Radiology

ORIGINAL RESEARCH • NEURORADIOLOGY

MRI-based Deep Learning Assessment of Amyloid, Tau, and Neurodegeneration Biomarker Status across the Alzheimer Disease Spectrum

Christopher O. Lew, MD • Longfei Zhou, PhD • Maciej A. Mazurowski, PhD • P. Murali Doraiswamy, MBBS • Jeffrey R. Petrella, MD • for the Alzheimer's Disease Neuroimaging Initiative[1]

From the Department of Radiology, Division of Neuroradiology, Alzheimer Disease Imaging Research Laboratory (C.O.L., J.R.P.), and Neurocognitive Disorders Program, Departments of Psychiatry and Medicine (P.M.D.), Duke University Medical Center, DUMC-Box 3808, Durham, NC 27710-3808; and Duke Institute for Brain Sciences (P.M.D.) and Department of Electrical and Computer Engineering, Department of Computer Science, Department of Biostatistics and Bioinformatics (L.Z., M.A.M.), Duke University, Durham, NC . Received September 27, 2022; revision requested November 28; final revision received September 3, 2023; accepted September 11. **Address correspondence to** J.R.P. (email: *jeffrey.petrella@duke.edu*).

Supported in part by the National Institutes of Health (grant U01AG024904) and a Charles Putman Seed Grant from the Department of Radiology, Duke University.

[1] Alzheimer's Disease Neuroimaging Initiative investigators are listed at *https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf*.

Conflicts of interest are listed at the end of this article.

Radiology 2023; 309(1):e222441 • https://doi.org/10.1148/radiol.222441 • Content codes: NR MR

Background: PET can be used for amyloid-tau-neurodegeneration (ATN) classification in Alzheimer disease, but incurs considerable cost and exposure to ionizing radiation. MRI currently has limited use in characterizing ATN status. Deep learning techniques can detect complex patterns in MRI data and have potential for noninvasive characterization of ATN status.

Purpose: To use deep learning to predict PET-determined ATN biomarker status using MRI and readily available diagnostic data.

Materials and Methods: MRI and PET data were retrospectively collected from the Alzheimer's Disease Imaging Initiative. PET scans were paired with MRI scans acquired within 30 days, from August 2005 to September 2020. Pairs were randomly split into subsets as follows: 70% for training, 10% for validation, and 20% for final testing. A bimodal Gaussian mixture model was used to threshold PET scans into positive and negative labels. MRI data were fed into a convolutional neural network to generate imaging features. These features were combined in a logistic regression model with patient demographics, *APOE* gene status, cognitive scores, hippocampal volumes, and clinical diagnoses to classify each ATN biomarker component as positive or negative. Area under the receiver operating characteristic curve (AUC) analysis was used for model evaluation. Feature importance was derived from model coefficients and gradients.

Results: There were 2099 amyloid (mean patient age, 75 years ± 10 [SD]; 1110 male), 557 tau (mean patient age, 75 years ± 7; 280 male), and 2768 FDG PET (mean patient age, 75 years ± 7; 1645 male) and MRI pairs. Model AUCs for the test set were as follows: amyloid, 0.79 (95% CI: 0.74, 0.83); tau, 0.73 (95% CI: 0.58, 0.86); and neurodegeneration, 0.86 (95% CI: 0.83, 0.89). Within the networks, high gradients were present in key temporal, parietal, frontal, and occipital cortical regions. Model coefficients for cognitive scores, hippocampal volumes, and *APOE* status were highest.

Conclusion: A deep learning algorithm predicted each component of PET-determined ATN status with acceptable to excellent efficacy using MRI and other available diagnostic data.

Supplemental material is available for this article.

An earlier incorrect version appeared online. This article was corrected on April 30, 2025.

Alzheimer disease (AD), once a clinicopathologic entity, is now commonly defined according to markers of its three hallmark pathologies, which are amyloid protein plaques, tau protein neurofibrillary tangles, and neurodegeneration (1). The combination of markers that reflect these pathologies define an individual's amyloid-tau-neurodegeneration (ATN) status, a research construct that provides a biologic definition of AD and enables characterization of the pathophysiologic events and multifactorial etiology leading to dementia (2). With the rising use of biomarkers for the selection and triage of patients in AD clinical research, there is an increasing need for widespread deployment of an unbiased classification system.

The ATN system is operationalized by combining assessments of positive or negative marker status in each of the three pathologic categories with the presence of β-amyloid using cerebrospinal fluid (CSF) Aβ or amyloid PET; the presence of hyperphosphorylated tau using CSF phosphorylated tau or tau PET; and the presence of neurodegeneration (ie, neuronal dysfunction or loss) using fluorodeoxyglucose (FDG) PET, structural MRI, or CSF total tau (2). Many of these tests are expensive and/or require invasive procedures that are not included in routine dementia assessment. PET has potential advantages as it is more sensitive than CSF biomarkers and noninvasive; however, it exposes patients to ionizing radiation and is expensive (3). Amyloid PET is not considered cost-effective for diagnostic purposes in patients with mild cognitive impairment; tau PET is currently a research tool; and FDG PET is only reimbursed by Medicare in the United States for distinguishing frontotemporal dementia from AD or in the context of a clinical trial (4,5). Thus, ATN status is not routinely assessed in patient care (2).

Abbreviations

AD = Alzheimer disease, ADAS13 = Alzheimer's Disease Assessment Scale–Cognitive Subscale with 13 items, ADNI = Alzheimer's Disease Neuroimaging Initiative, ATN = amyloid-tau-neurodegeneration, AUC = area under the receiver operating characteristic curve, CSF = cerebrospinal fluid, FDG = fluorodeoxyglucose, Grad-CAM = gradient-weighted class activation mapping, MMSE = Mini-Mental State Examination

Summary

A deep learning algorithm predicted PET-determined amyloid, tau, and neurodegeneration status in Alzheimer disease with good efficacy using high-resolution T1-weighted MRI scans and other readily available cognitive, genetic, volumetric, and demographic data.

Key Results

- This retrospective study used MRI volumes and clinical data to predict PET-determined amyloid, tau, and neurodegeneration biomarker status using 2099 amyloid, 557 tau, and 2768 fluorodeoxyglucose (FDG) PET scans, achieving areas under the receiver operating characteristic curve of 0.79, 0.73, and 0.86, respectively, at internal validation.
- Important features of all the amyloid, tau, and neurodegeneration biomarker prediction models included key temporal, parietal, frontal, and occipital cortical regions on MRI scans, age, cognitive scores, hippocampal volumes, and *APOE* status.
- The study included MRI volumes in patients with diagnoses across the cognitive spectrum at the time of imaging (2059 cognitively normal, 2843 mild cognitive impairment, and 522 Alzheimer disease).

Applications of deep learning techniques to medical imaging have grown due to increased computational power, improved methodology, and improved availability of data. Prediction of an individual's clinical disease status or progression from mild cognitive impairment to AD has been performed with high accuracy using MRI (6). Additionally, there have been numerous machine learning studies that have predicted amyloid status using clinical and genetic data, either alone (7–12), or combined with MRI data (13–22). The aim of the current study was to use deep learning to predict PET-determined ATN status using MRI and other readily available patient data.

Materials and Methods

This was a retrospective study approved by the health system institutional review board. Patients were enrolled according to state and federal Health Insurance Portability and Accountability Act regulations, and written informed consent was obtained.

Database

The data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, launched in 2003 as a public-private partnership led by the principal investigator, Michael W. Weiner, MD. The database is a longitudinal collection of biomarkers, imaging, and other patient data with the goal of advancing AD prevention and intervention. The investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. Additional up-to-date information can be found at *www.adni-info.org*.

Study Sample

To assemble the study sample, MRI data were paired with PET data gathered within 30 days of the PET acquisition date, from August 2005 to September 2020. The following radiotracers were used for each marker, and all other radiotracers were excluded: amyloid, fluorine 18 (^{18}F) florbetapir (^{18}F-AV-45); tau, flortaucipir (^{18}F-AV-1451); and neurodegeneration, ^{18}F-FDG. MRI and PET scans that were not acquired within 30 days of each other were excluded. In some cases, an individual might contribute more than one PET and MRI pair. An overview of this pairing is shown in Figure 1. The data were randomly split into three subsets without overlap, with 70% for training, 10% for validation, and 20% for testing. If an individual contributed multiple image pairs, all pairs were kept within one subset to prevent data leakage. The training set was used for model development. The validation set was used for assessing model performance during development. An internal holdout test set was used for final model evaluation to best assess model performance on a new sample of unseen data and provide a parallel to model performance when implemented at an outside institution. The test set was used once, after model development was completed, to evaluate model performance on new data.

Processing and Labeling of PET Scans

Amyloid marker values were calculated by averaging the standard uptake value ratio (SUVR) in frontal, anterior and posterior cingulate, lateral parietal, and lateral temporal cortical regions and dividing by the average of all reference regions (cerebellum, brainstem and pons, subcortical white matter regions) (23). Tau marker values were calculated by dividing the metatemporal SUVR (entorhinal, amygdala, fusiform, inferior, and middle temporal cortices) by the inferior cerebellum SUVR (24). Neurodegeneration marker values were calculated by dividing uptake in composite regions of interest (posterior cingulate gyrus, angular gyrus, inferior temporal gyrus) by uptake in pons and vermis reference regions (25,26). Because the ATN system classifies each marker as positive or negative, the continuous marker values were converted into discrete binary values using a bimodal Gaussian mixture model fit on the training set marker data. The model aimed to split the data into two separate Gaussian distributions and was then used to categorize markers.

Preprocessing of MRI Data

The MRI scans were preprocessed using sMRIPrep *(https://zenodo.org/record/6585099)* (27). The T1-weighted images were corrected for intensity nonuniformity, spatially conformed, and skull-stripped. The images were then intensity normalized by limiting the minimum and maximum 0.1% of values, and empty edges of the volume were excluded. A schematic of the preprocessing steps is shown in Figure 2.

Clinical Data

Structured clinical data that included age, sex, apolipoprotein E *(APOE)* gene polymorphism status, hippocampal volumes, Mini-Mental State Examination (MMSE) score, Alzheimer's

Disease Assessment Scale–Cognitive Subscale with 13 items (ADAS13) score, and clinical diagnosis were incorporated into model inputs. Hippocampal volume was used as a single feature due to its high utility in predicting AD progression in the absence of other volumetric measurements (28). All continuous features were scaled 0–1 based on the training set minimum and maximum to improve training stability and interpretability. For missing categorical data points, the chronologically nearest value within 1 year was used. For any continuous values that remained missing, imputation was done using training set medians for each feature.

Model Architecture

The deep learning model (hereafter, called the combined model) had two inputs and one output. The MRI input was passed through five convolutional blocks to generate 100 features that were concatenated with the structured clinical data input, which was then used in a logistic regression model to generate the final prediction score. An overview of the model architecture is shown in Figure 3. Separate models were trained for each biomarker task to improve model specificity. To improve model performance and reduce overfitting, a grid search was performed to tune hyperparameters and the number of imaging features on the validation set. For comparison with the combined model, a model that used MRI data alone (hereafter, called the MRI model) and structured clinical data alone (hereafter, called the structured model) were tested.

The code and trained models used in this study can be found at GitHub *(https://github.com/chris-lew/predicting_ATN_markers_using_MRI)*.

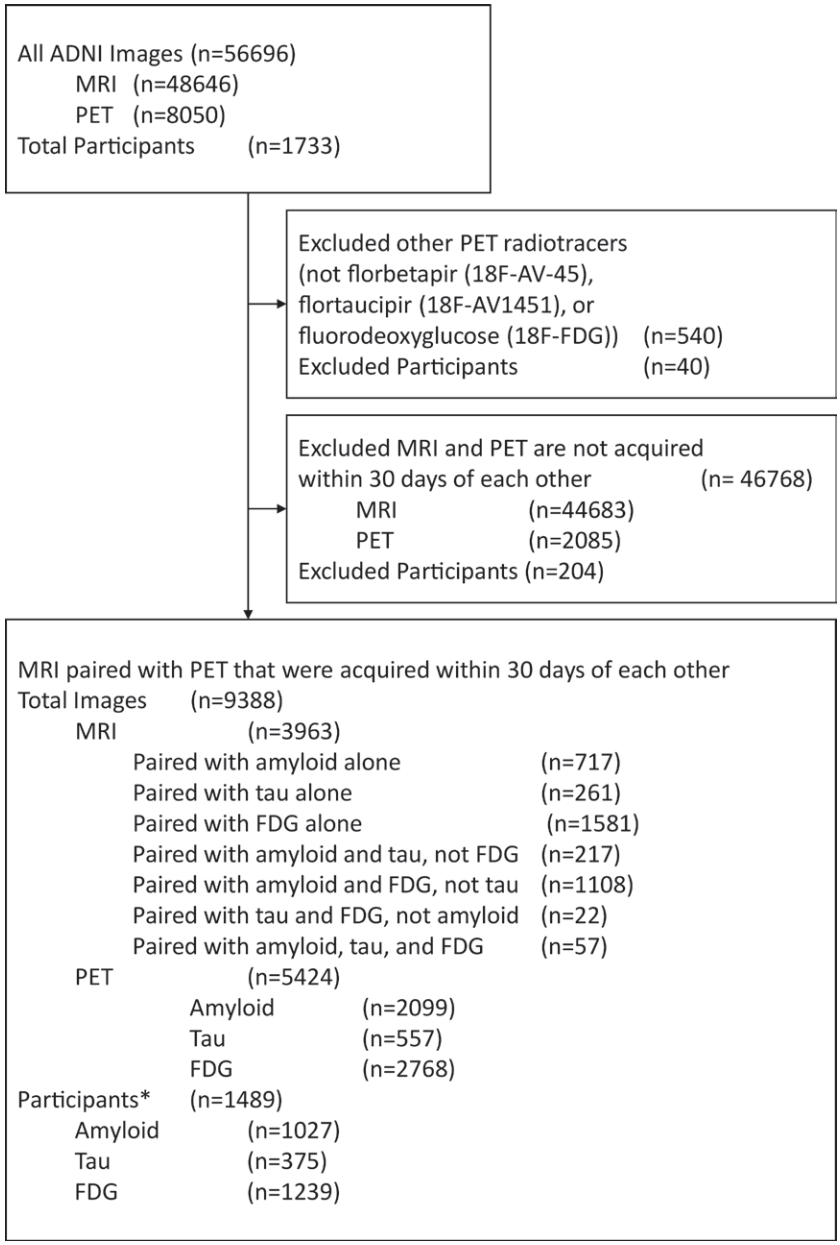

Figure 1: Flow diagram shows inclusion of study patients in the amyloid, tau, and neurodegeneration groups. Because the Alzheimer's Disease Neuroimaging Initiative (ADNI) database contains longitudinal data for each participant, one participant may contribute multiple MRI and PET scans. Additionally, each MRI scan may be paired with one or more PET scans. Fluorine 18 (18F) florbetapir (18F-AV-45), flortaucipir (18F-AV1451), and 18F fluorodeoxyglucose (18F-FDG) were the radiotracers used for amyloid, tau, and neurodegeneration assessments, respectively. * numbers do not add up to 1489 due to overlap between the sets.

Statistical Analysis

Summary statistics were calculated by fitting linear mixed models for continuous variables and generalized linear mixed models with a logit link for dichotomous variables, whereby the individual was included as a random intercept effect. The significance of the fixed effect variables on the dependent variable of interest was evaluated using a likelihood ratio test between the full model and a model with only the fixed effect. Other variables were evaluated using the χ^2 test. Model performance was evaluated on the validation and test sets by using accuracy and area under the receiver operating characteristic curve (AUC) values, and gradient-weighted class activation mapping (Grad-CAM) was used on the MRI model to identify important areas of the image for classification (29). A threshold of 0.5 was used to convert model predictions (ranging 0–1) to binary labels when calculating statistics. The DeLong method was used to calculate 95% CIs and to compare the AUCs of different nested models (30). $P < .05$ was considered indicative of a statistically significant difference. Finally, feature importance analysis of the logistic regression model was performed by examining the absolute value of coefficients of each feature.

The sample size of MRI and PET pairs was the maximum number of pairs in the ADNI database that fit our inclusion and exclusion criteria, aiming to maintain the best match in time between each MRI and PET study.

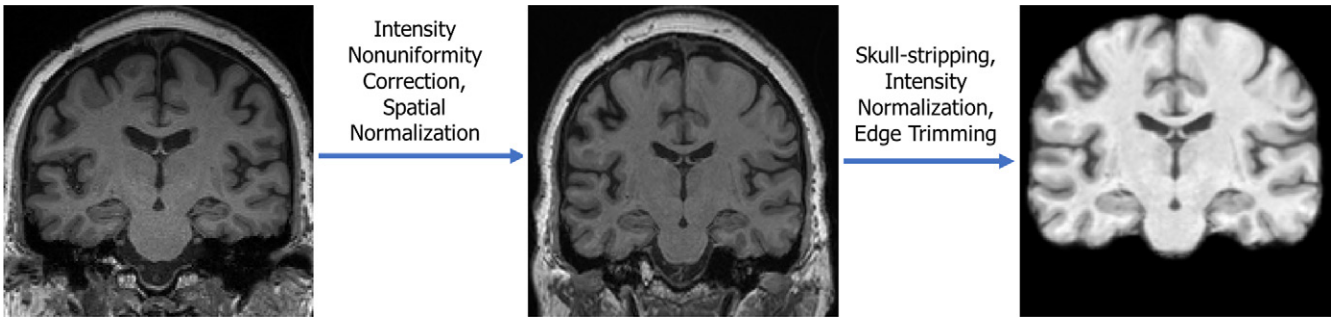

Figure 2: Schematic shows the image preprocessing pipeline. An MRI scan (left) is used as the original MRI data and, after processing, the preprocessed image data (right) is input into the model.

Results

Characteristics of the Study Sample

Biomarker and structured clinical data are summarized in Table 1 for the amyloid sample, Table 2 for the tau sample, and Table 3 for the neurodegeneration sample. There were 2099 amyloid PET and MRI pairs (mean patient age, 75 years ± 10 [SD]; 1110 male), 557 tau PET and MRI pairs (mean patient age, 75 years ± 7; 280 male), and 2768 FDG PET and MRI pairs (mean patient age, 75 years ± 7; 1645 male). There was a total of 1027 unique patients in the amyloid sample, 375 unique patients in the tau sample, and 1239 unique patients in the neurodegeneration sample (Fig 1). There were 217 unique MRI scans that were paired with amyloid and tau PET scans with no FDG PET scans, 1108 MRI scans that were paired with amyloid and FDG PET scans with no tau PET scans, and 22 MRI scans that were paired with tau and FDG PET scans with no amyloid PET scans. There were 57 unique MRI scans that were paired with amyloid, tau, and FDG PET scans. Scans were labeled positive in 46% (975 of 2099 pairs) of the amyloid sample, 15% (83 of 557 pairs) of the tau sample, and 40% (1094 of 2768 pairs) of the neurodegeneration sample. Figure 4 shows the distribution of continuous values for each biomarker and the cutoff determined when fitting a bimodal Gaussian mixture model on the training set distribution. Additional data describing characteristics of the training, validation, and test sets for each of the amyloid, tau, and neurodegeneration cohorts can be found in Tables S1–S3, respectively.

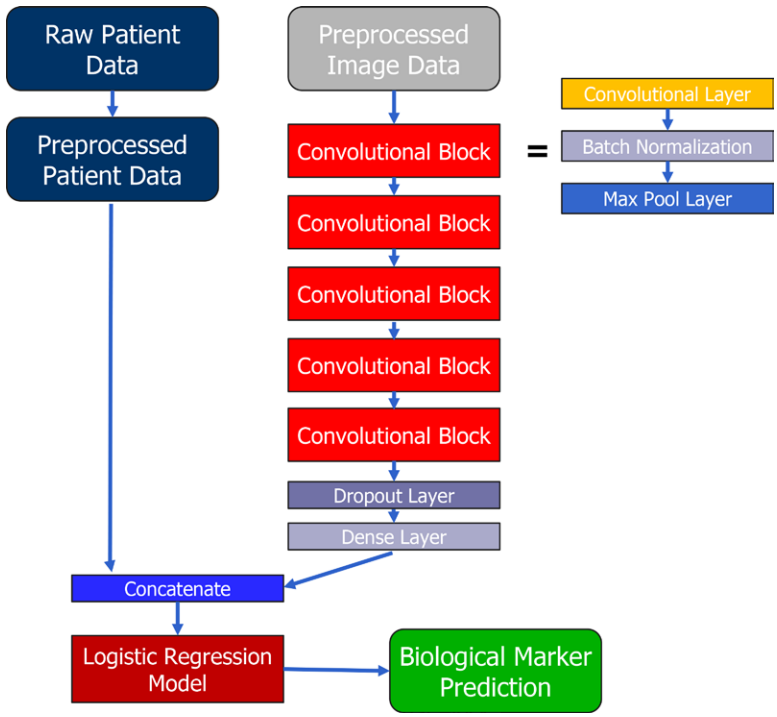

Figure 3: Diagram shows the model architecture. The model features two different inputs that include preprocessed image data and preprocessed patient data. The model outputs a biomarker prediction score within the range of 0–1. A combined model was trained separately for each of the amyloid, tau, and neurodegeneration biomarkers.

Performance of the Combined Model, MRI Model, and Structured Model

Performance of the structured model, MRI model, and combined model for each biomarker is summarized in Table 4. For the classifiers for each biomarker, the DeLong *P* value was used to compare AUCs of either the structured model or MRI model with the combined model. In the test set, there was a difference in performance between the structured model and the combined model when considering AUCs for amyloid (structured model, 0.71; combined model, 0.79; DeLong $P < .001$), tau (structured model, 0.54; combined model, 0.73; DeLong $P = .002$), and neurodegeneration (structured model, 0.76; combined model, 0.86; DeLong $P < .001$). There was no detectable difference in performance between the MRI model and the combined model when considering AUCs for amyloid (MRI model, 0.73; combined model, 0.79; DeLong $P = .07$), tau (MRI model, 0.80; combined model, 0.73; DeLong $P = .46$), or neurodegeneration (MRI model, 0.85; combined model, 0.86; DeLong $P = .28$).

Grad-CAM Analysis for the Amyloid, Tau, and Neurodegeneration MRI Models

Grad-CAM analysis was performed by examining the gradients of the final convolutional layer of the MRI model for

Table 1: Patient Characteristics at Time of Imaging for Scan Pairs in the Amyloid Study Sample according to Label

Characteristic	Amyloid Positive (*n* = 975)	Amyloid Negative (*n* = 1124)	*P* Value
Biomarker value*	1.40 ± 0.17	1.01 ± 0.06	<.001
Age (y)*	77 ± 7	75 ± 8	<.001
Sex			
M	501 (51)	609 (54)	.63
F	474 (49)	515 (46)	
MMSE score*†	25.7 ± 4.5	28.5 ± 2.3	<.001
ADAS13 score*†	19.3 ± 12.3	10.6 ± 7.1	<.001
Hippocampal volume (mm^3)*†	7375 ± 1040	7827 ± 990	<.001
APOEE4			
Noncarrier	373 (38)	880 (78)	<.001
Carrier	602 (62)	244 (22)	
Diagnosis			
Cognitively normal	301 (31)	588 (52)	<.001
Mild cognitive impairment	569 (58)	519 (46)	
Alzheimer disease	105 (11)	17 (2)	
Interval between MRI and PET (d)*	10.3 ± 9.6	9.8 ± 9.2	.39

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. Biomarker values are unitless standard uptake value ratios. Imputed values are excluded. The Alzheimer's Disease Neuroimaging Initiative (ADNI) gathers data on patients at intervals; therefore, one patient may contribute multiple unique samples with varying data to a cohort. In the ADNI database, the cognitively normal category includes both cognitively normal and subjective memory concern, and the mild cognitive impairment category includes early and late mild cognitive impairment, as well as mild cognitive impairment. ADAS13 = Alzheimer's Disease Assessment Scale–Cognitive Subscale with 13 items, *APOEE4* = allele producing the E4 type of *APOE*, MMSE = Mini-Mental State Examination.

* Data are means ± SDs for continuous variables.

† Less than 0.1% of all values were missing and required imputation.

Table 2: Patient Characteristics at Time of Imaging for Scan Pairs in the Tau Study Sample according to Label

Characteristic	Tau Positive (*n* = 83)	Tau Negative (*n* = 474)	*P* Value
Biomarker value*	1.75 ± 0.32	1.20 ± 0.09	<.001
Age (y)*	75 ± 8	76 ± 7	.06
Sex			
M	42 (51)	238 (50)	.72
F	41 (49)	236 (50)	
MMSE score*†	24 ± 4.5	28.6 ± 1.7	<.001
ADAS13 score*†	24.2 ± 12.5	10.7 ± 6.9	<.001
Hippocampal volume (mm^3)*†	6919 ± 1146	7787 ± 1014	<.001
APOEE4			
Noncarrier	28 (34)	301 (64)	<.001
Carrier	55 (66)	173 (36)	
Diagnosis			
Cognitively normal	24 (29)	308 (65)	<.001
Mild cognitive impairment	44 (53)	159 (34)	
Alzheimer disease	15 (18)	7 (2)	
Interval between MRI and PET (d)*	9.1 ± 9.8	9.6 ± 9.5	.54

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. Biomarker values are unitless standard uptake value ratios. Imputed values are excluded. The Alzheimer's Disease Neuroimaging Initiative (ADNI) gathers data on patients at intervals; therefore, one patient may contribute multiple unique samples with varying data to a cohort. In the ADNI database, the cognitively normal category includes both cognitively normal and subjective memory concern, and the mild cognitive impairment category includes early and late mild cognitive impairment, as well as mild cognitive impairment. ADAS13 = Alzheimer's Disease Assessment Scale–Cognitive Subscale with 13 items, *APOEE4* = allele producing the E4 type of *APOE*, MMSE = Mini-Mental State Examination.

* Data are means ± SDs for continuous variables.

† Less than 0.1% of all values were missing and required imputation.

each biomarker. High gradients in a region, relative to the rest of the image, indicate that the region was salient for performing predictions (29). For the amyloid MRI model, the perisylvian, inferior frontal, temporo-occipital, and frontal regions toward the vertex had high gradients. For the tau MRI model, the temporal, inferior frontal, perisylvian, and occipital regions, and the cerebellum, had high gradients. For the neurodegeneration MRI model, the perisylvian, parietal, medial occipital, and anterior cingulate regions had high gradients. The Grad-CAM analysis is illustrated in Figure 5. It is important to note that these maps were derived from the MRI model rather than the combined model, which has a more complex interplay between the structured data and MRI data that cannot be easily illustrated.

Feature Importance of the Various Models

Feature importance of the structured model for each biomarker is shown using bar graphs in Figure 6, depicting the weight of each structured data feature when a logistic regression model was fit on the training data. Because the model is deterministically fit and all features were scaled 0–1 and imputed, feature weights can be compared within the model. Weights from the combined model were not obtained because complex relations between structured data and MRI features would be difficult to interpret. For amyloid, the features with the greatest effect on positive predictions were the ADAS13 score, age, MMSE score, and *APOE* E4 allele. For tau, the features were ADAS13 score, hippocampal volumes, a cognitively normal diagnosis, and MMSE score. For neurodegeneration, the features were ADAS13 score, hippocampal volumes, age, and MMSE score. For all models, sex and a diagnosis of early mild cognitive impairment, late mild cognitive

Table 3: Patient Characteristics at Time of Imaging for Scan Pairs in the Neurodegeneration Study Sample according to Label

Characteristic	Neurodegeneration Positive (n = 1094)	Neurodegeneration Negative (n = 1674)	P Value
Biomarker value*	1.01 ± 0.10	1.29 ± 0.10	<.001
Age (y)*	77 ± 7	75 ± 7	<.001
Sex			
F	420 (38)	703 (42)	.50
M	674 (62)	971 (58)	
MMSE score*†	24.6 ± 4.5	28.1 ± 2.1	<.001
ADAS13 score*†	24.5 ± 11.7	12.8 ± 7.4	<.001
Hippocampal volume*‡ (mm^3)	6941 ± 1011	7747 ± 953	<.001
APOEE4			
Noncarrier	467 (43)	1053 (63)	<.001
Carrier	627 (57)	621 (37)	
Diagnosis			
Cognitively normal	150 (14)	688 (41)	<.001
Mild cognitive impairment	643 (59)	909 (54)	
Alzheimer disease	301 (28)	77 (5)	
Interval between MRI and PET (d)*	8.7 ± 9.1	10.3 ± 9.5	<.001

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. Biomarker values are unitless standard uptake value ratios. Imputed values are excluded. The Alzheimer's Disease Neuroimaging Initiative (ADNI) gathers data on patients at intervals; therefore, one patient may contribute multiple unique samples with varying data to a cohort. In the ADNI database, the cognitively normal category includes both cognitively normal and subjective memory concern, and the mild cognitive impairment category includes early and late mild cognitive impairment, as well as mild cognitive impairment. ADAS13 = Alzheimer's Disease Assessment Scale–Cognitive Subscale with 13 items, *APOEE4* = allele producing the E4 type of *APOE*, MMSE = Mini-Mental State Examination.

* Data are means ± SDs for continuous variables.

† Less than 0.1% of all values were missing and required imputation.

‡ For the neurodegeneration study sample, 52% of hippocampal volume measurements were missing and required imputation.

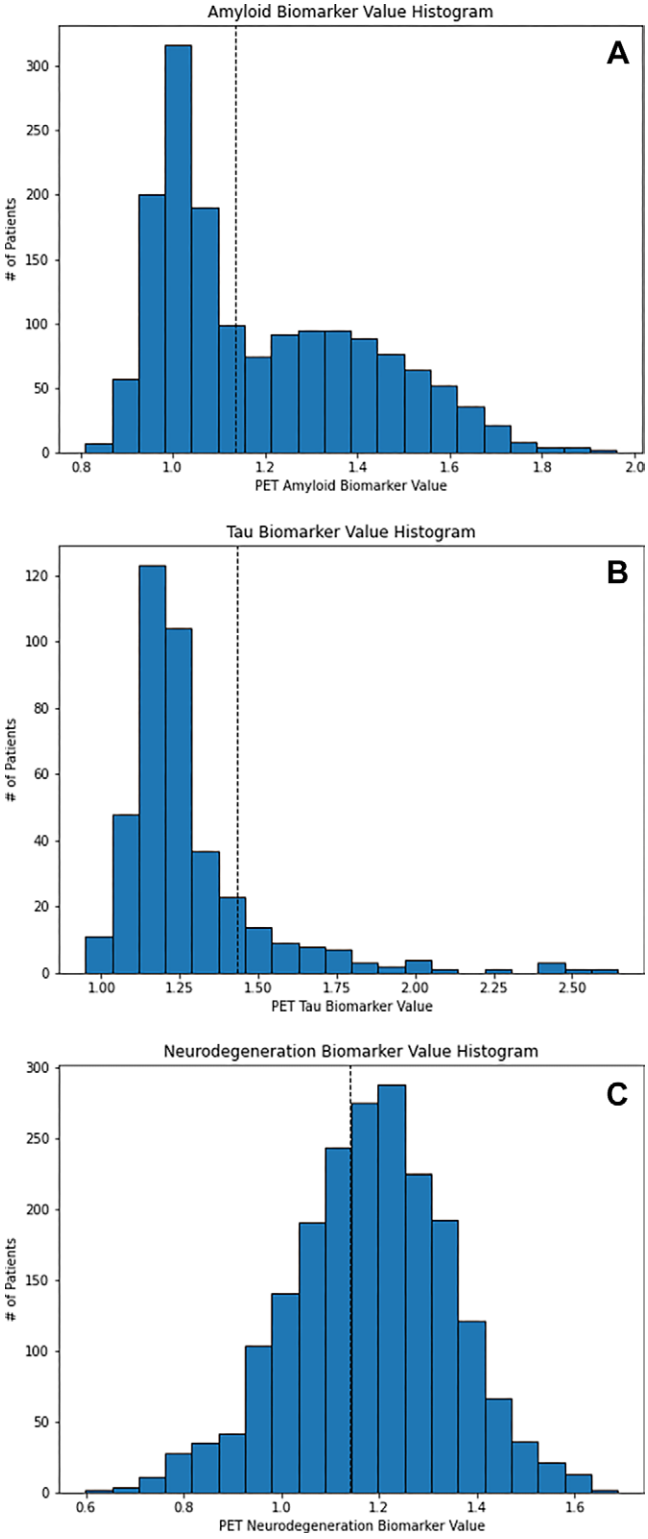

Figure 4: Histograms show PET biomarker data for **(A)** amyloid, **(B)** tau, and **(C)** neurodegeneration in the training set. The dashed vertical line in each histogram is the cutoff value determined by the bimodal Gaussian mixture model (1.134 for amyloid, 1.432 for tau, and 1.141 for neurodegeneration). Biomarker values are unitless standard uptake value ratios.

impairment, mild cognitive impairment, or subjective memory concern were weighted lower and thus less useful for predictions.

Discussion

The amyloid-tau-neurodegeneration (ATN) classification system aims to establish biomarker status based on a biologic definition of Alzheimer disease (AD); however, it requires costly or invasive procedures to obtain data for classification. In our study, PET scans for each ATN biomarker were paired with MRI scans acquired within 30 days, to yield 2099 amyloid PET and MRI pairs, 557 tau PET and MRI pairs, and 2768 FDG PET and MRI pairs. We assessed the value of combining deep learning features from MRI with readily

Table 4: Performance for Determining Positivity of Amyloid, Tau, and Neurodegeneration Biomarkers Using the Structured Model, MRI Model, and Combined Model on the Validation and Test Sets

	Validation Set			Test Set		
Biomarker and Metric	Structured Model	MRI Model	Combined Model	Structured Model	MRI Model	Combined Model
Amyloid						
AUC	0.67 (0.60, 0.75)	0.63 (0.53, 0.72)	0.74 (0.66, 0.82)	0.71 (0.67, 0.76)	0.73 (0.68, 0.78)	0.79 (0.74, 0.83)
Accuracy*	67 (60, 75) [95/141]	62 (54, 70) [88/141]	63 (55, 71) [89/141]	72 (68, 77) [273/378]	66 (61, 71) [249/376]	71 (67, 76) [268/376]
NPV*	68 (60, 75) [48/71]	60 (52, 69) [52/86]	63 (55, 71) [45/71]	71 (67, 76) [168/235]	65 (61, 70) [162/247]	72 (67, 76) [159/222]
PPV*	67 (59, 75) [47/70]	65 (58, 73) [36/55]	63 (55, 71) [44/70]	73 (69, 78) [105/143]	67 (63, 72) [87/129]	71 (66, 75) [109/154]
Specificity*	68 (60, 75) [48/71]	73 (66, 81) [52/71]	63 (55, 71) [45/71]	82 (78, 85) [168/206]	79 (75, 83) [162/204]	78 (74, 82) [159/204]
Sensitivity*	67 (59, 75) [47/70]	51 (43, 60) [36/70]	63 (55, 71) [44/70]	61 (56, 66) [105/172]	51 (46, 56) [87/172]	63 (59, 68) [109/172]
P value†	.004	.04	NA	<.001	.07	NA
Tau						
AUC	0.70 (0.54, 0.86)	0.76 (0.60, 0.93)	0.83 (0.64, 1.00)	0.54 (0.46, 0.61)	0.80 (0.68, 0.92)	0.73 (0.58, 0.88)
Accuracy*	88 (79, 97) [43/49]	86 (75, 95) [41/48]	88 (78, 97) [41/48]	89 (83, 95) [96/108]	86 (81, 93) [93/107]	89 (83, 95) [95/107]
NPV*	87 (77, 96) [39/45]	84 (74, 95) [38/45]	88 (79, 97) [37/42]	89 (83, 95) [95/107]	89 (83, 95) [91/102]	91 (86, 96) [91/100]
PPV*	100 (100, 100) [4/4]	100 (100, 100) [3/3]	84 (73, 94) [5/6]	100 (100, 100) [1/1]	40 (31, 49) [2/5]	57 (48, 67) [4/7]
Specificity*	100 (100, 100) [39/39]	100 (100, 100) [38/38]	97 (93, 100) [37/38]	100 (100, 100) [95/95]	97 (93, 100) [91/94]	97 (93, 100) [91/94]
Sensitivity*	40 (26, 54) [4/10]	30 (17, 43) [3/10]	50 (36, 64) [5/10]	8 (3, 13) [1/13]	15 (9, 22) [2/13]	31 (22, 40) [4/13]
P value†	.02	.43	NA	.002	0.46	NA
Neurodegeneration						
AUC	0.70 (0.64, 0.77)	0.81 (0.75, 0.86)	0.84 (0.79, 0.89)	0.76 (0.72, 0.79)	0.85 (0.80, 0.87)	0.86 (0.83, 0.89)
Accuracy*	73 (67, 79) [153/210]	73 (67, 79) [153/210]	74 (68, 80) [156/210]	77 (63, 75) [398/515]	74 (70, 78) [379/512]	75 (72, 79) [386/512]
NPV*	69 (63, 75) [49/71]	66 (59, 72) [56/85]	73 (67, 79) [47/64]	78 (74, 81) [141/181]	67 (63, 71) [167/250]	69 (65, 73) [163/235]
PPV*	75 (69, 81) [104/139]	78 (72, 83) [97/125]	75 (69, 81) [109/146]	77 (74, 81) [257/334]	81 (78, 84) [212/262]	81 (77, 84) [223/277]
Specificity*	58 (52, 65) [49/84]	67 (60, 73) [56/84]	56 (49, 63) [47/84]	65 (61, 69) [141/218]	77 (73, 81) [167/217]	75 (71, 79) [163/217]
Sensitivity*	83 (77, 88) [104/126]	77 (71, 83) [97/126]	87 82, 91) [109/126]	87 (84, 89) [257/297]	72 (68, 76) [212/295]	76 (72, 79) [223/295]
P value†	<.001	.009	NA	<.001	.28	NA

Note.—Data in parentheses are 95% CIs. The structured model used structured data only; the MRI model used MRI data only; and the combined model used both structured and MRI data. A threshold of 0.5 was used to convert model predictions (ranging 0–1) to binary labels. The total number of patients in the validation and test sets for each biomarker are as follows: amyloid, $n = 141$ and $n = 378$; tau, $n = 49$ and $n = 108$; and neurodegeneration, $n = 210$ and $n = 515$, respectively. AUC = area under the receiver operator characteristic curve, NA = not applicable, NPV = negative predictive value, PPV = positive predictive value.

* Data are percentages, with numbers of patients in brackets.

† DeLong *P* values were calculated by comparing predictions between the combined model and the selected model.

available diagnostic patient data (age, sex, *APOE* gene polymorphism status, hippocampal volumes, Mini-Mental State Examination [MMSE] score, Alzheimer's Disease Assessment Scale–Cognitive Subscale with 13 items [ADAS13] score, and clinical diagnosis) to predict the PET-determined ATN status. The method performed well, with AUCs of 0.79, 0.73,

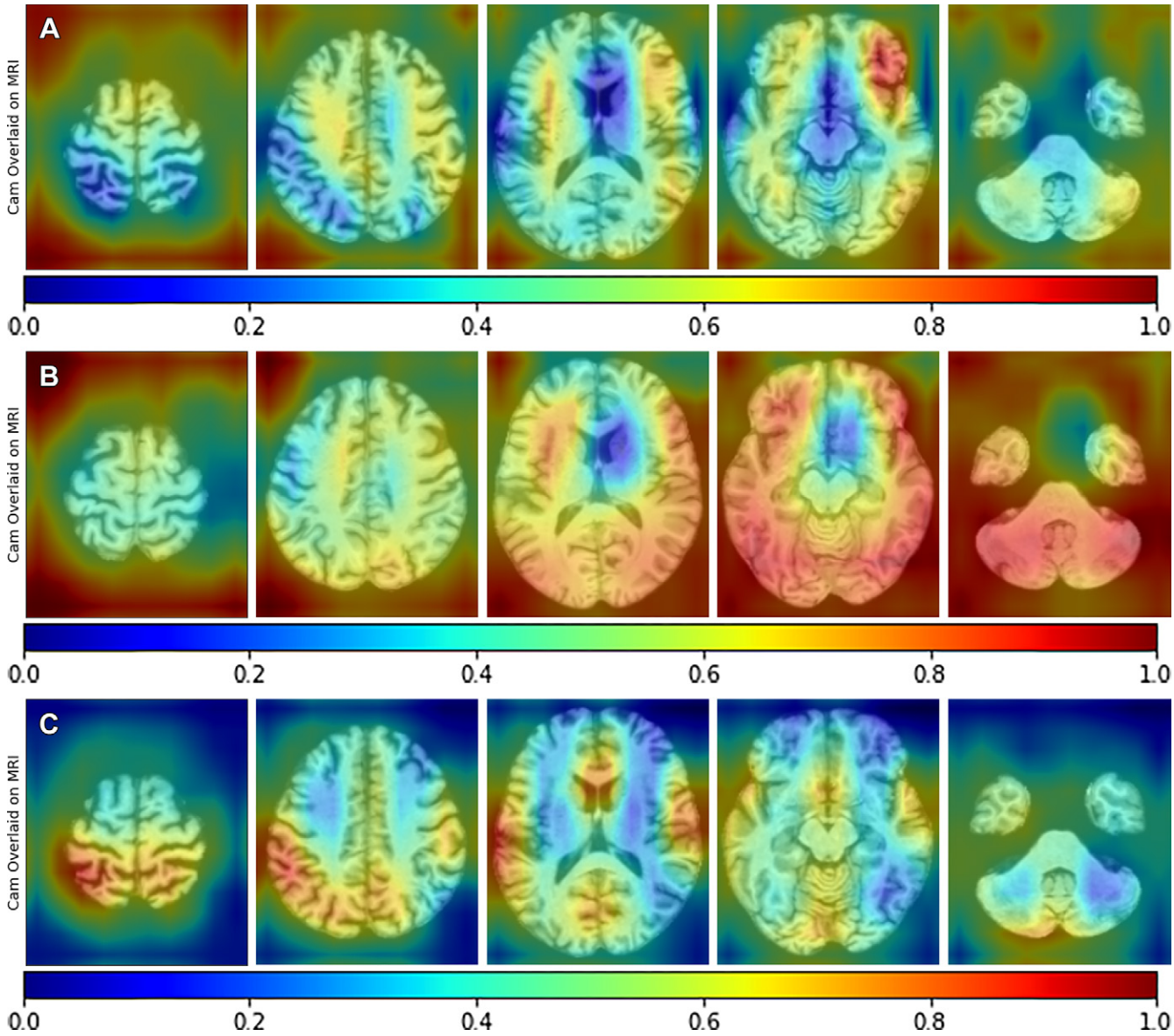

Figure 5: Image sequences show gradient-weighted class activation mapping (Grad-CAM, or Cam) superimposed on T1-weighted template images for the **(A)** amyloid, **(B)** tau, and **(C)** neurodegeneration MRI models that were trained on MRI data alone. This mapping method is used to analyze gradients of a convolutional neural network to identify salient areas of the image for classification. A higher gradient, relative to other regions, indicates that the region was salient for performing predictions. Gradients were scaled from 0 to 1, as indicated by the color bars.

and 0.86 for amyloid, tau, and neurodegeneration, respectively. Comparing the model that used MRI data alone (MRI model) against the model that used MRI and other diagnostic data (combined model), there were no detectable differences in performance in the test set for any of the biomarkers; although, there were improvements in the validation set for amyloid and neurodegeneration, which were the markers with the highest sample sizes. Our results support a strong relationship between diagnostic data features and ATN status, in particular for age, hippocampal volumes, ADAS13 score, and MMSE score, with ADAS13 and MMSE scores as the features of highest importance in the logistic regression model. These findings are consistent with our current understanding of how these features relate to AD markers, disease progression, and staging. However, with our current sample, we were unable to demonstrate an incremental gain in performance over the model with MRI data only.

Grad-CAM showed brain regions that are considered salient in the MRI model. The amyloid MRI model favored broad areas across the association cortex with less emphasis on the sensorimotor and visual cortex, brainstem, and basal ganglia regions. The tau MRI model favored the temporal and inferior frontal cortical areas toward the base of the brain with less emphasis at the vertex. The neurodegeneration MRI model strongly favored the temporal and parietal cortex, which are among the earliest affected by atrophy. The class activation maps did not necessarily match the known distributions of neuropathology and hypometabolism that are well described in AD (31). This may be because the algorithm is merely trying to predict binary labels from the data-driven MRI features

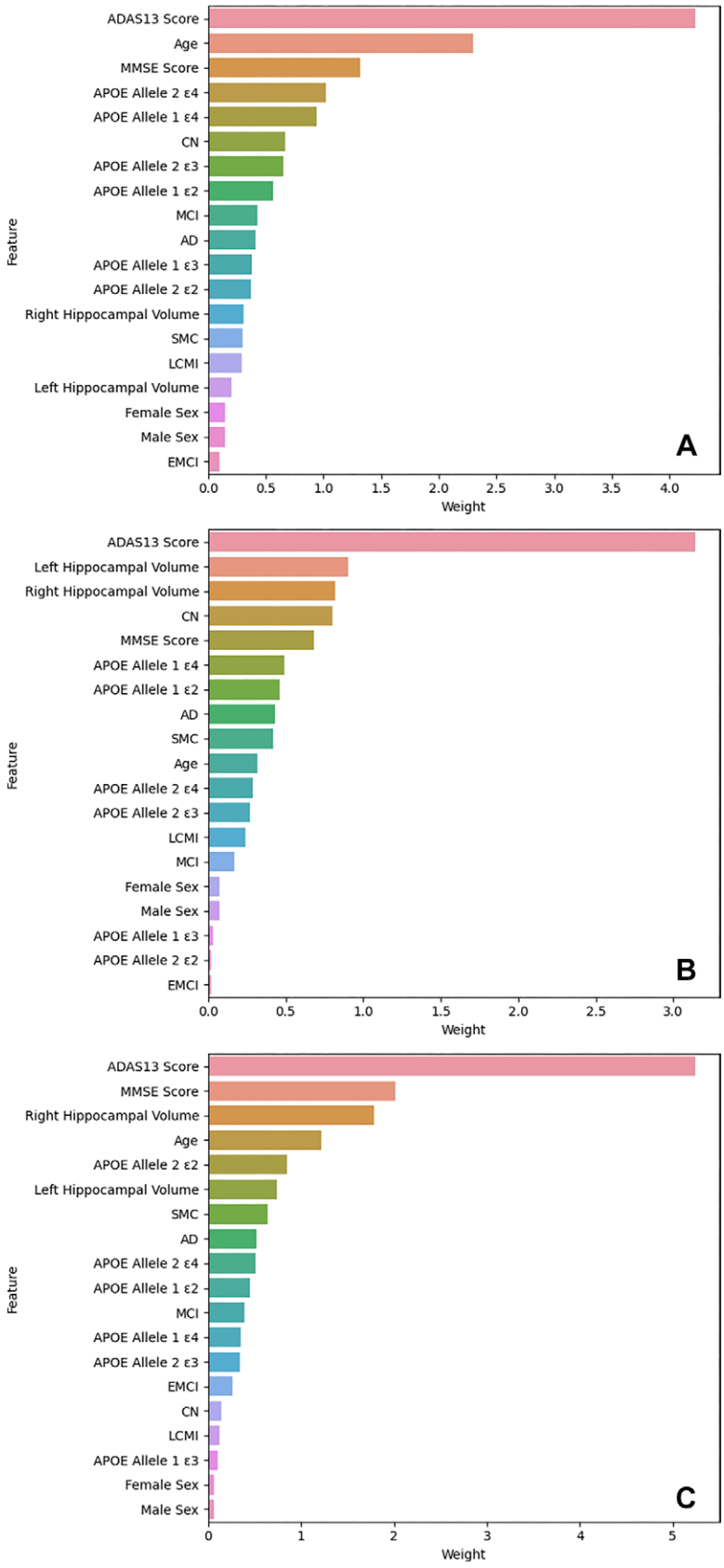

Figure 6: Bar graphs show absolute values of feature weights of the logistic regression models for the **(A)** amyloid, **(B)** tau, and **(C)** neurodegeneration structured models, which used structured clinical data only. Larger weights have an increasing impact on model predictions. AD = Alzheimer disease, ADAS13 = Alzheimer's Disease Assessment Scale–Cognitive Subscale with 13 items, CN = cognitively normal, EMCI = early mild cognitive impairment, LMCI = late mild cognitive impairment, MCI = mild cognitive impairment, MMSE = Mini-Mental State Examination, SMC = subjective memory concern.

rather than predict the distribution of pathology, which is known to vary according to disease stage. Alternatively, the map patterns may be an artifact of limited data, suggesting the need for further model validation and refinement in a larger sample prior to deployment.

Several other studies involving AD have applied machine learning to MRI data to predict PET outcomes; however, these studies have been limited to prediction of amyloid status. There have been studies aiming to classify amyloid status without using MRI data, with AUCs ranging 0.73–0.81 (8–12), and with MRI data, with AUCs ranging 0.80–0.90 (7,13–22). However, these studies often required additional information, such as difficult-to-obtain blood protein markers, as well as specific MRI findings scored by radiologists (7,9). Our study used amyloid PET values for labels, rather than CSF values, because of their higher sensitivity, which can be useful in screening applications (3,32). Our study also used a sample larger than those of the strongest performing studies, which demonstrated AUCs ranging 0.85–0.90 and used clinically defined samples of less than 500 individuals (15,17,18,33). This suggests there may be inherent limitations in using a single algorithm to predict amyloid status from imaging and structured clinical data in a broad sample across different cognitive stages.

Our study had limitations. First, due to the lack of standardized cutoffs for standard uptake value ratios (SUVRs) to separate positive and negative biomarker labels for all three ATN components, we used Gaussian mixture models to estimate cut points for consistency and objectivity. This can be a drawback for unimodal distributions, as in the case of neurodegeneration biomarker values, where patient diagnoses were predominantly in the heterogeneous mild cognitive impairment category. Second, the tau data set had 557 scan pairs with 15% (83 of 557) of these labeled positive. This positively skewed class imbalance reduced confidence in the tau results and explains the variability seen in model performance. Third, our training population is a biased sample, and may not match the disease prevalence of potential target populations. For example, the presence of

non-AD pathology is likely underrepresented in the ADNI database compared with community settings, limiting generalizability. Fourth, we had many patients contribute multiple data points thereby increasing our sample size, recognizing that MRI data, biomarker values, cognitive scores, and more could change at the time of different imaging examinations, even within the same patient. To remove potential individual biases, we used spatial normalization to a standardized template. Furthermore, we prevented data leakage by keeping different scans from the same patient within the same subset (training, validation, or test). Fifth, our neurodegeneration marker was based on FDG PET metabolism rather than the more commonly used MRI measures of atrophy, to avoid circular reasoning. Nevertheless, our FDG PET measurements were likely affected by atrophy. Atrophy correction was not incorporated into the composite region-of-interest SUVR measures used to label individual images. Moreover, although hypometabolic findings on PET scans are known to precede atrophy on MRI scans by years, they may coexist at later stages of disease (34). Therefore, our FDG PET measures may be affected by both atrophy and hypometabolism. Lastly, we treated the status of each ATN biomarker independently, rather than as an interdependent multivariate composite. The latter would require eight (ie, 2^3) different possible outcome combinations with markedly smaller cell sizes for model training. Furthermore, to provide such training data, we would need to include patients with all different combinations of ATN classification status and require patients to undergo imaging for all three PET markers within a short interval of an MRI examination, thus greatly reducing our sample size.

In conclusion, the deep learning method described herein was able to predict PET-determined amyloid, tau, and neurodegeneration classification across a wide cognitive spectrum with moderate to high diagnostic efficacy by using deep learning–based MRI features combined with other readily available patient data. Further work examining the utility of other MRI sequences that reflect additional relevant comorbidities for dementia, such as white matter disease and microhemorrhage, may prove useful, in addition to exploring where in the diagnostic algorithm this technique would add maximum value.

Acknowledgments: 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 no. 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.

Author contributions: Guarantors of integrity of entire study, **C.O.L., J.R.P.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **C.O.L., L.Z., J.R.P.**; clinical studies, **C.O.L., L.Z., P.M.D.**; experimental studies, **C.O.L., L.Z., M.A.M.**; statistical analysis, **C.O.L., L.Z., M.A.M.**; and manuscript editing, all authors

Data sharing: Data analyzed during the study were provided by a third party. Requests for data should be directed to the provider indicated in the Acknowledgments.

Disclosures of conflicts of interest: **C.O.L.** No relevant relationships. **L.Z.** No relevant relationships. **M.A.M.** No relevant relationships. **P.M.D.** Grant or contracts from National Institutes of Health, US Highbush Blueberry Council, Cure Alzheimer's Fund, Avid, and Office of Naval Research; consulting fees from Lumos, Alzheon, Nutricia, Sermo, Otsuka, UMethod, Clearview, Transposon, and Compass; meeting and/or travel support from Lululemon, Apollo, and Live Love Laugh; patents planned, issued, or pending related to diagnosis and treatment of Alzheimer disease as well as personalized prediction; advisory board member for Alzheon, Lumos, Umethod, Clearview, and Transposon; leadership role at Apollo, Live Love Laugh, and Goldie Hawn Foundation; stockholder with Alzheon, Umethod, Transposon, and Evidation; recipient of in-kind support from Lumos. **J.R.P.** Grants from National Institutes of Health and National Science Foundation; advisory board member for cortechs.ai, Biogen, and icometrix.

References

1. DeTure MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. Mol Neurodegener 2019;14(1):32.
2. Jack CR Jr, Bennett DA, Blennow K, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology 2016;87(5):539–547.
3. Rubin R. New Test to Help Diagnose Alzheimer Disease. JAMA 2022;327(23):2281.
4. Lee YS, Youn H, Jeong HG, et al. Cost-effectiveness of using amyloid positron emission tomography in individuals with mild cognitive impairment. Cost Eff Resour Alloc 2021;19(1):50.
5. Positron Emission Tomography (FDG) and Other Neuroimaging Devices for Suspected Dementia. Medicare Coverage Database: Centers for Medicare & Medicaid Services. https://www.cms.gov/medicare-coverage-database/view/ncacal-decision-memo.aspx?proposed=N&NCAId=104. Accessed August 9, 2022.
6. Yamanakkanavar N, Choi JY, Lee B. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey. Sensors (Basel) 2020;20(11):3243.
7. Apostolova LG, Hwang KS, Avila D, et al. Brain amyloidosis ascertainment from cognitive, imaging, and peripheral blood protein measures. Neurology 2015;84(7):729–737.
8. Shan G, Bernick C, Caldwell JZK, Ritter A. Machine learning methods to predict amyloid positivity using domain scores from cognitive tests. Sci Rep 2021;11(1):4822.
9. 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. Alzheimers Dement 2019;15(2):194–204.
10. Kim SE, Woo S, Kim SW, et al. A Nomogram for Predicting Amyloid PET Positivity in Amnestic Mild Cognitive Impairment. J Alzheimers Dis 2018;66(2):681–691.
11. Petersen KK, Lipton RB, Grober E, Davatzikos C, Sperling RA, Ezzati A. Predicting Amyloid Positivity in Cognitively Unimpaired Older Adults: A Machine Learning Approach Using A4 Data. Neurology 2022;98(24):e2425–e2435.
12. Langford O, Raman R, Sperling RA, et al. Predicting Amyloid Burden to Accelerate Recruitment of Secondary Prevention Clinical Trials. J Prev Alzheimers Dis 2020;7(4):213–218.
13. Li C, Liu M, Xia J, et al. Predicting brain amyloid-β PET phenotypes with graph convolutional networks based on functional MRI and multi-level functional connectivity. medRxiv 2021.2008.2026.21262325 [preprint] https://doi.org/10.1101/2021.08.26.21262325. Posted August 29, 2021. Accessed August 9, 2022.
14. Ten Kate M, Redolfi A, Peira E, et al. MRI predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study. Alzheimers Res Ther 2018;10(1):100.
15. Kang SH, Cheon BK, Kim JS, et al. Machine Learning for the Prediction of Amyloid Positivity in Amnestic Mild Cognitive Impairment. J Alzheimers Dis 2021;80(1):143–157.

16. 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. Stat Methods Med Res 2020;29(1):151–164.
17. 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. Alzheimers Res Ther 2019;11(1):72.
18. Joo SH, Lee CU. Cerebral Amyloid Positivity Prediction Models Using Clinical Data in Subjects With Mild Cognitive Impairment and Dementia. Psychiatry Investig 2021;18(9):864–870.
19. Tosun D, Joshi S, Weiner MW; Alzheimer's Disease Neuroimaging Initiative. Neuroimaging predictors of brain amyloidosis in mild cognitive impairment. Ann Neurol 2013;74(2):188–198.
20. 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 Commun 2021;3(2):fcab008.
21. Pekkala T, Hall A, Ngandu T, et al. Detecting Amyloid Positivity in Elderly With Increased Risk of Cognitive Decline. Front Aging Neurosci 2020;12:228.
22. Kim JP, Kim J, Jang H, et al. Predicting amyloid positivity in patients with mild cognitive impairment using a radiomics approach. Sci Rep 2021;11(1):6954.
23. Landau SM, Fero A, Baker SL, et al. Measurement of longitudinal β-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. J Nucl Med 2015;56(4):567–574.
24. Maass A, Landau S, Baker SL, et al. Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. Neuroimage 2017;157:448–463.
25. 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.
26. Durand-Martel P, Tremblay D, Brodeur C, Paquet N. Autopsy as gold standard in FDG-PET studies in dementia. Can J Neurol Sci 2010;37(3):336–342.
27. Esteban O, Markiewicz CJ, Blair R, Poldrack RA, Gorgolewski KJ. sMRIPrep: Structural MRI PREProcessing workflows. Version 0.7.1. Zenodo. https://doi.org/10.5281/zenodo.4313270. Published December 9, 2020. Accessed August 9, 2022.
28. Tanpitukpongse TP, Mazurowski MA, Ikhena J, Petrella JR; Alzheimer's Disease Neuroimaging Initiative. Predictive Utility of Marketed Volumetric Software Tools in Subjects at Risk for Alzheimer Disease: Do Regions Outside the Hippocampus Matter? AJNR Am J Neuroradiol 2017;38(3):546–552.
29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Int J Comput Vis 2020;128(2):336–359.
30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–845.
31. Petrella JR, Coleman RE, Doraiswamy PM. Neuroimaging and early diagnosis of Alzheimer disease: a look to the future. Radiology 2003;226(2):315–336.
32. Palmqvist S, Zetterberg H, Mattsson N, et al. Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. Neurology 2015;85(14):1240–1249.
33. Chun MY, Kim GH, Park HK, et al. Predictive Scale for Amyloid PET Positivity Based on Clinical and MRI Variables in Patients with Amnestic Mild Cognitive Impairment. J Clin Med 2022;11(12):3433.
34. Barthel H, Schroeter ML, Hoffmann KT, Sabri O. PET/MR in dementia and other neurodegenerative diseases. Semin Nucl Med 2015;45(3):224–233.

Radiology

COMMUNICATIONS • ERRATUM

Erratum for: MRI-based Deep Learning Assessment of Amyloid, Tau, and Neurodegeneration Biomarker Status across the Alzheimer Disease Spectrum

Originally published in:

https://doi.org/10.1148/radiol.222441

MRI-based Deep Learning Assessment of Amyloid, Tau, and Neurodegeneration Biomarker Status across the Alzheimer Disease Spectrum

Christopher O. Lew, Longfei Zhou, Maciej A. Mazurowski, P. Murali Doraiswamy, Jeffrey R. Petrella, for the Alzheimer's Disease Neuroimaging Initiative

Erratum in:

https://doi.org/10.1148/radiol.259008

Corrections were made in **Table 4**: Two sets of row labels (**NPV** and **PPV**; **Specificity** and **Sensitivity**) were transposed for each biomarker (amyloid, tau, neurodegeneration); minor changes were made in 10 performance values; and the numbers of patients (in brackets) were corrected for all values in the table.

Corrections were made in **Tables S1–S3** in the supplemental material: the total numbers of scan pairs and the numbers of scan pairs according to sex, *APOE4,* and diagnosis were corrected, and the titles of the tables were modified for consistency with Table 4 in the main article.