Neurobiology of Aging 99 (2021) 53–64

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Transfer learning for predicting conversion from mild cognitive impairment to dementia of Alzheimer's type based on a three-dimensional convolutional neural network

Jinhyeong Bae[a,*], Jane Stocks[b], Ashley Heywood[b], Youngmoon Jung[c], Lisanne Jenkins[d], Virginia Hill[a], Aggelos Katsaggelos[e], Karteek Popuri[f], Howie Rosen[g], M. Faisal Beg[f], Lei Wang[a,d], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Department of Radiology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[b] Department of Psychology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[c] School of Engineering, KAIST, Daejeon, South Korea
[d] Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[e] School of Engineering, Northwestern University, Evanston, IL, USA
[f] School of Engineering Science, Simon Fraser University, Burnaby, Canada
[g] School of Medicine, University of California, San Francisco, CA, USA

ARTICLE INFO

Article history:
Received 16 June 2020
Received in revised form 9 November 2020
Accepted 5 December 2020
Available online 13 December 2020

Keywords:
Convolutional neural network
Dementia of Alzheimer's type
Magnetic resonance imaging
Mild cognitive impairment
Predictive modeling

ABSTRACT

Dementia of Alzheimer's type (DAT) is associated with devastating and irreversible cognitive decline. Predicting which patients with mild cognitive impairment (MCI) will progress to DAT is an ongoing challenge in the field. We developed a deep learning model to predict conversion from MCI to DAT. Structural magnetic resonance imaging scans were used as input to a 3-dimensional convolutional neural network. The 3-dimensional convolutional neural network was trained using transfer learning; in the source task, normal control and DAT scans were used to pretrain the model. This pretrained model was then retrained on the target task of classifying which MCI patients converted to DAT. Our model resulted in 82.4% classification accuracy at the target task, outperforming current models in the field. Next, we visualized brain regions that significantly contribute to the prediction of MCI conversion using an occlusion map approach. Contributory regions included the pons, amygdala, and hippocampus. Finally, we showed that the model's prediction value is significantly correlated with rates of change in clinical assessment scores, indicating that the model is able to predict an individual patient's future cognitive decline. This information, in conjunction with the identified anatomical features, will aid in building a personalized therapeutic strategy for individuals with MCI.

1. Introduction

Dementia of Alzheimer's type (DAT) is a common and severe neurodegenerative disorder (Alzheimer's Association, 2019; Heun et al., 1997). Mild cognitive impairment (MCI), which is characterized by noticeable cognitive decline, precedes DAT and 10%–12% of individuals with MCI convert to DAT every year (Petersen, 2000). Predicting patients who will progress from MCI to DAT is important for patient care as well as in patient selection for clinical trials aimed at treating and preventing Alzheimer's disease (AD) (Roberson and Mucke, 2006). However, current diagnostic tools for predicting conversion to DAT rely heavily on clinical interviews and neuropsychological evaluations and may not be sensitive to the earliest changes required to predict future disease development. Thus, new methodology is needed to better predict disease progression.

With the development of computational methods such as machine learning and deep learning, there is increased utility of biomarker-based diagnosis for disease prediction. Numerous computational methodologies have been proposed to tackle the problem of predicting which MCI patients will convert to DAT (MCI-

This article is submitted to the Neurobiology of Aging on June 7, 2020. This research was funded by the following grants from the National Institute on Aging: AG055121 and AG045333, and by grants from Brain Canada, CIHR, NSERC, and Compute Canada.

* Corresponding author at: Psychiatry and Behavioral Sciences, Northwestern University, Chicago, IL 60611 USA. Tel.: (224) 713-4318.
E-mail address: jinhyeongbae2017@u.northwestern.edu (J. Bae).

[1] 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.

0197-4580/$ – see front matter
https://doi.org/10.1016/j.neurobiolaging.2020.12.005

converters or MCI-C) versus those who do not (MCI-nonconverters or MCI-NC) (Basaia et al., 2019; Cheng et al., 2015; Li et al., 2014; Suk et al., 2017). Of those published, reported accuracy of models is around 75%–80%. There are, however, several limitations of existing studies. First, many failed to assess their model using a separate, independent test data set. It is important to randomly reserve a portion of the whole data set to be included in the independent test set (Kuhn and Johnson, 2013). This is the gold standard practice in the field to evaluate a model's effectiveness and generalizability (Russell and Norvig, 2016), particularly in the absence of feature visualization. In addition, previous research has relied on specific features (e.g., cortical thickness and hippocampal volume) extracted from raw data to train the model (Zheng and Casari, 2018). This approach assumes that the chosen feature is the most informative and may neglect important information inherent in the raw data.

CNN is a deep-learning approach that has evolved in recent years to produce better classification performance and feature visualization than conventional machine learning methods across several fields (Borji et al., 2019). A CNN trained on raw, whole brain data can automatically extract the important imaging features and can offer insights beyond current methods in predicting disease progression. An end-to-end system, which places the model's input and output on each end of the model, requires minimal or no feature extraction, producing features that are not biased. This approach has not yet been actively implemented to predict conversion from MCI to DAT.

In the present study, we implemented an end-to-end 3D-CNN model with transfer learning (Torrey and Shavlik, 2010) to classify MCI-NC versus MCI-C patients using structural magnetic resonance images (sMRI). Transfer learning improves the model's performance by training the model through 2 classification tasks: the source task and the target task. At the source task, the model is pretrained using visual information similar to that used in the target task. Through this task, the model learns generic knowledge that will be helpful in target classification. At the target task, the model is retrained with the resource that is directly relevant to the classification objective using the previously established generic knowledge.

The present study aimed to predict which individuals with MCI converted (MCI-C) and did not covert (MCI-NC), using a CNN model that has been first trained on sMRI scans of healthy individuals (NC) and those with suspected AD (DAT). Using the terminology described previously, we used NC and DAT scans in the source task. The model learns features that most strongly distinguish healthy from diseased brains. The generic knowledge obtained from the source task is transferred to the target task in which scans from patients with MCI are used. The model is then retrained with MCI-NC and MCI-C patients' scans, to extract features that can predict conversion to DAT. Previous research suggests that the classification task of NC versus DAT is similar to the classification task of MCI-NC versus MCI-C (Coupé et al., 2012; Da et al., 2014; Young et al., 2013) and has been used to pretrain machine learning models in previous studies (Basaia et al., 2019; Cheng et al., 2015). In this project, we used a classification task of NC versus DAT as the source task for transfer learning to our target task model.

Furthermore, we used a novel occlusion map method (Zeiler and Fergus, 2014) to visualize the features significantly contributing to our model. Finally, we demonstrate the model's clinical relevance through the association of the model's prediction output to rate of cognitive decline.

2. Materials and methods

2.1. Subjects

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). 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, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

The source task used 1406 DAT and 2084 NC scans from 1080 subjects. At the source task only, scans from multiple time points are included, if available. In the target task, we examined MCI-C patients with a conversion time of up to 3 years (longer conversion times are examined later) to MCI-NC patients with a clinical diagnosis that remains MCI for a duration of at least 3 years. MCI subjects with a duration of MCI less than 3 years without conversion were excluded due to the potential possibility of near-future conversion. This resulted in 228 MCI-C patients and 222 MCI-NC patients.

Included in the target task is the single time point sMRI scan at which an individual first received a diagnosis of MCI. Group differences in demographic and clinical history information were evaluated with one-way analyses of variance and chi-square tests. Demographic information and clinical scores for the sample are shown in Table 1. The distributions of conversion time of MCI-C patients and duration of diagnosis for MCI-NC patients are shown in Fig. 1. Additional clinical information is provided in Supplemental Table 11.

2.2. Structural MRI data setup for transfer learning

1.5 T and 3T sMRI data were downloaded from ADNI. (Detailed MRI scanner protocols for T1-weighted sequences by vendor are available online: http://adni.loni.usc.edu/methods/documents/mri-protocols/). Preprocessing included skull stripping (Wang et al., 2011), re orientation, cropping, and padding. This resulted in images with $158 \times 196 \times 170$ voxels. The FMRIB Software Library (FSL; https://fsl.fmrib.ox.ac.uk) was then used to correct intensity in homogeneity by using an N3 algorithm (Sled et al., 1998) and to coregister the scans to the Montreal Neurological Institute 152 atlas by using affine linear alignment.

For the source task, DAT and NC scans were randomly selected and divided into train, validation, and test sets (Fig. 2). To provide diverse generic knowledge, 90% of the data (3143 scans) were assigned to the train set while the validation and test set each contained 5% of the data (172 and 175 scans, respectively). Analysis of variances were calculated to confirm that groups within the train, validation, and test set did not differ significantly in demographic and clinical characteristics: sex, race, ethnicity, marital status, age, years of education, clinical scores, and genetic information (Clinical scores and genetic information include CDR, ADAS11, ADAS13, MMSE, RAVLT Immediate, RAVLT Learning, RAVLT Forgetting, RAVLT Percent Forgetting, FAQ, APGN1, APGN2, APOE2, APOE3, and APOE4).

For the target task, MCI-C and MCI-NC scans were randomly split into training, validation, and test sets by following the conventional ratio of 70% versus 15% versus 15% (314, 68, and 68 scans, respectively). Analysis of variances were also calculated to confirm the no significant group difference between train, validation, and test set. To avoid data leakage (Wen et al., 2020), which exposes the information of the test set to the train and validation set, thereby falsely producing a higher test set classification accuracy, a single time point scan was used for each subject. The test portion of the target task was also ensured to be fully independent from the data used in both the source task and the training/validation portion of the target task. Therefore, no subjects in the target task test set overlapped with the rest of the samples. This step has been overlooked

Table 1
Demographic and clinical information within subjects for the source and target tasks

	Source task			Target task		
	NC	DAT	*p* value	MCI-NC	MCI-C	*p* value
N_{total}	2084	1406	–	222	228	–
Age	76.49 (±5.92)	76.18 (±7.22)	ns	72.25 (±7.32)	74.18 (±6.96)	$p < 0.05$
% Male	49.80%	60.10%	ns	63.10%	57.00%	ns
Education	16.35 (±2.74)	15.35 (±2.90)	$p < 0.05$	15.97 (±2.85)	15.87 (±2.78)	ns
CDRSB	0.09 (±0.30)	5.22 (±2.41)	$p < 0.05$	1.18 (±0.63)	1.97 (±0.98)	$p < 0.05$
ADAS11	5.56 (±2.85)	20.47 (±7.85)	$p < 0.05$	8.61 (±3.41)	13.17 (±5)	$p < 0.05$
ADAS13	8.7 (±1.32)	31.03 (±9.43)	$p < 0.05$	13.77 (±5.33)	21.27 (±6.04)	$p < 0.05$
MMSE	29.04 (±1.21)	22.31 (±3.68)	$p < 0.05$	28 (±1.69)	26.77 (±1.72)	$p < 0.05$

Results are reported as mean ± standard deviation. Age and education are reported in years.
Key: CDR, Clinical Dementia Rating Scale; ADAS11, Alzheimer's Disease Assessment Scale 11; ADAS13, Alzheimer's Disease Assessment Scale 13; MMSE, Mini Mental State Examination.

in previous research and is crucial for avoiding biased learning and increasing the generalizability of the model.

2.3. Architecture of convolutional neural network

A base model for transfer learning was developed by benchmarking Residual Network 50 (ResNet50) (He et al., 2016). ResNet50 is composed of 5 residual blocks. The first block contains one convolutional and pooling layer, and the following blocks consist of 3, 4, 6, and 3 bottleneck layers, respectively. Each bottleneck layer has 3 convolutional layers interconnected through a skip connection that can smooth the loss landscape and is beneficial in achieving global optima (Li et al., 2018; Orhan and Pitkow, 2017). Each convolutional layer receives representations from the previous layers and transforms them to the deeper level of feature maps. These feature maps then contribute to the model's classification decision.

ResNet50, however, has over 23 million trainable parameters, which is complex enough to cause high variance for the MCI-C versus MCI-NC classification task. Therefore, we tailored ResNet50 to this task by reducing the number and width of convolutional layers. The resulting model had narrower and shorter network architecture than ResNet50 and was named ResNet29 (Fig. 3). The number of filters of the convolutional layer in the first convolutional block was reduced from 64 to 32. The number of bottleneck layers in the following residual blocks was reduced from 3, 4, 6, and 3 to 2, 2, 2, and 2, respectively. Finally, one additional residual block which consists of one bottleneck layer was added at the end. The number of filters of each residual block was divided by 4. In the end, the model has about 4 million trainable parameters.

ResNet29 was developed as an end-to-end binary classification model. The model produces 2 prediction scores: the probability that the scan is classified as an MCI-C subject and the inverse probability that the scan is an MCI-NC subject. The sum of these 2 prediction scores is always one; if the prediction score for MCI-C is higher than the prediction score for MCI-NC, then the model decides the given brain scan is from an MCI-C patient. Similarly, if the prediction score for MCI-NC is higher than the prediction score for MCI-C, then the model predicts the given brain scan belongs to the MCI-NC patient group.

All codes were built in Python Keras as a TensorFlow backend. Experiments were conducted by using 4 NVIDIA P100 Pascal (12G HBM2 memory). The training time for the source and target task was 9 and 3 hours, respectively.

2.4. Hyperparameters

Hyperparameters are variables set before training which determine the network structure and how the network is trained. We evaluated multiple hyperparameters with the objective of improving classification accuracy. At the source task, the model was trained with a cyclically changing learning rate to avoid the model being stuck in local optima and to promote the model to reach to the global optima (Loshchilov and Hutter, 2016). The maximum learning rate and minimum learning rate were set as 1e-2 and 1e-4. The learning rate was cyclically changing through the entire epoch of 75 with a unit epoch of 25. To reduce overfitting, ridge regression and weight constraint with the value of 4e-4 and 2 were used throughout every convolutional layer, and a batch normalization layer was also used (Ioffe and Szegedy, 2015). To prevent gradient exploding, gradient clipping was set as 1 (Philipp et al., 2018).

The model and the weight matrix obtained from NC versus DAT classification task were transferred to the target task of classifying MCI-NC versus MCI-C. At the target task, the first 127 of 155 layers were frozen, which resulted in 2,767,106 trainable parameters. The

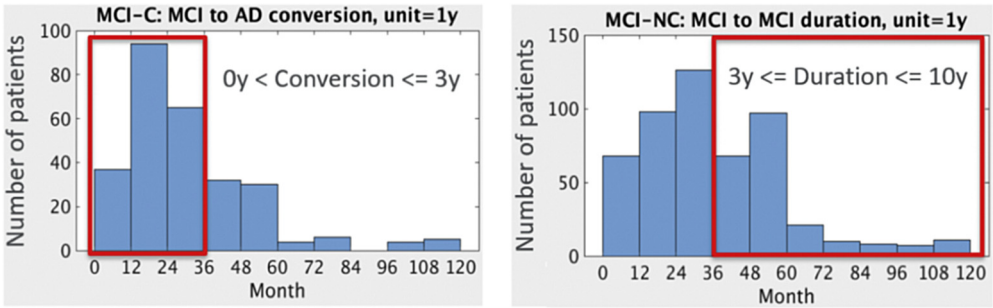


Fig. 1. Distribution of MCI-C (N = 277) and MCI-NC (N = 514) patients in accordance with years until conversion and duration of diagnosis, respectively: Upper histogram red box: MCI-C patients who converted to DAT within 3 y were selected for the target task (N = 228). As a comparison with this group, lower histogram red box: MCI-NC patients whose duration of MCI diagnosis is at least 3 y (N = 222) are included in this study. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

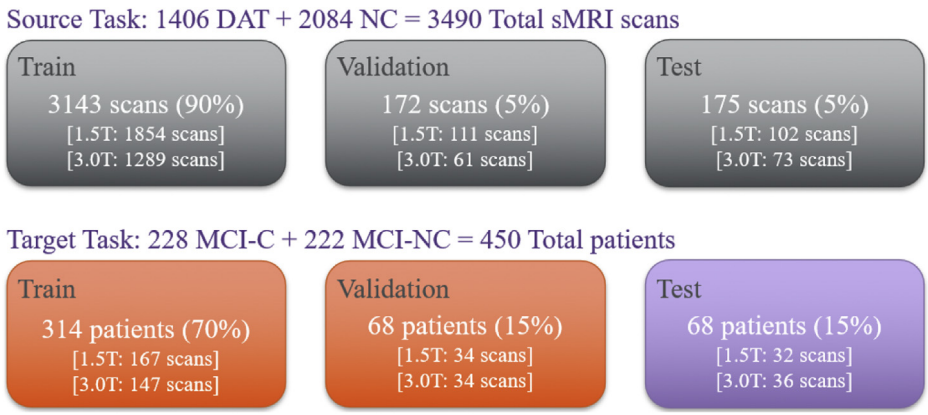


Fig. 2. Graphical layout of data division for the source task and target task. The number of scans and patients used for train, validation, and test set is shown. The percentage in parenthesis indicates the ratio of data size compared with the whole data set, that is, either source or target data set. The number of 1.5 T and 3.0 T scans are also shown in brackets.

model was retrained with a cyclically changing learning rate from 1e-3 to 1e-5 with a unit epoch of 25 through the entire epoch of 125. Ridge regression, weight constraint, and gradient clipping were set as 7e-4, 2, and 1, and a batch normalization layer was also used.

All convolutional layers were initialized with "*he_normal*" (He et al., 2015). In addition, the "*elu*" activation function, proposed by Clevert et al. (2015), was used with the intention of increasing training speed. Finally, the output layer used the "*softmax*" activation function, which produces the output probabilities between 0 and 1, with the sum of the probabilities been equal to 1 (Nwankpa et al., 2018). Categorical cross entropy was used as a loss function, and stochastic gradient descent was used as an optimizer.

2.5. Feature visualization method: occlusion map

Feature visualization was completed using an occlusion method implemented on all sMRI scans that had been included in the test set of the target task. After model training, each sMRI scan was fed into the model with a $2 \times 2 \times 2$ voxel patch (intensity 0) "occluded." The patch position was iterated through each voxel with a stride of 2 of the whole 3D brain. Prediction scores were extracted from the model of each iterated occluded brain, and the prediction score (for either class, MCI-C or MCI-NC) was recorded at the occluded brain region. This visualization creates a heatmap of brain regions that significantly alter the model prediction.

A degree of change in prediction score due to the occluded portion represents the importance of that region for the model's classification decision. The brain regions where the prediction score decreased when occluded versus unoccluded were colored as blue (blue occlusion map). In contrast, the brain regions where the prediction score increased when occluded were colored as red (red occlusion map). The blue regions contribute to a higher prediction score of the predicted class in the unoccluded image, and the red regions contribute to produce higher prediction score of the class

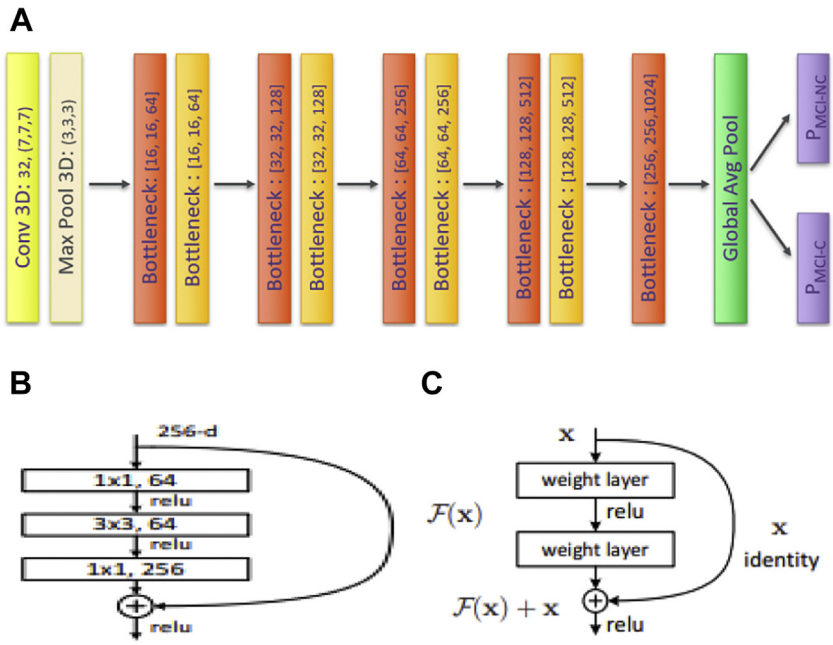


Fig. 3. (A) Architecture of convolutional neural network (CNN). The original ImageNet Model, that is, ResNet50 was scaled down by narrowing and shortening the model. (B) Bottleneck layers were set to reduce the model's complexity and thereby improve the classification performance (He et al., 2016). (C) Skip connection was used to enable the model to reach a global optima (He et al., 2016).

not predicted in the unoccluded image. To the best of our knowledge, this method has yet to be implemented for classifying MCI-NC versus MCI-C.

2.6. Relating mean intensity values of gray matter beneath the occlusion maps to neuropsychological and cerebrospinal fluid measures

To validate the model, we calculated the mean intensity values of gray matter beneath the occlusion map (MIGMBO). Atrophy in gray matter (both cortex and deep nuclei) is related to the accumulation of amyloid beta plaques and neurofilament tangles (Bejanin et al., 2017; Jack et al., 2019; Sepulcre et al., 2016). To identify the meaningful information that contributes to the conversion from MCI to DAT, MIGMBO for all patients in the test set of the target task were calculated and regressed with measures of clinical change, neuropsychological test performance, and cerebrospinal fluid (CSF) markers.

For clinical measures, we included the rate of change in the Clinical Dementia Rating-Sum of Boxes (CDRSB), Alzheimer's Disease Assessment Scale-cognitive 11 item (ADAS 11) and ADAS-cognitive 13 item (ADAS 13), Mini Mental State Examination (MMSE), Rey Auditory Verbal Learning Test (RAVLT)—RAVLT Immediate, RAVLT Learning, RAVLT Forgetting, RAVLT Percent Forgetting, and Functional Activities Questionnaire (FAQ) (Folstein et al., 1975; Mayo, 2012; Rosen et al., 1984; Samtani et al., 2014; Schmidt, 1996; Skinner et al., 2012). For CSF measures, we included Aβ, tau, phosphorylated tau (P-tau), Aβ/tau, and Aβ/P-tau (Detailed description of CSF acquisition protocols can be found on the ADNI website: http://adni.loni.usc.edu/data-samples/data-types/).

The output of our generated occlusion maps was divided into 3 bins based on strength of change in the model's prediction, and each bin was used as a predictor for change in variables listed previously in a multiple regression analysis. The “low” occlusion bin is those in which there was only a small change in the prediction score, that is, a change within one standard deviation. The “medium” occlusion bin is those in which the prediction score changed between one and 2 standard deviations. The “high” occlusion bin is those in which prediction score changed greater than 2 standard deviations.

$$Y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$Y = \begin{cases} \textit{Rate of } 9 \textit{ Neuropsychological measures} \\ 5 \textit{ CSF measures} \end{cases}$$

$$x_1 = m < |\text{occlusion mask}| \leq m + 1\sigma$$

$$x_2 = m + 1\sigma < |\text{occlusion mask}| \leq m + 2\sigma$$

$$x_3 = m + 2\sigma < |\text{occlusion mask}|$$

Where m indicates mean of whole occlusion map, and σ indicates standard deviation of whole occlusion map.

To determine the gray matter regions that contribute to the DAT progression, the blue occlusion map for MCI-NC—predicted patients and the red occlusion map for MCI-C—predicted patients were used. The blue occlusion map of MCI-NC patients identifies the brain regions that make the MCI-NC brain scan look more similar to the MCI-C brain scan. The red occlusion map of MCI-C patients shows the brain regions that make the MCI-C brain scan look more dissimilar to the MCI-NC brain scan. Therefore, the brain regions covered by these occlusion maps provide information about DAT progression.

The other occlusion map, that is, the red occlusion map for MCI-NC and the blue occlusion map for MCI-C, represents the brain regions that are indicative of MCI-NC that does not progress to DAT. Therefore, these regions were not used to examine brain regions that associate conversion to DAT to clinical and CSF measures.

2.7. Relating CNN's prediction score to neuropsychological measures

To evaluate the clinical validity of the 3D-CNN model, we examined the prediction score from the earliest sMRI scan of an MCI patient to the rate of cognitive decline. Of 514 MCI-NC and 277 MCI-C subjects throughout the whole conversion and duration years (Fig. 1), patients whose brain scan was used in training/validating the model were excluded. This resulted in a sample of 323 MCI-NC and 86 MCI-C patients. The longitudinal clinical scores (i.e., CDRSB, ADAS11, ADAS13, MMSE, RAVLT—RAVLT Immediate, RAVLT Learning, RAVLT Forgetting, RAVLT Percent Forgetting, and FAQ) from the first MCI-diagnosed time point to the end of clinical history were used to obtain the month-wise rate of change in clinical assessment scores. Pearson's correlation between the CNN prediction score from the baseline sMRI scan and the clinical scores' month-wise rate of change obtained through the first to the last clinical history were also examined.

3. Results

3.1. 3D-CNN classification results

Classifying MCI-C versus MCI-NC through transfer learning with a base model of ResNet29 was successful. The loss value of training and validation set decreased throughout the training epochs (Fig. 4A). This indicates that the model was well optimized with a set of well-defined hyperparameters. It produced a test set classification accuracy of 82.4% and 0.827 area under the curve as well as 0.189 equal error rate value (Fig. 4B).

The test set was composed of MCI-C patients whose conversion time was between 0 and 3 years. To further look at the models' prediction performance over a longer conversion time, a separate MCI-C data set with a conversion time longer than 3 years was used. In conversion times from 0 to 3 years, 3 to 6 years, and 6 to 10 years, there were 37, 39, and 9 MCI-C subjects, and the same model and its weight matrix were implemented to predict these patients. The model's sensitivity for these 3 groups was 81.08%, 71.79%, and 55.56%, respectively. The results showed that prediction score decreases with longer conversion times (Fig. 5).

Furthermore, we provide the model's separate accuracy on 1.5 T and 3.0 T sMRI scanner (Table 2). For 1.5 T scanner, 25 of 32 scans are correctly predicted and report 78.13% accuracy. For 3.0 T scanner, 31 of 36 scans are correctly predicted at 86.11% accuracy.

3.2. Feature visualization

Using occlusion mapping, we identified structural features recognized by the model. As seen in Figures 6 and 7, the occlusion of the hippocampus, parahippocampal gyrus, amygdala, and pons increased the probability score for MCI-C; the hippocampus, parahippocampal gyrus, amygdala, and pons were covered by the blue occlusion map for MCI-NC and the red occlusion map for MCI-NC. On the other hand, the occlusion of the nucleus accumbens, caudate nucleus, globus pallidus, thalamus, cerebellum, and midbrain increased the probability score for MCI-NC; these regions were covered by the red occlusion map for MCI-NC and the blue occlusion map for MCI-C. We note that the occlusion maps for MCI-NC and MCI-C are complementary.

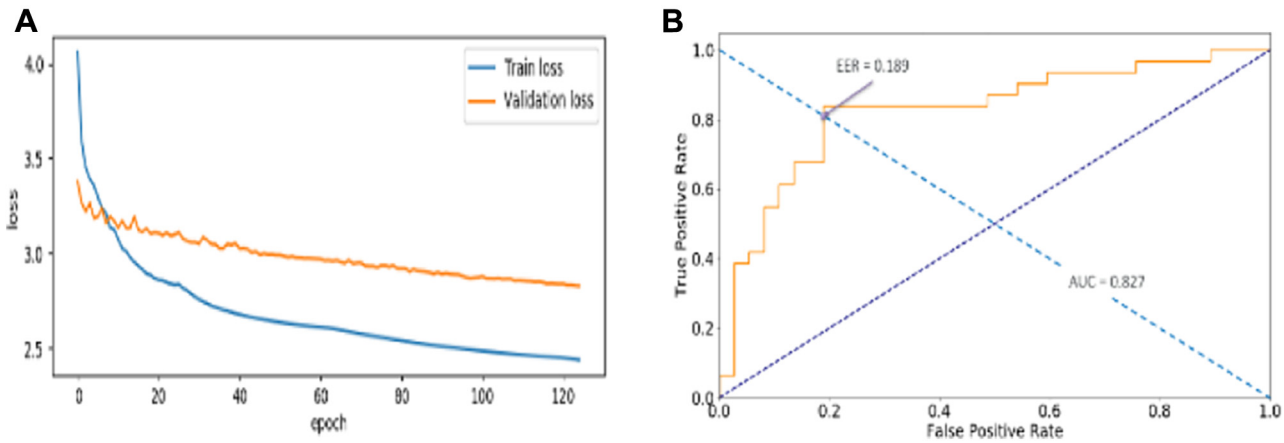


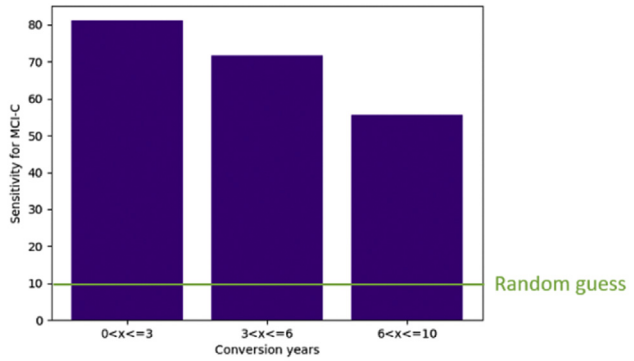
Fig. 4. Loss history of train and validation data (A) and classification performance (B), that is, area under the curve (AUC) and equal error rate (EER) on test data. Train and validation loss continuously decreased along the epochs, indicating that the model was learning. The weight matrix that was restored and used to evaluate the test classification accuracy was where the validation loss showed the minimum. Test classification accuracy reported 82.4%. AUC and EER values are 0.827 and 0.189, respectively.

3.3. Relating MIGMBO to rate of change in neuropsychological (clinical) measures

The MIGMBO score from the high occlusion bin predicted the rate of change in CDRSB, ADAS11, ADAS13, MMSE, and FAQ score at a significance level of 0.05 (Table 3). MIGMBO was negatively correlated with the rate of MMSE decline and positively correlated with the rate of increase in CDRSB, ADAS11, ADAS13, and FAQ score. In contrast, the MIGMBO score did not show a significant correlation with rate of change in RAVLT scores.

3.4. Relating MIGMBO to CSF measures

MIGMBO scores showed strong correlation with Aβ, Aβ/Tau, and Aβ/P-Tau with a *p*-value below 0.05, indicating statistical significance (Table 4). For relating Aβ/Tau and Aβ/P-Tau, all predictors showed a significant relationship. The low, medium, and high occlusion maps all played a crucial role in predicting the dependent variables. With Tau and P-Tau, MIGMBO showed *p*-values of 0.0683 and 0.0707, respectively, which approximate statistical significance.



Years	N participants	Sensitivity
0-3	37	81.08%
3-6	39	71.79%
6-10	9	55.56%

Fig. 5. The sensitivity to predict patients with conversion years from 0 to 10. The random guess is 10% as it is the chance of one of 10 different conversion years.

3.5. Relating CNN-based prediction score to rate of cognitive decline

The CNN-based prediction score at the first MCI-diagnosed time point showed significant correlation with the rate of change in CDRSB, FAQ, MMSE, and RAVLT forgetting (Table 5). The CNN prediction score was positively correlated with the rate of change in the CDRSB and FAQ scores and was negatively correlated with the rate of change in the MMSE and RAVLT forgetting scores. On the other hand, RAVLT immediate, RAVLT learning, ADAS11, and ADAD13 did not show a significant correlation with the 3D-CNN–based prediction scores.

The prediction score produced by the baseline sMRI scan also showed significant correlation with the rate of cognitive decline (Table 5). It was positively correlated with the rate of change in CDRSB and the FAQ score and negatively correlated with MMSE, RAVLT Forgetting, and RAVLT Percent Forgetting scores.

4. Discussion

Leveraging ADNI data, we aimed to predict MCI conversion to DAT using a CNN model trained on sMRI scans of healthy individuals and those with DAT. In so doing, we developed ResNet29, an end-to-end 3D-CNN which trained, through transfer learning of these sMRI scans of healthy versus DAT subjects, to predict MCI patients who either remained stable in their diagnosis or progressed to DAT. Our model achieved this with a 82.4% accuracy and also showed the most significant prediction increase from random guess, 31.7%.

ResNet 29 trained through a novel transfer learning meets the level of complexity that is required to interpret the heterogeneous nature of DAT development. Most biomarkers of DAT, including

Table 2
Prediction accuracy on 1.5 T and 3.0 T scans in the test set

Scanner	N_{total}	$N_{correctly\ predicted}$	Accuracy (%)
1.5 T	32	25	78.13
3.0 T	36	31	86.11

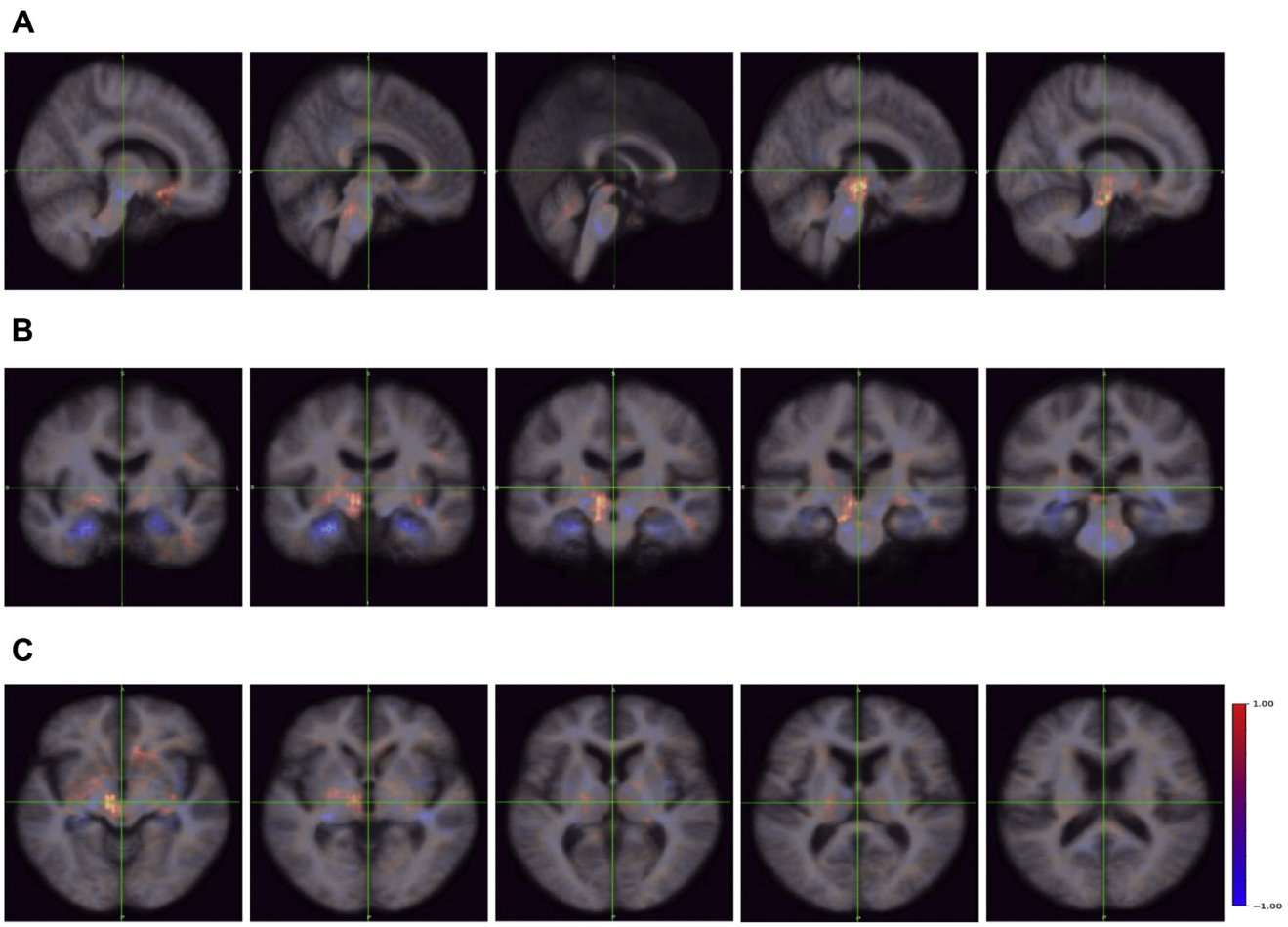


Fig. 6. Occlusion maps (A) sagittal plane, (B) coronal plane, and (C) transverse plane across all correctly predicted MCI-NC patients. The red color indicates a higher prediction score, whereas the blue color indicates a lower prediction score. The blue regions indicate importance in predicting MCI-NC and include the pons, amygdala, hippocampus, and parahippocampal gyrus. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

atrophy on sMRI scans, are known to nonlinearly worsen with increased disease severity (Jack Jr et al., 2010). We recognize this is a limitation of machine learning models which are trained on the final stage of disease outcomes. However, we note that our base model, ResNet29, is constructed with a series of convolutional layers, so that it may extract complex patterns through a series of nonlinear transformations. The model learns generic knowledge through NC and DAT scans during the source task, so that it is finely optimized to determine the degree (expressed as a probability of belonging to the MCI-C) to which a subject's baseline scan resembles a DAT scan. We additionally note that our model is designed to individually classify ultimate disease outcome, rather than reflect the nuances of disease progression.

Compared with previous studies (Table 6), our model achieved the highest accuracy in classifying MCI-NC from MCI-C. Li et al. (2014) used a random forest method with weak hierarchical lasso feature selection to achieve 74.8% classification accuracy using 161 MCI-NC and 132 MCI-C sMRI scans. Cheng et al. (2015) produced 79.4% classification accuracy by using domain transfer feature selection and domain transfer sample selection for extracting features and support vector machine model for classifying 43 MCI-NC and 56 MCI-C patients. Similarly, Suk et al. (2017) had 74.8% classification accuracy in classifying 226 MCI-NC and 167 MCI-C patients by using a 2D-CNN based on 93 regions of interest as features (93 ROI for each sMRI and PET and 3 features from CSF are used). By using 3D-CNN, Basaia et al. (2019) showed 74.9% classification accuracy in classifying 533 MCI-NC and 280 MCI-C patients based on gray matter tissue probability maps, and Yee et al. (2020) recorded 74.7% accuracy in classifying 871 MCI-NC and 362 MCI-C scans.

While most previous research did not use an independent test set, Basaia et al. (2019) assigned a relatively small portion (10%) of the whole data set as a test set to verify the model's generalized performance. The most effective splitting ratio of the training, validation, and test sets is still under discussion, although we set the ratio as 70:15:15 which is traditionally accepted and successfully demonstrated the generalizability of our model.

The ability to predict DAT conversion based on a single time point MRI is advantageous for the clinical field. While some previous studies include multimodal biomarkers in their prediction models, such as positron emission tomography and CSF biomarkers of disease (Cheng et al., 2015), our model outperformed these models with high accuracy by using a single time point sMRI scan. sMRI is often included in routine assessment of those at risk for AD. It is less expensive than other imaging scans and is minimally invasive, therefore reducing patient risk. Our model showed 8% higher accuracy in prediction with 3T than 1.5 T sMRI scans. This finding is consistent with the known accuracy advantages of higher resolution images for CNN (Sabottke and Spieler, 2020). Nevertheless, the accuracy of 78% with 1.5 T scanners, which are more common in hospital settings, is still high, indicating that our model is clinically implementable. In comparison with studies using combined modalities, our model produces more accurate

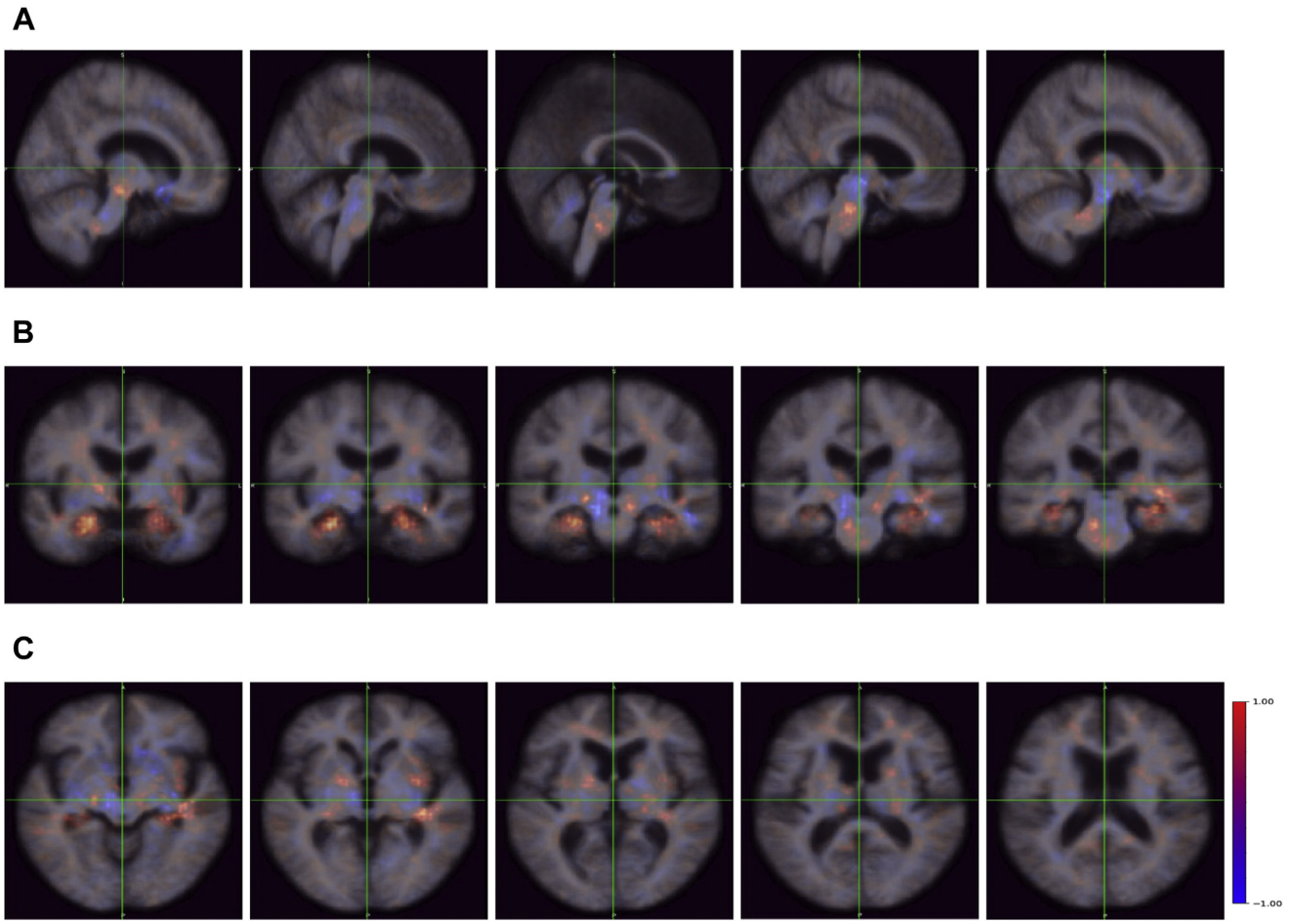


Fig. 7. Occlusion maps (A) sagittal plane, (B) coronal plane, and (C) transverse plane across all correctly predicted MCI-C patients. The red color indicates a higher prediction score, whereas the blue color indicates a lower prediction score. The blue regions indicate importance in predicting MCI-C and include the midbrain, nucleus accumbens, caudate nucleus, cerebellum, globus pallidus, and thalamus. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

predictions on DAT progression with less economic burden and infection risk to the patients.

Several factors contributed to the improvement of the classification performance by our model. It was largely empowered by the architecture of our deep learning model, specifically tailored for the MCI-C versus MCI-NC classification task. These factors include our novel transfer learning pipeline using healthy versus AD subjects that can produce diverse generic knowledge while avoiding data leakage, along with various engineering techniques such as a cyclically changing learning rate (Loshchilov and Hutter, 2016). Finally, we carefully tuned a set of hyperparameters including the type of activation layers, number of convolutional layers, and their size and learning rates, through numerous experimental condition until the model achieved the highest accuracy reported here. Further, these identified sets of hyperparameters that produced the best prediction results were validated using the test set which also

Table 3
Correlation between MIGMBO and rate of cognitive decline: red MCI-C and blue MCI-NC

Clinical measures	Constant	*Low* x_1	*Medium* x_2	*High* x_3	N	R^2	F-statistic	*p*-value
CDRSB	1.181 (0.771)	0.046 (0.677)	−0.073 (0.507)	[c]0.007 (0.000)	68	0.275	8.07	0.0001
ADAS11	−2.795 (−0.627)	0.303 (0.943)	−0.282 (−0.882)	[c]0.014 (3.251)	68	0.156	3.95	0.0120
ADAS13	−4.007 (−0.770)	0.405 (1.080)	−0.372 (−0.994)	[c]0.018 (3.529)	68	0.182	4.74	0.0048
MMSE	3.735 (1.475)	0.013 (0.073)	−0.048 (−0.264)	[c]−0.008 (−3.402)	68	0.204	5.47	0.0021
RAVLT immediate	4.529 (0.783)	0.008 (0.016)	−0.060 (−0.144)	−0.008 (−1.476)	68	0.053	1.19	0.3215
RAVLT learning	−3.392 (−1.938)	0.137 (0.282)	−0.088 (0.486)	−0.002 (0.233)	68	0.084	1.95	0.1306
RAVLT forgetting	−0.365 (−0.164)	−0.176 (−1.092)	0.183 (1.144)	−0.003 (−1.398)	68	0.039	0.86	0.4645
RAVLT percent forgetting	16.104 (0.585)	[b]−4.596 (0.024)	[b]4.382 (0.031)	−0.019 (0.493)	68	0.079	1.84	0.1493
FAQ	5.463 (1.333)	−0.046 (−0.156)	−0.057 (−0.194)	[c]0.018 (4.551)	68	0.260	7.51	0.0002

Figures in parentheses are t statistics.
MIGMBO score could predict the rate of change in CDRSB, ADAS11, ADAS13, MMSE, and FAQ score. MIGMBO score from the most significant occlusion map is important in predicting rate of cognitive decline.
[a] $p < 0.10$.
[b] $p < 0.05$.
[c] $p < 0.01$.

Table 4
Correlation between MIGMBO and CSF measures: red MCI-C and blue MCI-NC

CSF measures	Constant	*Low* x_1	*Medium* x_2	*High* x_3	N	R^2	F-statistic	*p*-value
Aβ	[c]4029.0 (3.0)	[c]480.7 (3.4)	[c]−507.3 (−3.5)	[b]−2.3 (−2.2)	26	0.494	7.15	0.00159
Tau	−791.4 (−1.3)	−76.5 (−1.2)	87.3 (1.3)	0.8 (1.6)	29	0.244	2.69	0.0683
P-tau	−93.1 (−1.3)	−10.1 (−1.3)	11.3 (1.5)	0.1 (1.6)	29	0.241	2.65	0.0707
Aβ/Tau	[c]30.8 (4.1)	[c]3.3 (4.2)	[c]−3.6 (−4.4)	[c]−0.0 (−2.9)	26	0.634	12.7	4.98e-05
Aβ/P-tau	[c]388.0 (4.4)	[c]42.1 (4.5)	[c]−45.3 (−4.7)	[c]−0.2 (−3.0)	26	0.665	14.6	1.92e-05

Figures in parentheses are t statistics.
MIGMBO score could predict accumulation of Aβ, and ratio of Aβ/Tau and Aβ/P-Tau. All 3 predictors, that is, MIGMBO score from the least, medium, and the most significant occlusion map, contribute to the prediction.
[a] $p < 0.10$.
[b] $p < 0.05$.
[c] $p < 0.01$.

showed the highest accuracy from the numerous experiments: 82.4%.

In addition, our model was provided with a whole 3D brain scan without specification of any particular feature for training. Previous studies have limited the input resource through feature engineering. For example, studies that selected gray matter as the feature for model training (Basaia et al., 2019) did not consider CSF space or white matter changes known to also play an important role in DAT and the pathologic process of AD (Jack et al., 2010; Li et al., 2014; Weiler et al., 2015). In addition, Cheng et al. (2015) selected subjectively defined "useful" features using domain transfer feature selection and domain transfer sample selection. Machines trained with these samples could be biased and thus may not be generalizable to independent populations. Therefore, unlike feature engineering which limits the information able to be learned based on a researcher's preassumption of what may be important in classification, the presented model learned from every possible feature available in the image.

It should be noted that the conversion times in previous studies range from 1.5 (Suk et al., 2017) to 4 years (Li et al., 2014), while one of the latest experiments uses a 3-year conversion time window (Basaia et al., 2019). We chose this conversion time for the present study to directly compare performance with the most current research. Further, setting the conversion time at 3 years provided a well-balanced data set between MCI-NC (N = 222) and MCI-C (N = 228) (Buda et al., 2018). It allowed unbiased learning by the model on MCI-C and MCI-NC patients' brains.

To the best of our knowledge, a deep learning model that can identify anatomical brain regions critical for predicting the conversion from MCI to DAT has not been demonstrated previously. Feature visualization methods are able to highlight regions in an input image with strong influence on the classification decision. It is important as it enables us to understand and validate the reasoning that has driven the model's classification. Especially in the study of neurodegenerative disease, it is critical to explain the behavior of a machine/deep learning model to elucidate the neuroimaging biomarkers that contribute to conversion from MCI to DAT. State-of-the-art visualization techniques include Gradient Class Activation Map and Guided Gradient Class Activation Map (Selvaraju et al., 2017; Yee et al., 2020). However, in the medical field, these methods are unable visualize the features that contribute to disease-negative samples (Ardila et al., 2019). Further, feature maps that directly contribute to the classification decision often have too low resolution to show fine structural features within the brain.

An occlusion method to feature visualization avoids these problems and produces finer feature maps (Zeiler and Fergus, 2014). We implemented an occlusion method and identified key brain structures that contribute to DAT conversion. The occlusion method is critical in this research as the occlusion patch could represent structural alteration. A major strength of the presented model is that the input is naïve to specified brain regions. As the model uses whole 3D sMRI scan as an input without limiting itself to predefined regions of interest or features that are obtained from feature engineering, the occlusion map, too, examined the level of contribution at the voxel level ($2 \times 2 \times 2$) in the progression of disease. The results that the occlusion map shows are completely driven by statistical calculations from the ResNet29.

The blue occlusion map presented brain regions that decreased the probability of being MCI-NC or MCI-C patients when such structural alteration occurs. For example, the hippocampus was covered by the blue occlusion map for MCI-NC–predicted patients. Thus, when information from the hippocampus was missing (occluded), the model recognized MCI-NC–predicted scans as more similar to an MCI-C scan; the probability for MCI-NC was decreased while the probability for MCI-C was increased. Therefore, morphologic alteration in the hippocampus contributes to the DAT classification. This aligns with our understanding of the importance of the hippocampus during MCI stages and progression into DAT (Ferrarini et al., 2009; Gupta et al., 2019; Lee et al., 2020; Li et al., 2007). In contrast, the thalamus was covered by the blue occlusion map for MCI-C–predicted patients—meaning that when information from the thalamus was missing in the model, and MCI-C–predicted patients looked more similar to MCI-NC patients. Therefore, as far as we can measure, structural change in the thalamus does not promote the DAT development within a 3-year window.

Table 5
Correlation between CNN prediction score and clinical assessment scores' rate of change

	CDRSB	ADAS11	ADAS13	MMSE	RAVLT immediate	RAVLT learning	RAVLT forgetting	RAVLT percent forgetting	FAQ
First MCI sMRI (N = 409)	[c]0.264	0.062	0.039	[c]−0.146	0.000	[a]0.089	[b]−0.117	−0.078	[c]0.243
Baseline sMRI (N = 409)	[c]0.242	0.008	0.005	[c]−0.177	−0.011	[a]0.094	[b]−0.106	[c]−0.143	[c]0.144

Correlation between CNN prediction score from first MCI-diagnosed sMRI scan and clinical assessment scores' rate of change (the first row). CNN prediction score is positively correlated with the rate of change in CDRSB and FAQ score and negatively correlated with rate of change in MMSE and RAVLT Forgetting score. Correlation between CNN-based score from the baseline sMRI scan and clinical assessment scores' rate of change. CNN prediction score is positively correlated with the rate of change in CDRSB and FAQ score and negatively correlated with the rate of change in MMSE, RAVLT Forgetting, and RAVLT Percent Forgetting (the second row).
[a] $p < 0.10$.
[b] $p < 0.05$.
[c] $p < 0.01$.

Table 6
Summary of MCI-C versus MCI-NC classification research

	Biomarker	Conversion time (y)	Random guess (%)	Accuracy (%)	Increase (%)
Proposed model	*sMRI*	*3*	*50.7*	*82.4*	*31.7*
Yee et al. (2020)	FDG	3	70.6	74.7	4.1
Basaia et al. (2019)	sMRI	3	65.6	74.9	9.3
Suk et al. (2017)	sMRI, Clinical Score	1.5	57.5	74.8	17.3
Cheng et al. (2015)	sMRI, PET, CSF	2	56.6	79.4	22.8
Li et al. (2014)	MRI, Meta features[a]	4	54.9	74.8	19.9

[a] MRI features indicate average cortical thickness, standard deviation in cortical thickness, volumes of cortical parcellations, volumes of specific white matter parcellations, and the total surface area of the cortex, and meta features include demographic, genetic information, baseline cognitive scores, and laboratory tests. 305 MRI features and 52 meta features are used.

Many subcortical white matter and deep gray structures were detected as features. As both MCI-C and MCI-NC patients have MCI, they do not yet manifest significant cortical atrophy on sMRI. These patients experience cognitive decline related to atrophy in these regions. (For DAT patients, cortical regions are recognized as a feature (Supplementary Fig. S14)).

For the quantitative voxel analysis, we segmented subcortical regions by using FMRIBs Integrated Registration and Segmentation Tool and count the number of voxels of the blue occlusion map, the red occlusion map, and whole brain structure. The sizes of 2 groups' subcortical structures presented similarly; they did not show a difference between their size at the significance level of 0.1. Surprisingly, however, the blue occlusion map was dominant in the hippocampus, amygdala, and pons for MCI-NC patients, while was dominant in the nucleus accumbens, caudate nucleus, globus pallidus, and thalamus for MCI-C patients (Table 7). Therefore, within a 3-year time window until DAT diagnosis, structural changes in the hippocampus, amygdala, and pons promote DAT development, rather than the nucleus accumbens, caudate nucleus, globus pallidus, putamen, and thalamus. We note that the occlusion patch color (black) used in the occlusion map did not alter the visualization results, as we found identical results using a white colored patch.

Brain structures recognized by our deep learning model were consistent with previous research. Previous research has been published that structural alteration of subcortical brain structures reflects DAT progression. Many research studies indicate that morphological changes in the hippocampus and amygdala (Ball et al., 1985; Convit et al., 1993; Gupta et al., 2019; Lee et al., 2020; Lehericy et al., 1994; Li et al., 2007; Poulin et al., 2011; Zanchi et al., 2017) are significant. In addition, the pons was recognized as a significant biomarker in predicting AD progression. Olivieri et al. (2019) suggested that structural alteration occurs in the pons before AD develops (Olivieri et al., 2019).

To further provide content validity on regions identified by the occlusion map, we used the mean intensity value of gray matter beneath the 3 occlusion maps. Gray matter is known to be associated with the biomarkers of AD pathology (Bejanin et al., 2017; Jack et al., 2019; Sepulcre et al., 2016). Therefore, by showing the relationship between MIGMBOs and rate of cognitive decline, verified that our occlusion maps captured clinically meaningful brain regions. The MIGMBO score of the high occlusion map showed positive correlation with the rate of change in CDRSB, ADAS11, ADAS13, and FAQ and negative correlation with the rate of change in MMSE. Therefore, the higher the mean intensity of gray matter is, CDRSB, ADAS11, ADAS13, and FAQ scores increase and MMSE scores decrease more quickly.

In addition, by showing a significant correlation with MIGMBO and accumulation of Aβ, Aβ/Tau and, Aβ/P-Tau, we confirmed that the model's prediction aligns with neuropathologic markers of AD by solely utilizing information from a sMRI scan. All 3 occlusion maps, which were split based on their degree of prediction change, could be a useful resource in predicting Aβ, Aβ/tau, and Aβ/P-tau.

Finally, we showed that the prediction scores from our model were related to worsening of neuropsychological performance measures over time. Because all scans received a prediction score of being classified as MCI-C, we calculated the Pearson's correlation between this score and the rate of cognitive decline for all patients with MCI. The rate of change in CDRSB and FAQ was positively correlated with an MCI-C classification. This indicates that as the confidence in a scan being classified as MCI-C increases, the faster the increase in CDRSB and FAQ score. In addition, the rate of change in MMSE, RAVLT Forgetting, and RAVLT Percent Forgetting was negatively correlated with the predicted MCI-C score. This indicates that these scores decrease more quickly as the confidence that the scan should be classified as MCI-C increases.

Aside from the sMRI scan from the first MCI-diagnosed time point, we used the baseline sMRI scan of all MCI-C and MCI-NC patients in showing these correlations. Therefore, regardless of conversion time and duration time of MCI-C and MCI-NC patients, the model could predict the future cognitive decline of an individual patient by solely utilizing a sMRI scan from the first visit to the clinic. Considering that we do not know which patients will suffer from

Table 7
Mean number of voxels beneath the occlusion map

	$N_{Blue\ occlusion\ map}$		$N_{Red\ occlusion\ map}$		N_{Total}	
	MCI-NC	MCI-C	MCI-NC	MCI-C	MCI-NC	MCI-C
Brain stem	***13,927 ± 4200	17,025 ± 2814	17,051 ± 3728	16,218 ± 2332	30,979 ± 7256	33,247 ± 3501
Accumbens	***305 ± 279	526 ± 276	***655 ± 303	357 ± 255	960 ± 324	883 ± 266
Amygdala	***2608 ± 1422	1303 ± 1096	***955 ± 991	2261 ± 1228	3562 ± 1053	3564 ± 830
Caudate	***4567 ± 1157	5408 ± 1330	5056 ± 1632	4721 ± 1047	9623 ± 2221	10,129 ± 1637
Hippocampus	***7218 ± 3455	3505 ± 2633	***3122 ± 2373	6128 ± 2768	10,340 ± 2882	9632 ± 1686
Pallidus	***1655 ± 749	2174 ± 697	***2363 ± 695	1899 ± 622	4017 ± 1121	4073 ± 606
Putamen	5293 ± 1717	5680 ± 1336	5164 ± 1837	4918 ± 1312	10,457 ± 3086	10,598 ± 1890
Thalamus	***6878 ± 1901	8210 ± 1836	9585 ± 2203	8577 ± 2066	16,463 ± 3361	16,787 ± 1884

The number shows mean ± standard deviation.
The symbol * shows p-value from t-statistics, which indicate the difference between the number of voxels in the occlusion map for MCI-C and MCI-NC patients; *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$. For MCI-NC patients, the blue occlusion map is dominant in the amygdala and hippocampus. For MCI-C patients, the blue occlusion map is dominant in the accumbens, caudate, pallidus, and thalamus. There is no significant difference in the whole volume of brain structure between MCI-NC and MCI-C patients.

cognitive deficit in clinical practice, these results provide evidence that the model could be used to foretell future cognitive decline.

In future research, we plan to include all subcortical brain structures, including substructures of the brain stem and cerebellum, as well as parcellated cortical regions to predict the DAT progression throughout different conversion years of MCI-C patients and duration years of MCI-NC patients. This regression model could show the contribution of each brain region in promoting conversion to DAT and improve personalized preventive medicine.

In conclusion, the current clinical evaluation protocols cannot accurately predict which patients with MCI will progress to DAT (Ward et al., 2013). An automated classification system for MCI-NC versus MCI-C, such as the method presented in this study, offers promise for informing the clinical prognosis of these patients. Furthermore, the methods presented here will be useful for identifying which patients would benefit most from participating in clinical trials by providing individualized information on the disease progression, that is, brain regions that cause cognitive deficit and future cognitive decline. Our methods not only produced the highest performance in the field, but also avoided problems previously neglected such as data shortage, high variance, and data leakage. Our research showed high accuracy in predicting conversion as well as novel visualization features, both critical to advancing our understanding of conversion from MCI to DAT and personalized preventive medicine.

Disclosure statement

This manuscript has nothing to disclose for actual or potential conflict and interest.

CRediT authorship contribution statement

Jinhyeong Bae: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Jane Stocks:** Formal analysis, Writing - review & editing. **Ashley Heywood:** Software, Validation, Formal analysis, Writing - review & editing. **Youngmoon Jung:** Conceptualization, Writing - review & editing. **Lisanne Jenkins:** Writing - review & editing. **Virginia Hill:** Validation, Writing - review & editing. **Aggelos Katsaggelos:** Conceptualization. **Karteek Popuri:** Resources, Data curation. **Howie Rosen:** Funding acquisition. **M. Faisal Beg:** Resources, Data curation, Funding acquisition. **Lei Wang:** Conceptualization, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

This research was funded by grant AG055121 and AG045333 from the National Institute on Aging and by grants from Brain Canada, CIHR, NSERC, and Compute Canada.

ADNI 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.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.neurobiolaging.2020.12.005.

References

Alzheimer's Association, 2019. 2019 Alzheimer's disease facts and figures. Alzheimers Dement. 15, 321–387.

Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat. Med. 25, 954–961.

Ball, M., Hachinski, V., Fox, A., Kirshen, A., Fisman, M., Blume, W., Kral, V., Fox, H., Merskey, H., 1985. A new definition of Alzheimer's disease: a hippocampal dementia. Lancet 325, 14–16.

Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., Filippi, M., Initiative AsDN, 2019. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. Neuroimage Clin. 21, 101645.

Bejanin, A., Schonhaut, D.R., La Joie, R., Kramer, J.H., Baker, S.L., Sosa, N., Ayakta, N., Cantwell, A., Janabi, M., Lauriola, M., 2017. Tau pathology and neurodegeneration contribute to cognitive impairment in Alzheimer's disease. Brain 140, 3286–3300.

Borji, A., Cheng, M.-M., Hou, Q., Jiang, H., Li, J., 2019. Salient Object Detection: A Survey. Comput Vis Media arXiv, pp. 1–34.

Buda, M., Maki, A., Mazurowski, M.A., 2018. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. 106, 249–259.

Cheng, B., Liu, M., Zhang, D., Munsell, B.C., Shen, D., 2015. Domain transfer learning for MCI conversion prediction. IEEE Trans. Biomed. Eng. 62, 1805–1817.

Clevert, D.A., Unterthiner, T., Hochreiter, S., 2015. Fast and Accurate Deep Network Learning by Exponential Linear Units (Elus) arXiv preprint arXiv:151107289.

Convit, A., De Leon, M., Golomb, J., George, A., Tarshish, C., Bobinski, M., Tsui, W., De Santi, S., Wegiel, J., Wisniewski, H., 1993. Hippocampal atrophy in early Alzheimer's disease: anatomic specificity and validation. Psychiatr. Q. 64, 371–387.

Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Pruessner, J.C., Allard, M., Collins, D.L., Initiative AsDN, 2012. Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. Neuroimage Clin. 1, 141–152.

Da, X., Toledo, J.B., Zee, J., Wolk, D.A., Xie, S.X., Ou, Y., Shacklett, A., Parmpi, P., Shaw, L., Trojanowski, J.Q., 2014. Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. Neuroimage Clin. 4, 164–173.

Ferrarini, L., Frisoni, G.B., Pievani, M., Reiber, J.H., Ganzola, R., Milles, J., 2009. Morphological hippocampal markers for automated detection of Alzheimer's disease and mild cognitive impairment converters in magnetic resonance images. J. Alzheimers Dis. 17, 643–659.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.

Gupta, Y., Lee, K.H., Choi, K.Y., Lee, J.J., Kim, B.C., Kwon, G.R., Dementia, N.R.C.f., Initiative AsDN, 2019. Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images. PLoS One 14, e0222446.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, pp. 1026–1034. https://doi.org/10.1109/ICCV.2015.123.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Heun, R., Mazanek, M., Atzor, K.-R., Tintera, J., Gawehn, J., Burkart, M., Gänsicke, M., Falkaic, P., Stoeter, P., 1997. Amygdala-hippocampal atrophy and memory performance in dementia of Alzheimer type. Dement. Geriatr. Cogn. Disord. 8, 329–336.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift arXiv preprint arXiv:150203167.

Jack Jr., 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. Lancet Neurol. 9, 119–128.
Jack, C.R., Wiste, H.J., Therneau, T.M., Weigand, S.D., Knopman, D.S., Mielke, M.M., Lowe, V.J., Vemuri, P., Machulda, M.M., Schwarz, C.G., 2019. Associations of amyloid, tau, and neurodegeneration biomarker profiles with rates of memory decline among individuals without dementia. JAMA 321, 2316–2325.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, Berlin, Germany.
Lee, S., Lee, H., Kim, K.W., 2020. Magnetic resonance imaging texture predicts progression to dementia due to Alzheimer disease earlier than hippocampal volume. J. Neurosci. 45, 7.
Lehericy, S., Baulac, M., Chiras, J., Pierot, L., Martin, N., Pillon, B., Deweer, B., Dubois, B., Marsault, C., 1994. Amygdalohippocampal MR volume measurements in the early stages of Alzheimer disease. Am. J. Neuroradiol. 15, 929–937.
Li, S., Shi, F., Pu, F., Li, X., Jiang, T., Xie, S., Wang, Y., 2007. Hippocampal shape analysis of Alzheimer disease based on machine learning methods. Am. J. Neuroradiol 28, 1339–1345.
Li, H., Liu, Y., Gong, P., Zhang, C., Ye, J., Initiative, A.D.N., 2014. Hierarchical interactions model for predicting mild cognitive impairment (MCI) to Alzheimer's disease (AD) conversion. PLoS One 9, e82450.
Li, H., Xu, Z., Taylor, G., Studer, C., Goldstein, T., 2018. Visualizing the loss landscape of neural nets. Adv. Neural Inf. Process. Syst. 6389–6399.
Loshchilov, I., Hutter, F., 2016. Sgdr: Stochastic Gradient Descent with Warm Restarts arXiv preprint arXiv:160803983.
Mayo, A.M., 2012. Use of the functional Activities Questionnaire in older adults with dementia. In: Try This: Best Practices in Nursing Care to Older Adults with Dementia, p. 13.
Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S., 2018. Activation Functions: Comparison of Trends in Practice and Research for Deep Learning arXiv preprint arXiv:181103378.
Olivieri, P., Lagarde, J., Lehericy, S., Valabrègue, R., Michel, A., Macé, P., Caillé, F., Gervais, P., Bottlaender, M., Sarazin, M., 2019. Early alteration of the locus coeruleus in phenotypic variants of Alzheimer's disease. Ann. Clin. Trans. Neurol. 6, 1345–1351.
Orhan, A.E., Pitkow, X., 2017. Skip Connections Eliminate Singularities arXiv preprint arXiv:170109175.
Petersen, R., 2000. Mild cognitive impairment: transition between aging and Alzheimer's disease. Neurologia 15, 93–101.
Philipp, G., Song, D., Carbonell, J.G., 2018. Gradients Explode-Deep Networks Are Shallow-Resnet Explained.
Poulin, S.P., Dautoff, R., Morris, J.C., Barrett, L.F., Dickerson, B.C., Initiative AsDN, 2011. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. Psychiatry Res. Neuroimaging 194, 7–13.
Roberson, E.D., Mucke, L., 2006. 100 years and counting: prospects for defeating Alzheimer's disease. Science 314, 781–784.
Rosen, W.G., Mohs, R.C., Davis, K.L., 1984. A new rating scale for Alzheimer's disease. Am. J. Psychiatry 141, 1356–1364.
Russell, S.J., Norvig, P., 2016. Artificial Intelligence: A Modern Approach. Pearson Education Limited, Malaysia.
Sabottke, C.F., Spieler, B.M., 2020. The effect of image resolution on deep learning in radiography. Radiol. Artif. Intelligence 2, e190015.
Samtani, M.N., Raghavan, N., Novak, G., Nandy, P., Narayan, V.A., 2014. Disease progression model for clinical dementia rating–sum of boxes in mild cognitive impairment and Alzheimer's subjects from the Alzheimer's disease Neuroimaging initiative. Neuropsychiatr. Dis. Treat 10, 929.
Schmidt, M., 1996. Rey Auditory Verbal Learning Test: A Handbook. Western Psychological Services, Los Angeles, CA.
Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.
Sepulcre, J., Schultz, A.P., Sabuncu, M., Gomez-Isla, T., Chhatwal, J., Becker, A., Sperling, R., Johnson, K.A., 2016. In vivo tau, amyloid, and gray matter profiles in the aging brain. J. Neurosci. 36, 7364–7374.
Skinner, J., Carvalho, J.O., Potter, G.G., Thames, A., Zelinski, E., Crane, P.K., Gibbons, L.E., Initiative AsDN, 2012. The Alzheimer's disease assessment scale-cognitive-plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI. Brain Imaging Behav. 6, 489–501.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.
Suk, H.I., Lee, S.W., Shen, D., Initiative AsDN, 2017. Deep ensemble learning of sparse regression models for brain disease diagnosis. Med. Image Anal. 37, 101–113.
Torrey, L., Shavlik, J., 2010. Transfer learning, Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI Glob. 242–264.
Wang, Y., Nie, J., Yap, P.-T., Shi, F., Guo, L., Shen, D., 2011. Robust deformable-surface-based skull-stripping for large-scale studies. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin, Germany, pp. 635–642.
Ward, A., Tardiff, S., Dye, C., Arrighi, H.M., 2013. Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: a systematic review of the literature. Dement. Geriatr. Cogn. Dis. Extra 3, 320–332.
Weiler, M., Agosta, F., Canu, E., Copetti, M., Magnani, G., Marcone, A., Pagani, E., Balthazar, M.L.F., Comi, G., Falini, A., 2015. Following the spreading of brain structural changes in alzheimer's disease: a longitudinal, multimodal MRI study. J. Alzheimers Dis. 47, 995–1007.
Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O., 2020. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. Med. Image Anal. 101694.
Yee, E., Popuri, K., Beg, M.F., Initiative AsDN, 2020. Quantifying brain metabolism from FDG-PET images into a probability of Alzheimer's dementia score. Hum. Brain Mapp. 41, 5–16.
Young, J., Modat, M., Cardoso, M.J., Mendelson, A., Cash, D., Ourselin, S., Initiative AsDN, 2013. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. Neuroimage Clin. 2, 735–745.
Zanchi, D., Giannakopoulos, P., Borgwardt, S., Rodriguez, C., Haller, S., 2017. Hippocampal and amygdala gray matter loss in elderly controls with subtle cognitive decline. Front Aging Neurosci. 9, 50.
Zeiler, M.D., Fergus, R., 2014. Visualizing and Understanding Convolutional Networks, European Conference on Computer Vision. Springer, Berlin, Germany, p. 818.
Zheng, A., Casari, A., 2018. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. O'Reilly Media, Inc, Newton, MA.