Aβ) deposition is one of the earliest indicators of disease. In many clinical trials for AD, amyloid deposition has been treated as a surrogate marker representing disease,[2–4] since it is related to more rapid progression to dementia.[5,6] Amyloid positron emission tomography (PET) and cerebrospinal fluid (CSF) sampling can directly identify the presence of Aβ; however, their limited accessibility, radioactive dose, and cost restrict their widespread use.

Brain magnetic resonance imaging (MRI) is routinely obtained during the workup of memory patients because it is required to exclude alternative explanations for patient symptoms. Amyloid deposition has been implicated in the structural alteration of the brain. Such alterations, manifesting as cerebral atrophy or hippocampal volume loss, can be quantitatively assessed utilizing structural MRI scans.[5–8] Furthermore, several machine learning-based studies have highlighted the utility of MRI data to predict amyloid status when integrated with clinical and genetic information.[9–16] Most of these studies utilized T1-weighted (T1w) image-based hand-crafted features due to its excellent tissue contrast, resolution, and widespread availability in public datasets. A few specifically investigated a deep learning-based approach using T1w images, avoiding the use of subjectively determined features.[9,15,16] Although T1w images are useful to visualize brain anatomy, other imaging sequences may provide additional information. For example, T2-weighted fluid attenuated inversion recovery (T2-FLAIR) better delineates white matter abnormalities, which may be associated with amyloid deposition or cognitive impairment.[17–21] More generally, amyloid prediction may benefit from multiple MRI image contrasts as inputs, given that different sequences may highlight different aspects of the disease. Finally, none of these previous studies report their model's performance in a truly external test set, an essential test of model generalizability.

In this study, we trained and tested a deep learning algorithm to predict amyloid positivity in a large cohort of patients using both single and multi-contrast MRI and then further examined its generalizability in a separate external test group.

MATERIALS AND METHODS

Datasets

The data were obtained from the three largest publicly available datasets and one in-house dataset (hereafter referred to as the Stanford dataset). The publicly available datasets were the Alzheimer's Disease Neuroimaging Initiative (ADNI-2 and ADNI-3), the Open Access Series of Imaging Studies 3 (OASIS3), and the Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4). Only subjects who underwent T1w, T2-FLAIR, and amyloid-PET were included in this study. For the public datasets, subjects with MRI scans and amyloid-PET imaging using either 18F-florbetapir (FBP) or 18F-florbetaben (FBB) within 30 days were identified (Figure 1). Given changes in MRI technology over the past decades, we included data acquired between 2010 and 2023.

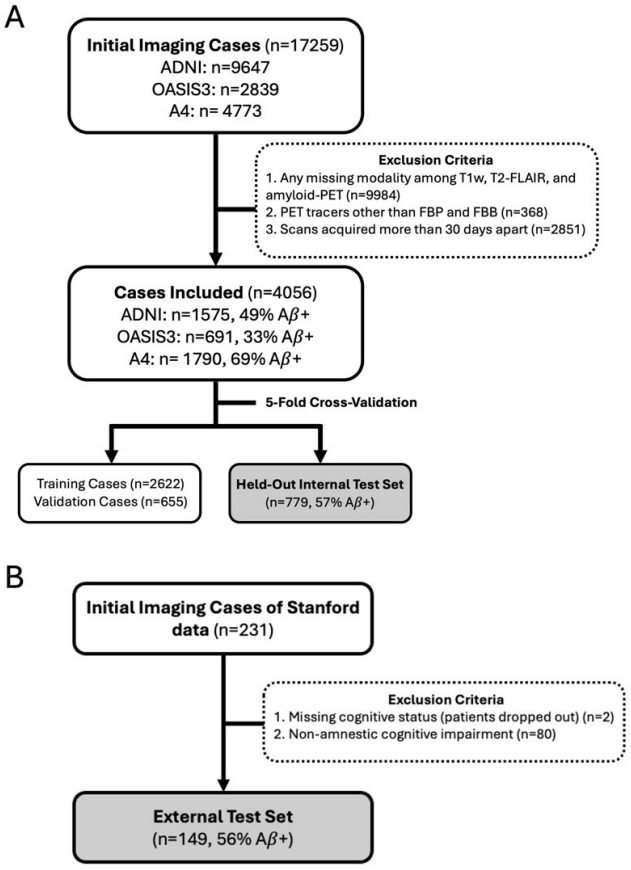

FIG 1. Flow diagrams show inclusion of study participants. (A) describes the public datasets, and (B) describes the external Stanford ADRC dataset. ADNI is the Alzheimer's Disease Neuroimaging Initiative. OASIS3 is the Open Access Series of Imaging Studies 3, and A4 is the Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease Study. For the public datasets in (A), 5-fold cross-validation was used to train models, ensuring robust evaluation and generalizability across the entire cohort.

The Stanford external test set was acquired from subjects enrolling in our Alzheimer's Disease Research Center (ADRC) between 2022 and 2024. The study protocols were approved by the University's Institutional Review Board, and informed consent was obtained from participants or their legal representatives. Imaging was performed using a 3.0T PET/MRI scanner (Signa, GE Healthcare). All subjects underwent T1w, T2-FLAIR, and amyloid-PET scans with FBB. Amyloid-PET acquisition details are described in a previous study.[22] Cognitive status and etiology were determined by a clinical panel from the ADRC-affiliated neurologists, neuropsychologists, and study staff. Subjects with non-amnestic cognitive impairment or missing cognitive status were excluded (Figure 1B).

MRI Data Acquisition and Post-Processing

For the Stanford dataset, 3D T1w images were acquired using a sagittal spoiled gradient recalled sequence with echo time (TE) = 3.06-3.23 ms, repetition time (TR) = 7.65-8.02 ms, inversion time (TI) = 400 ms, and flip angle = 11°. 3D T2-FLAIR images were acquired with TE = 116–165 ms, TR = 4800–6000 ms, TI = 1440–1800 ms, and a flip angle of 90°. T2-FLAIR images were co-registered to T1w images using rigid registration with Advanced Normalization Tools (ANTS).[23] Brain masks were generated using HD-BET[24] on the T1w images, and images were skull-stripped. All images were reoriented to the left-posterior-inferior (LPI) orientation and resampled to an isotropic voxel size of 1 mm using trilinear interpolation. Intensity normalization was performed based on the 5th to 95th percentiles.

Amyloid Status Determination

Amyloid status was determined using the centiloid cutoff values specific to each dataset and radiotracer. For the OASIS3 dataset, a cutoff of 20.6 was used based on its documentation.[25] The ADNI, A4, and Stanford datasets were processed using an optimized MRI-free Aβ quantification approach.[26] A brain atlas and a whole cerebellum mask, obtained from the Global Alzheimer's Association Interactive Network (GAAIN; gaain.org/centiloid-project), were used to calculate standard uptake value ratio (SUVR) values, which were then converted to centiloid values. Using this method, the ADNI imaging core developed tracer-specific centiloid cutoff values, which were 12 for FBB and 18 for FBP.[27]

Deep Learning Model and Experiments

The public datasets were independently and randomly split into a training set (comprising 65% of the cases), an internal validation set (16% of the cases), and a held-out internal test set (19% of the cases). The training and internal validation datasets were used during training, and the held-out internal test set was used for internal evaluation. Subjects with longitudinal data were included in the training or internal validation sets but excluded from the internal test set unless they were not used elsewhere, to prevent overfitting. A 3D EfficientNet-B3[28] was adapted from the MONAI framework[29] as the deep learning method for this study. EfficientNet is a convolutional neural network-based architecture that employs a compound scaling strategy, optimizing model width, depth, and input resolution simultaneously to achieve enhanced performance while maintaining computational efficiency. Additionally, given the high dimensionality of 3D MRI data and the computational challenges associated with processing multi-contrast inputs from a large dataset, this model was selected for its resource-efficient training capabilities. The systemically increased input resolution by the compound scaling strategy enables EfficientNet to process large 3D inputs and extract fine-grained patterns more effectively than conventional architectures. Since amyloid-related changes can be subtle and spatially distributed, the ability to model detailed volumetric features is essential. Furthermore, EfficientNet's integration of depth-wise separable convolutions and squeeze-and-excitation blocks allows it to capture fine-grained patterns in large, high-resolution imaging data, making it well-suited for this application.[28]

Two separate models were trained for amyloid positivity prediction: a single image contrast model using T1w images and a multi-contrast model using both T1w and T2-FLAIR images as inputs (Figure 2). All models were trained for 500 epochs with a batch size of 8 with the Adam optimizer,[30] a cosine annealing learning rate scheduler[31] with a 20-epoch warm up period, and a learning rate of 0.0005 after warmup. Binary cross entropy was used as the loss function. The input images were augmented by randomly rotating them along each of the three spatial axes within a range of ±11 degrees, with a 30% chance of rotation per axis. 5-fold cross-validation within the training set was performed to enhance the robustness and generalizability of the findings. The output of the model is a number between 0 and 1 representing the likelihood that the scan is amyloid positive. For binary predictions, a threshold was defined using Youden's *J* index.[32] The held-out internal test set was evaluated by each of the five trained models from the cross-validation.

Additional analyses were conducted in the internal test set, with evaluation both in the entire group as well as in patients with different cognitive statuses, including cognitively normal (CN), MCI, and dementia, using each of 5 trained models. Note that in these cohorts, the dementia group may include both AD and dementia from other causes (i.e., not all of the dementia cases are amyloid positive).

To further evaluate generalizability, we tested both models on the Stanford external dataset, derived from a different imaging protocol and study population from those used to train the network. The external test set was evaluated using each of the five trained models from the cross-validation, and their predictions were aggregated for evaluation.

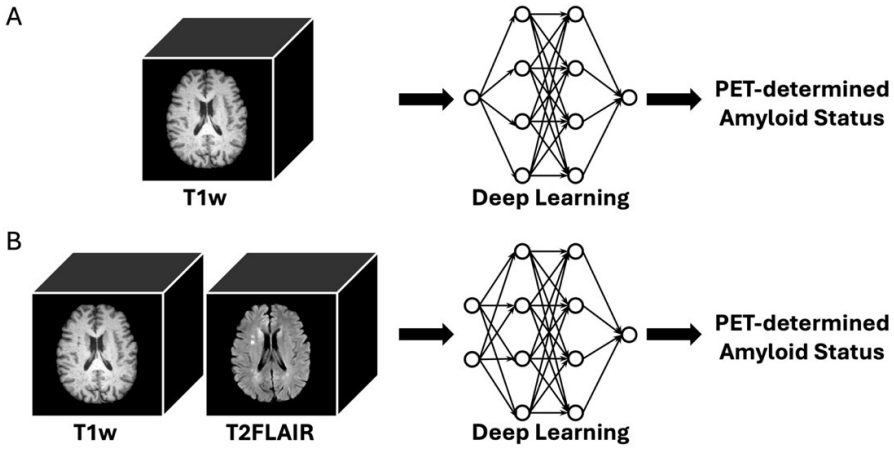

FIG 2. Schematic showing the two separate 3D EfficientNet models that were trained for amyloid positivity prediction: (A) a single contrast model using only T1w images, and (B) a multi-contrast model using both T1w and T2-FLAIR images as inputs.

Statistical Evaluation

Study subjects' demographics were stratified by amyloid-beta status into Aβ+ and Aβ- groups. The demographic characteristics were compared using chi-squared tests and t-tests for categorical variables and continuous variables, respectively. Predictions from the five-fold cross-validation were aggregated for performance evaluation and statistical analysis. The primary performance metric was the area under the receiver operating characteristic curve (AUC) to predict amyloid positivity. DeLong's test (AUC) and McNemar's test (accuracy) were used to compare the performance of the single and multi-contrast models. Binary predictions were assessed using accuracy, sensitivity, and specificity using a threshold based on Youden's *J* index. All statistical analyses were conducted using Python version 3.11.5. DeLong's test was implemented in Python, and the McNemar test was performed using the *statsmodels* package (version 0.14.2).

RESULTS

Study Sample Characteristics

Figure 1 presents flowcharts illustrating the selection process of study subjects, starting from the initial number of imaging participants, followed by the application of specific exclusion criteria, and resulting in the final cohort used for training, internal validation, internal testing, and external testing. For network development, 4,056 subjects were included from the public datasets (age 71.6 ± 6.3 yrs, 55% female, 55% amyloid positive) as shown in Table 1A. There were significant differences between the Aβ+ and Aβ- groups in terms of age, sex distribution, and cognitive status, with the amyloid positive subjects being more likely to be older, male, and to have MCI or dementia. A total of 149 subjects (age 72.1 ± 9.6 yrs, 58% female, 56% amyloid positive) were included in the external test set, derived from normal subjects and those with amnestic cognitive impairment who were participants in the Stanford ADRC. Aβ+ subjects were more likely to have MCI or dementia (Table 1B).

Table 1: Subjects' characteristics at time of amyloid PET imaging: (A) Included cohorts from ADNI, OASIS3, and A4 studies; (B) External test set. Values are presented as mean ± s.d. or as the numbers of subjects. P-values in the rightmost column reflect comparisons between amyloid-positive and amyloid-negative groups. Specifically, t-tests were used for comparing age, and Chi-squared tests were used for comparing the distributions of sex and cognitive status.

(A)

Public Datasets	Total (n=4056)	Amyloid Positive (n=2234)	Amyloid Negative (n=1822)	P-value
Age (years)	71.6 ± 6.3	72.6 ± 5.9	70.3 ± 6.6	<0.001
Sex				0.025
Male	1839 (45.3%)	1048 (46.9%)	791 (43.4%)	
Female	2217 (54.7%)	1186 (53.1%)	1031 (56.6%)	
Cognitive Status				<0.001
Normal	2808 (69.2%)	1536 (68.8%)	1272 (69.8%)	
MCI	812 (20.0%)	429 (19.2%)	383 (21.0%)	
Dementia	165 (4.1%)	153 (6.8%)	12 (0.7%)	

(B)

External Dataset	Total (n=149)	Amyloid Positive (n=83)	Amyloid Negative (n=66)	P-value
Age (years)	72.1 ± 9.6	73.1 ± 9.3	70.8 ± 9.9	0.153
Sex				0.343
Male	64 (43.0%)	39 (47.0%)	25 (37.9%)	
Female	85 (57.0%)	44 (53.0%)	41 (62.1%)	
Cognitive Status				<0.001
Normal	95 (63.8%)	40 (48.2%)	55 (83.3%)	

MCI	27 (18.1%)	18 (21.7%)	9 (13.6%)
Dementia	27 (18.1%)	25 (30.1%)	2 (3.0%)

Prediction Performance in the Validation and Internal Test Sets

In both the internal validation and held-out internal test sets, multi-contrast models incorporating T2-FLAIR in addition to T1w images outperformed single contrast models using only T1w imaging (Table 2A). Specifically, in the internal test set, the prediction of amyloid positivity was as follows: AUCs were 0.61 (95% CI: 0.59-0.63) and 0.67 (95% CI: 0.65-0.70), accuracies were 0.59 (95% CI: 0.58-0.61) and 0.63 (95% CI: 0.62-0.65), sensitivities were 0.61 (95% CI: 0.59-0.62) and 0.67 (95% CI: 0.66-0.69), and specificities were 0.57 (95% CI: 0.56-0.59) and 0.58 (95% CI: 0.56-0.60) for the single contrast and multi-contrast models, respectively. These differences were statistically significant, as indicated by DeLong's test for AUC (P<0.001) and McNemar's test for accuracy (P<0.001). The receiver operating characteristic (ROC) curves from the internal test set is shown in Figure 3A. Figure 4 presents the T1w, T2-FLAIR, and amyloid-PET images corresponding to the maximum and minimum predicted scores of the multi-contrast model from the internal test set. The case with the maximum predicted score showed cortical atrophy and severe white matter hyperintensities (WMH), in contrast to the case with the minimum predicted score. Youden's *J* index values were found to be 0.45 and 0.43 for T1w-only model and the T1w+T2-FLAIR model, respectively.

Table 2: Performance metrics for amyloid positivity prediction from (A) internal validation and internal held-out test set, and (B) the external test set. Values in parentheses represent 95% confidence intervals. Accuracy, sensitivity, and specificity metrics are presented, calculated at the Youden's *J* index operating points.

(A)

	Internal Validation Set		Internal Test Set	
	Single contrast	**Multi-contrast**	**Single contrast**	**Multi-contrast**
All				
AUC	**0.61** (0.59, 0.64)	**0.70** (0.67, 0.73)	**0.61** (0.59, 0.63)	**0.67** (0.65, 0.70)
DeLong's test	P<0.001		P<0.001	
Accuracy	**0.59** (0.57, 0.61)	**0.66** (0.64, 0.67)	**0.59** (0.58, 0.61)	**0.63** (0.62, 0.65)
McNemar's test	P<0.001		P<0.001	
Sensitivity	**0.52** (0.50, 0.54)	**0.68** (0.66, 0.69)	**0.61** (0.59, 0.62)	**0.67** (0.66, 0.69)
Specificity	**0.68** (0.66, 0.69)	**0.63** (0.62, 0.65)	**0.57** (0.56, 0.59)	**0.58** (0.56, 0.60)

(B)

	External Test Set	
	Single contrast	**Multi-contrast**
All		
AUC	**0.54** (0.49, 0.60)	**0.65** (0.60, 0.71)
DeLong's test	P=0.014	
Accuracy	**0.53** (0.49, 0.56)	**0.62** (0.58, 0.65)
McNemar's test	P<0.001	
Sensitivity	**0.57** (0.53, 0.60)	**0.75** (0.72, 0.79)
Specificity	**0.48** (0.44, 0.51)	**0.45** (0.42, 0.49)

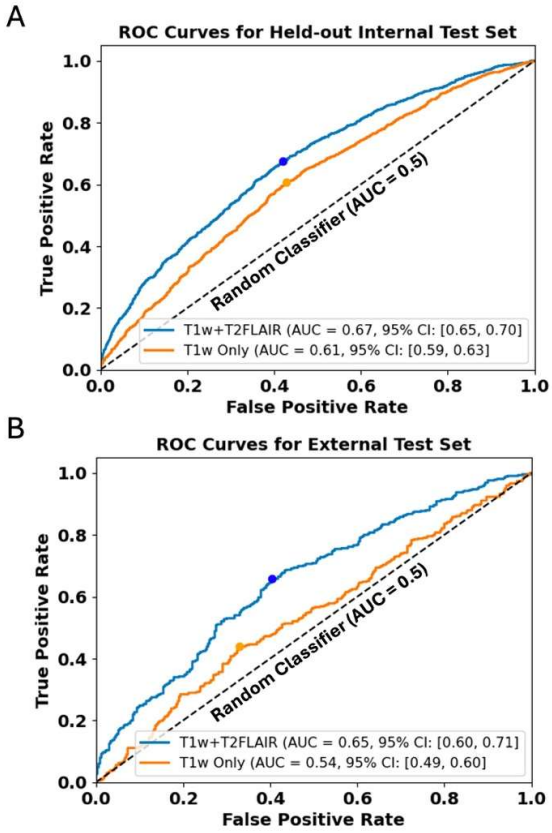

FIG 3. ROC curves of the single contrast (T1w only) and multi-contrast (T1w+T2-FLAIR) models, evaluated using 5-fold cross-validation in (A) the held-out internal test set and (B) the external test set. In all settings, adding multi-contrast MRI information improved performance. The blue and orange circles on the ROC curves represent the Youden's *J* index points that maximize combined sensitivity and specificity.

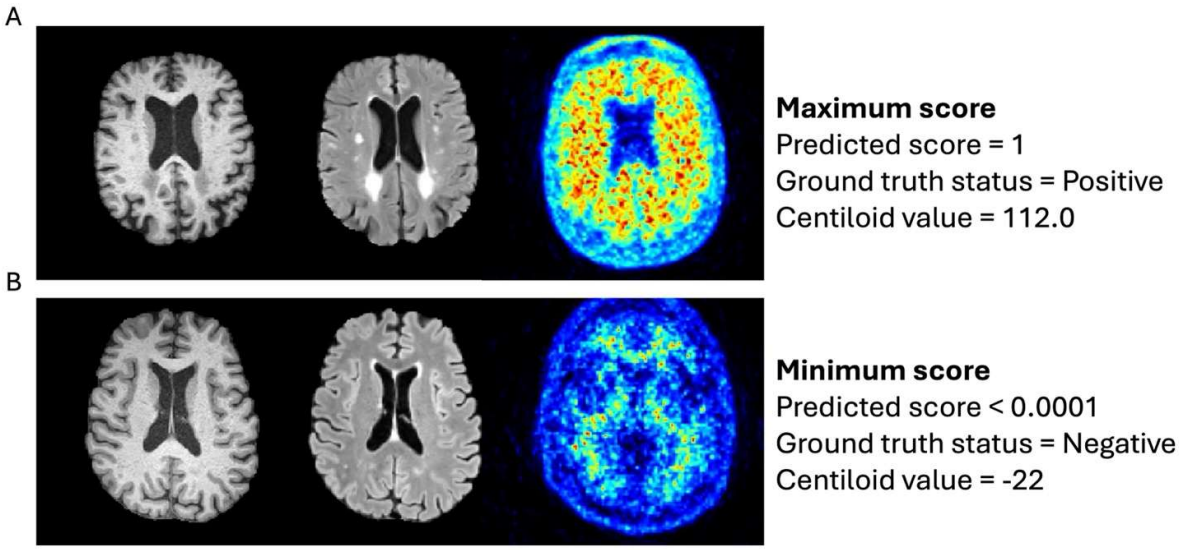

FIG 4. Representative T1w, T2-FLAIR, and amyloid-PET images from subjects with (A) the maximum and (B) the minimum predicted scores from multi-contrast deep learning model.

Prediction Based on Different Cognitive Statuses

When analyzing performance based on cognitive status, the single contrast model provided moderate AUC values in the CN and MCI groups, 0.61 (95% CI: 0.58-0.64) and 0.68 (95% CI: 0.62-0.73), respectively. However, the dementia group showed an AUC of 0.47 (95%

CI: 0.33-0.62), which was no better than chance. The multi-contrast model had higher AUC (CN: 0.66, 95% CI: 0.60-0.66; MCI: 0.71, 95% CI: 0.64-0.79; dementia: 0.61, 95% CI 0.35-0.88) and accuracy (CN: 0.63, 95% CI: 0.61-0.65; MCI: 0.66, 95% CI: 0.62-0.70; dementia: 0.66, 95% CI: 0.58-0.73) across all cognitive groups (Table 3). Notably, the highest AUC value was observed in the MCI group for the multi-contrast model.

Table 3: Performance metrics for amyloid positivity prediction by cognitive status in the internal held-out test sets. Values in parentheses represent 95% confidence intervals. Accuracy, sensitivity, and specificity metrics are presented, calculated at the Youden's *J* index operating points.

	Internal Test Set	
	Single contrast	**Multi-contrast**
CN (n=502, 59% Aβ+)		
AUC	**0.61** (0.58, 0.64)	**0.66** (0.63, 0.70)
DeLong's test	P=0.018	
Accuracy	**0.60** (0.58, 0.62)	**0.63** (0.61, 0.65)
McNemar's test	P=0.028	
Sensitivity	**0.65** (0.63, 0.67)	**0.68** (0.66, 0.70)
Specificity	**0.53** (0.51, 0.55)	**0.55** (0.53, 0.57)
MCI (n=97, 55% Aβ+)		
AUC	**0.68** (0.62. 0.73)	**0.71** (0.64. 0.79)
DeLong's test	P=0.435	
Accuracy	**0.66** (0.62, 0.71)	**0.66** (0.62, 0.70)
McNemar's test	P=0.999	
Sensitivity	**0.76** (0.72, 0.80)	**0.70** (0.66, 0.74)
Specificity	**0.55** (0.50, 0.59)	**0.62** (0.57, 0.66)
Dementia (n=31, 94% Aβ+)		
AUC	**0.47** (0.33, 0.62)	**0.61** (0.35, 0.88)
DeLong's test	P=0.366	
Accuracy	**0.70** (0.63, 0.76)	**0.66** (0.58, 0.73)
McNemar's test	P=0.349	
Sensitivity	**0.74** (0.68, 0.81)	**0.68** (0.61, 0.76)
Specificity	**0.10** (0.05, 0.15)	**0.30** (0.23, 0.37)

Prediction Performance in the External Test Set

Performance on the external test set was comparable to that of the internal test set, as shown in Table 2. As in the internal test set, the multi-contrast model outperformed the single-contrast model, exhibiting superior performance in terms of AUC (0.65, 95% CI: 0.60–0.71 versus 0.54, 95% CI: 0.49–0.60; P=0.014) and accuracy (0.62, 95% CI: 0.58–0.65 versus 0.53, 95% CI: 0.49–0.56; P<0.001) (Table 2 and Figure 3B). This demonstrates that the models, trained on publicly available datasets, generalize effectively to data from a different center and population.

DISCUSSION

Amyloid-PET is a gold standard in vivo imaging method to detect amyloid deposition in the brain. Despite its precision, the technique has drawbacks such as limited accessibility, high costs, radioactive dose, and the need for invasive procedures such as intravenous access. In this study, we sought to estimate amyloid status as determined by PET using readily accessible, non-contrast MRI images through an entirely data-driven deep learning approach, without requiring hand-crafted features such as cortical segmentations. Using the largest cohort yet reported with validation on an external test set, we evaluated a compute-efficient deep learning model that incorporated either a single contrast (T1w) or multiple contrasts (T1w and T2-FLAIR) as inputs. The multi-contrast model showed the best performance, achieving an AUC of 0.65, accuracy of 0.61, sensitivity of 0.78, and specificity of 0.39 in an external test set, significant improvements over the model trained with a single contrast for all test groups evaluated. These findings suggest that adding multi-contrast MR information significantly enhances the ability of a deep learning model to distinguish amyloid status.

T2-FLAIR images highlight regions of abnormality with positive contrast, making them sensitive to detect subtle pathology, including WMH that are thought to represent small vessel ischemic disease in elderly subjects. The prediction results across different cognitive

statuses demonstrated a clear advantage when T2-FLAIR images were added to T1w images, regardless of cognitive status. The reason for the improvement could relate to the better WMH delineation, as a previous study showed regional associations between WMH and amyloid accumulation.[19]

Previous studies have demonstrated the potential of using MRI data to predict amyloid status as determined by amyloid-PET scans through machine learning methods.[9–16] The majority of these studies were based on single contrast T1w images. Most of the previous works used hand-crafted features such as segmented cortical volumes as inputs to the machine learning model, a step that can be prone to error and require manual correction. Additionally, many of these previous studies also used demographic, genetic, and clinical information, which is not always available, and which may be subjective, such as distinguishing between MCI and dementia. In our study, we excluded any demographic, genetic, or clinical information, in order to focus only on the information in the images and the impact of using multiple contrasts. Our model using only T1w images yielded results comparable to those in a previous whole-image deep learning study.[9] Other previous studies that utilized MRI volumes reported an AUC of 0.74[15] and balanced accuracy of 0.76,[16] though limited details were provided in these reports. However, all of these previous studies[9,15,16] used only T1w images from the ADNI cohort. Although some studies have incorporated hand-crafted imaging features derived from T2-FLAIR,[10,33] such as the volume of WMH, none have directly employed T2-FLAIR images themselves. Also, this study is the first to report performance in an external test set, a critical step to determine a model's generalizability across different populations and imaging protocols.

Plasma biomarkers of AD have shown promising performance in predicting amyloid deposition, and are less invasive and more cost-effective than PET imaging or CSF sampling.[34,35] However, these methods rely on laboratory-based analytical systems, and the high cost and limited accessibility of ultra-sensitive platforms still restrict their widespread use.[35] In contrast, MRI scans are routinely acquired as part of the work-up of patients with memory concerns, primarily in order to exclude alternative pathologies, and both T1w and T2-FLAIR sequences are part of most standard protocols. While MRI is more costly than a blood test, it is invariably obtained as part of the workup and, therefore, does not add additional cost to the typical dementia evaluation. Also, for this reason, MRI scans may enable opportunistic screening, given that many MRI studies are performed in elderly patients for non-dementia reasons, such as headache and dizziness. With a very specific threshold, models such as the ones detailed in this report could be used to flag patients at high likelihood of amyloid positivity. Although not yet intended for standalone clinical use, such models could serve as one component within a broader, multimodal screening or risk stratification framework, potentially increasing access for individuals who might not present to neurology clinics or who live in areas where neurology clinics are scarce or overburdened.

There are several limitations to this study. While the models predicted amyloid status at better than chance levels, the AUCs were typically in the 0.6-0.7 range, reflecting moderate performance. There are probably several reasons for this, the most important of which is that amyloid changes are known to precede major structural changes in the brain by decades.[36] Very early cases of amyloid positivity may not show sufficient MRI changes for prediction. Next, although large by neuroimaging standards, the dataset size may still be insufficient to fully extract all of the relevant information from the images. Also, this study does not specifically evaluate how the addition of T2-FLAIR images improved the model predictions. Although we hypothesize that better visualization of WMHs on T2-FLAIR contributed to more accurate amyloid status estimations, the precise mechanisms are not fully explained. Based on the comparison between the maximum and minimum predicted score cases (Figure 4), the network seems capable of capturing such abnormalities. Additionally, just using dual-modality input (T1w + T2-FLAIR) may reduce the risk of relying on noise or artifacts from a single modality, thereby enhancing model robustness. We have not employed saliency maps to identify regions driving predictions as used in some prior works,[9] given the growing understanding of their limitations with respect to variability and explainability.[37,38] Furthermore, the developed model may perform less well in non-AD pathology cases, as our training cohort was enriched for AD pathology. The premise of predicting amyloid status using imaging features from T1w and T2-FLAIR, which are primarily associated with neurodegeneration and vascular burden, is also a limitation. While amyloid pathology is a hallmark of Alzheimer's disease, it is generally believed to precede neurodegeneration. Moreover, the consistently lower specificity compared to sensitivity across cognitive groups underscores the challenge of accurately identifying amyloid-negative individuals based solely on structural MRI imaging. Also, we have not added demographic and clinical variables into the model, which suggests that the performance presented represents a lower bound. Multiple prior studies have shown that adding this information improves predictions.[9–12,14,15] However, we chose to focus on the question of whether using multiple image contrasts improved performance; there are many methods to incorporate clinical with imaging data, which is an area for future research. Finally, we defined amyloid positivity using tracer-specific centiloid thresholds. The centiloid scale was developed to standardize amyloid PET measures across tracers, but some variability still exists. To reduce this, we locally processed the majority of the datasets (ADNI, A4, and Stanford datasets) using a pipeline aligned with the ADNI imaging core's and applied thresholds recommended by the ADNI imaging core.[27] For the OASIS dataset, which was processed separately, we used the threshold published by the OASIS group. While visual reads are often used in clinical trials, we focused on centiloid-based definitions to ensure consistency across datasets. We acknowledge that some residual variability remains despite these efforts.

CONCLUSIONS

In conclusion, we found that deep learning models with volumetric MRI data only as inputs can be used to predict amyloid status with an AUC of about 0.7. The use of multi-contrast MRI significantly improved the predictions, and this generalized to an external data set. Given the ubiquity of MRI scanning of elderly subjects, such a model may have value to identify potential amyloid positive patients in an opportunistic screening paradigm, flagging subjects who do not typically have access to specialty neurological care. Future work will explore whether additional MRI sequences and/or clinical and demographic features may further enhance amyloid prediction from MRI.

ACKNOWLEDGMENTS

We thank Duygu Tosun-Turgut, Ph.D., at UCSF for her invaluable insights and feedback on this work. We also thank Tie Liang at Stanford

University for her support with statistical analyses. De-identified data are available to qualified researchers upon approved request to the senior author and Stanford ADRC. The codes for the proposed models and pretrained weights will be made available at https://github.com/dknkim/Amyloid upon publication.

For ADNI data: Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.;Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.;Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

For OASIS-3 data: Data were provided by OASIS (OASIS-3, Longitudinal Multimodal Neuroimaging: Principal Investigators: T. Benzinger, D. Marcus, J. Morris, were supported by NIH Grants: NIH P30 AG066444, P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, and R01 EB009352).

For A4 data: The A4 Study is a secondary prevention trial in preclinical Alzheimer's disease, aiming to slow cognitive decline associated with brain amyloid accumulation in clinically normal older individuals. The A4 Study is funded by a public-private-philanthropic partnership, including funding from the National Institutes of Health-National Institute on Aging, Eli Lilly and Company, Alzheimer's Association, Accelerating Medicines Partnership, GHR Foundation, an anonymous foundation and additional private donors, with in-kind support from Avid and Cogstate. The companion observational Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) Study is funded by the Alzheimer's Association and GHR Foundation. The A4 and LEARN Studies are led by Dr. Reisa Sperling at Brigham and Women's Hospital, Harvard Medical School and Dr. Paul Aisen at the Alzheimer's Therapeutic Research Institute (ATRI), University of Southern California. The A4 and LEARN Studies are coordinated by ATRI at the University of Southern California, and the data are made available through the Laboratory for Neuro Imaging at the University of Southern California. The participants screening for the A4 Study provided permission to share their de-identified data in order to advance the quest to find a successful treatment for Alzheimer's disease. We would like to acknowledge the dedication of all the participants, the site personnel, and all of the partnership team members who continue to make the A4 and LEARN Studies possible. The complete A4 Study Team list is available on: www.actcinfo.org/a4-study-team-lists.

REFERENCES

1. Rasmussen J, Langerman H. Alzheimer's disease–why we need early diagnosis. *Degenerative neurological and neuromuscular disease* 2019. [Epub ahead of print].
2. Van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in Early Alzheimer's Disease. *N Engl J Med* 2023;388:9–21.
3. Rezai Ali R., D'Haese Pierre-Francois, Finomore Victor, et al. Ultrasound Blood–Brain Barrier Opening and Aducanumab in Alzheimer's Disease. *New England Journal of Medicine* 2024;390:55–62.
4. Budd Haeberlein S, Aisen PS, Barkhof F, et al. Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. *J Prev Alz Dis* https://doi.org/10.14283/jpad.2022.30.
5. Jack Jr CR, Wiste HJ, Vemuri P, et al. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain* 2010;133:3336–48.
6. Huijbers W, Mormino EC, Schultz AP, et al. Amyloid-β deposition in mild cognitive impairment is associated with increased hippocampal activity, atrophy and clinical progression. *Brain* 2015;138:1023–35.
7. Insel PS, Mattsson N, Donohue MC, et al. The transitional association between β-amyloid pathology and regional brain atrophy. *Alzheimer's & Dementia* 2015;11:1171–9.
8. Kang KM, Sohn C-H, Byun MS, et al. Prediction of Amyloid Positivity in Mild Cognitive Impairment Using Fully Automated Brain Segmentation Software. *NDT* 2020;Volume 16:1745–54.
9. Lew CO, Zhou L, Mazurowski MA, et al. MRI-based Deep Learning Assessment of Amyloid, Tau, and Neurodegeneration Biomarker Status across the Alzheimer Disease Spectrum. *Radiology* 2023;309:e222441.
10. Ten Kate M, Redolfi A, Peira E, et al. MRI predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study. *Alzheimer's research & therapy* 2018;10:1–12.
11. Kang SH, Cheon BK, Kim J-S, et al. Machine learning for the prediction of amyloid positivity in amnestic mild cognitive impairment. *Journal of Alzheimer's Disease* 2021;80:143–57.
12. Ansart M, Epelbaum S, Gagliardi G, et al. Reduction of recruitment costs in preclinical AD trials: validation of automatic pre-screening algorithm for brain amyloidosis. *Statistical methods in medical research* 2020;29:151–64.
13. Petrone PM, Casamitjana A, Falcon C, et al. Prediction of amyloid pathology in cognitively unimpaired individuals using voxel-wise analysis of longitudinal structural brain MRI. *Alzheimer's Research & Therapy* 2019;11:1–13.
14. Pekkala T, Hall A, Ngandu T, et al. Detecting amyloid positivity in elderly with increased risk of cognitive decline. *Frontiers in Aging Neuroscience* 2020;12:228.
15. Tosun D, Veitch D, Aisen P, et al. Detection of β-amyloid positivity in Alzheimer's Disease Neuroimaging Initiative participants with demographics, cognition, MRI and plasma biomarkers. *Brain communications* 2021;3:fcab008.
16. Chattopadhyay T, Ozarkar SS, Buwa K, et al. Comparison of deep learning architectures for predicting amyloid positivity in Alzheimer's disease, mild cognitive impairment, and healthy aging, from T1-weighted brain structural MRI. *Front Neurosci* 2024;18.
17. Gordon BA, Najmi S, Hsu P, et al. The effects of white matter hyperintensities and amyloid deposition on Alzheimer dementia. *NeuroImage: Clinical* 2015;8:246–52.
18. Provenzano FA, Muraskin J, Tosto G, et al. White matter hyperintensities and cerebral amyloidosis: necessary and sufficient for clinical expression of Alzheimer disease? *JAMA neurology* 2013;70:455–61.

19. Graff-Radford J, Arenaza-Urquijo EM, Knopman DS, et al. White matter hyperintensities: relationship to amyloid and tau burden. *Brain* 2019;142:2483–91.
20. Bakshi R, Ariyaratana S, Benedict RH, et al. Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions. *Archives of neurology* 2001;58:742–8.
21. Gawne-Cain ML, O'Riordan JI, Thompson AJ, et al. Multiple sclerosis lesion detection in the brain: a comparison of fast fluid-attenuated inversion recovery and conventional T2-weighted dual spin echo. *Neurology* 1997;49:364–70.
22. Johns E, Vossler HA, Young CB, et al. Florbetaben amyloid PET acquisition time: Influence on Centiloids and interpretation. *Alzheimer's & Dementia* 2024;20:5299–310.
23. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight j* 2009;2:1–35.
24. Isensee F, Schell M, Pflueger I, et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Human brain mapping* 2019;40:4952–64.
25. LaMontagne PJ, Benzinger TL, Morris JC, et al. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. https://doi.org/10.1101/2019.12.13.19014902.
26. Landau SM, Ward TJ, Murphy A, et al. Quantification of amyloid beta and tau PET without a structural MRI. *Alzheimers Dement* 2023;19:444–55.
27. Chadwick T, Murphy AE, Lee J, et al. SCAN Amyloid PET MRI-free Processing. 2024. [Epub ahead of print].
28. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International conference on machine learning*. PMLR; 2019:6105–14.
29. Cardoso MJ, Li W, Brown R, et al. Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:221102701* 2022. [Epub ahead of print].
30. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980* 2014. [Epub ahead of print].
31. Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:160803983* 2016. [Epub ahead of print].
32. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
33. Palmqvist S, Insel PS, Zetterberg H, et al. Accurate risk estimation of β-amyloid positivity to identify prodromal Alzheimer's disease: Cross-validation study of practical algorithms. *Alzheimer's & Dementia* 2019;15:194–204.
34. Nakamura A, Kaneko N, Villemagne VL, et al. High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nature* 2018;554:249–54.
35. Pais MV, Forlenza OV, Diniz BS. Plasma Biomarkers of Alzheimer's Disease: A Review of Available Assays, Recent Developments, and Implications for Clinical Practice. *J Alzheimers Dis Rep* 2023;7:355–80.
36. Hampel H, Hardy J, Blennow K, et al. The Amyloid-β Pathway in Alzheimer's Disease. *Mol Psychiatry* 2021;26:5481–503.
37. Saporta A, Gui X, Agrawal A, et al. Benchmarking saliency methods for chest X-ray interpretation. *Nat Mach Intell* 2022;4:867–78.
38. Zhang J, Chao H, Dasegowda G, et al. Revisiting the Trustworthiness of Saliency Methods in Radiology AI. *Radiol Artif Intell* 2023;6:e220221.