

RESEARCH

Open Access



A deep learning MRI approach outperforms other biomarkers of prodromal Alzheimer's disease

Xinyang Feng^{1,2}, Frank A. Provenzano^{3,4}, Scott A. Small^{3,4*}  and for the Alzheimer's Disease Neuroimaging Initiative

Abstract

Background: The three core pathologies of Alzheimer's disease (AD) are amyloid pathology, tau pathology, and neurodegeneration. Biomarkers exist for each. Neurodegeneration is often detected by neuroimaging, and we hypothesized that a voxel-based deep learning approach using structural MRI might outperform other neuroimaging methods.

Methods: First, we implement an MRI-based deep learning model, trained with a data augmentation strategy, which classifies Alzheimer's dementia and generates class activation maps. Next, we tested the model in prodromal AD and compared its performance to other biomarkers of amyloid pathology, tau pathology, and neuroimaging biomarkers of neurodegeneration.

Results: The model distinguished between controls and AD with high accuracy (AUROC = 0.973) with class activation maps that localized to the hippocampal formation. As hypothesized, the model also outperformed other neuroimaging biomarkers of neurodegeneration in prodromal AD (AUROC = 0.788) but also outperformed biomarkers of amyloid (CSF A β = 0.702) or tau pathology (CSF tau = 0.682), and the findings are interpreted in the context of AD's known anatomical biology.

Conclusions: The advantages of using deep learning to extract biomarker information from conventional MRIs extend practically, potentially reducing patient burden, risk, and cost.

Keywords: Neuroimaging, Prodromal Alzheimer's disease, Biomarkers, Deep learning

Introduction

Biomarkers can aid in the clinical evaluation of Alzheimer's disease (AD), and biomarkers currently exist for AD's three core neuropathologies—amyloid pathology, tau pathology, and neurodegeneration [1, 2]. The first two can be estimated from CSF levels of A β and tau or by direct visualization using PET-sensitive radioligands. Neurodegeneration, a term currently used to encompass

neuronal or synaptic loss [3], can be estimated from PET-based measures of parietal cortex metabolism or from MRI-based measurements that reflect the structural integrity of the hippocampal formation.

Deep learning is a subset of machine learning that, in principle, holds promise for MRI-based classification of neurodegenerative diseases, including AD [4, 5]. Furthermore, while some studies have examined classifying MCI conversion using machine learning frameworks, they have largely done so using other architectures like SVM⁶, examining only up to 36 months [6–10], using clinical or other biomarker information in the model [7,

*Correspondence: sas68@columbia.edu

⁴Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY 10032, USA

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

8, 11], and few have examined the performance independently directly against existing biomarkers [9, 10]. We hypothesized that designing a deep learning model that captures AD's known pathophysiology and anatomy would be accurately comparable or better than existing biomarkers. For example, because "cell sickness" occurs before dramatic neuronal loss in AD's pathophysiological course [3, 12, 13], a classifier sensitive to subtle intensity difference, not necessarily volume shrinkage, might be most sensitive in the disease's early stages. Additionally, because of the brain's anatomical complexity, particularly the areas targeted by AD, a three-dimensional classifier seems most suitable for AD detection.

One challenge with a 3D classifier based solely on voxel signal intensity is that its training is estimated to require an unusually large number of scans from cases and controls, more than is typically available for AD. Having access to large-scale datasets is a common challenge for deep learning in all fields, and strategies have been developed for data augmentation [14]. In one study [5], we develop and implement a deep learning strategy to classify AD. We employ a data augmentation strategy that is particularly well suited for MRI-only datasets, by including scans acquired from the same patient across multiple visits. By training, validating, and testing the classifier at the level of individual subjects, instead of individual scans, we minimize the potential limitations of this approach, namely data leakage.

We elected not to augment data by traditional methods of image perturbation, like rotating or applying transformations, since structural MRI data have well-known preprocessing pipelines to spatially align images. We did not include available clinical information, as studies have done prior [7], to avoid a model dependent on information that might be sparse or unavailable, as might be the case of clinical evaluation outside of a carefully controlled and harmonized setting, like ADNI.

AD progresses through a prodromal stage before causing dementia, presenting clinically as mild cognitive impairment (MCI) [15]. Only a subset of patients with MCI have prodromal AD, and in contrast to AD dementia, where a clinical evaluation is often sufficient to diagnose the disease, our ability to diagnose prodromal AD when presented with an MCI patient is currently inadequate. With increased awareness and concern over AD, a growing number of MCI patients are presenting to clinicians wanting to know whether they have prodromal AD, and if so, how quickly they will progress to dementia. Showing that the deep learning algorithm can address the clinical questions that relate to prodromal AD would not only better validate its classification capabilities, but since derived from conventionally acquired MRI scans, would potentially expand its potential utility as a screening tool.

Accordingly, in the second series of studies, we set out to test how well the deep learning MRI scores, derived from the deep learning model trained on AD dementia, perform in detecting prodromal AD and in predicting time to dementia progression. Additionally, we compared its performance to other established biomarkers of amyloid pathology, tau pathology, and neurodegeneration. Based on the premise of deep learning's classification abilities, we hypothesized that deep learning MRI scores would outperform other MRI-based biomarkers of neurodegeneration. At the same time, given the proposed temporal profile of AD's neuropathology [16], we hypothesized that amyloid or tau biomarkers would outperform the deep learning MRI score in classifying prodromal AD. Additionally, we investigated the link of deep learning MRI scores to amyloid and tau pathology, using cross-sectional, longitudinal, premortem, and postmortem data, providing a mechanistic explanation for the deep learning MRI score.

The diagnostic cutoffs for all AD biomarkers are traditionally derived from patients in the dementia stage, and biomarkers shift over the disease's progressive course, particularly dynamic during its early stages. Since cutoffs for prodromal AD have not yet been established for any of the biomarkers, the best experimental design with which to test these hypotheses is to clinically follow a large group of MCI patients as they do or do not progress to dementia, so that the patients can be retroactively dichotomized into those with and without prodromal AD, respectively, at baseline. Biomarkers can then be tested to determine which best classifies prodromal AD and which best predicts progression. The challenge with this design is that, based on current estimates, approximately 5 years of clinical follow-up is needed in order to allot sufficient time for the majority of prodromal AD patients to clinically manifest as dementia [17, 18]. Here, we were able to implement this experimental design thanks to the Alzheimer's Disease Neuroimaging Initiative (ADNI), which has been acquiring biomarker data in a large population of MCI patients since 2005, and to test the two hypotheses about which biomarker best classifies prodromal AD and which predicts progression to dementia.

Results

Classifying the dementia stage of Alzheimer's disease

The deep learning model was trained, validated, and tested on 975 MRI scans repeatedly acquired in patients in the dementia stage of AD, versus 1943 MRI scans repeatedly acquired from healthy controls. In the test set, a "deep learning MRI" score was derived for each scan from the model, with the score reflecting the probability of each scan having AD. A receiver operating

characteristic (ROC) analysis revealed that the deep learning MRI scores accurately classified AD dementia vs. healthy controls with an area under the receiver operating characteristics curve (AUROC) of 0.973 (Fig. 1a).

Next, we generated an AD “class activation map” to determine whether the deep learning MRI scores derived from the model were regionally dominated. We find that the deep learning MRI scores are dominated by alterations in voxel signal intensity that localized to the anterior medial temporal lobe, in the vicinity of the anterior entorhinal cortex and hippocampus (Fig. 1b). We note that while the class activation map localized to the left more than the right anterior medial temporal lobe, in agreement with previous findings [19–21], contralateral areas emerged with lowered thresholding (Fig. S2). This anatomical profile supports the biological premise of our classification, potentially placing our deep learning MRI scores within the “neurodegeneration” biomarker category.

Classifying the prodromal stage of Alzheimer's disease

From ADNI, we identified a cohort of participants who were diagnosed with MCI at baseline and who had a complete set of CSF amyloid and tau biomarkers and structural MRI ($N = 582$; the inclusionary and exclusionary algorithm is illustrated in Fig. S1). Among these, 205 participants progressed to AD dementia at follow-up (“MCI progression” group), and thus had prodromal AD at baseline, while 179 participants remained MCI stable for at least 4 years (“MCI stable” group) (Fig. 2). The dementia-derived deep learning classifier was used to generate deep learning MRI scores on each individual case.

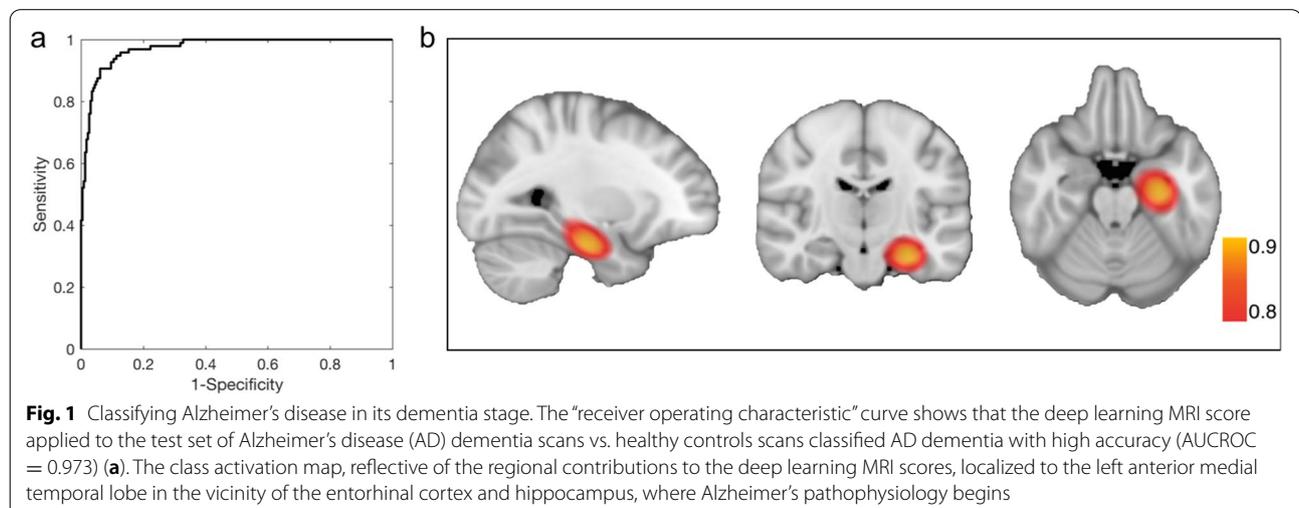
ROC analyses revealed that the deep learning MRI score outperformed all other biomarkers in classifying

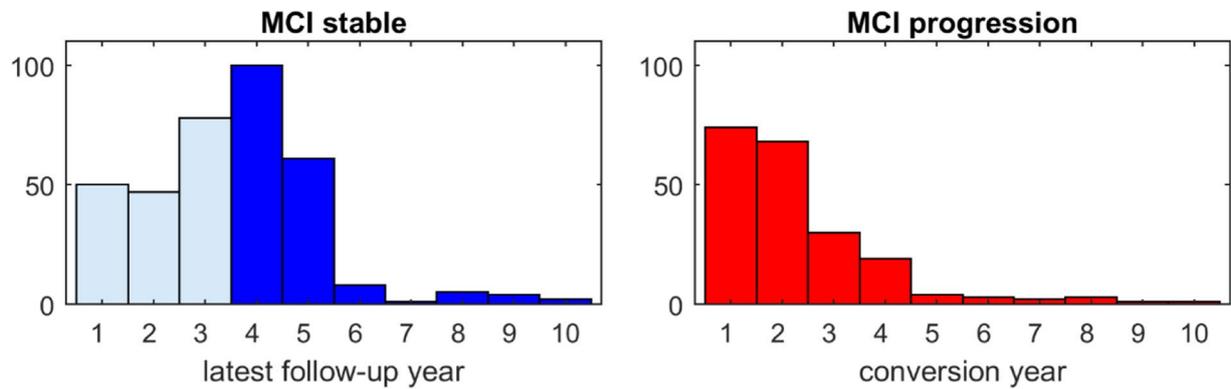
the MCI stable from the MCI progression group (Fig. 3). The AUROC of deep learning MRI score was 0.788 (accuracy at Youden (ACC) = 75%), superior to CSF A β (AUROC = 0.702, ACC = 66.7%, significantly lower than the deep learning MRI score, $p = 0.0141$), CSF tau (AUROC = 0.682, ACC = 66.4%, $p = 0.0161$), and CSF tau/A β (AUROC = 0.703, ACC = 68.5%, $p = 0.0161$); superior to MRI-based measures of hippocampal volume (AUROC = 0.733, ACC = 67.7%, $p = 0.0484$), entorhinal cortex volume (AUROC = 0.64, ACC = 62.5%, $p = 2.01E-6$), and entorhinal cortex thickness (AUROC = 0.685, ACC = 64.1%, $p = 1.71E-4$); and, finally, superior to Mini-Mental State Exam (AUROC = 0.648, ACC = 63.3%, $p = 6.70E-5$) and to neuropsychological measure most sensitive to the early stages of AD, the RAVLT retention score [22] (AUROC = 0.686, ACC = 67.7%, $p = 2.28E-3$).

Additionally, the deep learning MRI score was found to outperform or perform as well when tested in a subset of participants in whom additional PET-based biomarkers were available—FDG-PET that by measuring the parietal cortex metabolism is considered a biomarker of neurodegeneration [23], and AV45-PET, which by using an amyloid radioligand is a biomarker of amyloid pathology [24]. In this subset, the deep learning MRI score classified prodromal AD with an AUROC = 0.815 (ACC = 78.6%), compared to the AUROC of 0.782 for PDG-PET (ACC = 75.4%) and 0.751 (ACC = 71.4%) for amyloid-PET, although the differences were not statistically significant (Fig. 3, bottom panel).

Predicting progression to Alzheimer's disease dementia

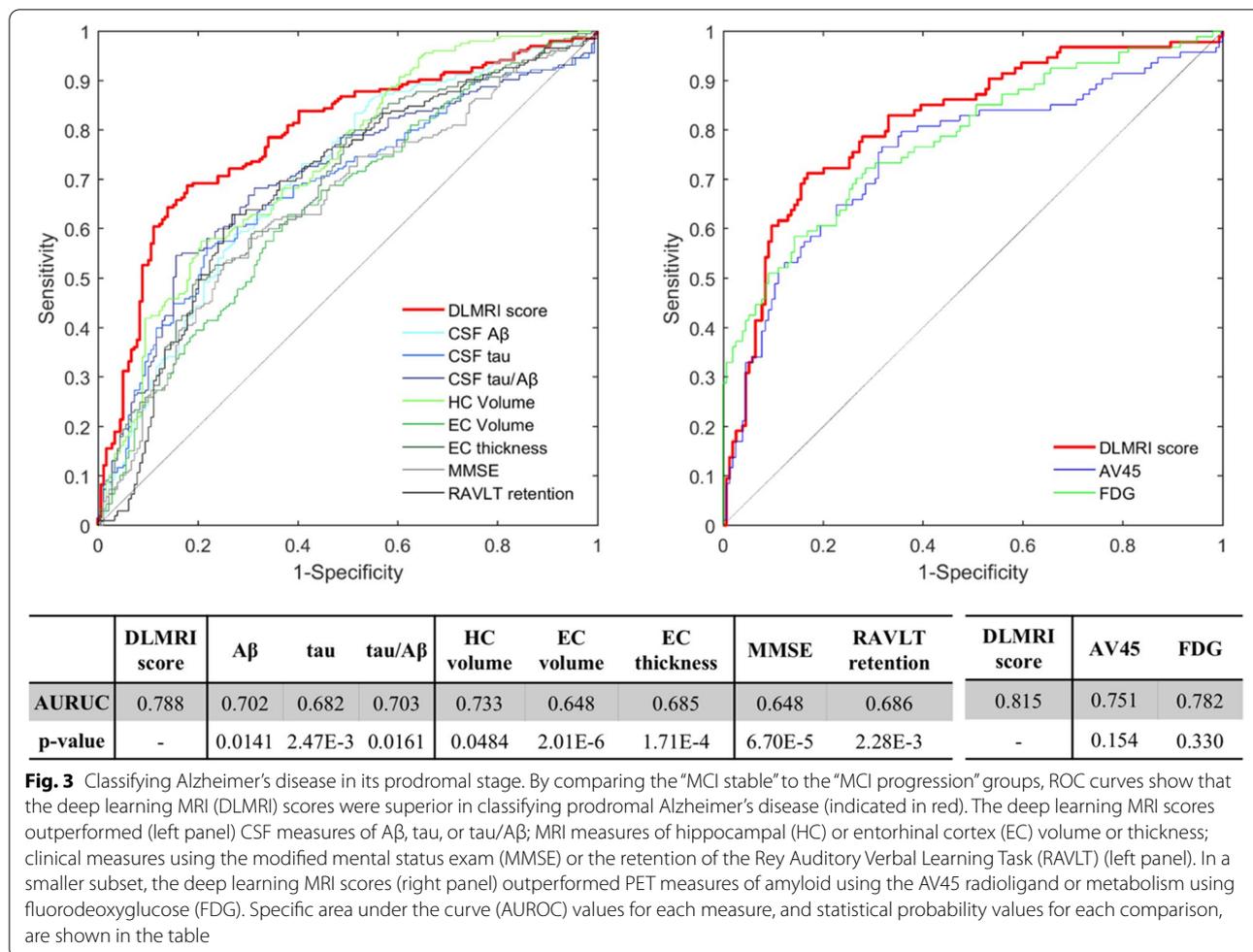
Survival analyses were performed to determine which biomarker best predicted progression to AD dementia among the MCI groups. The results revealed that





Main dataset	MCI stable N = 179	MCI progression N = 205	Total N = 384
age	71.3 ± 7	73.5 ± 7.1	72.5 ± 7.1
sex M/F (%M)	107/72 (59.8)	124/81 (60.5)	231/153 (60.2)
APOE ε4 frequency (2/1/0)	13/52/114	33/101/71	46/153/185
Last visit FU year	4.74 ± 1.2	-	-
conversion year	-	2.2 ± 1.7	-
DLMRI score	0.385 ± 0.165	0.622 ± 0.196	0.511 ± 0.217
Aβ	190.3 ± 51.3	144 ± 39.8	165.6 ± 51
tau	74.1 ± 43.0	115.3 ± 56.4	96.1 ± 54.6
tau/Aβ	0.457 ± 0.395	0.881 ± 0.532	0.684 ± 0.518
EC thickness	3.34 ± 0.4	3.03 ± 0.43	3.18 ± 0.44
EC volume (/ICV)	2.62E-03 ± 5.15E-04	2.31E-03 ± 4.95E-04	2.46E-03 ± 5.27E-04
HC volume (/ICV)	4.81E-03 ± 6.88E-04	4.14E-03 ± 6.00E-04	4.45E-03 ± 7.25E-04
MMSE	28.1 ± 1.7	26.9 ± 1.8	27.5 ± 1.8
RAVLT retention	49.6 ± 30.8	24.6 ± 29.3	36.2 ± 32.5
<hr/>			
PET subset	MCI stable N = 154	MCI progression N = 94	Total N = 248
age	70.9 ± 7	72.8 ± 6.6	71.6 ± 6.9
sex M/F (%M)	88/66 (57.1)	52/42 (55.3)	140/108 (56.5)
APOE ε4 frequency (2/1/0)	11/44/99	17/50/27	28/94/126
Last visit FU year	4.43 ± 0.55	-	-
conversion year	-	1.91 ± 1.16	-
DLMRI score	0.376 ± 0.163	0.635 ± 0.190	0.474 ± 0.214
FDG	1.31 ± 0.11	1.16 ± 0.11	1.26 ± 0.13
AV45	1.15 ± 0.19	1.4 ± 0.21	1.24 ± 0.23

Fig. 2 Distribution and demographics of subjects in the “mild cognitive impairment” study. Distribution frequencies of the participants with amnesic mild cognitive impairment (MCI) at baseline, who either remained stable (MCI stable) or progressed to Alzheimer’s dementia (MCI progression), organized by the latest follow-up years and conversion years. The dark blue bars indicate participants included in the study. Demographic and baseline biomarker data are listed in the table for the MCI stable and MCI progression groups



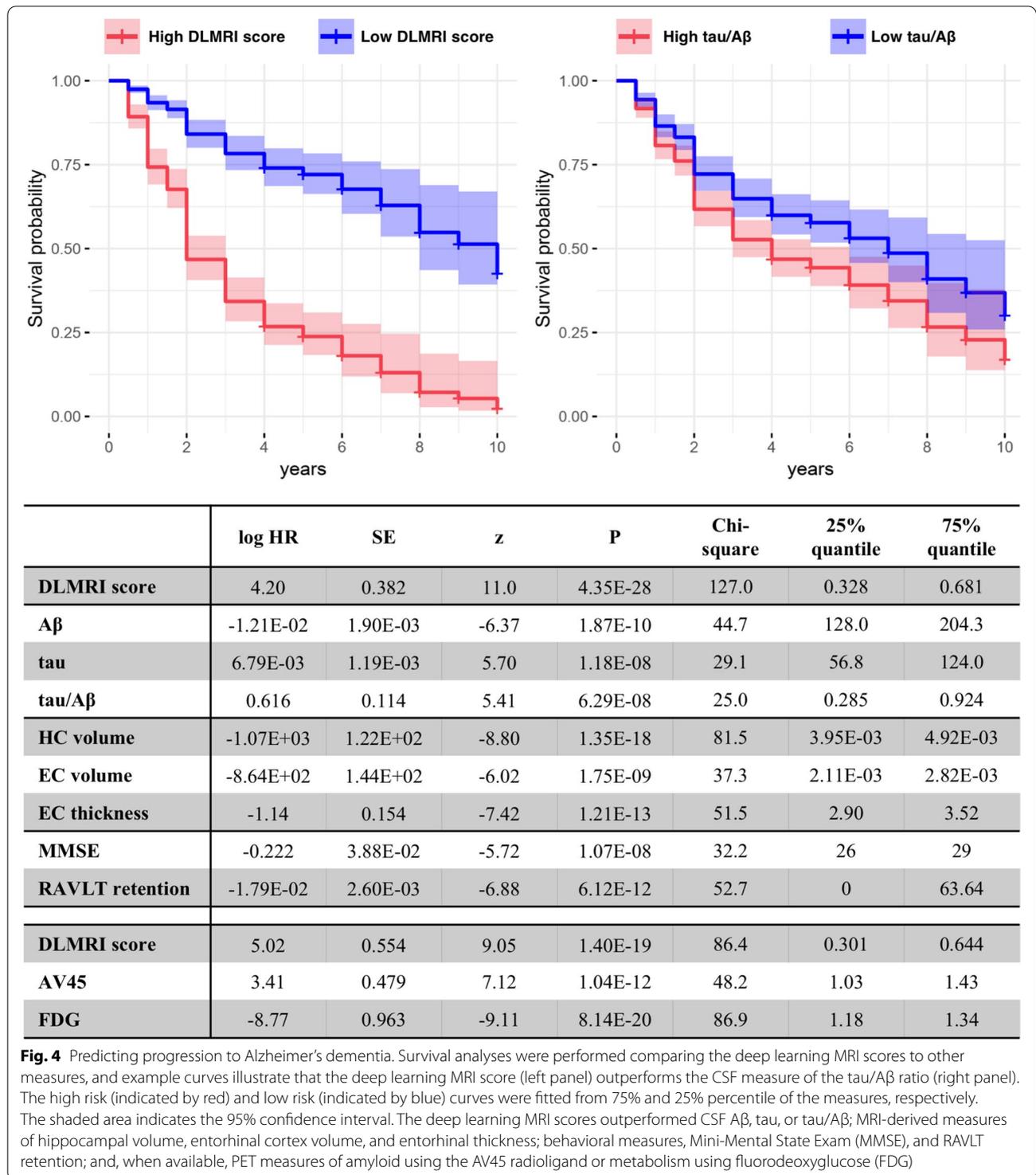
compared to other biomarkers, the deep learning MRI score best predicted the time to conversion to AD dementia, as illustrated by the survival curves of high and low deep learning MRI scores and tau/Aβ ratios (Fig. 4). The deep learning MRI scores showed better prediction capability ($|z| = 11.0, p = 4.35E-28$) than CSF biomarkers of amyloid and tau pathology (Aβ $|z| = 6.37, p = 1.87E-10$, tau $|z| = 5.70, p = 1.18E-08$, tau/Aβ $|z| = 5.41, p = 6.29E-08$) than MRI-based biomarkers of neurodegeneration (hippocampal volume $|z| = 8.80, p = 1.35E-18$, entorhinal volume $|z| = 6.02, p = 1.75E-09$, and entorhinal thickness $|z| = 7.42, p = 1.21E-13$) and than behavioral measures (MMSE $|z| = 5.72, p = 1.07E-08$ and RAVLT retention $|z| = 6.88, p = 6.12E-12$). Similarly, in the subset in whom the additional PET biomarkers were available, the deep learning MRI score ($|z| = 9.04, p = 1.40E-19$) outperformed or performed as well as

FDG-PET ($|z| = 9.11, p = 8.14E-20$) and AV45-PET ($|z| = 7.12, p = 1.04E-12$).

Correlations with amyloid pathology and tau pathology

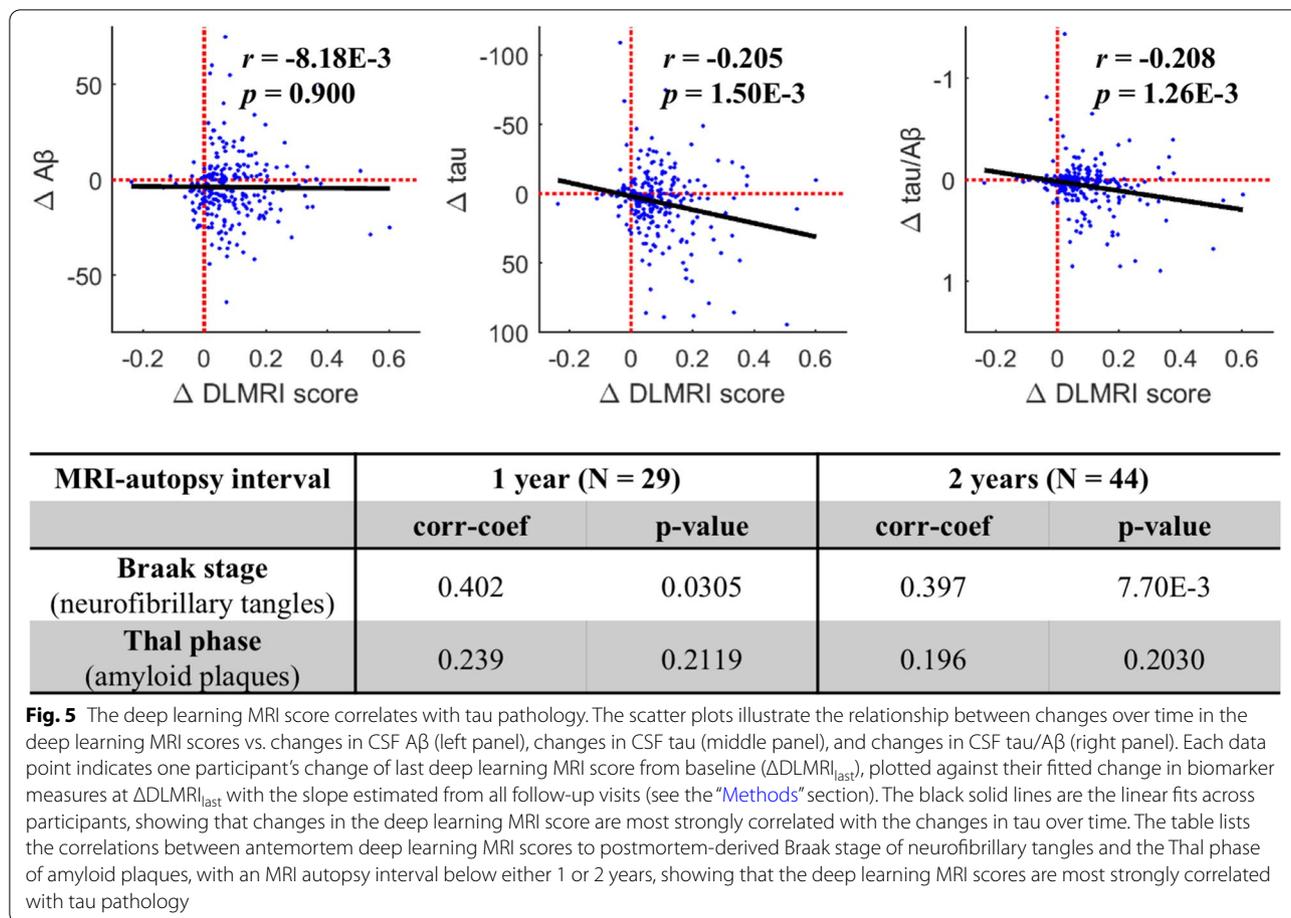
Correlational analyses were performed to determine whether the deep learning MRI score was correlated more with amyloid pathology or tau pathology. Cross-sectionally, we found that while the deep learning MRI score showed a stronger correlation with CSF tau ($r = 0.225, p = 9.00E-6$), it also correlated with CSF Aβ ($r = -0.190, p = 1.86E-4$). Longitudinally, however, changes in the deep learning MRI scores over time were significantly associated with the changes in CSF tau ($r = -0.205, p = 1.50E-3$), but not with the changes in CSF Aβ ($r = -8.18E-3, p = 0.900$).

Next, in a subsample with available postmortem data, we correlated the deep learning MRI score with neuropathological evidence of amyloid pathology, as indicated



by the Thal staging [25], or tau pathology indicated by Braak staging [26]. The deep learning MRI scores were found to associate more with tau pathology (with an MRI-autopsy interval below 2 years, Braak staging: $r = 0.397$, $p = 7.70e-3$; Thal staging: $r = 0.196$, $p = 0.203$)

(Fig. 5, bottom panel). To further explore the regionality of this relationship, we found that the deep learning MRI score correlated with tau levels mapped by tau-PET, with strong correlations observed with tau pathology in the entorhinal cortex ($r = 0.449$, $p = 1.66E-15$).



Discussion

The level of performance achieved by our deep learning model in classifying AD dementia supports the hypothesis that this approach of neuroimaging machine learning outperforms the traditional methods of measuring neurodegeneration. Further validating the assumptions, design, and implementation of our model is the fact that, despite incorporating information from the whole brain, the class activation map was dominated by a signal in the anterior entorhinal cortex and hippocampus, precisely where AD pathophysiology begins [3, 19–21, 26].

Stronger validation of the deep learning model was provided by the second series of studies when the dementia-derived classifier was applied to the prodromal stages of AD. Supporting the first hypothesis of this study, we found that our deep learning MRI scores outperformed other MRI-based measures of neurodegeneration in both classifying prodromal AD and predicting progression to dementia. Refuting the second hypothesis, we found that the deep learning MRI scores typically outperformed the biomarkers of amyloid and tau pathology.

We do not consider this unexpected finding a challenge to the primacy of amyloid and tau pathology in the neuropathological progression of AD [27]. The deep learning MRI scores were found strongly linked to tau pathology in the entorhinal cortex, a region where AD pathology begins [26], and its performance likely reflects this sensitivity. It is possible, therefore, that tau-PET may outperform the deep learning MRI score and other biomarkers. Future analyses from ADNI and other long-term PET studies will be able to test this prediction.

The observation that the deep learning MRI scores outperformed biomarkers of amyloid and tau pathology in predicting the time to dementia is less surprising. As a biomarker of neurodegeneration, this finding agrees with prior studies [28] and with the current model for the temporal sequence of AD's neuropathology [27]. Since in this scheme neurodegeneration occurs last, accurate biomarkers of it are more proximal to the development of dementia. Alternatively, these results might imply that neurodegeneration as a categorical for AD diagnosis may be more granular and focal, reliably detectable only by

newer analytic or measurement techniques, in the course of disease pathogenesis.

The strength of our prodromal AD study is that by relying on progression to AD dementia as a way to retroactively identify patients with prodromal AD, we overcame the limitation that precise biomarker cutoffs for prodromal AD have not yet been established. We designed the analysis based on prior studies that suggest that the majority of MCI patients with prodromal AD will progress to dementia within 4–5 years [17], an assumption confirmed in our study. Furthermore, approximately half of the MCI cohort ended up having prodromal AD, which agrees with the previous approximations [29].

Although the primary focus of the study is to demonstrate MRI information extracted via deep learning as an accurate and feasible biomarker for prodromal AD, we also show DLMRI can be used as an individual biomarker in combination with other categories of biomarkers to further boost the prodromal AD classification accuracy. We show the 5-fold cross-validation analysis results using individual and combined categories of data in Table S1.

Limitations

A potential weakness of our study is the possibility that a minority of patients in the stable MCI category are harboring prodromal AD at baseline. The number of misclassified patients is likely to be low [29], and so, this potential imprecision would not be expected to significantly alter our results. Tracking stable MCI patients for longer periods might address this concern but would in fact raise a new one: when tracking patients for a decade or more, particularly given the high incidence of AD in older populations, some are expected to develop AD de novo after the baseline evaluation. We can conclude that our findings and their conclusions are beyond reproach for a 5-year time window after initial evaluation, a clinically meaningful epoch for both patients and clinicians.

Conclusions

Our study provides the proof-of-principle that imaging-based deep learning models that are examined in concert with a disease's pathophysiology will yield a highly accurate model and improve performance in prognosticating disease. Showing that deep learning can enhance the utility of MRI in prodromal AD is the more important clinical implication of this study. Ordering "neuroimaging studies" [30] is the current standard of care when evaluating a patient with MCI suspected of having AD, most typically the conventional MRIs from which the deep learning MRI scores were derived. The rationale for this recommendation and its routine clinical implementation is not to "rule in" AD but rather to exclude other non-neurodegenerative causes of dementia, such as strokes,

bleeds, and tumors. Machine learning techniques, such as these, that can extract useful information for the purposes of prodromal AD detection, from conventional MRIs that have in any case been acquired, have the additional advantages of reducing patient burden and cost incurred by lumbar punctures, injection of radioactive ligands, or another additional testing.

Methods

Participants in the Alzheimer's disease dementia study

All data were obtained from ADNI, a multi-site observational study, which were acquired in accordance with each site's respective Institutional Review Board, including obtaining written consent acquired from each participant. We included 2918 scans ($N_{\text{healthy control}} = 1943$, $N_{\text{AD}} = 975$) from 626 subjects as training set, 382 scans ($N_{\text{healthy control}} = 251$, $N_{\text{AD}} = 131$) from 80 subjects as validation set, and 325 scans ($N_{\text{healthy control}} = 229$, $N_{\text{AD}} = 96$) from 80 subjects as test set.

Our data augmentation method of using scans from multiple visits of the same participant requires dealing with two problems: data leakage and disease progression. Data leakage is the problem of including different scans from the same participant in the training and test set; the trained model might make the prediction by matching the subject instead of extracting disease-relevant patterns. In this study, the training, validation, and test sets were partitioned at the subject level to ensure non-overlapping subjects. Disease progression is the problem that the diagnosis status of subjects might change during follow-up visits, and the diagnosis at scan time might be different from the baseline label. In this study, we labeled all the scans with their cross-sectional diagnosis at scan time, and although one participant's diagnostic labels may change, and therefore appear in both groups, there are few such cases.

Participants in the "Mild Cognitive Impairment" study

From ADNI, we identified a cohort of participants who were diagnosed with MCI at baseline and who had a complete set of CSF amyloid and tau biomarkers and structural MRI ($N = 582$; the inclusionary and exclusionary algorithm is illustrated in Fig. S1). Among these, 205 participants progressed to AD dementia at follow-up ("MCI progression" group), and 179 participants remained MCI stable for at least 4 years ("MCI stable" group). The time distribution and demographics of these two groups are shown in Fig. 2.

The deep learning MRI score

The deep learning model used in this study is a three-dimensional convolutional neural network (3D CNN) model with five convolutional stages and one fully

connected layer with sigmoid output [5]. Each convolutional stage consists of two convolutional layers with rectified linear unit (ReLU) activation function, a batch normalization operation and a max pooling layer. The model was optimized using the ADAM method with cross-entropy loss, using a learning rate of $2e-5$ determined through a grid search. The model was trained on the brain-extracted T1-weighted structural MRI scans from the ADNI cohort to classify patients in the dementia stage of AD versus healthy control subjects. To evaluate the regional contribution to AD classification, we generated a 3D class activation map, which visualizes the predictive regions in deep learning classification models [31, 32].

We applied the model trained to classify AD dementia versus healthy controls to the baseline scans of patients diagnosed with MCI. The continuous output from the model is reflective of the progressive structural patterns of AD pathology. We refer to it as a “deep learning MRI” (DLMRI) score, where a value of 0 is likely to be cognitively normal and 1 is likely to be AD. All analyses were performed using this score.

Amyloid and tau biomarkers

CSF biomarkers

CSF tau levels, reflective of neurofibrillary tangle, and CSF A β levels, reflective of amyloid pathology, were included in the analysis [33]. Additionally, the tau/A β ratio, which has been shown to best capture AD [34], was also included [35]. CSF was acquired at individual ADNI sites in accordance with the ADNI acquisition protocols and analyzed as previously described [35], using the multiplex xMAP Luminex platform. The median values provided by ADNI were used.

PET measures

In a subset of participants ($N_{\text{MCI progression}} = 94$, $N_{\text{MCI stable}} = 154$), amyloid pathology was also estimated with PET, mapping amyloid burden with the amyloid-binding radioligand AV45. The composite AV45-PET score provided by ADNI [36] was used in the analyses, which is based on the average AV45 SUVR (standard uptake value ratio) of the frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum [37].

Neurodegeneration biomarkers

MRI morphometry

FreeSurfer 6.0 [38, 39] was used to segment the structural MRI scans and derive regional morphometric measures. Hippocampal (HC) volume, entorhinal cortex (EC) volume, and entorhinal cortex thickness were used as structural integrity measures of the hippocampal formation.

Hippocampal and entorhinal cortex volumes were normalized by the intra-cranial volume (ICV).

PET measures

In a subset of participants ($N_{\text{MCI progression}} = 94$, $N_{\text{MCI stable}} = 154$), neurodegeneration was also estimated with PET using fluorodeoxyglucose (FDG). The composite FDG score provided by ADNI [36] was used in the analyses, which is based on the average FDG uptake of angular, temporal, and posterior cingulate [23].

Additional measures

Behavioral and neuropsychological measures

The Mini-Mental State Examination (MMSE) score and Rey Auditory Verbal Learning Test (RAVLT) retention scores were used in the analysis. The RAVLT retention score measures the number of delayed recalled words divided by the number of words learned in the last learning trial (trial 5) and has been found to be one of the most sensitive to AD²³.

Neuropathology

Among subjects with postmortem neuropathology data, 44 cases were identified who had an MRI within 2 years prior to death, and 29 cases were identified who had MRI within 1 year prior to death. DLMRI scores were derived from the last antemortem MRI scans in this cohort. An association was investigated between the DLMRI score and the neuropathologically derived Braak stage, which reflects neurofibrillary tangles [26], and the Thal phase, which reflects amyloid plaques [25].

Tau-PET

ADNI began acquiring PET scans using the AV1451 radioligand, which binds neurofibrillary tangles [40], in the late phase of ADNI2 and resumed in ADNI3. Due to the smaller number of subjects with available longitudinal tau-PET data or follow-up visits, cross-sectional analyses on these subjects ($N = 296$) using the regional AV1451 retention levels provided by ADNI [36] were performed.

Statistical analysis

ROC analysis

A receiver operating characteristic (ROC) analysis was used to determine the accuracy of the deep learning MRI score in prodromal AD classification, i.e., MCI stable and MCI progression classification, using standardized residuals controlling for age, sex, and APOE $\epsilon 4$ frequency with linear regression. The DeLong test [41] was used to test for the significance of the differences in the area under the ROC curve (AUROCs) between DLMRI score and other measures using the pROC R package [42].

Survival analysis

Cox proportional hazards regression models were fit to examine the association between each baseline measure and time to conversion to AD dementia from MCI, controlling for age, sex, and APOE $\epsilon 4$ frequency, using the survival R package [43]. MCI-stable participants are included in the models as censored data with the last visit as the censored point. The high-risk and low-risk survival curves were generated with the 75% percentile and 25% percentile of the observed measures, respectively.

Longitudinal analysis

The longitudinal association between DLMRI score and CSF biomarkers was studied by examining the deviation from baseline measurements for each participant over time. From the “MCI progression” and “MCI stable” groups, we further identified participants that had at least one follow-up of both MRI and CSF and collapsed them into a group for longitudinal analysis ($n = 238$). The changes in either CSF biomarker or DLMRI score of all follow-up visits from baseline were used to estimate the slope β of the change in tau ($\Delta\tau$), $A\beta$ ($\Delta A\beta$), and tau/ $A\beta$ ratio ($\Delta\tau/A\beta$) versus the change in DLMRI score (ΔDLMRI) for each participant using linear regression through the origin. Each participant was represented by the point based on the last follow-up visit's $\Delta\text{DLMRI}_{\text{last}}$ (x -coordinate) and the fitted change $\beta\Delta\text{DLMRI}_{\text{last}}$ (y -coordinate) of the respective measure. The last follow-up visit was used to anchor the representation of the participant in order to reflect the full follow-up. A correlation analysis was performed across participants. A linear regression model was fit across participants and illustrated.

Correlational analysis

A partial correlation was performed between baseline DLMRI score and CSF biomarkers, regional tau-PET measures, controlling for age, sex, and APOE $\epsilon 4$ frequency. As the Braak stage of neurofibrillary tangles and the Thal phase of amyloid plaques are both rank ordinal measures, we correlated the DLMRI score with the neuropathological measures using Spearman correlation.

Multivariate analysis of biomarkers from multiple categories

Linear SVM analyses were performed using individual and combined categories of data for prodromal AD classification in the MCI group. Fivefold cross-validation was performed, and the average AUROC scores on the test splits were reported.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13195-022-00985-x>.

Additional file 1: Fig. S1. Participant selection flow-chart. **Fig. S2.** Class activation maps with lowered threshold. The class-activation map with a relaxed thresholding with a focus on the right medial temporal region. **Table S1.** Cross-validation analysis of multi-variate prodromal AD classification using individual and combined categories of data including DLMRI score.

Acknowledgements

Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging and 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.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Authors' contributions

XF, FAP, and SAS contributed to the literature search, figures, study design, data interpretation, and writing. XF contributed to data collection and data analysis. The authors read and approved the final manuscript.

Funding

This study was not sponsored by a funding source; however, ADNI funding information can be found in the “Acknowledgements” section.

Availability of data and materials

All data used in this analysis can be obtained from the Alzheimer's Disease Neuroimaging Initiative (<http://adni.loni.usc.edu>). All code is available at <https://github.com/fengcls/Deep-learning-MRI-AD-prediction/>.

Declarations

Ethics approval and consent to participate

This study was approved by the respective institutional review boards of the contributing sites with all participants providing written consent. This analysis was found to be exempt by the Columbia University IRB (Protocol Number AAAS6975).

Consent for publication

Not applicable.

Competing interests

FAP is a consultant for and equity holder of Imij Technologies. SAS serves on the scientific advisory board of Meira GTX and is an equity holder in Imij Technologies. XF, FAP, and SAS have applied for a provisional patent on neuroimaging-based diagnosis.

Author details

¹Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA. ²Current address: Meta Platforms, Inc., Menlo Park, CA, USA. ³Department of Neurology, Columbia University, 630 West 168th Street, P&S Box 16, New York, NY 10032, USA. ⁴Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY 10032, USA.

Received: 10 February 2021 Accepted: 7 March 2022

Published online: 29 March 2022

References

- Jack CR, Holtzman DM. Biomarker modeling of Alzheimer's disease. *Neuron*. 2013;80:1347–58.
- Olsson B, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol*. 2016;15:673–84.
- Khan UA, et al. Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nat Neurosci*. 2014;17:304–11.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
- Feng X, Yang J, Lipton ZC, Small SA, Provenzano FA. Deep learning on MRI affirms the prominence of the hippocampal formation in Alzheimer's disease classification bioRxiv. 2018. p. 456277. <https://www.biorxiv.org/content/10.1101/456277v1>. <https://doi.org/10.1101/456277>.
- Moradi E, et al. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage*. 2015;104:398–412.
- Spasov S, et al. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage*. 2019;189:276–87.
- Basaia S, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Neuroimage Clin*. 2019;21:101645.
- Popescu SG, Whittington A, Gunn RN, Matthews PM, Glocker B, Sharp DJ, Cole JH. Alzheimer's Disease Neuroimaging Initiative. Nonlinear biomarker interactions in conversion from mild cognitive impairment to Alzheimer's disease. *Hum Brain Mapp*. 2020;41(15):4406–18. <https://doi.org/10.1002/hbm.25133>.
- Platero C, Tobar MC. Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's conversion in mild cognitive impairment patients using longitudinal neuroimaging and clinical markers. *Brain Imaging Behav*. 2021;15(4):1728–38. <https://doi.org/10.1007/s11682-020-00366-8>.
- Tong T, et al. A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. *IEEE Trans Biomed Eng*. 2017;64:155–65.
- Small SA. Alzheimer disease, in living color. *Nat Neurosci*. 2005;8:404–5.
- Selkoe DJ. Alzheimer's disease is a synaptic failure. *Science*. 2002;298:789–91.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems (NIPS)*; 2012. p. 1097–105.
- Petersen RC, et al. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. 1999;56:303–8.
- Jack CR, et al. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*. 2016;87:539–47.
- Visser PJ, Kester A, Jolles J, Verhey F. Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology*. 2006;67:1201–7.
- Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand*. 2009;119:252–65.
- Yushkevich PA, et al. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum Brain Mapp*. 2015;36:258–87.
- Maass A, et al. Entorhinal tau pathology, episodic memory decline, and neurodegeneration in aging. *J Neurosci*. 2018;38:530–43.
- Miller Ml, et al. The diffeomorphometry of temporal lobe structures in preclinical Alzheimer's disease. *NeuroImage*. 2013;3:352–60.
- Chang YL, et al. Brain substrates of learning and retention in mild cognitive impairment diagnosis and progression to Alzheimer's disease. *Neuropsychologia*. 2010;48:1237–47.
- Landau SM, et al. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging*. 2011;32:1207–18.
- Clark CM, et al. Use of florbetapir-PET for imaging β -amyloid pathology. *JAMA*. 2011;305:275–83.
- Thal DR, Rüb U, Orantes M, Braak H. Phases of A β -deposition in the human brain and its relevance for the development of AD. *Neurology*. 2002;58:1791–800.
- Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol*. 1991;82:239–59.
- Jack CR, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol*. 2013;12:207–16.
- Vemuri P, et al. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurol*. 2009;73:294–301.
- Vos SJ, et al. Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. *Brain*. 2015;138:1327–38.
- Albert MS, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:270–9.
- Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep features for discriminative localization. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016.
- Selvaraju RR, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *IEEE International Conference on Computer Vision (ICCV)*; 2017. p. 618–26.
- Tapiola T, et al. Cerebrospinal fluid β -amyloid 42 and tau proteins as biomarkers of Alzheimer-type pathologic changes in the brain. *Arch Neurol*. 2009;66:382–9.
- Fagan AM, et al. Cerebrospinal fluid tau/ β -amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults. *Arch Neurol*. 2007;64:343–9.
- Shaw LM, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*. 2009;65:403–13.
- Jagust WJ, et al. The Alzheimer's Disease Neuroimaging Initiative 2 PET Core: 2015. *Alzheimer's Dementia*. 2015;11:757–71.
- Landau, S. & Jagust, W. Florbetapir processing methods. 2015 <http://adni.loni.usc.edu>.
- Fischl B, et al. Whole brain segmentation: automated labeling of neuro-anatomical structures in the human brain. *Neuron*. 2002;33:341–55.
- Fischl B, et al. Automatically parcellating the human cerebral cortex. *Cereb Cortex*. 2004;14:11–22.
- Marquié M, et al. Validating novel tau positron emission tomography tracer [F-18]-AV-1451 (T807) on postmortem brain tissue. *Ann Neurol*. 2015;78:787–800.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. *Biometrics*. 1988;44(3):837–45.
- Robin X, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
- Therneau TM, Grambsch PM. *Modeling survival data: extending the Cox model*: Springer Science & Business Media; 2013.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.