




¹⁸F-FDG-PET-based deep learning for predicting cognitive decline in non-demented elderly across the Alzheimer's disease clinical spectrum

Beomseok Sohn , MD, PhD¹, Seok Jong Chung, MD, PhD², Jeong Ryong Lee, BSc³,
Dosik Hwang, PhD³, Wanying Xie, MBBS⁴, Ling Ling Chan , MD^{5,6},
Yoon Seong Choi , MD, PhD^{7,8,*}, for the Harvard Brain Imaging Study[#];
for the Japanese Alzheimer's Disease Neuroimaging Initiative[#];
for the Alzheimer's Disease Neuroimaging Initiative[#]

¹Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, 06351, Korea

²Department of Neurology, Yonsei University College of Medicine, Seoul, 03722, Korea

³Medical Artificial Intelligence Lab, School of Electrical and Electronic Engineering, Yonsei University, Seoul, 03722, Korea

⁴Department of Nuclear Medicine and Molecular Imaging, Singapore General Hospital, Singapore, 169608, Singapore

⁵Department of Diagnostic Radiology, Singapore General Hospital, Singapore, 169608, Singapore

⁶Duke-NUS Medical School, Singapore, 169857, Singapore

⁷Department of Diagnostic Radiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 117597, Singapore

⁸Clinical Imaging Research Centre, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 117597, Singapore

*Corresponding author: Yoon S. Choi, MD, PhD, Department of Diagnostic Radiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119074, Singapore (yoonseong.choi07@gmail.com)

[#]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, the Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) database deposited in the National Bioscience Database Center Human Database, Japan (Research ID: hum0043.v1, 2016), and Harvard Aging Brain Study (HABS). As such, the investigators of ADNI, J-ADNI and HABS provided data but did not participate in analysis or writing of this report. A complete listing of ADNI, J-ADNI, and HABS investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf (ADNI), <https://humandbs.biosciencedbc.jp/en/hum0043-j-adni-authors> (J-ADNI), and <https://habs.mgh.harvard.edu/> (HABS).

Abstract

Background: With disease-modifying treatments for Alzheimer's disease (AD), prognostic tools for the pre-dementia stage are needed. This study aimed to evaluate the prognostic value of an ¹⁸F-fluorodeoxyglucose-positron emission tomography (¹⁸F-FDG-PET)-based deep-learning (DL) model in the pre-dementia stage of mild cognitive impairment (MCI) and normal cognition (NC).

Materials and Methods: A ¹⁸F-FDG-PET-based DL model was developed to classify diagnosis of AD-dementia vs NC using AD Neuroimaging Initiative (ADNI) and Japanese-ADNI (J-ADNI) datasets ($n = 756$), which provided the degree of similarity to AD-dementia. The prognostic value of the DL output for cognitive decline was assessed in the ADNI MCI ($n = 663$), J-ADNI MCI ($n = 129$), and Harvard Aging Brain Study (HABS) NC ($n = 274$) participants using Cox regression and calculating the integrated area under the time-dependent ROC curves (iAUC), along with clinical information and ¹⁸F-FDG-PET standardized uptake value ratio (SUVR). Subgroup analysis in the amyloid-positive ADNI MCI participants was performed using Cox regression and calculating the area under the time-dependent ROC (tdAUC) curves at 4-year follow-up to assess prognostic value of DL output over clinical information, ¹⁸F-FDG-PET SUVR, and amyloid PET Centiloids.

Results: DL output remained independently prognostic among other factors in all three datasets ($P < .05$ for all by Cox regression). By adding DL output to other prognostic factors, prediction significantly improved in ADNI-MCI (iAUC differences 0.020 [0.007-0.034] before and after adding DL output) and improved without statistical significance in J-ADNI (0.020 [-0.005 to 0.044], and HABS-NC sets (0.059 [-0.003 to 0.126]). DL output showed independent ($P = .002$ by Cox regression) and significant added prognostic value (tdROC difference 0.019 [< 0.001 -0.036]) over clinical information, ¹⁸F-FDG-PET SUVR, and Centiloids in the amyloid-positive ADNI MCI participants.

Conclusion: The ¹⁸F-FDG-PET-based DL model demonstrated the potential to improve cognitive decline prediction beyond clinical information, and conventional measures from ¹⁸F-FDG-PET and amyloid PET and may prove useful for clinical trial recruitment and individualized management.

Keywords: Alzheimer's disease, cognitive decline, deep learning, dementia, ¹⁸F-FDG-PET, prognosis

Abbreviations ^{18}F -FDG = ^{18}F -fluorodeoxyglucose; AD = Alzheimer's disease; ADNI = Alzheimer's Disease Neuroimaging Initiative; APOE4 = apolipoprotein E4; AUC = area under the receiver operating characteristic curve; DL = deep learning; FAQ = Functional Activities Questionnaire; HABS = Harvard Aging Brain Study; iAUC = integrated area under the time-dependent receiver operating characteristic curve; J-ADNI = Japanese Alzheimer's Disease Neuroimaging Initiative; MCI = mild cognitive impairment; MMSE = Mini-Mental State Examination; NC = normal cognition; PACC = Preclinical Alzheimer's Cognitive Composite; PET = positron emission tomography; ROC = receiver operating characteristic; SUV = standardized uptake value; SUVR = standardized uptake value ratio; T1W1 = T1-weighted; tdAUC = area under the time-dependent receiver operating characteristic curve

Summary

^{18}F -FDG-PET-based deep learning model may help improve the prediction of cognitive decline in non-demented elderly individuals beyond conventional ^{18}F -FDG-PET measures as well as clinical characteristics.

Key Results

- A deep learning (DL) model captured the hypometabolic patterns related to Alzheimer's disease (AD) on ^{18}F -FDG-PET images and provided the degree of similarity to AD-dementia.
- The DL output showed independent and added prognostic value over clinical characteristics and conventional ^{18}F -FDG-PET for predicting cognitive decline in non-demented elderly.
- The DL output also showed independent and added prognostic value in the amyloid-positive subgroup even beyond amyloid PET measures.

Introduction

Alzheimer's disease (AD) is the leading cause of dementia in the elderly. The introduction of the ATN (amyloid, tau, neurodegeneration) biomarker-driven diagnosis of AD, coupled with recent FDA-approved treatments and clinical trials targeting early stages of AD, underscores the importance of identifying individuals with pre-dementia stages of mild cognitive impairment (MCI) or normal cognition (NC) before the onset of overt dementia symptoms.^{1–3} However, due to the heterogeneous prognosis of the non-demented elderly, with only 10%-15% of individuals with MCI progressing to AD-dementia,⁴ there is a critical need for accurate prognostic tools to identify potential treatment beneficiaries and to aid in the interpretation of ongoing AD clinical trials.^{5–8}

^{18}F -fluorodeoxyglucose (^{18}F -FDG)-positron emission tomography (PET), used as a measure of N (neurodegeneration) in ATN, provides earlier detection of neuronal degeneration than MRI changes.⁹ Although conventional metrics such as standardized uptake value (SUV) ratio (SUVR) of ^{18}F -FDG PET have been evaluated for prognostic prediction in AD, their accuracy remains variable.^{10–14}

Deep learning (DL) is capable of extracting relevant features directly from images, outperforming human visual cognition and conventional imaging measures.¹⁵ Previous studies have applied DL to ^{18}F -fluorodeoxyglucose-positron emission tomography (^{18}F -FDG-PET) for cross-sectional diagnosis, such as differentiating between NC and AD-dementia.^{16,17} However, the prognostic potential of such DL models for predicting future cognitive decline remains unexplored. Several studies have used FDG-PET-based machine learning for prognosis,^{18–20} but they mainly trained models on MCI subjects only. This approach did not take into account the broader, continuous disease trajectory from NC to AD-dementia, thereby limiting its utility to MCI subjects alone. In addition, previous studies faced challenges including small sample sizes, limited ethnic diversity, and lack of integration of conventional imaging measures such as SUVR.

We hypothesized that a DL model trained on NC and AD-dementia, originally designed to discriminate AD-dementia from NC, could be repurposed to predict cognitive trajectories in the pre-dementia stages, providing insight into where

predementia individuals are on the spectrum of cognitive decline from NC to AD-dementia. In this study, we attempted to assess the prognostic potential of ^{18}F -FDG-PET-based DL in a clinical context. We predicted cognitive decline in non-demented elderly individuals of different stages and ethnicities in the presence of conventional ^{18}F -FDG-PET measures (ie, SUVR) and clinical information for the entire follow-up period. We also assessed our DL within the biomarker-based AD framework in a subgroup where amyloid burden was considered. Cognitive decline in our study was defined as either conversion from MCI to AD-dementia or conversion from NC to MCI. Our aim was to develop and test the prognostic value of an ^{18}F -FDG-PET-based DL model for the prediction of cognitive decline in elderly individuals with MCI and NC.

Materials and methods

Study population

The patient enrollment process is shown in Figure 1. Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI, <https://adni.loni.usc.edu>), the Japanese ADNI (J-ADNI, <https://www.j-adni.org>)²⁰ studies, and the Harvard Aging Brain Study (HABS, <https://habs.mgh.harvard.edu/>).²¹ The details of ADNI, J-ADNI, and HABS are described in Supplementary Material. Additionally, the inclusion criteria for the three datasets and the number of individuals included in the analysis are detailed in Figure 1 and the Supplementary Material.

For prognostic value assessment, follow-up duration and time to cognitive decline (ie, conversion from MCI to AD-dementia in the ADNI and J-ADNI MCI groups and from NC to MCI in the HABS NC group) were recorded in the ADNI MCI, J-ADNI MCI, and HABS NC groups. As the number of individuals progressing from MCI to AD-dementia decreases sharply after 9 years of follow-up (Figure S1), we used the first 9 years of follow-up data for the ADNI MCI group. In addition, amyloid PET quantification (Centiloid) and amyloid-positivity information of the ADNI MCI group were recorded from the ADNI website for the subgroup analysis.

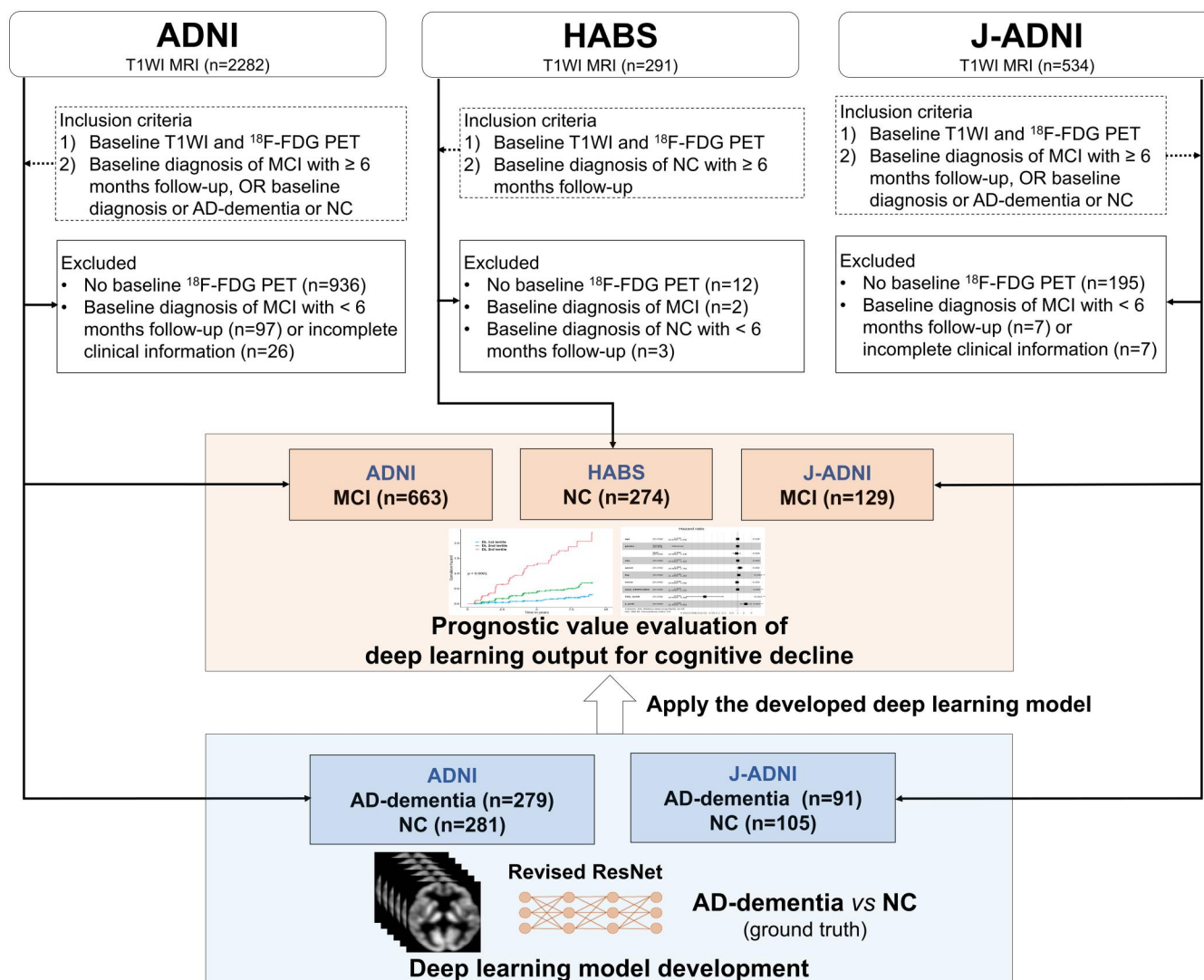


Figure 1. Participant enrollment flowchart. Abbreviations: ^{18}F -FDG = ^{18}F -fluorodeoxyglucose; AD = Alzheimer's disease; ADNI = Alzheimer's disease neuroimaging initiative; HABS = Harvard Aging Brain Study; J-ADNI = Japanese-Alzheimer's Disease Neuroimaging Initiative; MCI = mild cognitive impairment; NC = normal cognition; PET = positron emission tomography.

The characteristics of the AD-dementia/NC groups from the ADNI and J-ADNI datasets are detailed in Table S1; the number of participants was 281/279 and 105/91, respectively. Table 1 summarizes the characteristics of the ADNI MCI ($n=663$), J-ADNI MCI ($n=129$) and HABS NC ($n=274$) groups.

Study design

The study design is shown in Figure 1. First, a DL model for classifying AD-dementia vs NC was developed, which provided the degree of similarity to AD-dementia from ^{18}F -FDG-PET images (Figure S2). The ADNI AD-dementia/NC set was divided into the development ($n=487$) and test sets ($n=73$) with stratification for the AD-dementia/NC ratio. The DL model was developed using the ADNI AD-dementia/NC development set and tested on the ADNI AD-dementia/NC test set, and externally tested on the J-ADNI AD-dementia/NC set to assess the diagnostic performance on an Asian population. Saliency maps were generated and averaged from the ADNI AD-dementia/NC test set using layer-wise relevance

propagation, calculating relevance in each anatomical area.²² The DL model development and saliency maps are detailed in the Supplementary Material.

Second, after the DL model development, the prognostic value of DL output for cognitive decline was evaluated in conjunction with clinical features and ^{18}F -FDG-PET SUVR in the ADNI MCI, J-ADNI MCI, and HABS NC sets. The prognostic value of DL output was assessed in the three different datasets to assess the applicability of our model across different ethnicities and baseline diagnoses. In addition, the subgroup analysis for amyloid-positive participants from the ADNI MCI was performed to assess the applicability of our model in the biomarker-based framework of AD.

Image processing

The T1WI and ^{18}F -FDG-PET images were downloaded from the ADNI, J-ADNI, and HABS data inventory. The image processing for the DL model included registration of ^{18}F -FDG-PET images to T1WI and MNI standard space by using FSL FLIRT, background cropping, resampling to a $72 \times 72 \times 72$

Table 1. Study population characteristics.

Characteristic	ADNI MCI	J-ADNI MCI	HABS NC
Baseline diagnosis	MCI	MCI	NC
Number of subjects	663	129	274
Age in years (mean ± SD)	72.6 ± 7.5	72.4 ± 5.6	73.7 ± 6.1
Gender, number (%)			
Female	272 (41.0%)	66 (51.2%)	164 (59.9%)
Male	391 (59.0%)	63 (48.8%)	110 (40.1%)
Education in years (mean ± SD)	16.1 ± 2.7	13.3 ± 2.9	15.9 ± 3.0
APOE4 allele			
0	340 (51.3%)	59 (45.7%)	196 (71.5%)
1	252 (38.0%)	62 (48.1%)	74 (27.0%)
2	71 (10.7%)	8 (6.2%)	4 (1.5%)
FAQ (mean ± SD)	2.9 ± 3.8	3.9 ± 4.3	N/A
MMSE (mean ± SD)	27.8 ± 1.8	26.4 ± 1.7	N/A
PACC (mean ± SD)	N/A	N/A	−0.1 ± 0.7
Follow-up time (quartiles in years)	2.5-7.0	2.9-3.0	4.5-5.2
Progressors	224 (33.8%)	66 (51.2%)	22 (8.0%)
Time to progression (quartiles in years)	1.0-4.0	0.6-2.0	3.0-5.1

Abbreviations: ADNI = Alzheimer’s disease neuroimaging initiative; APOE4 = Apolipoprotein E4; FAQ = functional activities questionnaire; HABS = Harvard Aging Brain Study; J-ADNI = Japanese-Alzheimer’s disease neuroimaging initiative; MCI = mild cognitive impairment; MMSE = mini-mental state examination; NC = normal cognition; PACC = Preclinical Alzheimer’s Cognitive Composite.

matrix, and min–max normalization of voxel values, where the normalized voxel value is calculated as $\text{voxel}_{\text{normalized}} = (\text{voxel}_{\text{original}} - \text{voxel}_{\text{min}})/(\text{voxel}_{\text{max}} - \text{voxel}_{\text{min}})$, where $\text{voxel}_{\text{min}}$ and $\text{voxel}_{\text{max}}$ represent the minimum and maximum values of the ^{18}F -FDG-PET voxel values.

^{18}F -FDG-PET SUVR values of the ADNI MCI set were provided by the ADNI.²³ For the J-ADNI MCI and HABS NC sets, SUVRs were calculated by the author using the posterior cingulate and precuneus as targets and the pons as a reference.²⁴ The J-ADNI MCI T1WIs were segmented with FreeSurfer, and ^{18}F -FDG-PET images were registered to T1WI for SUVR calculations, while HABS provided SUVRs of each anatomical region to calculate the SUVRs. Amyloid positivity information and Centiloid values of the ADNI MCI set were provided by the ADNI side, as described elsewhere (<https://adni.loni.usc.edu/updated-uc-berkeley-amyloid-pet-methods/>), which was used for the subgroup analysis as detailed later.

Statistical analysis

Statistical tests were performed in R, version 4.2.1 (R Foundation for Statistical Computing). Performance of the DL output in classifying AD-dementia vs NC was assessed using the area under the receiver operating characteristic (ROC) curve (AUC) and accuracy. DL-ADprob prediction of ≥ 0.5 was classified as AD-dementia, while values below 0.5 were considered NC for calculating accuracy.

In assessing the prognostic value of the DL output, the main outcome was conversion from MCI to AD-dementia in the ADNI MCI and J-ADNI MCI sets, and from NC to MCI in the HABS NC set. The prognostic value of the DL output was assessed in three ways: First, the ADNI and J-ADNI MCI sets were divided into three groups based on the DL output tertiles, while the HABS NC set was divided into two groups based on the DL output median (low-risk: DL output < the median, high-risk: DL output \geq the median) due to the small number of progressors (Table 1). The prognosis was compared between different risk groups using the log-rank test. Additionally, the SUVR median value was used to divide the ADNI MCI, J-ADNI MCI, and HABS NC sets into two risk

groups (low-risk: SUVR \geq the median, high-risk: SUVR < the median). If the SUVR median-based risk groups significantly stratified prognosis, individuals were further stratified into four different risk groups based on the median values of both SUVR and DL output, with various combinations of low and high-risk groups based on SUVR and DL output. Specifically, the prognosis was compared between the two subgroups belonging to discordant risk groups of SUVR and DL output by log-rank test. Second, the independent prognostic value of DL output was assessed by Cox regression analysis with DL output along with clinical features (ie age, gender, education, APOE4, FAQ, and MMSE for the ADNI MCI and J-ADNI MCI sets, and age, gender, education, APOE4 and PACC for the HABS NC set) and SUVR. Third, the added prognostic value of DL output over clinical features and SUVR was assessed by calculating the integrated area under the time-dependent ROC curves (iAUC) for Cox regression models with clinical features and SUVR only (Model 1) and with DL output added (Model 2). The iAUC measures prediction accuracy over time, with an iAUC of 1 indicating perfect prediction and 0.5 indicating no better than random guessing.²⁵ The iAUC values were compared between Model 1 and 2 through 1000 resampling iterations. Statistical significance was determined if the 95% CI of iAUC difference did not include zero.

A subgroup analysis of ADNI MCI participants who were amyloid-positive on baseline amyloid PET scans was performed. Similarly, Kaplan-Meier curves were plotted for risk groups based on DL output (low risk: DL power < median, high risk: DL power \geq median), and prognoses were compared using the log-rank test. The independent prognostic value of DL output was assessed by Cox regression analysis. Clinical features, ^{18}F -FDG-PET SUVR, and amyloid PET Centiloids were included as covariates. Additionally, the added prognostic value of DL output was assessed by calculating the area under the time-dependent ROC curves at 4-year follow-up (tdAUC). The tdAUCs of Cox regression models before (Model 1) and after adding DL output (Model 2) to clinical features, ^{18}F -FDG-PET SUVR and Centiloids were compared by 1000 resampling iterations.

Deep learning model availability

The trained model is available at https://github.com/yoonychoi-neuro/FDG-PET_AD. All the data used in this study are available through the respective portals.

Results

Performance of the DL in classification of AD-dementia vs NC

^{18}F -FDG-PET-based DL could distinguish between AD-dementia and NC with AUCs of 0.97 and 0.98, and accuracies of 91.8% and 90.8% on the ADNI AD-dementia/NC test set and the J-ADNI AD-dementia/NC set, respectively. On the saliency map (Figure 2A and B), the precuneus cortex (37.5%) was the most relevant, followed by the posterior cingulate cortex (21.4%), intracalcarine cortex (11.3%), and cuneal cortex (10.6%).

Prognostic value of the DL output

The DL output-based risk groups significantly stratified the prognosis of the ADNI MCI, J-ADNI MCI, and HABS NC sets ($P < .001$ for the ADNI and J-ADNI MCI sets, and $P = .034$ for the HABS NC set by log-rank test, Figure 3). In contrast, the SUVR-based risk groups significantly stratified prognosis only in the ADNI MCI set only ($P < .001$ by log-rank test), but not in the J-ADNI MCI and HABS NC sets ($P = .072$ and $.340$ by log-rank test, respectively, Figure S3). In the ADNI MCI set, the four risk groups split by the medians of SUVR and DL output significantly stratified prognoses ($P < .001$ by log-rank test, Figure 3B). Among the two groups with discordant risk from SUVR and DL output, the group with DL output-based high-risk and SUVR-based low-risk tended to have a worse prognosis than the group with DL output-based low-risk and SUVR-based high-risk, despite lower risk based on SUVR ($P = .095$ by log-rank test, Figure 3B). The representative PET images of the four groups in the ADNI MCI set are shown in Figure 4A-D.

In Cox regression analyses, the DL output remained independently prognostic among clinical features and SUVR in all datasets (HR 3.45 [2.39-4.98], $P < .001$ for the ADNI-MCI; HR 2.28 [1.16-4.46], $P = .016$ for the J-ADNI MCI; and HR 4.57 [1.52-13.75], $P = .007$ for the HABS NC sets, Table 2). The performance of the Cox regression models before and after adding DL output is shown in Table 3. The addition of DL output to clinical features and SUVR significantly improved the prognosis prediction in the ADNI MCI set (iAUC 0.768 [0.744-0.791] and 0.787 [0.766-0.811] for Model 1 and 2, respectively, iAUC difference 0.020 [0.007-0.034]), while it increased prognostic performance without statistical significance in the J-ADNI MCI (iAUC 0.694 [0.639-0.752] and 0.716 [0.657-0.772] for Model 1 and 2, respectively, iAUC difference 0.020 [-0.005 to 0.044]) and HABS NC sets (iAUC 0.817 [0.692-0.909] and 0.902 [0.751-0.951] for Model 1 and 2, respectively, iAUC difference 0.059 [-0.003 to 0.126]).

Subgroup analysis in the amyloid-positive participants

In the ADNI MCI group, 486 out of 663 participants had a baseline amyloid PET scan, of whom 258 were amyloid positive. Of the 258 amyloid-positive participants, 144 and 92 participants had progressed to AD-dementia during the entire follow-up period and at 4 years, respectively. The DL output-based risk groups significantly stratified prognosis of cognitive decline in the ADNI MCI set ($P = .001$ by log-rank test, Figure 5A). The DL output was independently prognostic in the presence of clinical features, ^{18}F -FDG-PET SUVR and Centiloids (HR 2.60 [1.40-4.80], $P = .002$ by Cox regression, Table 4). The addition of DL output to clinical information, ^{18}F -FDG-PET SUVR and Centiloids significantly improved the prediction of cognitive decline at 4-year follow-up in the amyloid-positive participants from the ADNI MCI set (tdAUC 0.720 [0.678-0.760] and 0.739 [0.691-0.783] for

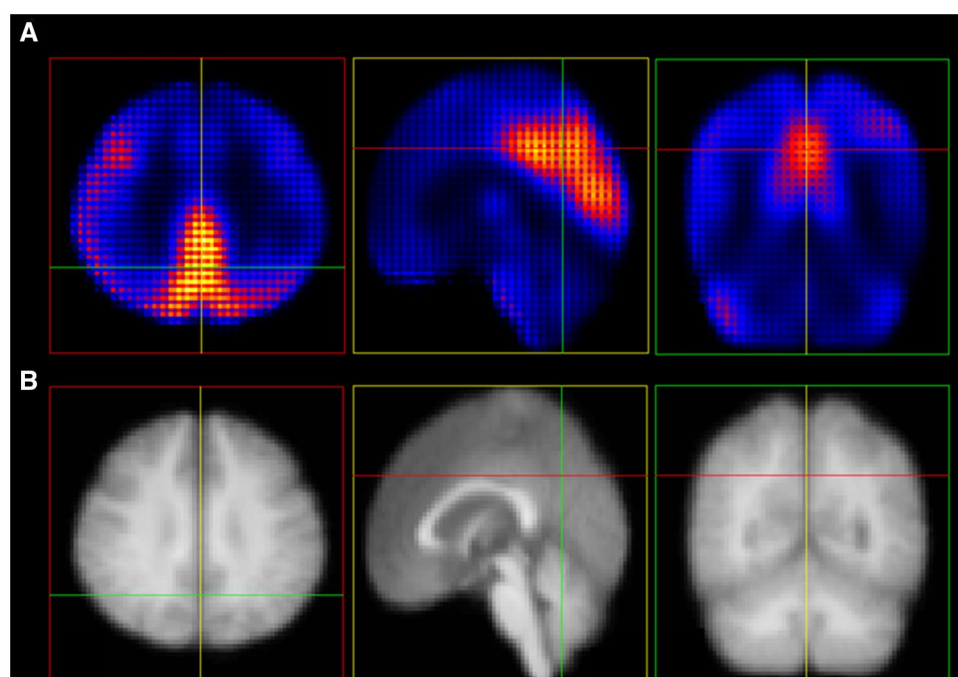


Figure 2. Saliency map of the ^{18}F -FDG-PET-based deep learning model for Alzheimer's dementia.

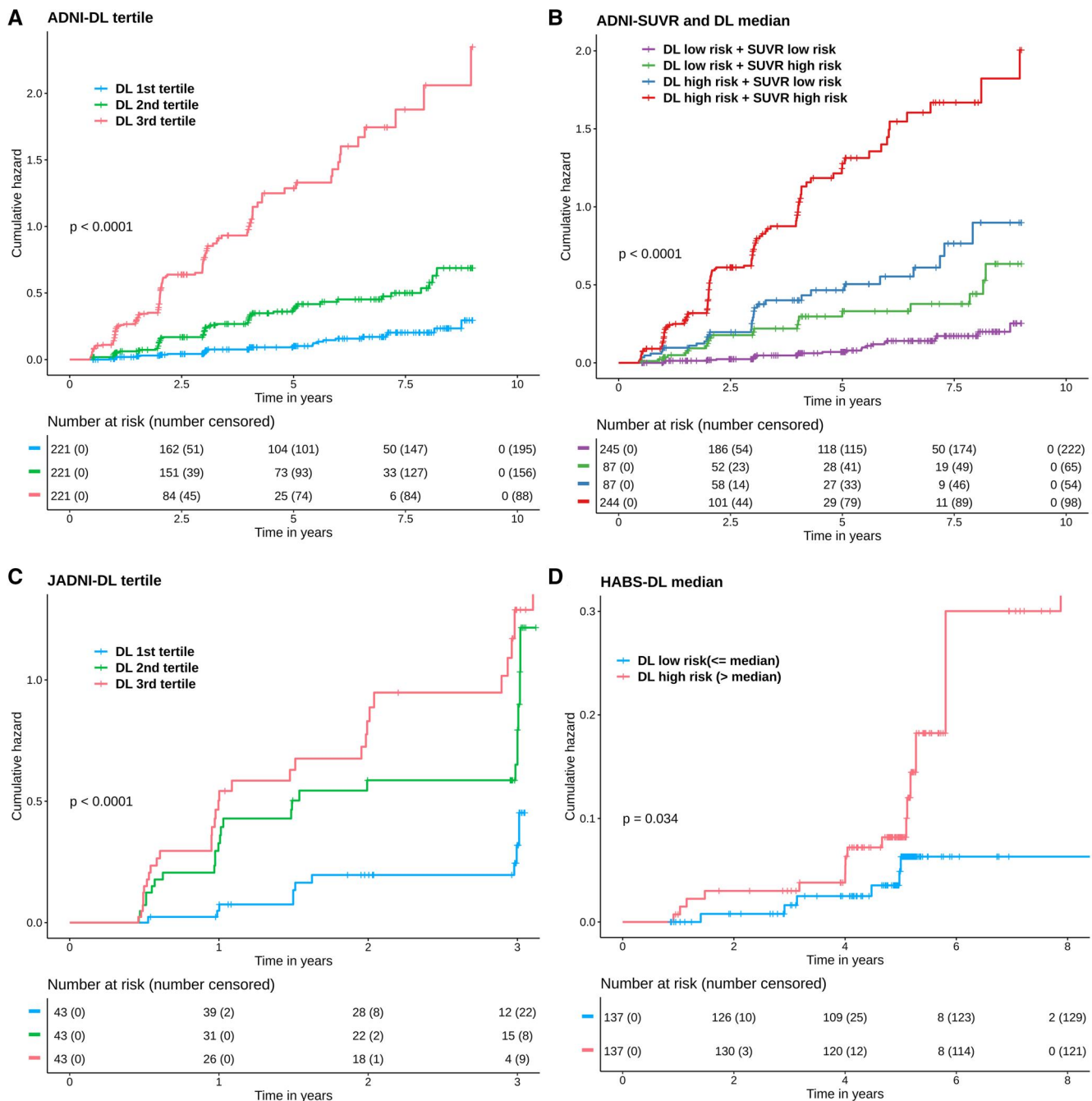


Figure 3. Kaplan-Meier curves showing the cumulative incidence of cognitive decline in the ADNI (A and B), J-ADNI (C), and HABS (D) data, stratified by DL tertiles (A and C), DL median (D) and the median of DL and ^{18}F -FDG-PET SUVR (B). Abbreviations: ADNI = Alzheimer's disease neuroimaging initiative; DL = deep-learning model for classification of AD-dementia vs NC; HABS = Harvard Aging Brain Study; J-ADNI = Japanese-Alzheimer's Disease Neuroimaging Initiative; SUVR = standardized uptake value ratio.

Model 1 and 2, respectively, tAUC difference 0.019 [<0.001 -0.036]), **Figure 5B**.

Discussion

We investigated the prognostic value of the ^{18}F -FDG-PET-based DL model, initially developed for AD-dementia diagnosis. In the predementia stage of MCI and NC, the DL output provided prognostic information, beyond ^{18}F -FDG-PET SUVR as well as clinical information. Moreover, the DL output showed prognostic potential even beyond amyloid PET measures.

Previous studies reported excellent diagnostic performance of ^{18}F -FDG-PET-based DL for AD-dementia; one study reported AUCs of 0.92-0.98 for diagnosing AD-dementia.¹⁶ However, such DL models have been tested only for diagnostic purposes, and their prognostic ability remains largely unexplored. While some research has evaluated the prognostic value of ^{18}F -FDG-PET-based machine learning such as DL-based radiomic features¹⁹ or principal component analysis,^{14,26} their model training was typically based on classification of stable vs progressive MCI based on a single follow-up time, which may not adequately capture the continuous and progressive nature of AD as it may provide limited insights

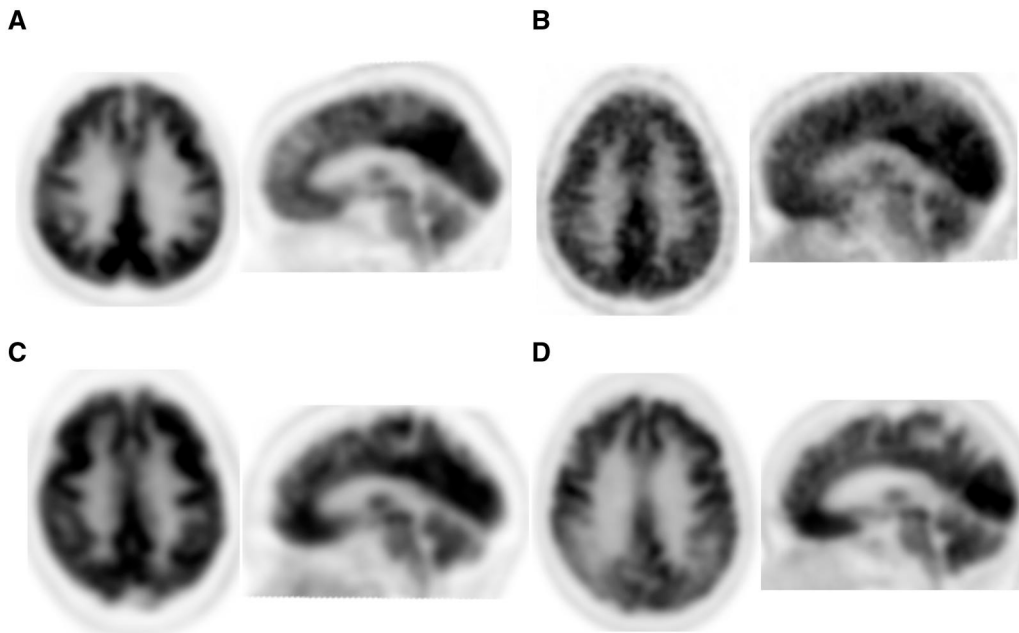


Figure 4. Representative brain PET images of the four groups in the ADNI MCI set, categorized by combinations of DL and SUVR high- and low-risk predictions. (A) DL low risk & SUVR low risk, no progression during 7.98 years of follow-up. (B) DL low risk & SUVR high risk, no progression within 8.19 years of follow-up. (C) DL high risk and SUVR low risk, progression at 1.00 years of follow-up. (D) DL high risk & SUVR high risk, progression at 0.99 years of follow-up. Abbreviations: SUVR = standardized uptake value ratio; ADNI = Alzheimer’s disease neuroimaging initiative; MCI = mild cognitive impairment.

Table 2. Cox regression model of cognitive decline with DL model output, clinical features, and ¹⁸F-FDG-PET in the ADNI MCI, J-ADNI MCI, and HABS NC sets.

	ADNI MCI		J-ADNI MCI		HABS NC	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Age (years)	1.01 (0.99-1.03)	.227	1.01 (0.96-1.06)	.786	1.09 (1.00-1.20)	.053
Gender (male vs female)	0.85 (0.65-1.12)	.254	1.07 (0.59-1.94)	.818	0.20 (0.06-0.62)	.006
Education (years)	1.01 (0.96-1.06)	.666	0.99 (0.89-1.09)	.772	1.30 (1.07-1.59)	.010
APOE4 status	1.72 (1.42-2.08)	<.001	1.31 (0.85-2.03)	.219	1.92 (0.95-3.87)	.070
MMSE	0.90 (0.84-0.98)	.013	0.91 (0.77-1.07)	.245	NA	
FAQ	1.12 (1.09-1.15)	<.001	1.15 (1.08-1.23)	<.001	NA	
PACC	NA		NA		0.19 (0.08-0.45)	<.001
SUVR	0.04 (0.01-0.14)	<.001	1.62 (0.13-20.22)	.706	2.30 (0.07-73.95)	.639
DL	3.45 (2.39-4.98)	<.001	2.28 (1.16-4.46)	.016	4.57 (1.52-13.75)	.007

Abbreviations: ADNI = Alzheimer’s disease neuroimaging initiative; APOE4. apolipoprotein E4; DL = deep-learning model for classification of AD-dementia vs NC; FAQ = Functional Activities Questionnaire; J-ADNI = Japanese-Alzheimer’s disease neuroimaging initiative; HABS = Harvard Aging Brain Study; MCI = mild cognitive impairment; MMSE = Mini-Mental State Examination; NC = normal cognition; PACC = Preclinical Alzheimer’s Cognitive Composite; SUVR = standardized uptake value ratio.

Table 3. Performance of Cox regression models in the prediction of cognitive decline.

Dataset	Predictors	iAUC (95% CI)	iAUC difference (95% CI)
ADNI MCI	Clinical features ^a + SUVR	0.768 (0.744-0.791)	NA
	Clinical features + SUVR + DL ^b	0.787 (0.766-0.811)	0.020 (0.007-0.034)
J-ADNI MCI	Clinical features + SUVR	0.694 (0.639-0.752)	NA
	Clinical features + SUVR + DL	0.716 (0.657-0.772)	0.020 (–0.005 to 0.044)
HABS NC	Clinical features + SUVR	0.817 (0.692-0.909)	NA
	Clinical features + SUVR + DL	0.902 (0.751-0.951)	0.059 (–0.003 to 0.126)

^aClinical features consisted of age, gender, years of education, APOE4 allele, FAQ and MMSE scores for the ADNI and J-ADNI data, and age, gender, years of education, APOE4 allele, and PACC score for HABS data.
^bThe DL output obtained from the ¹⁸F-FDG-PET images.
Abbreviations: ADNI = Alzheimer’s disease neuroimaging initiative; DL = deep-learning model for classification of AD-dementia vs NC; HABS = Harvard Aging Brain Study; iAUC integrated area under the time-dependent receiver operating characteristic curve J-ADNI = Japanese-Alzheimer’s disease neuroimaging initiative; MCI = mild cognitive impairment; NC = normal cognition; SUVR = standardized uptake value ratio.

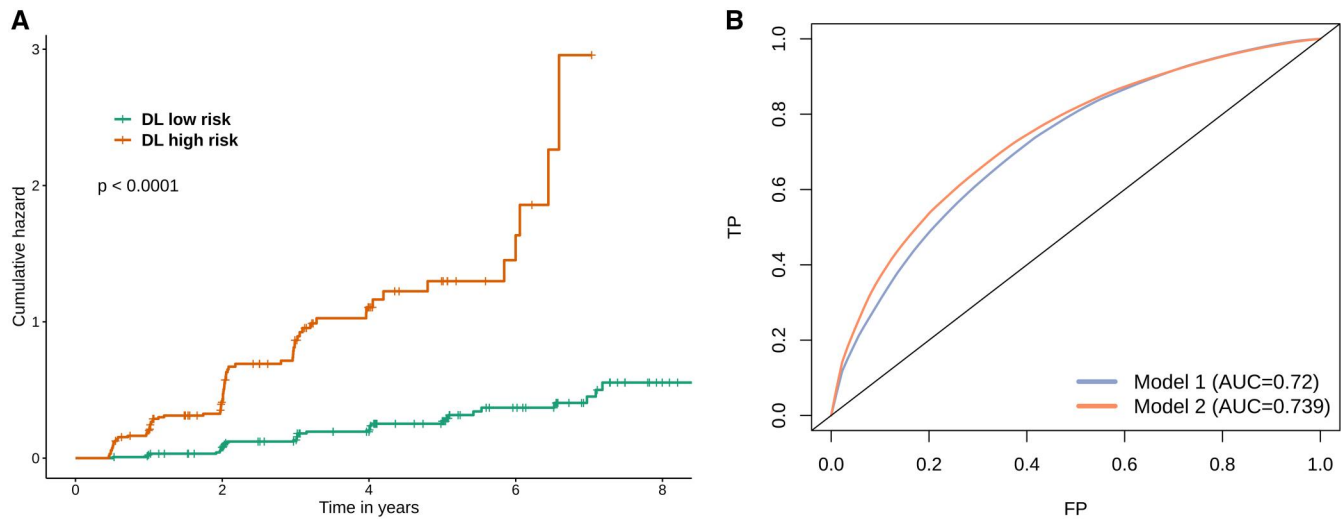


Figure 5. Cox regression models for predicting cognitive decline in the amyloid-positive participants from the ADNI MCI set. (A) Kaplan-Meier curves stratified by the DL tertiles, (B) time-dependent ROC curve at 4-year follow-up before (Model 1) and after (Model 2) adding DL output to clinical features, ¹⁸F-FDG-PET SUVR, and amyloid PET Centiloids. Abbreviations: AUC = area under the receiver operating characteristic curve; DL = deep-learning model for classification of AD-dementia vs NC.

Table 4. Cox regression models of cognitive decline in the amyloid-positive participants from the ADNI MCI set with deep learning output, clinical features, ¹⁸F-FDG-PET, and amyloid PET.

	Hazard ratio (95% CI)	P value
Age (years)	1.01 (0.98-1.04)	.538
Gender	0.89 (0.59-1.34)	.568
Education (years)	1.02 (0.94-1.10)	.668
APOE4 status	1.31 (0.97-1.77)	.082
MMSE	0.94 (0.83-1.05)	.264
FAQ	1.15 (1.10-1.20)	<.001
¹⁸ F-FDG-PET SUVR	0.02 (0.00-0.18)	<.001
Amyloid PET Centiloids	1.01 (1.00-1.01)	.005
DL	2.60 (1.40-4.80)	.002

Abbreviations: ¹⁸F-FDG = ¹⁸F-fluorodeoxyglucose; APOE4 = apolipoprotein E4; DL = deep-learning model for classification of AD-dementia vs NC; FAQ = Functional Activities Questionnaire; MMSE = Mini-Mental State Examination; PET = positron emission tomography; SUVR = standardized uptake value ratio.

by grouping MCI subjects with heterogeneous prognosis into the same group, limiting the model application to MCI subjects only. In this study, we hypothesized that a DL model trained on NC and AD-dementia could be repurposed to predict cognitive trajectories in the predementia stages, providing insight into where predementia individuals are on the broader spectrum of cognitive decline from NC to AD-dementia, benefiting both NC and MCI predementia stages. For practicality, we integrated ¹⁸F-FDG PET SUVR, clinical information, and amyloid PET measures, and used various datasets spanning different countries and ethnicities.

The excellent diagnostic performance of our model indicates its ability to detect AD-dementia-related patterns in ¹⁸F-FDG PET images across ethnicities. The saliency map is consistent with previously reported hypometabolic patterns of AD on ¹⁸F-FDG-PET, with the precuneus and posterior cingulate gyrus being one of the earliest areas of hypometabolic change.²⁷ In the assessment of prognostic value, DL output maintained its independent prognostic value among SUVR and clinical information in all datasets. However, there were variances in the hazard ratios of DL output and in the improvement in prognostic performance by adding DL

output between datasets. Differences in sample size, follow-up duration, and number of progressions are one possible explanation for such variance; the J-ADNI MCI set has the smallest sample size and shortest follow-up duration, and the HABS NC set has the smallest number of progressors, making it more difficult to show statistically significant improvement than the ADNI MCI set. Other explanations may include race-specific intrinsic differences in AD biomarkers,²⁸ and extrinsic differences between different countries and cognitive stages, such as follow-up intervals.^{29,30} Nevertheless, despite the excellent diagnostic performance in the external J-ADNI dataset, we cannot completely exclude the possibility of limited generalizability to datasets other than the ADNI. Although our model is applicable across a broad range of predementia stages, further research is needed to account for this inevitable diversity in the clinical settings.

A notable finding from our study is the stronger stratification by DL output than ¹⁸F-FDG-PET SUVR in the groups with discordant risk between DL output and SUVR in the ADNI MCI set. In addition, the DL output showed the potential to improve prediction of cognitive decline even beyond amyloid PET measures in the amyloid-positive participants. This suggests that DL output may provide neurodegeneration-related insight beyond ¹⁸F-FDG PET SUVR which also works within the ATN framework.

Our study has limitations. First, we used only ¹⁸F-FDG-PET images and amyloid PET for subgroup analysis, not including MRI and tau PET. A more comprehensive approach with diverse imaging modalities may be needed. Second, although we attempted to address the ethnic imbalance in the ADNI and HABS data by including the J-ADNI dataset, future studies are needed to generalize to broader ethnic groups. Third, although a substantial number of NC participants were enrolled from the HABS dataset, there were a relatively small number of progressors to MCI among them, resulting in less robust validation for NC subjects compared to MCI subjects. Future research is needed with larger sample size and longer follow-up.

In conclusion, our study developed an ¹⁸F-FDG-PET-based deep learning model that outputs the degree of similarity to

AD-dementia and tested its prognostic value. The DL output showed independent and added prognostic value over clinical information, ^{18}F -FDG-PET SUVR, and amyloid PET measures, suggesting the potential to aid effective clinical trial recruitment and individualized treatment in clinical practice.

Author contributions

Beomseok Sohn (Data curation, Investigation, Resources, Visualization, Writing—original draft), Seok J. Chung (Data curation, Writing—original draft), Jeong R. Lee (Visualization, Writing—review & editing), Dosik Hwang (Funding acquisition, Visualization, Writing—review & editing), Wanying Xie (Writing—review & editing), Ling L. Chan (Writing—review & editing), Yoon S. Choi (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing—review & editing)

Supplementary material

Supplementary material is available at *Radiology Advances* online.

Funding

This study was supported by Samsung Research Funding Center of Samsung Electronics (SRFC-TF2103-01), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT, 2022-31-0220), and Start-up Grant funded by National University of Singapore, Yong Loo Lin School of Medicine (A-0009827-01-00).

Conflicts of interest

Please see ICMJE form(s) for author conflicts of interest. These have been provided as [supplementary materials](#). There is no conflict of interest.

Data availability

The data used in this study are available in the Alzheimer's Disease Initiative, Japanese Alzheimer's Disease Initiative, and HABS databases, where the data can be obtained upon request.

References

- Seigny J, Chiao P, Bussiere T, et al. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature*. 2016;537(7618):50-56. <https://doi.org/10.1038/nature19323>
- Salloway S, Chalkias S, Barkhof F, et al. Amyloid-related imaging abnormalities in 2 phase 3 studies evaluating aducanumab in patients with early Alzheimer disease. *JAMA Neurol*. 2022;79(1):13-21. <https://doi.org/10.1001/jamaneurol.2021.4161>
- Cummings J, Aisen P, Apostolova LG, Atri A, Salloway S, Weiner M. Aducanumab: appropriate use recommendations. *J Prev Alzheimers Dis*. 2021;8(4):398-410. <https://doi.org/10.14283/jpad.2021.41>
- Parnetti L, Chipi E, Salvadori N, D'Andrea K, Eusebi P. Prevalence and risk of progression of preclinical Alzheimer's disease stages: a systematic review and meta-analysis. *Alzheimers Res Ther*. 2019;11(1):7. <https://doi.org/10.1186/s13195-018-0459-7>
- Farias ST, Mungas D, Reed BR, Harvey D, DeCarli C. Progression of mild cognitive impairment to dementia in clinic- vs community-based cohorts. *Arch Neurol*. 2009;66(9):1151-1157. <https://doi.org/10.1001/archneurol.2009.106>
- Dubois B, Villain N, Frisoni GB, et al. Clinical diagnosis of Alzheimer's disease: recommendations of the International Working Group. *Lancet Neurol*. 2021;20(6):484-496. [https://doi.org/10.1016/S1474-4422\(21\)00066-1](https://doi.org/10.1016/S1474-4422(21)00066-1)
- Schneider LS, Mangialasche F, Andreasen N, et al. Clinical trials and late-stage drug development for Alzheimer's disease: an appraisal from 1984 to 2014. *J Intern Med*. 2014;275(3):251-283. <https://doi.org/10.1111/joim.12191>
- Ezzati A, Abdulkadir A, Jack CR Jr, et al. Predictive value of ATN biomarker profiles in estimating disease progression in Alzheimer's disease dementia. *Alzheimers Dement*. 2021;17(11):1855-1867. <https://doi.org/10.1002/alz.12491>
- Zimmer ER, Parent MJ, Souza DG, et al. [^{18}F]FDG PET signal is driven by astroglial glutamate transport. *Nat Neurosci*. 2017;20(3):393-395. <https://doi.org/10.1038/nn.4492>
- Trzepacz PT, Yu P, Sun J, et al. Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer's dementia. *Neurobiol Aging*. 2014;35(1):143-151. <https://doi.org/10.1016/j.neurobiolaging.2013.06.018>
- Grimmer T, Wutz C, Alexopoulos P, et al. Visual versus fully automated analyses of 18F-FDG and amyloid PET for prediction of dementia due to Alzheimer disease in mild cognitive impairment. *J Nucl Med*. 2016;57(2):204-207. <https://doi.org/10.2967/jnumed.115.163717>
- Frings L, Hellwig S, Bormann T, Spehl TS, Buchert R, Meyer PT. Amyloid load but not regional glucose metabolism predicts conversion to Alzheimer's dementia in a memory clinic population. *Eur J Nucl Med Mol Imaging*. 2018;45(8):1442-1448. <https://doi.org/10.1007/s00259-018-3983-6>
- Smailagic N, LaFortune L, Kelly S, Hyde C, Brayne C. ^{18}F -FDG PET for prediction of conversion to Alzheimer's disease dementia in people with mild cognitive impairment: an updated systematic review of test accuracy. *J Alzheimers Dis*. 2018;64(4):1175-1194. <https://doi.org/10.3233/JAD-171125>
- Blazhenets G, Ma Y, Sorensen A, et al. Predictive value of (18)F-Florbetapir and (18)F-FDG PET for conversion from mild cognitive impairment to Alzheimer dementia. *J Nucl Med*. 2020;61(4):597-603. <https://doi.org/10.2967/jnumed.119.230797>
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. <https://doi.org/10.1038/nature14539>
- Ding Y, Sohn JH, Kawczynski MG, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using (18)F-FDG PET of the brain. *Radiology*. 2019;290(2):456-464. <https://doi.org/10.1148/radiol.2018180958>
- Duan J, Liu Y, Wu H, Wang J, Chen L, Chen CLP. Broad learning for early diagnosis of Alzheimer's disease using FDG-PET of the brain. *Front Neurosci*. 2023;17:1137567. <https://doi.org/10.3389/fnins.2023.1137567>
- Zhang W, Zhang T, Pan T, et al. Deep learning with 18F-fluorodeoxyglucose-PET gives valid diagnoses for the uncertain cases in memory impairment of Alzheimer's disease. *Front Aging Neurosci*. 2021;13:764272. <https://doi.org/10.3389/fnagi.2021.764272>
- Zhou P, Zeng R, Yu L, et al. Deep-learning radiomics for discrimination conversion of Alzheimer's disease in patients with mild cognitive impairment: a study based on (18)F-FDG PET imaging. *Front Aging Neurosci*. 2021;13:764872. <https://doi.org/10.3389/fnagi.2021.764872>
- Iwatsubo T, Iwata A, Suzuki K, et al. Japanese and North American Alzheimer's disease neuroimaging initiative studies: harmonization for international trials. *Alzheimers Dement*. 2018;14(8):1077-1087. <https://doi.org/10.1016/j.jalz.2018.03.009>
- Dagley A, LaPoint M, Huijbers W, et al. Harvard aging brain study: dataset and accessibility. *Neuroimage*. 2017;144(Pt B):255-258. <https://doi.org/10.1016/j.neuroimage.2015.03.069>

22. Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One*. 2015;10(7):e0130140. <https://doi.org/10.1371/journal.pone.0130140>
23. Landau SM, Harvey D, Madison CM, et al. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging*. 2011;32(7):1207-1218. <https://doi.org/10.1016/j.neurobiolaging.2009.07.002>
24. Nugent S, Croteau E, Potvin O, et al. Selection of the optimal intensity normalization region for FDG-PET studies of normal aging and Alzheimer's disease. *Sci Rep*. 2020;10(1):9261. <https://doi.org/10.1038/s41598-020-65957-3>
25. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med*. 2006;25(20):3474-3486. <https://doi.org/10.1002/sim.2299>
26. Blazhenets G, Ma Y, Sorensen A, et al. Principal components analysis of brain metabolism predicts development of Alzheimer dementia. *J Nucl Med*. 2019;60(6):837-843. <https://doi.org/10.2967/jnumed.118.219097>
27. Brown RK, Bohnen NI, Wong KK, Minoshima S, Frey KA. Brain PET in suspected dementia: patterns of altered FDG metabolism. *Radiographics*. 2014;34(3):684-701. <https://doi.org/10.1148/rgr.343135065>
28. Morris JC, Schindler SE, McCue LM, et al. Assessment of racial disparities in biomarkers for Alzheimer disease. *JAMA Neurol*. 2019;76(3):264-273. <https://doi.org/10.1001/jamaneurol.2018.4249>
29. Wammes JD, Nakanishi M, van der Steen JT, MacNeil Vroomen JL. Japanese national dementia plan is associated with a small shift in location of death: an interrupted time series analysis. *J Alzheimers Dis*. 2021;83(2):791-797. <https://doi.org/10.3233/JAD-210521>
30. Fortinsky RH, Downs M. Optimizing person-centered transitions in the dementia journey: a comparison of national dementia strategies. *Health Aff (Millwood)*. 2014;33(4):566-573. <https://doi.org/10.1377/hlthaff.2013.1304>