Received: 8 March 2022 | Revised: 2 June 2022 | Accepted: 8 July 2022

DOI: 10.1002/hbm.26026

RESEARCH ARTICLE

WILEY

Identifying the regional substrates predictive of Alzheimer's disease progression through a convolutional neural network model and occlusion

Kichang Kwak[1] | William Stanford[2] | Eran Dayan[1,2,3] | for the Alzheimer's Disease Neuroimaging Initiative

[1]Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2]Neuroscience Curriculum, Biological and Biomedical Sciences Program, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[3]Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Correspondence
Eran Dayan, Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, 130 Mason Farm Road, CB 7513, Chapel Hill, NC 27599, USA.
Email: eran_dayan@med.unc.edu

Funding information
University of Southern California; Northern California Institute for Research and Education; Foundation for the National Institutes of Health; Canadian Institutes of Health Research; Transition Therapeutics; Takeda Pharmaceutical Company; Piramal Imaging; Servier; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Neurotrack Technologies; NeuroRx Research; Meso Scale Diagnostics, LLC.; Merck & Co., Inc; Lundbeck; Lumosity; Johnson & Johnson Pharmaceutical Research & Development LLC; Janssen Alzheimer Immunotherapy Research & Development, LLC.; IXICO Ltd.; GE Healthcare; Fujirebio; Genentech, Inc.; F. Hoffmann-La Roche Ltd; EuroImmun; Eli Lilly and Company; Elan Pharmaceuticals, Inc.; Eisai Inc.; Cogstate; CereSpir, Inc.; Bristol-Myers Squibb Company; Biogen; BioClinica, Inc; Araclon Biotech; Alzheimer's Drug Discovery Foundation; Alzheimer's Association; AbbVie; National Institute of Biomedical Imaging and Bioengineering; National Institute on Aging; Department of Defense, Grant/Award Number: W81XWH-12-2-0012; National Institutes of Health, Grant/Award Number:

Abstract

Progressive brain atrophy is a key neuropathological hallmark of Alzheimer's disease (AD) dementia. However, atrophy patterns along the progression of AD dementia are diffuse and variable and are often missed by univariate methods. Consequently, identifying the major regional atrophy patterns underlying AD dementia progression is challenging. In the current study, we propose a method that evaluates the degree to which specific regional atrophy patterns are predictive of AD dementia progression, while holding all other atrophy changes constant using a total sample of 334 subjects. We first trained a dense convolutional neural network model to differentiate individuals with mild cognitive impairment (MCI) who progress to AD dementia versus those with a stable MCI diagnosis. Then, we retested the model multiple times, each time occluding different regions of interest (ROIs) from the model's testing set's input. We also validated this approach by occluding ROIs based on Braak's staging scheme. We found that the hippocampus, fusiform, and inferior temporal gyri were the strongest predictors of AD dementia progression, in agreement with established staging models. We also found that occlusion of limbic ROIs defined according to Braak stage III had the largest impact on the performance of the model. Our predictive model reveals the major regional patterns of atrophy predictive of AD dementia progression. These results highlight the potential for early diagnosis and stratification of individuals with prodromal AD dementia based on patterns of cortical atrophy, prior to interventional clinical trials.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in analysis or writing of this article. 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.

U01 AG024904; Alzheimer's Disease Neuroimaging Initiative

KEYWORDS
Alzheimer's disease, brain atrophy, deep learning, mild cognitive impairment, neurodegeneration, occlusion analysis

1 | INTRODUCTION

Alzheimer's disease (AD) dementia is a complex progressive neurodegenerative disease and the leading cause of dementia (Jack et al., 2018; McKhann et al., 1984). The phenotype of AD dementia is characterized by brain atrophy (Frisoni et al., 2010), and progressive volume loss, as evident with structural magnetic resonance imaging (MRI) (Dallaire-Théroux et al., 2017; Halliday, 2017; Pini et al., 2016). The pattern of brain atrophy displayed by single individuals is highly complex and variable (Jack, Knopman, et al., 2010), generally starting in the medial temporal lobes (Bobinski et al., 1997; Jack et al., 1999), and later progressing to neocortical regions (Fox et al., 2001; Whitwell et al., 2007). The rate of atrophy in AD dementia is nonlinear (Jack et al., 2013), with considerable variability observed in the spatial pattern of atrophy displayed by individuals (Noh et al., 2014). Thus, simple univariate measures of regional volume loss may not adequately capture and quantify the spatiotemporal complexities of progressive brain atrophy in AD dementia.

To properly handle the nonlinearities in the spatiotemporal evolution of biomarkers along the AD dementia continuum (Jack et al., 2013), many studies have employed machine (and deep) learning modeling solutions (Stamate et al., 2019; Suk et al., 2014; Yang et al., 2021). In particular, studies utilizing convolutional neural network architecture, with its established capability of extracting complex feature representations from large datasets, have been employed to accurately predict progression from mild cognitive impairment (MCI) to AD dementia (Huang et al., 2019; Li et al., 2019; Li & Liu, 2019; Spasov et al., 2019). However, most previous studies focused on improving the predictive accuracy of models, rather than providing interpretable findings in clinical settings. Thus, the spatiotemporal patterns of brain atrophy that are predictive of progression from MCI to AD dementia, tested in a model that can properly capture complex feature representations, remain largely unknown.

In the current study, we aimed to identify the major spatial patterns of brain atrophy underlying the progressive neurodegenerative cascade in AD dementia. To that effect, we deployed a convolutional neural network model to differentiate progressive versus stable MCI based on whole brain gray matter (GM) density maps derived from structural MRI. We then performed occlusion analysis (Kwak et al., 2022), retesting the model while removing single regions from its input, to estimate the contribution of regional cortical atrophy to the progression of AD dementia. Finally, we studied the spatial pattern of atrophy in more localized substates (the medial temporal lobe) and evaluated our methods against a well-established staging scheme for AD dementia progression (Braak & Braak, 1991).

2 | MATERIALS AND METHODS

2.1 | Participant characteristics

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and AD dementia. For up-to-date information, see www.adni-info.org. All subjects provided written informed consent and the study protocol was approved by the local Institutional Review Boards. We used 219 subjects from the ADNI database to train the deep learning model to differentiate between subjects with AD dementia and cognitively normal (CN) control participants. In accordance with recent recommendations (Jack et al., 2018) all subjects with AD dementia had abnormal levels of cerebrospinal fluid (CSF) Amyloid beta $A\beta_{42}$ and CSF p-tau_{181} (henceforth, A + T+), based on established cut-offs (Ewers et al., 2019), while all CN subjects displayed no signs of such pathology (henceforth, A − T−). Cut-off values were <976.6 pg/ml for CSF $A\beta_{42}$ and >21.8 pg/ml for p-tau_{181}. The model was optimized during training based on five-fold cross validation (Kwak et al., 2021). We subsequently tested the model with an independent dataset composed of 115 A+ MCI subjects, obtained from the ADNI-2/GO cohort. This testing set was partitioned into subjects who showed progressive MCI (pMCI) or stable MCI (sMCI), depending on whether they progressed to A + T+ AD dementia or not over an 18 months period. We excluded 13 additional subjects who reverted to normal cognition over the same time period. Demographic characteristics of each subsample used in this study are presented in Table 1.

2.2 | Image acquisition

Structural MRIs were acquired in the ADNI study using 3 T scanners and were based on either an inversion recovery-fast spoiled gradient recalled or a magnetization-prepared rapid gradient-echo sequences with sagittal slices and voxel size of 1.0 × 1.0 × 1.2 mm (Jack, Bernstein, et al., 2010). Full details of the T1 acquisition parameters

TABLE 1 Demographics

	Training data		Test data	
	AD (A + T+)	CN (A − T−)	pMCI	sMCI
N	110	109	34	81
Age	72.9 ± 8.0	71.0 ± 5.3	72.3 ± 5.1	72.8 ± 6.9
Gender, female	46 (41.8%)	57 (52.3%)	14 (41.2%)	24 (29.6%)
MMSE	22.9 ± 2.0	29.1 ± 1.1	27.2 ± 1.7	28.0 ± 1.8

Note: Continuous variables are presented as mean ± SD and categorical variable is presented as %.
Abbreviations: AD, Alzheimer's disease; CN, cognitively normal; MMSE, Mini-Mental State Examination; *N*, number of subjects; pMCI, progressive mild cognitive impairment; sMCI, stable mild cognitive impairment.

and imaging processing steps are listed on the ADNI website (http://adni.loni.usc.edu/methods/documents/).

2.3 | Image processing

T1-weighted images were downloaded from the ADNI database. T1-weighted images were analyzed using Statistical Parametric Mapping 12 (SPM12; Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm) running on MATLAB 9.8.0 (Math-Works, Natick, MA, USA). Briefly, all MR images were aligned such that the origin was located at the anterior commissure and segmented into GM, white matter, and CSF (Ashburner & Friston, 2005). We then used the diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL) registration (Ashburner, 2007) to normalize the six tissue probability maps to Montreal Neurological Institute space, with a resolution of 2 mm isotropic voxels, and produce the final GM density maps. Subsequently, all GM density images were used in the original intact models, as well as during occlusion analysis (see below). Cortical volumetric parcellation was performed as an initial step prior to occlusion analysis with the automated parcellation tool available in FreeSurfer v6.0 (https://surfer.nmr.mgh.harvard.edu/), based on the Desikan–Killiany protocol (Desikan et al., 2006).

2.4 | Procedure overview

While the progression from MCI to AD dementia is a continuous process, the binary task of classifying pMCI versus sMCI relies on clinical diagnoses made by trained clinicians and established criteria, coupled here by the NIA-AA research framework (Jack et al., 2018). As such, it provides a useful avenue for studying progression along the AD continuum. To delineate the contribution of regional atrophy patterns to AD dementia progression, we extended an approach we have introduced recently in a study on the involvement of hippocampal subfields in AD dementia progression (Kwak et al., 2022). This approach is based on a combination of a prognostic deep learning model with occlusion learning. Our approach begins by training and testing the model in the task of differentiating pMCI versus sMCI, based on whole-brain GM volume. The model is then retested multiple times, wherein specific brain regions are occluded from the model's testing set's input data, while all other model inputs remain unchanged. The performance of the models with occluded input is then compared to that of the intact, full model, and ranked according to differences with respect to the intact model. Confidence intervals for the performance metrics of the model were estimated via bootstrapping (10,000 permutations, 80% of the data).

2.5 | Deep learning model

Our model is based on the 3D Densely connected convolutional neural network (DenseNet) architecture (Huang et al., 2017), which has been designed such that all layers are directly connected to ensure maximum information flow in the network. The model's proposed architecture is shown in Figure 1. The 3D volumes of GM density with size of 91 × 109 × 91 were used as input to the our model. Data augmentation was performed on the training set by flipping volumes left to right and randomly shifting by up to 10% and rotating by up to 20° in any direction. For each layer, the feature maps resulting from of all preceding layers are used as inputs. The DenseNet model alleviates the vanishing gradient problem and has a substantially lower number of parameters than many other models (Huang et al., 2017). DenseNet consists of a convolutional layer, four dense blocks, three transition layers, a global averaging pooling layer, and a fully connected layer. Each dense block consists of several convolutional layers and a transition layer. Each dense block includes a composite function of multiple consecutive operations, including batch normalization layer, leaky rectified linear unit, a 1 × 1 × 1 convolutional layer, a 5 × 5 × 5 convolutional layer and a dropout layer. A transition layer between two dense blocks performs a down-sampling operation, which consists of batch normalization, 3 × 3 × 3 convolutional, leaky rectified linear unit, a dropout layer, and 2 × 2 × 2 average pooling layer. Global averaging pooling layers are then connected by a fully connected layer. The last fully connected layer generates a probability distribution over two labels with a sigmoid function.

2.6 | Implementation

The deep learning model was implemented in the Keras library with the TensorFlow 2.0 backend. All training and testing were performed on an Ubuntu system 18.04.3, with 16 GB RAM, Intel® Xeon® CPU

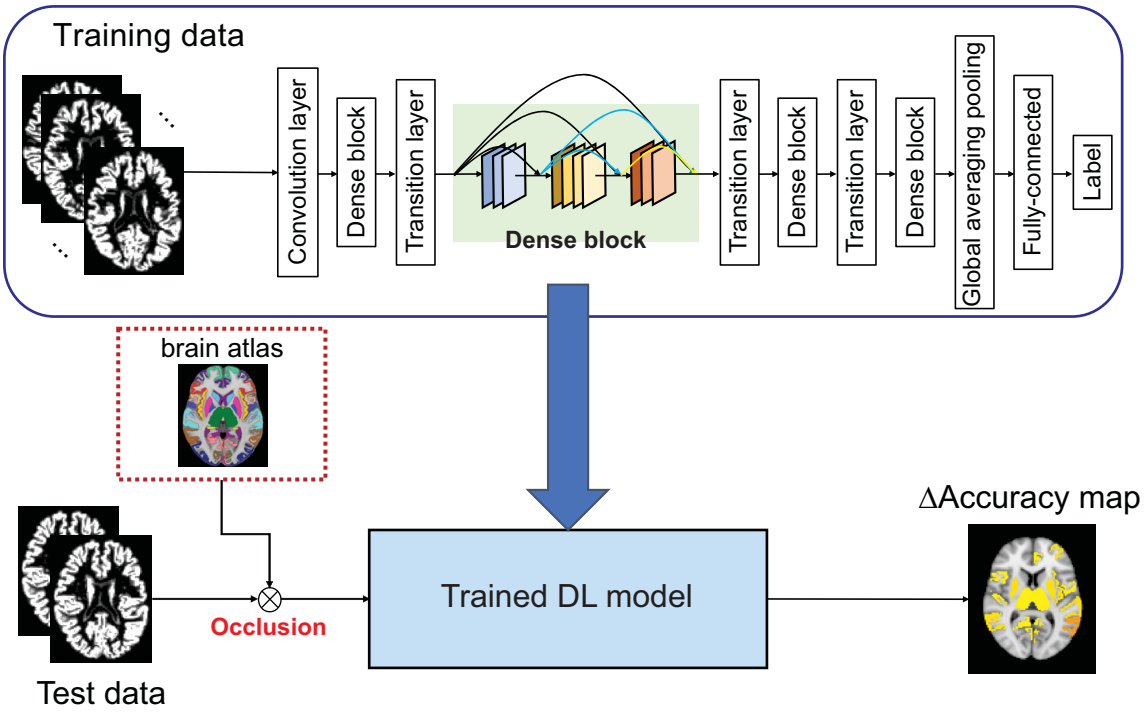

FIGURE 1 Study design. An illustration of the proposed deep learning model used in this study. A 3D-DenseNet convolutional neural network model was trained to differentiate AD and CN subjects based on whole brain GM density maps. The model was then tested in the task of classifying stable and progressive MCI. The trained model was subsequently retested multiple times in conjunction with occlusion analysis, wherein brain regions, defined based on a whole brain atlas, were removed from the model's input. The effect of occlusion on the performance of the model, relative to that of an intact model, was then calculated and visualized. AD, Alzheimer's disease; CN, cognitively normal; GM, gray matter; 3D-DenseNets, 3D densely convolutional neural network

@2.4 GHz, and 16 GB Nvidia Tesla V100 graphic cards. The weights of our model were randomly initialized from a Gaussian distribution. The model was trained for 200 epochs with a batch size of 24 and optimized using stochastic gradient descent based on adaptive estimation of first- and second-order moments (Kingma & Ba, 2015) and an exponentially decaying learning rate. The initial learning rate was set at 0.0001 and decayed by a factor of 0.9 after every 10,000 steps. A dropout layer was included in the dense block, with the dropout rate set to 0.2. In the batch normalization step, beta and gamma weights were initialized with L2 regularization set at 1×10^{-4} and epsilon set to 1.1×10^{-5}. An L2 regularization penalty coefficient, included in the fully connected layer, was set at 0.01. The model was stable after an iteration of 150 epochs. Training time was about 10 h.

2.7 | Occlusion analysis

Occlusion analysis was used as in our previous studies (Kwak, Giovanello, et al., 2021; Kwak et al., 2022). Here, occlusion was used to identify the relative contribution of each brain region to prediction of AD dementia progression, while holding all other complex, multivoxel atrophy patterns constant. Cortical regions were identified based on FreeSurfer's cortical parcellation routine (see Section 2.3). To perform occlusion analysis on each single brain region, we masked it out from each input image by setting the values of all voxels corresponding to that brain region to zero. We then retested the model's performance with occlusion and quantified the contribution of the occluded region to the model's performance in predicting regional atrophy as the following:

$$\Delta \mathrm{Acc}_i = \frac{\mathrm{Acc} - \mathrm{Acc}_i}{\mathrm{Acc}} \quad (1)$$

with Acc referring to the performance of the intact model with unmodified input images, and Acc_i referring to the model's performance on input images with region *i* occluded.

We also implemented patch-based occlusion analysis to evaluate more specific regional contributions to the performance of the prediction model. We first defined the initialization mask around the structure of interest, for example, the medial temporal lobe, merging ROIs (i.e., entorhinal, fusiform, parahippocampal, hippocampus, and amygdala) to define the initial mask. We then masked out $5 \times 5 \times 5$ patches centered around each voxel of the initial mask from the input data. We then evaluated the ΔAcc for each occluded patch by retesting the trained model and assigning the ΔAcc to the central voxel of that patch. We iteratively performed this task for all voxels in the initial mask to generate a map depicting the contribution of voxels to the prediction of AD dementia progression within the target ROI.

2.8 | Contribution of Braak ROIs to prediction of AD dementia progression

We next explored the contribution of atrophy in ROIs defined based on Braak's staging scheme (Braak & Braak, 1991) to prediction of AD dementia progression. This influential staging scheme delineates neurofibrillary pathology in early, intermediate, and late AD dementia. We tested the extent to which atrophy in Braak ROIs contributed to the performance of our model in the task of differentiating pMCI versus sMCI. In the current study, we focused on Braak stages I/II, which correspond to the transentorhinal stage, Stages III and IV, which correspond to the limbic stage, and Stages V and VI, which correspond to the isocortical stages. Braak ROIs were created by compositing FreeSurfer-derived ROIs (Table S1). We then evaluated via occlusion analysis the ΔAcc for each Braak ROI/stage, as explained in previous sections. Results are displayed on a cortical surface using *Simple Brain plot* (https://github.com/dutchconnectomelab/Simple-Brain-Plot).

3 | RESULTS

3.1 | Participant demographics

This study included data from a total of 334 subjects, across the AD dementia, CN, pMCI, and sMCI groups, all obtained from the ADNI database. Demographic characteristics are presented in Table 1. Based on the NIA-AA guidelines (Jack et al., 2018), only AD dementia subjects with abnormal CSF Aβ and tau were included in the study, while CN subjects had to display no signs of abnormality on the same biomarkers. In the comparison between the AD dementia and CN groups, there were no significant differences in gender ($\chi^2 = 2.41$, $p = .12$) while age ($t_{217} = 2.19$, $p = .04$) was significantly different. All MCI subjects included in the analysis were amyloid positive (Jack et al., 2018). Comparison between the pMCI and sMCI groups revealed no significant differences in age ($t_{113} = 0.32$, $p = .75$) or gender distributions ($\chi^2 = 1.44$, $p = .23$).

3.2 | A deep learning model for differentiating between pMCI and sMCI

We first trained a 3D convolutional neural network with the DenseNet architecture (Huang et al., 2017) for differentiating between pMCI and sMCI. We adopted a fivefold cross validation framework in the training phase, in order to optimize the model and assess its stability, and applied data augmentation to the training dataset, as means to improve the performance of the model. As done in previous studies, we trained our model on the task of differentiating AD dementia and CN subjects and then tested the model's performance in classifying pMCI and sMCI (Huang et al., 2019; Li & Liu, 2019; Kwak et al., 2021; Falahati et al., 2014). We reasoned that training the model on the former task will allow it to learn the necessary representations needed in order to successfully complete the latter task. For the task of differentiating AD dementia and CN subjects, the model with the best performance achieved an accuracy of 93.75% in one of the folds and an area under the curve (AUC) of the receiver operating characteristic (ROC) curve of 0.98. In the task of classifying pMCI versus sMCI, the proposed deep learning model achieved an accuracy of 73.90% (95% CI: 69.57–77.17) and an AUC of 0.74 (95% CI: 0.65–0.76; Figure 2a). Accuracy may be misleading when used with an imbalanced dataset; therefore, we also evaluated the normalized confusion matrix for the classifier's performance (Figure 2b), finding that the model achieved sensitivity of 0.62 (95% CI: 0.54–0.70) for correct prediction of pMCI and specificity of 0.78 (95% CI: 0.73–0.83) for true negative prediction of pMCI (i.e., sMCI). In addition, retesting the performance of the deep learning model with age-matched training

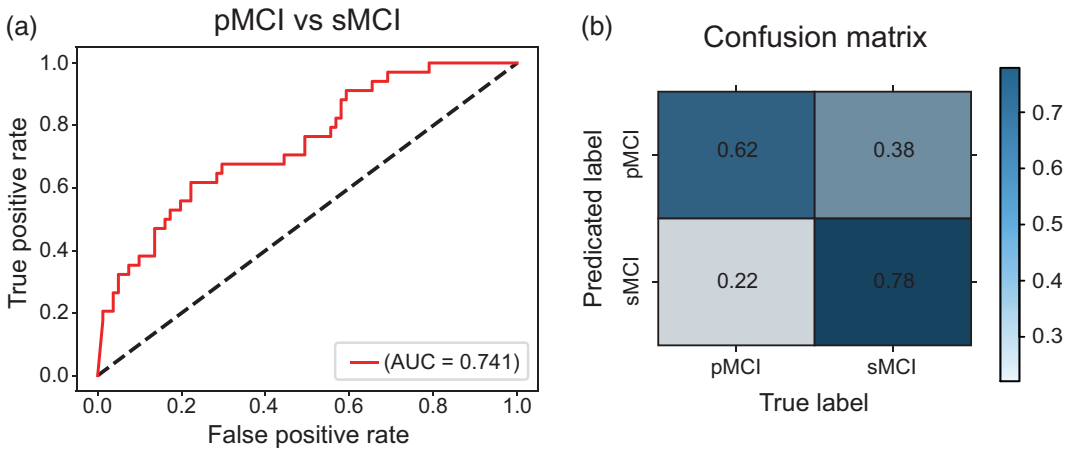

FIGURE 2 The performance of the proposed convolutional neural network model. The proposed convolutional neural network model was evaluated (a) ROC curve of the proposed model in the task of differentiating pMCI versus sMCI. (b) Confusion matrix evaluating sensitivity (correct prediction of pMCI) and specificity (true negative prediction of pMCI, i.e., sMCI) in the task of differentiating pMCI vs. sMCI. AUC, area under the (ROC) curve; pMCI, progressive mild cognitive impairment; ROC, receiver operating characteristic; sMCI, stable mild cognitive impairment

data resulted in similar accuracy and AUC values (Figure S1). We also evaluated the extent to which the performance of the proposed model differed from chance levels, with a permutation test under a null distribution generated with a random classifier ($p < .001$, Figure S2).

3.3 | Regional contribution to prediction of AD dementia progression

We then used occlusion analysis to identify which regions contributed to the classification of pMCI versus sMCI. In this analysis, we repeatedly tested the deep learning model while occluding different brain regions in the test dataset to evaluate the individual contribution of each brain region to model performance, while all other atrophy patterns are held constant. To occlude a region, we set the intensity values of voxels in that region to zero before inputting the full image into to the deep learning model. The delta accuracies for each brain region are defined by the change in the model's accuracy in classifying pMCI and sMCI after occlusion of a particular brain region normalized by the model's baseline performance prior to occlusion (See Figure 3a, b). Occlusion analysis was used in the current study as means to improve the explainability of the deep learning model's output. We next wished to assess whether the results of the occlusion analysis would match those obtained when using another widely used approach to improve explainability in machine/deep leaning—the use of class activation map (Zhou et al., 2016). More specifically, we have generated a gradient-weighted class activation map (Grad-CAM), generated by visualizing the gradients of the target class flowing back into the final convolutional layer. The Grad-CAM, however, was very diffuse (Figure S3), diminishing the potential to improve explainability via their use (Reyes et al., 2020; Saporta et al., 2021). Thus, in the context of the current study, occlusion analysis emerges as a favorable approach to improve explainability. We found that the left and right hippocampus, fusiform gyrus, inferior temporal gyrus, and precuneus had the largest influence on the performance of the model, and by

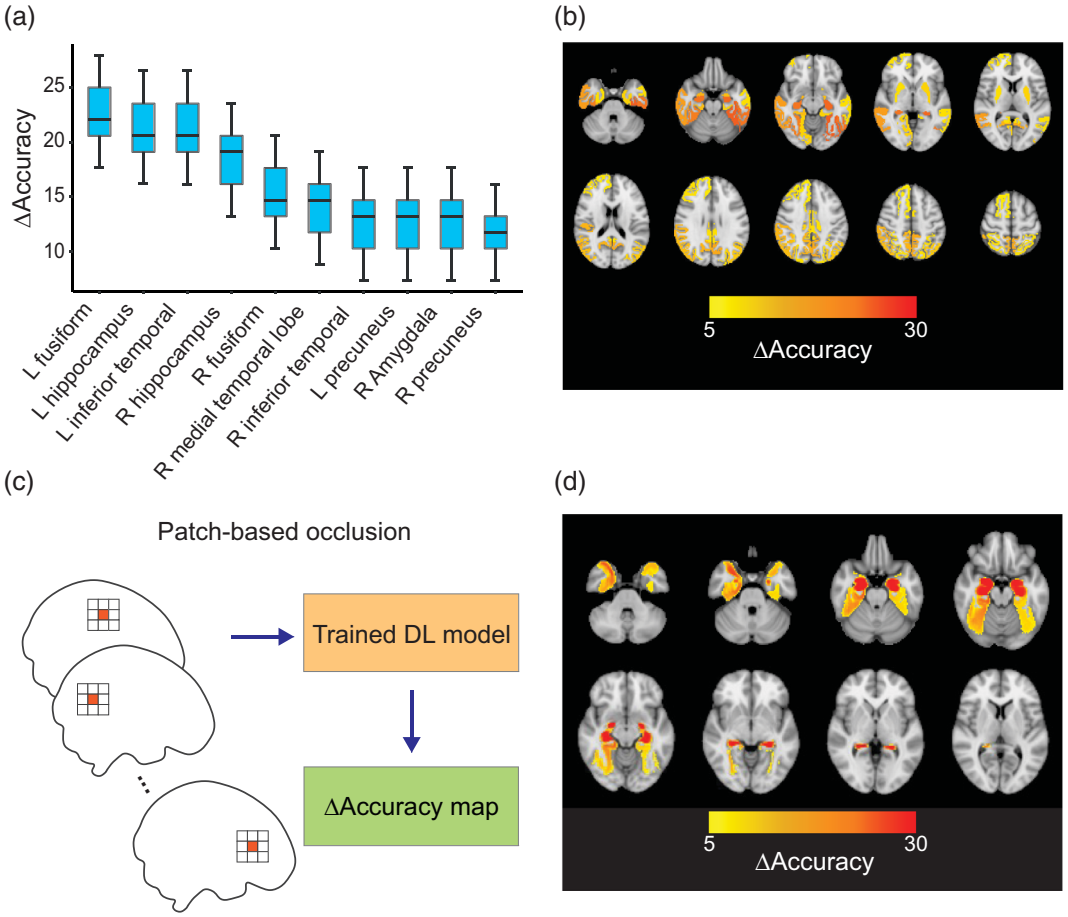

FIGURE 3 Occlusion analysis. (a) Δ accuracy (performance when regional-based occlusion is applied relative to that of an intact model) is shown for the top 10 regions with the highest contribution (highest accuracy loss) to prediction of AD progression. Confidence intervals were generated via bootstrapping (10,000 permutations) for Δ accuracy. (b) Occlusion maps depicting the results of regional occlusion analysis overlayed on top of an anatomical image. (c) An illustration of patch-based occlusion analysis. At each step, a single patch is masked out and then delta accuracy is mapped to the center voxel of that patch. (d) The results of patch-based occlusion analysis performed in the medial temporal lobe. pMCI, progressive mild cognitive impairment; sMCI, stable mild cognitive impairment

proxy to prediction of progression of MCI to AD dementia. We tested whether the size of the occluded ROIs related to accuracy in our proposed framework. No significant correlation was observed between the size of individual occluded ROI (Figure S4A) and accuracy loss ($R^2 = 0.01$, $p = .36$; Figure S4B). To further validate our proposed method, we next assessed the more precise and localized contribution of voxels within the medial temporal lobe to the prediction of AD dementia progression using patch-based occlusion analysis. Briefly, we retested the trained deep learning model, each time occluding a patch centered on each voxel of the medial temporal region and calculating differences in accuracy relative to the intact model (Figure 3c). This analysis showed that within the medial temporal region, the bilateral hippocampus, and amygdala are more central to the prediction of AD dementia progression (Figure 3d), validating the spatial precision of our proposed method.

3.4 | Regional occlusion of ROIs based on the Braak staging scheme

Finally, we validated our occlusion analysis approach by testing whether occlusion of ROIs, which are defined based on Braak's staging scheme (Braak & Braak, 1991), results in appropriate model loss across the different ROIs (Figure 4). We found that the occlusion of the ROI corresponding to the intermediate Braak stage III, which primarily consists of limbic areas, had the largest impact on the performance of the model, relative to other staging-related ROIs (Figure 4). However, we also found that occlusion of transentorhinal ROIs defined in Braak stages I/II and limbic ROIs defined in Braak stages IV also highly impacted the performance. On the other hand, loss in accuracy was lower when occluding regions associated with the later Braak stages V and VI. Therefore, the relationships learned by our model aligned well with existing studies indicating the importance of regions defined in early Braak stages for predicting progression from MCI to AD dementia (Braak & Braak, 1991).

4 | DISCUSSION

While progressive brain atrophy is a ubiquitous consequence of AD dementia, the major spatial patterns of atrophy observed when individuals progress from the prodromal (MCI) to the clinical phases of AD dementia are not understood well. The motivation of the current study was to identify the major spatial patterns of atrophy predictive of AD dementia progression by leveraging an interpretable deep learning modeling approach. To that end, we propose a framework based on a prognostic convolutional neural network model and occlusion analysis to investigate the regional contribution of atrophy patterns to the prediction of AD dementia progression. Our use of occlusion analysis, in particular, yielded explainable results wherein the contribution of specific brain regions to the performance of the model, and by proxy to AD dementia progression can be evaluated and visualized while holding the contribution of (atrophy in) other regions constant. We found that the hippocampus, fusiform and inferior temporal gyri and precuneus were the strongest contributors to the prediction of AD dementia progression. When considering the medial temporal lobe separately, our analysis revealed larger effects after hippocampal and amygdala occlusion. Finally, we validated our approach by occluding ROIs motivated by the influential Braak staging scheme (Braak & Braak, 1991), finding that the occlusion of a Stage III (limbic) ROI had the largest impact on the performance of the model.

We report that atrophy in bilateral hippocampus, fusiform and inferior temporal gyri, and precuneus played a more central role than other regions in the prediction of progression from MCI to AD dementia. These findings are consistent with studies reporting that hippocampus, fusiform, and inferior temporal gyri are affected in early stages of AD dementia (Convit et al., 1997; De Santi et al., 2001; Jack et al., 1999). Additionally, we evaluated the spatial precision of our method by performing patch-based occlusion analysis focused more locally on the medial temporal lobe. In line with previous studies (Convit et al., 2000; Grundman et al., 2002; Mitolo et al., 2019) voxels in the hippocampus and amygdala contributed the most to the

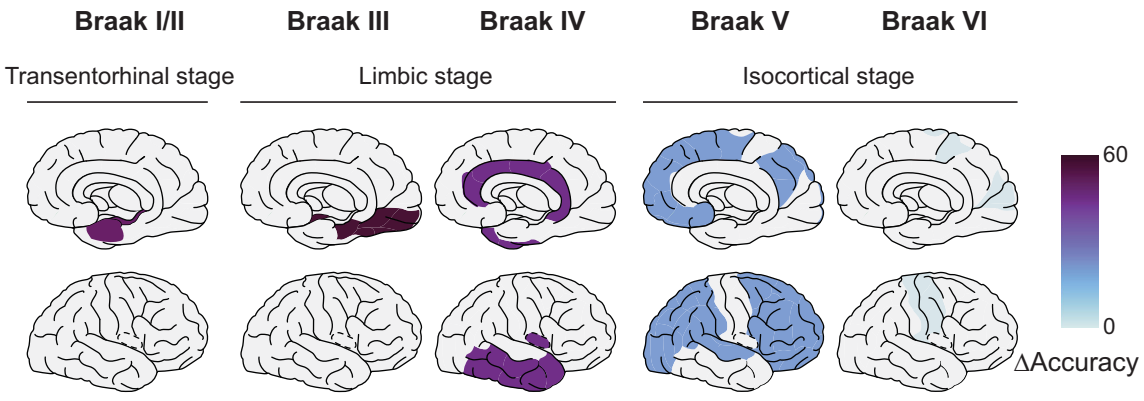

FIGURE 4 Occlusion of Braak staging ROIs. The results of the occlusion analysis are shown where composite ROIs corresponding to five Braak stages (Stages I/II, III, IV, V, and VI) were occluded, evaluating their contribution to differentiating between pMCI and sMCI. The ΔAcc of each composite Braak staging ROI superimposed on a medial and radial cortical surface model. pMCI, progressive mild cognitive impairment; ROI, region of interest; sMCI, stable mild cognitive impairment

performance of the model and by proxy to AD dementia progression. Combined, our findings demonstrate that patterns of cortical atrophy, particularly in medial temporal lobe regions, differentiate individuals with MCI into those more and less likely to progress to AD dementia. On the other hand, most of the other brain regions considered in the model were redundant, contributing minimally to its performance. It was previously shown that functional redundancy in the brain may constitute a form of brain reserve (Stern et al., 2020), offering protection against cognitive decline in normal and pathological aging (Langella, Mucha, et al., 2021; Langella, Sadiq, et al., 2021; Sadiq et al., 2021). While speculative at this stage, the current results may similarly reflect the existence of structural reserve/redundancy, which contributes to the differences between individuals with stable and progressive MCI. This could be addressed in future research.

We validated our approach of combining a prognostic deep learning model with occlusion analysis, by occluding ROIs based on the Braak staging scheme (Braak & Braak, 1991) and evaluating whether this step resulted in stage-appropriate results. According to this influential staging scheme neurofibrillary pathology in AD dementia starts in transentorhinal cortex, spreading into entorhinal cortex, hippocampus, amygdala and inferior temporal cortex and eventually to other neocortical regions (Braak & Braak, 1991). Studies utilizing MRI have demonstrated that the trajectory of atrophy changes in AD dementia is largely consistent with Braak's staging scheme (Burton et al., 2009; Jack et al., 2002; Silbert et al., 2003; Whitwell et al., 2012). In the current study, we found that occlusion of ROIs corresponding to Braak stage III had a larger effect on the performance of the prognostic model, relative to occlusion of Braak I/II ROIs. These results highlight the importance of limbic areas in the progression from MCI to AD dementia. Moreover, the effect of occlusion gradually decreased when ROIs from later Braak stages were occluded. These findings are consistent with previous studies, documenting brain atrophy in Braak III regions during the prodromal stages of AD dementia (Desikan et al., 2008; Yao et al., 2012), as well as with the early involvement of entorhinal cortex and hippocampus along the AD dementia continuum (Devanand et al., 2007; Jack et al., 1999). Altogether, our findings suggest that patterns of atrophy starting from the entorhinal cortex and accumulating up to Braak III regions differentiated subjects with progressive MCI from those with a more stable MCI diagnosis, consistent with the former group being composed of subjects who can be viewed as being at the prodromal stage of AD dementia.

Deep learning models have been applied in multiple biomedical image analysis tasks, including segmentation (Chen et al., 2018; Kamnitsas et al., 2017), reconstruction (Schlemper et al., 2018; Yang et al., 2018), and classification (Cheng et al., 2019; Hosseini-Asl et al., 2016; Suk et al., 2014). Emphasis on interpretability has increased in recent years, aimed at better understating how and why extracted features contribute to successful class prediction. However, studies have shown that methods such as class activation maps often fail to provide sufficient interpretable findings in models based on medical imaging data (Reyes et al., 2020; Saporta et al., 2021). On the other hand, occlusion analysis, as has been applied here, has been successfully applied as an interpretable approach in similar settings (Kermany et al., 2018; Lu et al., 2021). We suggest that, when combined with a diagnostic or prognostic machine learning model, occlusion analysis can serve as a useful approach for studying the contribution of complex image-based features to various disease states.

Several limitations of our study should be acknowledged. First, although we used one of the biggest publicly available AD dementia datasets, we acknowledge that future work could benefit from working with larger sample sizes. Second, our model is based solely on structural MRI data. Incorporating other imaging modalities or fluid biomarker data, to assess the contribution of amyloid and tau burden and their synergies with atrophy (Sadiq et al., 2022) may further benefit the performance of the model and should be considered in future studies. Third, our study evaluated the major regional patterns of atrophy predicative of subsequent progression from MCI to AD dementia. We recognize that incorporating longitudinal neuroimaging into a similar design could provide additional valuable information on the spatiotemporal patterns of atrophy seen in AD dementia progression. This could also be achieved in future research.

5 | CONCLUSION

In conclusion, we identified the major regional substrates of cortical atrophy, predictive of AD dementia progression using a combination of deep learning and occlusion. In agreement with previous results, we found that atrophy in the hippocampus, fusiform, and inferior temporal gyri was the strongest predictors of AD dementia progression. These results further establish the potential for early identification of individuals with MCI who will likely progress to AD dementia based on patterns of cortical atrophy.

AUTHOR CONTRIBUTIONS

Kichang Kwak and Eran Dayan conceived research; Kichang Kwak analyzed data; Kichang Kwak, William Stanford, and Eran Dayan interpreted results and wrote the paper.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health under Award Number R01AG062590 and by the UNC Idea Grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La

Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

All data reported in the current study were obtained from the ADNI (http://adni.loni.usc.edu/). Other materials are available from the authors upon reasonable request.

ORCID

Eran Dayan https://orcid.org/0000-0001-9710-9210

REFERENCES

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*, 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*, 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Bobinski, M., Wegiel, J., Tarnawski, M., Bobinski, M., Reisberg, B., de Leon, M. J., Miller, D. C., & Wisniewski, H. M. (1997). Relationships between regional neuronal loss and neurofibrillary changes in the hippocampal formation and duration and severity of Alzheimer disease. *Journal of Neuropathology and Experimental Neurology*, *56*(4), 414–420. https://doi.org/10.1097/00005072-199704000-00010

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, *82*, 239–259. https://doi.org/10.1007/BF00308809

Burton, E. J., Barber, R., Mukaetova-Ladinska, E. B., Robson, J., Perry, R. H., Jaros, E., Kalaria, R. N., & O'Brien, J. T. (2009). Medial temporal lobe atrophy on MRI differentiates Alzheimer's disease from dementia with Lewy bodies and vascular cognitive impairment: A prospective study with pathological verification of diagnosis. *Brain: A Journal of Neurology*, *132*(Pt 1), 195–203. https://doi.org/10.1093/brain/awn298

Chen, H., Dou, Q., Yu, L., Qin, J., & Heng, P. A. (2018). VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage.*, *170*, 446–455. https://doi.org/10.1016/j.neuroimage.2017.04.041

Cheng, B., Liu, M., Zhang, D., & Shen, D. (2019). Robust multi-label transfer feature learning for early diagnosis of Alzheimer's disease. *Brain Imaging and Behavior*, *13*(1), 138–153. https://doi.org/10.1007/s11682-018-9846-8

Convit, A., de Asis, J., de Leon, M. J., Tarshish, C. Y., De Santi, S., & Rusinek, H. (2000). Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. *Neurobiology of Aging*, *21*(1), 19–26. https://doi.org/10.1016/s0197-4580(99)00107-4

Convit, A., De Leon, M. J., Tarshish, C., De Santi, S., Tsui, W., Rusinek, H., & George, A. (1997). Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiology of Aging*, *18*(2), 131–138. https://doi.org/10.1016/s0197-4580(97)00001-8

Dallaire-Théroux, C., Callahan, B. L., Potvin, O., Saikali, S., & Duchesne, S. (2017). Radiological-pathological correlation in Alzheimer's disease: Systematic review of Antemortem magnetic resonance imaging findings. *Journal of Alzheimer's Disease: JAD*, *57*(2), 575–601. https://doi.org/10.3233/JAD-161028

De Santi, S., de Leon, M. J., Rusinek, H., Convit, A., Tarshish, C. Y., Roche, A., Tsui, W. H., Kandil, E., Boppana, M., Daisley, K., Wang, G. J., Schlyer, D., & Fowler, J. (2001). Hippocampal formation glucose metabolism and volume losses in MCI and AD. *Neurobiology of Aging*, *22*(4), 529–539. https://doi.org/10.1016/s0197-4580(01)00230-5

Desikan, R. S., Fischl, B., Cabral, H. J., Kemper, T. L., Guttmann, C. R. G., Blacker, D., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2008). MRI measures of temporoparietal regions show differential rates of atrophy during prodromal AD. *Neurology*, *71*(11), 819–825. https://doi.org/10.1212/01.wnl.0000320055.57329.34

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Devanand, D. P., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G. H., Honig, L. S., Mayeux, R., Stern, Y., Tabert, M. H., & de Leon, M. J. (2007). Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. *Neurology*, *68*(11), 828–836. https://doi.org/10.1212/01.wnl.0000256697.20968.d7

Ewers, M., Franzmeier, N., Suárez-Calvet, M., Morenas-Rodriguez, E., Caballero, M. A. A., Kleinberger, G., Piccio, L., Cruchaga, C., Deming, Y., Dichgans, M., Trojanowski, J. Q., Shaw, L. M., Weiner, M. W., & Haass, C. (2019). Increased soluble TREM2 in cerebrospinal fluid is associated with reduced cognitive and clinical decline in Alzheimer's disease. *Science Translational Medicine*, *11*(507), eaav6221. https://doi.org/10.1126/scitranslmed.aav6221

Falahati, F., Westman, E., & Simmons, A. (2014). Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. *Journal of Alzheimer's Disease.*, *41*, 685–708. https://doi.org/10.3233/JAD-131928

Fox, N. C., Crum, W. R., Scahill, R. I., Stevens, J. M., Janssen, J. C., & Rossor, M. N. (2001). Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *Lancet (London, England)*, *358*(9277), 201–205. https://doi.org/10.1016/S0140-6736(01)05408-3

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews. Neurology*, *6*(2), 67–77. https://doi.org/10.1038/nrneurol.2009.215

Grundman, M., Sencakova, D., Jack, C. R., Petersen, R. C., Kim, H. T., Schultz, A., Weiner, M. F., DeCarli, C., DeKosky, S. T., Van Dyck, C., Thomas, R. G., & Thal, L. J. (2002). Brain MRI hippocampal volume and prediction of clinical status in a mild cognitive impairment trial. *Journal of Molecular Neuroscience.*, *19*, 23–27. https://doi.org/10.1007/s12031-002-0006-6

Halliday, G. (2017). Pathology and hippocampal atrophy in Alzheimer's disease. *The Lancet Neurology*, *16*(11), 862–864. https://doi.org/10.1016/S1474-4422(17)30343-5

Hosseini-Asl, E., Keynton, R., & El-Baz, A. (2016). Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In *2016 IEEE*

international conference on image processing (ICIP), pp. 126–130. IEEE, 2016. https://doi.org/10.1109/ICIP.2016.7532332

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. IEEE. https://doi.org/10.1109/CVPR.2017.243

Huang, Y., Xu, J., Zhou, Y., Tong, T., & Zhuang, X. (2019). Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00509

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., & Snyder, H. M. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Bernstein, M. A., Borowski, B. J., Gunter, J. L., Fox, N. C., Thompson, P. M., Schuff, N., Krueger, G., Killiany, R. J., DeCarli, C. S., Dale, A. M., Carmichael, O. W., Tosun, D., Weiner, M. W., & Alzheimer's Disease Neuroimaging Initiative. (2010). Update on the magnetic resonance imaging core of the Alzheimer's Disease Neuroimaging Initiative. *Alzheimers Dementia*, *6*, 212–220. https://doi.org/10.1016/j.jalz.2010.03.004

Jack, C. R., Dickson, D. W., Parisi, J. E., Xu, Y. C., Cha, R. H., O'Brien, P. C., Edland, S. D., Smith, G. E., Boeve, B. F., Tangalos, E. G., Kokmen, E., & Petersen, R. C. (2002). Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology*, *58*(5), 750–757.

Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., Lesnick, T. G., Pankratz, V. S., Donohue, M. C., & Trojanowski, J. Q. (2013). Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, *12*(2), 207–216. https://doi.org/10.1016/S1474-4422(12)70291-0

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet. Neurology*, *9*(1), 119–128. https://doi.org/10.1016/S1474-4422(09)70299-6

Jack, C. R., Petersen, R. C., Xu, Y. C., O'Brien, P. C., Smith, G. E., Ivnik, R. J., Boeve, B. F., Waring, S. C., Tangalos, E. G., & Kokmen, E. (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, *52*(7), 1397. https://doi.org/10.1212/WNL.52.7.1397

Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis.*, *36*, 61–78. https://doi.org/10.1016/j.media.2016.10.004

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., Pei, J., Ting, M., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., … Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*, 1122–1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In *3rd international conference on learning representations, ICLR 2015 - conference track proceedings* (pp. 1–15).

Kwak, K., Giovanello, K. S., Bozoki, A., Styner, M., & Dayan, E. (2021). Subtyping of mild cognitive impairment using a deep learning model based on brain atrophy patterns. *Cell Reports Medicine*, *2*(12), 100467. https://doi.org/10.1016/j.xcrm.2021.100467

Kwak, K., Niethammer, M., Giovanello, K. S., Styner, M., Dayan, E., & for the Alzheimer's Disease Neuroimaging Initiative. (2022). Differential role for hippocampal subfields in Alzheimer's disease progression revealed with deep learning. *Cerebral Cortex*, *32*, 467–478. https://doi.org/10.1093/cercor/bhab223

Langella, S., Mucha, P. J., Giovanello, K. S., Dayan, E., & for the Alzheimer's Disease Neuroimaging Initiative. (2021). The association between hippocampal volume and memory in pathological aging is mediated by functional redundancy. *Neurobiology of Aging*, *108*, 179–188. https://doi.org/10.1016/j.neurobiolaging.2021.09.002

Langella, S., Sadiq, M. U., Mucha, P. J., Giovanello, K. S., Dayan, E., & Alzheimer's Disease Neuroimaging Initiative. (2021). Lower functional hippocampal redundancy in mild cognitive impairment. *Translational Psychiatry*, *11*(1), 61. https://doi.org/10.1038/s41398-020-01166-w

Li, F., & Liu, M. (2019). A hybrid convolutional and recurrent neural network for hippocampus analysis in Alzheimer's disease. *Journal of Neuroscience Methods*, *323*, 108–118. https://doi.org/10.1016/j.jneumeth.2019.05.006

Li, H., Habes, M., Wolk, D. A., & Fan, Y. (2019). A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal MRI. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *15*(8), 1059–1070. https://doi.org/10.1016/j.jalz.2019.02.007

Lu, B., Li, H.-X., Chang, Z.-K., Li, L., Chen, N.-X., Zhu, Z.-C., Zhou, H.-X., Li, X.-Y., Wang, Y.-W., Cui, S.-X., Deng, Z.-Y., Fan, Z., Yang, H., Chen, X., Thompson, P. M., Castellanos, F. X., & Yan, C.-G. (2021). A practical Alzheimer disease classifier via brain imaging-based deep learning on 85,721 samples. *BioRxiv*. https://doi.org/10.1101/2020.08.18.256594

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*, *34*(7), 939–944. https://doi.org/10.1212/wnl.34.7.939

Mitolo, M., Stanzani-Maserati, M., Capellari, S., Testa, C., Rucci, P., Poda, R., Oppi, F., Gallassi, R., Sambati, L., Rizzo, G., Parchi, P., Evangelisti, S., Talozzi, L., Tonon, C., Lodi, R., & Liguori, R. (2019). Predicting conversion from mild cognitive impairment to Alzheimer's disease using brain 1H-MRS and volumetric changes: A two-year retrospective follow-up study. *NeuroImage: Clinical*, *23*, 101843. https://doi.org/10.1016/j.nicl.2019.101843

Noh, Y., Jeon, S., Lee, J. M., Seo, S. W., Kim, G. H., Cho, H., Ye, B. S., Yoon, C. W., Kim, H. J., Chin, J., Park, K. H., Heilman, K. M., & Na, D. L. (2014). Anatomical heterogeneity of Alzheimer disease. *Neurology*, *83*(21), 1936–1944. https://doi.org/10.1212/WNL.0000000000001003

Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., & Frisoni, G. B. (2016). Brain atrophy in Alzheimer's disease and aging. *Ageing Research Reviews*, *30*, 25–48. https://doi.org/10.1016/j.arr.2016.01.002

Reyes, M., Meier, R., Pereira, S., Silva, C. A., Dahlweid, F.-M., von Tengg-Kobligk, H., Summers, R. M., & Wiest, R. (2020). On the interpretability of artificial intelligence in radiology: Challenges and opportunities. *Radiology: Artificial Intelligence*, *2*, e190043. https://doi.org/10.1148/ryai.2020190043

Sadiq, M. U., Kwak, K., Dayan, E., & for the Alzheimer's Disease Neuroimaging Initiative. (2022). Model-based stratification of progression along the Alzheimer disease continuum highlights the centrality of biomarker synergies. *Alzheimer's Research & Therapy*, *14*(1), 16. https://doi.org/10.1186/s13195-021-00941-1

Sadiq, M. U., Langella, S., Giovanello, K. S., Mucha, P. J., & Dayan, E. (2021). Accrual of functional redundancy along the lifespan and its effects on cognition. *NeuroImage*, *229*, 117737. https://doi.org/10.1016/j.neuroimage.2021.117737

Saporta, A., Gui, X., Agrawal, A., Pareek, A., Truong, S. Q., Nguyen, C. D., Ngo, V.-D., Seekins, J., Blankenberg, F. G., Ng, A. Y., Lungren, M. P., & Rajpurkar, P. (2021). Deep learning saliency maps do not accurately highlight diagnostically relevant regions for medical image interpretation. *medRxiv*. https://doi.org/10.1101/2021.02.28.21252634

Schlemper, J., Caballero, J., Hajnal, J. V., Price, A. N., & Rueckert, D. (2018). A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*., *37*, 491–503. https://doi.org/10.1109/TMI.2017.2760978

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W. S., Okonkwo, O., Scarmeas, N., Soldan, A., Udeh-Momoh, C., Valenzuela, M., Vemuri, P., & Vuoksimaa, E. (2020). The Reserve, Resilience and Protective Factors PIA Empirical Definitions and Conceptual Frameworks Workgroup. Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimers Dement*. *16*(9), 1305–1311. https://doi.org/10.1016/j.jalz.2018.07.219

Silbert, L. C., Quinn, J. F., Moore, M. M., Corbridge, E., Ball, M. J., Murdoch, G., Sexton, G., & Kaye, J. A. (2003). Changes in premorbid brain volume predict Alzheimer's disease pathology. *Neurology*, *61*(4), 487–492. https://doi.org/10.1212/01.wnl.0000079053.77227.14

Spasov, S., Passamonti, L., Duggento, A., Liò, P., & Toschi, N. (2019). A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *NeuroImage*, *189*, 276–287. https://doi.org/10.1016/j.neuroimage.2019.01.031

Stamate, D., Kim, M., Proitsi, P., Westwood, S., Baird, A., Nevado-Holgado, A., Hye, A., Bos, I., Vos, S. J. B., Vandenberghe, R., Teunissen, C. E., Kate, M. T., Scheltens, P., Gabel, S., Meersmans, K., Blin, O., Richardson, J., de Roeck, E., Engelborghs, S., … Legido-Quigley, C. (2019). A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: Results from the European medical information framework for Alzheimer disease biomarker discovery cohort. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *5*, 933–938. https://doi.org/10.1016/j.trci.2019.11.001

Suk, H.-I., Lee, S.-W., & Shen, D. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*, *101*, 569–582. https://doi.org/10.1016/j.neuroimage.2014.06.077

Whitwell, J. L., Dickson, D. W., Murray, M. E., Weigand, S. D., Tosakulwong, N., Senjem, M. L., Knopman, D. S., Boeve, B. F., Parisi, J. E., Petersen, R. C., Jack, C. R., & Josephs, K. A. (2012). Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: A case-control study. *The Lancet. Neurology*, *11*(10), 868–877. https://doi.org/10.1016/S1474-4422(12)70200-4

Whitwell, J. L., Przybelski, S. A., Weigand, S. D., Knopman, D. S., Boeve, B. F., Petersen, R. C., & Jack, C. R. (2007). 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain: A Journal of Neurology*, *130*, 1777–1786. https://doi.org/10.1093/brain/awm112

Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P. L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., & Firmin, D. (2018). DAGAN: Deep De-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging*, *37*, 1310–1321. https://doi.org/10.1109/TMI.2017.2785879

Yang, Z., Nasrallah, I. M., Shou, H., Wen, J., Doshi, J., Habes, M., Erus, G., Abdulkadir, A., Resnick, S. M., Albert, M. S., Maruff, P., Fripp, J., Morris, J. C., Wolk, D. A., & Davatzikos, C. (2021). A deep learning framework identifies dimensional representations of Alzheimer's disease from brain structure. *Nature Communications*, *12*(1), 7065. https://doi.org/10.1038/s41467-021-26703-z

Yao, Z., Hu, B., Liang, C., Zhao, L., & Jackson, M. (2012). A longitudinal study of atrophy in amnestic mild cognitive impairment and normal aging revealed by cortical thickness. *PLoS One*, *7*(11), e48973. https://doi.org/10.1371/journal.pone.0048973

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921–2929).

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kwak, K., Stanford, W., Dayan, E., & for the Alzheimer's Disease Neuroimaging Initiative (2022). Identifying the regional substrates predictive of Alzheimer's disease progression through a convolutional neural network model and occlusion. *Human Brain Mapping*, 1–11. https://doi.org/10.1002/hbm.26026