UC San Diego
UC San Diego Previously Published Works

Title
Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images

Permalink
https://escholarship.org/uc/item/552232zx

Journal
Scientific Reports, 8(1)

ISSN
2045-2322

Authors
Lu, Donghuan
Popuri, Karteek
Ding, Gavin Weiguang
et al.

Publication Date
2018

DOI
10.1038/s41598-018-22871-z

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

SCIENTIFIC REPORTS

OPEN

Received: 10 October 2017
Accepted: 2 March 2018
Published online: 09 April 2018

Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images

Donghuan Lu[1], Karteek Popuri[1], Gavin Weiguang Ding[1], Rakesh Balachandar[1], Mirza Faisal Beg[1] & Alzheimer's Disease Neuroimaging Initiative*

Alzheimer's Disease (AD) is a progressive neurodegenerative disease where biomarkers for disease based on pathophysiology may be able to provide objective measures for disease diagnosis and staging. Neuroimaging scans acquired from MRI and metabolism images obtained by FDG-PET provide in-vivo measurements of structure and function (glucose metabolism) in a living brain. It is hypothesized that combining multiple different image modalities providing complementary information could help improve early diagnosis of AD. In this paper, we propose a novel deep-learning-based framework to discriminate individuals with AD utilizing a multimodal and multiscale deep neural network. Our method delivers 82.4% accuracy in identifying the individuals with mild cognitive impairment (MCI) who will convert to AD at 3 years prior to conversion (86.4% combined accuracy for conversion within 1–3 years), a 94.23% sensitivity in classifying individuals with clinical diagnosis of probable AD, and a 86.3% specificity in classifying non-demented controls improving upon results in published literature.

Alzheimer's disease (AD), the most common dementia, affecting 1 out of 9 people over the age of 65 years[1]. Alzheimer's diseases involves progressive cognitive impairment, commonly associated with early memory loss, requiring assistance for activities of self care during advanced stages. Alzheimer's is posited to evolve through a prodromal stage which is commonly referred to as the mild cognitive impairment (MCI) stage and 10–15% of individuals with MCI, progress to AD[2] each year. With improved life expectancy, it is estimated that about 1.2% of global population will develop Alzheimer's disease by 2046[3] thereby affecting millions of individuals directly, as well as many more indirectly through the effects on their families and caregivers. There is an urgent need to develop biomarkers that can identify the changes in a living brain due to the pathophysiology of AD providing numerical staging scores, as well as identifying syndromal stages.

Neuroimaging modalities such as magnetic resonance imaging (MRI)[4] and fluorodeoxyglucose positron emission tomography (FDG-PET)[5] have been previously used to develop such pathophysiology-based biomarkers for diagnosis of AD, specially targeting the prodromal stage of AD, where the pathology has begun but the clinical symptoms have not yet manifested. Structural MRI provides measures of brain gray matter, white matter and CSF compartments enabling the quantification of volumes, cortical thickness and shape of various brain regions and utilize these in developing classifiers for AD[6–13]. FDG-PET provides measures of the resting state glucose metabolism[14], reflecting the functional activity of the underlying tissue[5] that has also been utilized for AD biomarker development[15–17]. Other published approaches have utilized a combination of modalities for developing neuroimaging AD biomarkers[4,18–24].

Recent advances in deep neural network approaches for developing classifiers have delivered astounding performance for many recognition tasks[25]. The application of deep neural networks in recognition of AD has also attracted application for AD[26–28]. By applying deep neural network to extract features, such as stacked

[1]School of Engineering Science, Simon Fraser University, Burnaby, V5A 1S6, Canada. *A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to M.F.B. (email: mfbeg@sfu.ca)

autoencoder (SAE) or Deep Boltzmann Machine (DBM), these approaches outperform other popular traditional machine learning methods, e.g., support vector machine (SVM) and random forest techniques. A major problem of deep neural network's application in AD diagnosis is that only a small amount of training data is available for learning discriminative patterns in very high dimensional feature spaces. Another issue is that the scale at which the discriminative signal resides is not a-priori known hence dimensionality reduction techniques need to be sensitive to multiple scales to increase the chances of extracting the discriminative signal.

In this paper, we are proposing a novel approach for combining multimodal information from both MRI and FDG-PET images at multiple scales within a deep neural network framework. Our proposed multiscale approach extracts features at coarse-to-fine structural scales[29,30]. This is achieved by segmenting the structural image into cortical and subcortical gray-matter compartments, and further subdividing each into patches of a hierarchical size, and extract features from each-sized patch[26–28] by averaging within the patch and use these multi-scale features taken from multiple modalities into a deep learning framework. Unlike the simple approach of down sampling, which could lead to the loss of discriminative information, our multi-scale approach preserves the structural and metabolism information at multiple scales and may potentially improve the classification accuracy for this diagnostic task[31]. To validate our proposed novel methodology, we performed cross validation experiments with all available ADNI data (subjects that include both a T1-structural MRI and an FDG-PET metabolism image). A comprehensive set of results of these experiments for the detection of controls and MCI that convert to AD as a function of years to conversion, as well as classification of controls, and AD subjects are presented for each modality separately and in combination, and compared to existing methods available in literature demonstrating superiority of the deep neural network framework in AD diagnosis and prognosis.

Methods

There are two major steps in the proposed framework: (1)image preprocessing: segment both MRI and FDG-PET images, subdivide the gray-matter segmentation into patches of a range of sizes, and extract features from each-sized patch; and, (2)classification: train a deep neural network to learn the patterns that discriminate AD individuals, and then use for individual classification.

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

For a comprehensive validation of the proposed method, it is emphasized that all the available ADNI subjects (N = 1242) with both a T1-weighted MRI scan and FDG-PET image at the time of preparation of this manuscript were used in this study. These subjects were categorized into 5 groups based on the individual's clinical diagnosis at baseline and future-timepoints. 1) Stable Normal controls (sNC): 360 subjects diagnosed to be NC at baseline and remained NC at the time of preparation of this manuscript. 2) Stable MCI (sMCI): 409 subjects diagnosed to be MCI at all time points (at least for 2 years). 3) Progressive NC (pNC): 18 subjects evaluated to be NC at baseline visit but progressed to clinical diagnosis of probable AD at the time of preparation of this manuscript. 4) Progressive MCI (pMCI): 217 subjects evaluated to be MCI at baseline visit and progressed to a clinical diagnosis of probable AD at some point in the future (data available for upto 8 years prior to conversion for some individuals). 5) Stable Alzheimer's disease (sAD): 238 subjects with a clinical diagnoses of probably AD. Subjects showing improvement in their clinical diagnosis during their follow up, i.e. those clinically diagnosed as MCI but reverted to NC or those clinically diagnosed as probable AD but reverted to MCI were excluded from the proposed study because of the potential uncertainty of clinical misdiagnosis considering AD is an irreversible form of dementia[1]. The progressive controls and progressive MCI subjects have some neuroimaging timepoints with a clinical diagnosis of probable AD. Hence, the subset of images from pNC and pMCI subjects that have a clinical diagnosis of probable AD will be identified as part of the sAD group for assessment of classifier accuracy, while the remaining images before the conversion to AD will be assessed as part of the pNC and pMCI groups. Demographic and clinical information of the subjects are shown in Table 1. Numbers in brackets are the number of male and female subjects in the second row, while in the rest 3 rows the two number represent the minimum and maximum value of age, education year and MMSE (Mini–Mental State Examination) score. In total, there are 2402 FDG-PET scans and 2402 MRI images including all the longitudinal time-points. Detailed descriptions of the ADNI subject cohorts, image acquisition protocols procedures and post-acquisition preprocessing procedures can be found at http://www.adni-info.org.

Image Processing. Unlike typical image recognition problems where deep learning has shown to be effective, our data set, although very large in a neuroimaging context, is relatively smaller. Hence directly using this smaller database of images to train the deep neural network is unlikely to deliver high classification accuracy. However, contrary to typical image recognition tasks, where the database of images contains large heterogeneity, the images in this database are all human brain images acquired with similar pose and scale which show relatively much less heterogeneity in comparison. Therefore we applied the following processing steps to extract patch-wise features as shown in Fig. 1: FreeSurfer 5.3[32] was used to segment each T1 structural MRI image into gray matter and white matter followed by subdivision of the gray matter into 87 anatomical regions of interest (ROI). The FreeSurfer segmentation were quality controlled by an expert neuroanatomist and any errors noted were manually corrected. Then, a T1 MRI image was chosen as the template. Each ROI of this template was further subdivided into smaller regions of varying sizes, denoted here as "patches". The voxels in each ROI were clustered into patches through *k*-means clustering based on Euclidean distance of their spatial coordinates[33], i.e. voxels spatially close to each other would belong to the same patch. Given that the size of FreeSurfer ROIs were different, we predefined the number of voxels in each patch instead of fixing the number of patches in each ROI

	sNC	sMCI	pNC	pMCI	sAD
Count (Male/Female)					
Number of Subjects	360 (167/193)	409 (239/170)	18 (11/7)	217 (126/91)	238 (141/97)
Number of Images	753 (399/354)	409 (239/170)	74 (51/23)	702 (422/280)	464 (270/194)
Mean (min-max)					
Age in years	73.4 (60–94)	74 (56–91)	77 (68–84)	74 (55–89)	75 (55–90)
Education in years	16.5 (6–20)	15.8 (7–20)	15.7 (12–20)	16.0 (8–20)	15.3 (4–20)
MMSE score	29.1 (24–30)	28.0 (22–30)	29.4 (27–30)	26.5 (9–30)	23.2 (18–27)

Table 1. Subject Demographics. Except for the count where the numbers in brackets are (male/female), age (years), education (years) and MMSE score are displayed in the mean (min-max) format.

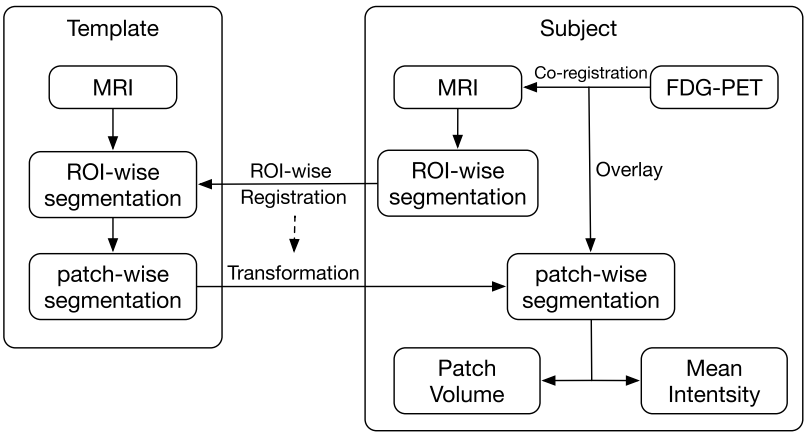


Figure 1. Flowchart of extracting patch-wise features from MRI scans and FDG-PET images. Each FreeSurfer ROI was segmented into "patches" through registration to a patch-segmented template. Patch-based volume and mean intensity of FDG-PET were extracted as features to represent each patch.

to keep uniform patch size density (patches in ROI/voxels in ROI) across the brain leading to signal aggregation at the same scale among the different ROIs. In this study, the size of patches was predefined to be 500, 1000 and 2000 voxels. Using these sizes, the number of patches in total across the brain gray matter ROIs segmented by FreeSurfer was found to be 1488, 705 and 343, respectively. The patch size chosen were designed to keep enough detailed information as well as avoiding too large feature dimension considering the limited number of available data samples. Subsequently, each ROI of the standard template MRI was registered to the same ROI of every target image via a high-dimensional non-rigid registration method (LDDMM[34]). The registration maps were then applied to the patch-wise segmentation of the standard template. This transformed the template patch segmentation into each target MRI space so the target images were subdivided into the same number of patches for their FreeSurfer ROIs. It is also worth mentioning that after the transformation, the size of a template patch in different images is not the same due to non-rigid registration encoding local expansion/contraction and hence is one of the features used to represent the regional information of a given structural brain scan. Then, for each target subject, the FDG-PET image of the subject was co-registered to its skull-stripped T1 MRI scan with a rigid transformation using FSL-FLIRT program[35] based on normalized mutual information. The degrees of freedom (DOF) was set as 12 and Normalized correlation was used as cost function. The mean intensity in the brainstem region of the FDG-PET image was the chosen reference to normalize the voxel intensities in that individual brain metabolism image, because brainstem region was most unlikely to be affected by AD. The mean intensity of each patch was used to form the feature vector representing the metabolism activity, and the volume of each patch was used to represent the brain structure.

Multimodal and Multiscale Deep Neural Network. With the features extracted from MRI and FDG-PET images, we trained a Multimodal and Multiscale Deep Neural Network (MMDNN) to perform the classification. As shown in Fig. 2, the network consists of two parts. The first part consisted of 6 independent deep neural networks (DNNs) corresponding to each scale of a single modality. The second part was another DNN used to fuse the features extracted from these 6 DNNs. The input data of this DNN was the concatenated latent representation learned from each single DNN. The DNNs in the two parts shared the same structure. For each DNN, the number of nodes for each hidden layer were set as $3N$, $\frac{3}{4}N$ and 100 respectively, where N denotes the dimension of input feature vector. The number of nodes was chosen to explore all possible hidden correlation across features from different patches in the first layer and gradually reduce the number of features in the following layers to avoid over-fitting. We trained each DNN with two steps, unsupervised pre-training and supervised fine-tuning, respectively. Then all the parameters of MMDNN were tuned together. The trained DNN output is a

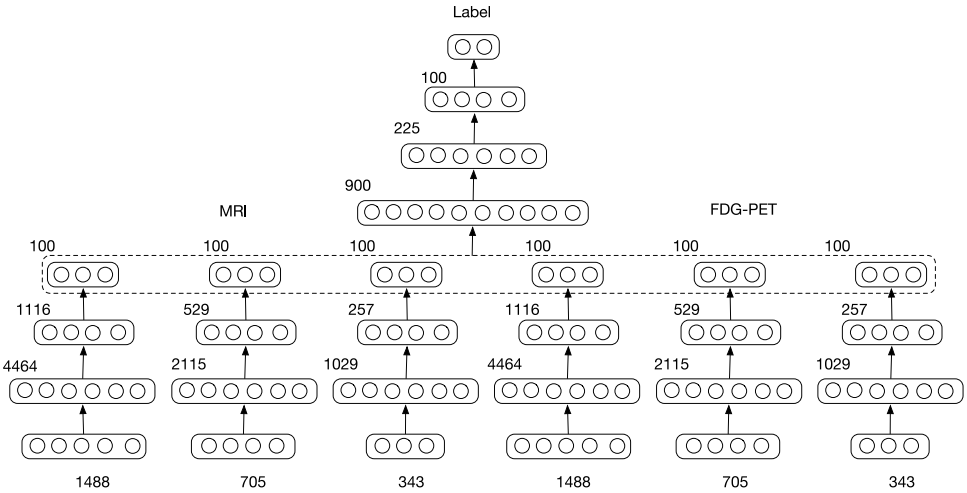


Figure 2. Multimodal and Multiscale Deep Neural Network. The input feature dimension (number of patches) extracted from different scales is 1488, 705 and 343. For each layer, its number of nodes is shown on the top left of the layer representation. For each scale of each image modality, its patch-wise measures were fed to a single DNN. The features from these 6 DNNs were fused by another DNN to generate the final probability score for each of the two classes being discriminated. Of the two classes, the class being the one with the highest probability (effectively a threshold of 0.5 for probability) is the assigned final classification. The probability output of the DNN can be interpreted as a staging score, with extreme value of 0 representing the highest probability of belonging to the sNC class, and extreme value of 1 representing the highest probability of belonging to the AD class.

probability value for each class, the final classification is to the label with the highest probability. The probability value can also be interpreted as a disease staging score, with extreme value of 0 representing the highest probability of belonging to the sNC class, and extreme value of 1 representing the highest probability of belonging to the AD class.

Unsupervised Pre-training. For the unsupervised pre-training step, each DNN was trained as a stacked-autoencoder (SAE). Autoencoder is an artificial neural network used for unsupervised learning of non-linear hidden patterns from input data. It consists of three layers, input layer, hidden layer and output layer, for which two nearby layers are fully-connected. Three functions are used to define an autoencoder, encoding function, decoding function and loss function. In this study, encoding function is defined as: $y = s(W_1x + b_1)$, where x is the input data, y is the latent representation, W_1 is the weight matrix, b_1 is the bias term and s is the activation function for which we used rectified linear function $max(0, x)$. Similarly, decoding function can be represented as: $z = s(W_2y + b_2)$, where we constrained it with tied weight $W_1 = W^T$ and z is the reconstructed data which is supposed to be close to input x. Squared error $\frac{1}{2}\|x - z\|^2$ is applied as loss function to optimize the network. The hypothesis is that the latent representation can capture the main factors of variation in the data. Comparing with another popular unsupervised feature learning method, the principle component analysis (PCA), the activation function enables the network to capture non-linear factors of data variation, especially when multiple encoders and decoders are stacked to form a SAE. To fully train the network, we applied greedy layer-wise training[36] approach where every hidden layer was trained separately.

Supervised Fine-tuning. After pre-training, the first three layers of a DNN were initialized with the parameters of encoders from pre-trained SAE followed by a softmax output layer. At first, we trained the output layer independently while fixing the parameters of first 3 layers. Then we fine-tuned the whole network as Multilayer Perceptron (MLP) with subject labels for criterion. The network outputs the probabilities of a subject belonging to each class and the class with highest probability determines the output label of the subject. If we use x^i, y^i to represent the input feature vector and label of the i_{th} sample, respectively, the loss function based on cross entropy can be displayed as:

$$H(i) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{2}[1\{y^i = j\}log(h(x^i)_j] \quad (1)$$

where N is the number of input samples, j represents the class of samples, and h represents the network function.

Optimization of Network. Training of the network was performed via back propagation with the Adam algorithm[37]. It is a first-order gradient-based optimization algorithm which has been proven to be computationally efficient and appropriate for training deep neural networks. During the training stage, the training set was randomly split into mini batches[38] where each split contains 50 samples in this study. At every iteration, only a single

mini batch was used for optimization. After every batch has been used once, the training set was reordered and randomly divided again so that each batch would have different samples in different epochs.

Dropout. In order to prevent the deep neural network from overfitting, regularization is necessary to reduce its generalization error. In this study, we used dropout[39] to learn more robust features and prevent overfitting. In the dropout layer, some units were randomly dropped, providing a way to combine many different neural networks. In this study, we inserted dropout layers after every hidden layer. In each iteration of training stage, only half of hidden units were randomly selected to feed the results to the next layer, while in the testing stage all hidden units were kept to perform the classification. By avoiding training all hidden units on every training sample, this regularization technique not only prevented complex co-adaptations on training data and decrease overfitting, but also reduced the amount of computation and improved training speed.

Early Stopping. Another approach we used to prevent overfitting is early stopping. Because deep architectures were trained with iterative back propagation, the networks were prone to be more adaptive to the training data after every epoch. At a certain point, improving the network's fit to the training set is likely to decrease generalization accuracy. In order to terminate the optimization algorithm before over-fitting, early stopping was used to provide guidance for how many iterations are needed. In the cross validation experiment, after dividing the data set into training and testing, we further split the training samples into a training set and a validation set. The networks were trained only with data in the former training set, while samples in the latter validation set were used to determine when to stop the algorithm: while the network has the highest generalization accuracy for validation set. In actual training, we stopped the optimization if the validation accuracy had ceased to increase for 50 epochs.

Ensemble Classifiers. Although early stopping has proven to be useful in most deep learning problems, relatively small data set limited the number of samples we could use for validation. And a small validation set may not able to represent the whole data set resulting in a biased network. Therefore, we resorted to ensemble multiple classifiers to perform more stable and robust classification. Instead of selecting a single validation set, we randomly divided the training set into 10 sets and used them to train 10 different networks to 'vote' for the classification. At the training stage, for network i, set i would be used for validation while the rest 9 sets were used for training. At the testing stage, the test samples were fed into all these networks resulting in 10 sets of probabilities. For each sample, the probabilities from 10 networks were added and the class with highest probability was the classification result of this sample. Although the performance of ensemble classifiers may not be greater than a single classifier on every occasion, the ensemble strategy can statistically improve the classification accuracy as well as the robustness and the stability of the classifier.

Ensemble Classifier Probability Distribution. The output of the DNN for each individual image is a pair of probability values representing the probabilities of the given input subject image features (or image pair features for multimodal images) as belonging to one of the two classes on which the DNN was trained. This probability score for belonging to the disease positive (AD) class can be interpreted as a disease severity staging score, since value of 1 represents the highest probability of being from the AD class, and 0 represents highest probability of being from the disease negative (NC) class.

Classifier Validation Experiment Setup. To validate the discriminant ability of proposed network, two kinds of binary classification experiments were performed. First, we performed discrimination between sMCI and pMCI to compare our results on this experiment directly with the published state-of-the-art methods[18,20,21,28,40–46]. Since the published literature typically used only baseline images, we also used a single baseline image for each of the 409 sMCI subjects. Hence, the number of sMCI images is the same as the number of sMCI subjects. For the 217 pMCI subjects, their earliest image within 3 years before conversion was selected. The data samples were randomly divided into 10 sets. For each iteration, 1 set was used for testing while the rest sets were all used for training. Therefore, all subjects were used for testing exactly once.

One potential issue with respect to the sMCI class is that some of these individuals may progress to AD or other dementias in the future and if some of these individuals convert to probable AD in the future, these earlier timepoints would become part of the pMCI group, whereas some other individuals may revert back to NC. Hence, although the sMCI vs. pMCI experiment is commonly used to assess classifier performance in recent studies, the classification of sMCI subjects may not be entirely accurate due to the potential uncertainty in the clinical diagnosis of the sMCI class. Therefore, we performed additional experiments that involved classifying individuals with known future progression to AD, namely the pNC, pMCI and sAD classes, denoted as the dementia positive class, against those that are stable normal controls (sNC), denoted as the dementia negative class.

We investigated the performance of the classifier by using various combinations of samples during training phase. At the first level, the classifier was trained soley on samples from the sNC subjects (the dementia negative class) and the sAD subjects (the dementia positive class). At the next level, the dementia positive class was enriched with pMCI subjects' images that represent an earlier stage in the evolution of AD. In the last level, the positive class was further enriched with adding pNC subjects' images representing an even earlier stage in the evolution of AD. For each level, the classifier training followed the standard 10-fold cross validation procedure (90% of data samples used for training and 10% of data used for testing in each iteration). The groups not used for training, if any, were utilized in the testing group. In these experiments, allocation into training or testing was done on the level of subjects, not images. If a subject was allocated into the training group, all the available baseline and longitudinal images for this subject would be used for training. Otherwise, all the available images of a subject would be used for testing.

Method	Modality	Year to conversion	#Subjects	Accuracy	Sensitivity	Specificity
Young *et al.*[18]	MRI	0–3	143	64.3	53.2	69.8
Liu *et al.*[40]	MRI	0–3	234	68.8	64.29	74.07
Suk *et al.*[28]	MRI	unknown	204	72.42	36.7	90.98
Cheng *et al.*[41]	MRI	0–2	99	73.4	74.3	72.1
Zhu *et al.*[42]	MRI	0–1.5	99	71.8	48.0	92.8
Huang *et al.*[46]	longitudinal MRI	0–3	131	79.4	86.5	78.2
Proposed	MRI	0–3	**626**	**75.44** (7.74)	73.27 (7.58)	76.19 (8.35)
Young *et al.*[18]	PET	0–3	143	65.0	66.0	64.6
Liu *et al.*[40]	PET	0–3	234	68.8	57.14	82.41
Suk *et al.*[28]	PET	unknown	204	70.75	25.45	96.55
Cheng *et al.*[41]	PET	0–2	99	71.6	76.4	67.9
Zhu *et al.*[42]	PET	0–1.5	99	71.2	47.4	93.0
Proposed	PET	0–3	**626**	**81.53** (7.42)	78.20 (7.72)	82.47 (9.30)
Young *et al.*[18]	MRI + PET + APOE	0–3	143	69.9	78.7	65.6
Liu *et al.*[40]	PET + MRI	0–3	234	73.5	76.19	70.37
Suk *et al.*[28]	PET + MRI	unknown	204	75.92	48.04	95.23
Cheng *et al.*[41]	PET + MRI + CSF	0–2	99	79.4	84.5	72.7
Zhu *et al.*[42]	MRI + PET	0–1.5	99	72.4	49.1	94.6
Moradi *et al.*[20]	MRI + Age + cognitive measure	0–3	264	82	87	74
Xu *et al.*[43]	MRI + PET + florbetapir PET	0–3	110	77.8	74.1	81.5
Zhang *et al.*[44]	longitudinal MRI + PET	0–2	88	78.4	79.0	78.0
An *et al.*[45]	MRI + SNP	0–2	362	80.8	71.5	85.4
Korolev *et al.*[21]	MRI + Plasma + cognitive measure	0–3	259	80.0	83.0	76.0
Proposed	PET + MRI	0–3	**626**	**82.93** (7.25)	79.69 (8.37)	83.84 (6.37)

Table 2. Classification performance for the sMCI vs. pMCI task: Accuracy (%), Sensitivity (%), and Specificity (%) of the proposed network compared with the published state-of-the-art methods for the task of discriminating between sMCI and pMCI subjects. 'PET' in this table represents FDG-PET neuroimaging. Our proposed approach using deep neural networks was performed using a single FDG-PET image and a single T1-MRI acquired from each of 409 sMCI subjects and 217 pMCI subjects (total 626 subjects) and the average (standard-deviation) of the accuracy, sensitivity and specificity of classifier performance are reported.

Sensitivity of the classifier is defined as the number of positive class images that are correctly classified, which in this case is the classification of the test subset of pNC, pMCI and sAD images as the positive class. Specificity of the classifier is the number of negative class images (the sNC class) that are correctly classified as sNC. Accuracy of the classifier is the fraction of images from both the positive and the negative classes that are correctly classified.

The proposed deep neural network (DNN) was built with Tensorflow[47], an open source deep learning toolbox provided by Google. For all the experiments, the number of nodes in each layer was predefined as shown in Fig. 2 and the learning rate was set as 10^{-4}. The deep network parameter space is very large, with a large range of choices from which to sample i.e. number of layers and number of nodes, testing all the possible parameter combinations exhaustively is computationally unrealistic. Instead of doing parameter selection for each of the 10-fold experiments, the parameters were selected based on the results of the first fold experiment.

Results

Discrimination between Stable and Progressive MCI (sMCI vs pMCI). We conducted the sMCI vs. pMCI experiment to be able to compare the classification accuracy of our proposed novel method with published and comparable state-of-the-art methods[18,20,21,28,40–46]. The FDG-PET image and MRI image acquired at a single time point for each subject were used for the 10-fold cross validation experiment. For sMCI subjects, the images acquired at the first time to visit, while for pMCI subjects, the images acquired at the earliest time point within 3 years before conversion were used. Results of this experiment and comparable results from published methods are shown in Table 2. These results reveal an accuracy of 82.9% for our MMDNN method over 626 subjects and both specificity (83.8%) and sensitivity (79.7%) are high. The results for single modality DNN are also found to improve upon the state-of-art. These results suggest that our proposed MMDNN network is promising for applications requiring classification between sMCI and pMCI individuals for the single modality T1-MRI and FDG-PET or the multimodal (T1-MRI and FDG-PET combined) neuroimaging approach.

Discrimination between disease negative (sNC) and disease positive (the pNC, pMCI, sAD) classes. The classifier was trained to discriminate the negative class (sNC) from the disease positive class (pNC, pMCI, sAD) using three different enrichments for the positive class samples, namely training with the positive class containing only sAD, or, pMCI and sAD, or, pNC and pMCI and sAD samples. Each subject was used for testing at least once in the 10-fold cross validation experiments. In each fold of the experiment, images of the same subject

	FDG-PET (Metabolism)			T1-MRI (Volume)			Multimodal (Metabolism + Volume)		
	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
Training Set									
sNC vs. sAD	84.5 (1.4)	79.9 (1.6)	91.9 (6.9)	81.9 (1.2)	75.5 (1.3)	92.3 (4.9)	84.6 (1.5)	80.2 (2.0)	91.8 (6.8)
sNC vs. (pMCI and sAD)	85.5 (2.8)	85.0 (2.9)	86.2 (10.1)	82.8 (3.4)	79.8 (4.1)	87.7 (6.3)	86.0 (2.5)	85.7 (3.2)	86.5 (8.6)
sNC vs. (pNC, pMCI and sAD)	85.9 (4.8)	85.6 (3.8)	86.3 (7.8)	82.5 (5.2)	80.2 (7.6)	86.1 (7.6)	86.4 (4.7)	86.5 (5.2)	86.3 (8.6)

Table 3. Accuracy (%), sensitivity (%) and specificity (%) of each modality and the multimodal combination using different training sets for the classification of AD pathology (discrimination of pNC, pMCI and sAD from sNC). The numbers in each cell are the average value and standard deviation from the 10-fold cross validation experiments. Sensitivity is the fraction of correctly classified pNC, pMCI and sAD images, specificity is the fraction of correctly classified sNC images and accuracy is correctly classified images taken together.

acquired at different time points were either all used for training or all used for testing to ensure the independence of training and testing at all times, as further detailed in the Classifier Validation Experiment Setup Section.

The classification result of these experiments are shown as Table 3. The DNN based on FDG-PET neuroimaging features (accuracy 85.9%) performs better than the DNN based on T1-MRI (accuracy 82.5%) neuroimaging features, and the combined MMDNN outperforms each of the single modality DNNs (accuracy 86.4%). As the positive class is enriched with samples from the pMCI and then further with the pNC samples, there is an increase in the sensitivity (correctly classified members of the dementia positive class i. e. pNC, pMCI and sAD). Since some of the early stage patterns of AD represented in pMCI and pNC may overlap the sNC group, there is a slight decrease in specificity, but overall an increase in accuracy.

The features extracted by the deep neural network are displayed in Fig. 3. Although difficult to interpret as these are extracted from multiple nonlinear transformations of data, they show that the patterns for the different classes appear to be distinct, whereas patterns within each class appear to be relatively similar.

Classification performance of pNC and pMCI as function of time (years) to conversion. We analyzed the accuracy of classification of pNC and pMCI as a function of the time (years) to conversion and the numbers of subjects available for the MMDNN classifier. These results are shown in Fig. 4 for each of the three training scenarios with progressive enrichment of the positive class. As the positive class training set of sAD (top row, left panel) is enriched with samples from pMCI (top row, middle panel) and with pNC and pMCI samples (top row, right panel), the accuracy of detection of the pMCI and pNC class increases, as well as an increase in accuracy for identifying AD in pNC and pMCI earlier. The numerical values of classifier performance for the pNC, pMCI and sAD enriched positive class (top row, third panel on the right) are provided in the table in the second row of this figure.

The MMDNN classifier accuracy in identifying pMCI individuals with future conversion to AD was 90%, 86.6% and 82.4%, for years 1, 2, and 3 away to conversion. The accuracy for all the years taken together for pMCI classification was 79.22%, and 86.4% total for conversion within 1–3 years. The neuroimaging scans farther away from conversion are likely more challenging to classify correctly leading to overall lowered accuracy. The classification accuracy for sAD group, i. e, those images associated with a clinical diagnosis of AD, is 94.25%. The accuracy for correctly classifying all pNC images is 41.1% with higher numbers of 100%, 60.0% and 66.7% for years 1, 2 and 3 from conversion to clinical diagnosis of probable AD.

Classification Probability score distribution. The probability score output by the MMDNN trained with the dementia negative (sNC) class and the three enrichment choices for the dementia positive class (namely, sAD, pMCI + sAD, and pNC + pMCI + sAD) class samples is visualized as histograms in the top row of Fig. 5. The fraction of images of each class is shown on the y axis, along with classifier probability score shown on the x axis. This distribution shows how the sNC, pNC, pMCI and sAD classes are scored by the classifier for their probability of being from the dementia positive class. Further, the bottom row of Fig. 5 shows aggregate values of the probability score with respect to each class with a box plot. As the training set for the dementia positive class is enriched with samples from pMCI and then additionally, pNC class, the probability score for these classes is shown to increase. Overall, the distribution generated by the MMDNN leads to good separation between the classes, and the threshold choice of 0.5 (highest class probability assignment) is visually shown to provide good classification between the classes.

Discussion

In this paper, we have proposed a novel deep neural network (DNN) based method that utilizes multi-scale and multi-modal information (MMDNN) combining metabolism (FDG-PET) and regional volume (T1-MRI) for the discrimination of AD, with a focus on assessing classification accuracy in those pNC and pMCI subjects with known future conversion to probable AD. In accordance with scale-space theory, our incorporation of multiscale approach was intended to capture the discriminant signals at multiple scales, and avoid apriori assumption of the scale at which the discriminant signals may reside.

The comparison between our novel proposed MMDNN method and state-of-the-art methods for the sMCI vs. pMCI classification task is shown in Table 2. Although the data used for the cited studies are not identical, they all come from the ADNI database and have comparable image acquisition and preprocessing procedures. One of the

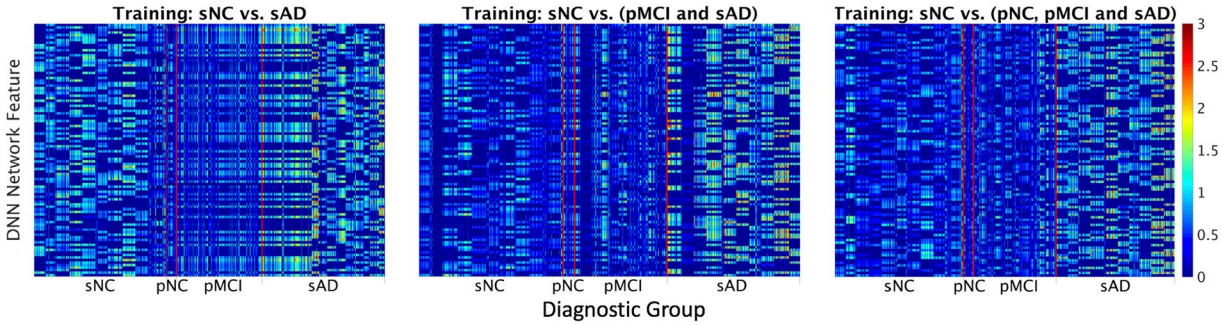
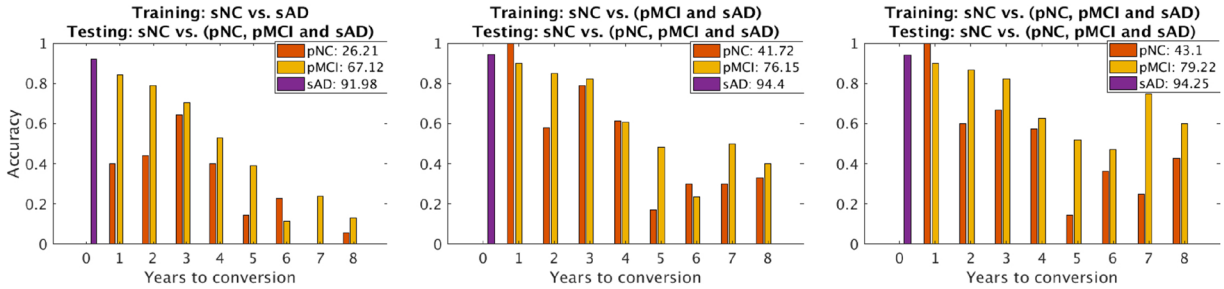


Figure 3. Features extracted from the input data by the deep neural network at the penultimate layer that are fed to the output layer for classification. From left to right the training set is sNC vs sAD, sNC vs (pMCI + sAD), and sNC vs (pNC + pMCI + sAD) respectively. The y axis represents the units of the second from last layer, while the x axis denotes the different data groups. The vertical red lines are added to enhance visual distinction between the boundaries of each group. This figure shows that while it is difficult to provide an interpretation of the features found by the deep neural network from the input neuroimaging features, the patterns as distilled by the deep learning network from the sNC, pNC, pMCI and sAD images are distinct with more uniformity within each class as compared to across classes.



Years To Conversion	1	2	3	4	5	6	7	8
pNC: Accuracy(#Subjects)	100.0(1)	60.0(5)	66.7(9)	57.1(7)	14.3(7)	36.4(11)	25.0(4)	42.9(7)
pMCI: Accuracy(#Subjects)	90.0(120)	86.6(134)	82.4(108)	62.8(51)	51.7(29)	47.1(17)	75.0(8)	60.0(10)

Figure 4. Accuracy of correctly identifying prodromal AD for the pNC and pMCI subjects as function of (years) to conversion. The top row shows the effect of enriching the training set of the dementia positive class, with sAD (top row, left panel), pMCI and sAD (top row, middle panel) and pNC and pMCI and sAD (top row, right panel). The y axis represents accuracy of classification, while x axis shows time (years) to AD conversion. The x axis value '0' indicates subjects with current clinical diagnosis of probable AD (sAD subjects). The number in legend is the classification accuracy taken over all time points for each group. The table in the second row shows the numeric accuracy value for pNC and pMCI subjects at different time (years) to conversion to AD corresponding to the top right panel (MMDNN with pNC, pMCI and sAD for training the dementia positive class).

strengths of our work is that we have analyzed all the available ADNI sMCI and pMCI subjects having both MRI and FDG-PET neuroimages at the time of preparation of this manuscript. When using only the T1-MRI modality, our method has better accuracy than most methods expect Huang *et al.*'s[46]. However, they used a longitudinal method with multiple MRI images acquired from different time points for the classification of each subject, whereas we classify each image separately, an approach consistent with the other published cross-sectional methods. For single modality-based classifiers using only FDG-PET, our method outperforms the published methods by a significant margin as shown in Table 2. Extension of our DNN for utilization of longitudinal timepoints for single subject classification is a direction for future work, and we anticipate that adding longitudinal measures explicitly could further improve the classifier performance.

When using multiple modalities for sMCI vs. pMCI classification, our MMDNN approach has the best performance specially compared with the methods that also used the same T1-MRI and FDG-PET modalities. The study of Chen *et al.*[41] performed domain transfer learning to exploit the auxiliary domain data (sAD/sNC subjects) to improve the classification whereas our proposed MMDNN method's performance was better even though we did not utilize domain transfer learning in our sMCI vs. pMCI classification task.

Further, we performed experiments to detect prodromal AD by training the MMDNN classifier with samples from the dementia positive class namely the pNC, pMCI and sAD subjects. The accuracy of correctly classifying pNC and pMCI subjects as having patterns indicative of AD improved when the classifier training included pMCI

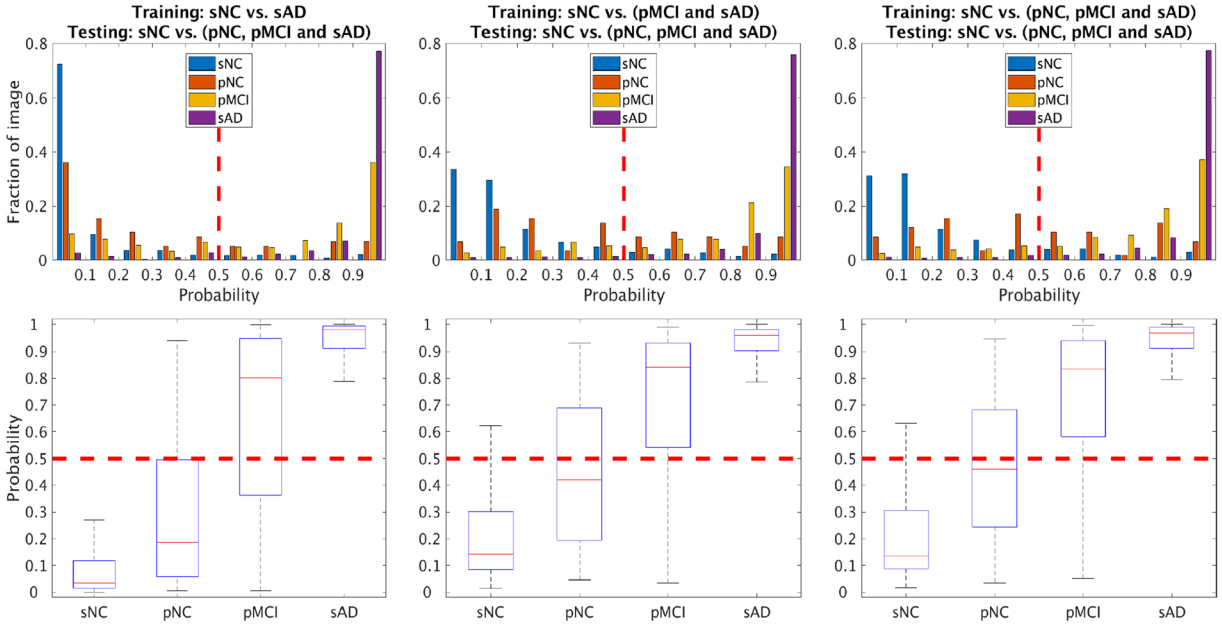


Figure 5. Multimodal classification probability distribution of different training sets. From left to right the training set is sAD, sAD and pMCI, sAD, pMCI and pNC versus sNC respectively. The y axis represents fraction of images, while x axis denotes the probability of class sAD, where 0 represents high likelihood of being from the sNC pattern, and 1 represents high likelihood of being from the sAD pattern.

and pNC images, as displayed in Table 3. Further, comparison of the DNN results for T1-MRI and FDG-PET classifiers as shown in Table 3 indicates that the sensitivity of detection of prodromal AD is higher with FDG-PET neuroimaging features as compared to T1-MRI neuroimaging features. This finding is consistent with previous studies[18,28,40,41] and could indicate support for the hypothesis that alterations in metabolism may precede changes in structure, and further, the altered metabolism measures could be detected with FDG-PET earlier than the detection of structural changes with T1-MRI.

Analysis of the accuracy of classifying prodromal AD i. e. detecting patterns corresponding to AD in pNC and pMCI individuals as function of time (years) to conversion is shown in Fig. 4. As the training set was enriched with samples from the pNC and the pMCI groups, the accuracy of detection of prodromal AD also increased. The MMDNN classifier delivered high accuracy upto three years prior to conversion and then performance was reduced for the timepoints 4–8 years prior to conversion. The number of subjects in 1–3 years before conversion are large (over 100 each), and there is also reduced numbers of available subject numbers 4–8 years away from conversion. The reduced sample for timepoints farther away from conversion to AD could potentially increase classification uncertainty. With more neuroimaging data corresponding to timepoints farther from conversion to AD becoming available, models such as the MMDNN proposed here could provide better classification performance for the earlier detection of prodromal AD.

The probability score output from the DNN is visualized in Fig. 5. The probability score is highest for the sAD class, and lowest for the sNC class, being the two extreme ends of the spectrum for the classifier. The probability score for the pNC and pMCI subjects is in between, and higher for pMCI than pNC generally in line with the expectation of progressive alterations detected with neuroimaging for subjects further along the disease trajectory. Further analysis of the classifier probability score could be an interesting avenue to develop a surrogate staging score for disease severity.

Despite the remarkable ability of DNN to discover patterns that may not be apparent on human visual examination, one major disadvantage of the DNN framework is that as a result of multiple non-linear transformations between the input in generating the output, it is not readily possible to map the output classification probability back to neuroimaging patterns in the input neuroimaging data that give rise to this output. The visualization of the output of the penultimate layer in the DNN for individual subject images is shown in Fig. 3 and except for observing a qualitative difference between the features of different classes, it is not possible to relate these to neuroimaging features from specific locations in the brain at the current time. Understanding how to provide pathophysiologically meaningful interpretation of the features extracted by the DNN for classificaion remains an unsolved problem and is an important future research direction.

A small number of subjects are awarded a probability score inconsistent with their clinical diagnosis. One of the main requirements of training DNNs are large quantities of well-characterized data[25]. It is therefore possible that as more comprehensive and homogeneous training databases are developed and become available for learning, the accuracy numbers may increase and these outliers will be reduced. It is also possible that there may be some uncertainty in the available clinical diagnosis. Despite the limitations, our findings indicate that the DNN

framework has considerable potential in learning the AD-related patterns for promising future applications in adding to the toolbox of clinical AD diagnosis.

Conclusion

In summary, we have proposed a novel deep neural network to identify individuals at risk of developing Alzheimer's disease. Our multi-scale and multi-modal deep neural network (MMDNN) was designed to incorporate multiple scales of information from multiple regions in the gray matter of the brain taken from multiple modalities (T1-MRI and FDG-PET). First we demonstrated the discriminant ability of the proposed MMDNN approach by comparing with state-of-the-art methods on the task of discriminating between sMCI vs. pMCI individuals. Then we trained the classifier to distinguish subjects on trajectory towards clinical diagnosis of probable AD (i. e. the pNC, pMCI subjects). We observed the performance of MMDNN classifier built with a combination of FDG-PET and structural MRI images was better than those built using either structural MRI or FDG-PET neuroimaging scans alone. Further the classifier trained with the combined sample of pNC, pMCI and sAD was found to yield the highest overall classification accuracy of 82.4% accuracy in the identifying the individuals with mild cognitive impairment (MCI) who will convert to AD at 3 years prior to conversion (86.4% combined accuracy for conversion within 1–3 years), a 94.23% sensitivity in classifying individuals with clinical diagnosis of probable AD, and a 86.3% specificity in classifying non-demented controls. These results suggest that deep neural network classifiers may be useful as a potential tool for providing evidence in support of the clinical diagnosis of probable AD.

References

1. Association, A. *et al.* Alzheimer's disease facts and figures. *Alzheimer's & dementia: journal Alzheimer's Assoc.* **7**, 208 (2011).
2. Petersen, R. C. *et al.* Mild cognitive impairment: ten years later. *Arch. neurology* **66**, 1447–1455 (2009).
3. Brookmeyer, R., Johnson, E., Ziegler-Graham, K. & Arrighi, H. M. Forecasting the global burden of Alzheimer's disease. *Alzheimer's & dementia* **3**, 186–191 (2007).
4. Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N. & Trojanowski, J. Q. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. aging* **32**, 2322–e19 (2011).
5. Landau, S. M. *et al.* Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol. aging* **32**, 1207–1218 (2011).
6. Farhan, S., Fahiem, M. A. & Tauseef, H. An ensemble-of-classifiers based approach for early diagnosis of Alzheimer's disease: Classification using structural features of brain images. *Comput. and mathematical methods medicine* **2014** (2014).
7. Korolev, S., Safiullin, A., Belyaev, M. & Dodonova, Y. Residual and Plain Convolutional Neural Networks for 3D Brain MRI Classification. *arXiv preprint arXiv:1701.06643* (2017).
8. Payan, A. & Montana, G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. *arXiv preprint arXiv:1502.02506* (2015).
9. Eskildsen, S. F. *et al.* Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *Neuroimage* **65**, 511–521 (2013).
10. Misra, C., Fan, Y. & Davatzikos, C. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *Neuroimage* **44**, 1415–1422 (2009).
11. Wolz, R. *et al.* Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. *PLoS One* **6**, e25446 (2011).
12. Cuingnet, R. *et al.* Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* **56**, 766–781 (2011).
13. Cho, Y. *et al.* Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage* **59**, 2217–2230 (2012).
14. Mosconi, L. *et al.* Pre-clinical detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging. *J. Alzheimer's Dis.* **20**, 843–854 (2010).
15. Gray, K. R. *et al.* Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. *NeuroImage* **60**, 221–229 (2012).
16. Toussaint, P.-J. *et al.* Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses. *Neuroimage* **63**, 936–946 (2012).
17. Illán, I. *et al.* 18 F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis. *Inf. Sci.* **181**, 903–916 (2011).
18. Young, J. *et al.* Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *NeuroImage: Clin.* **2**, 735–745 (2013).
19. Zhang, D. *et al.* Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856–867 (2011).
20. Moradi, E. *et al.* Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* **104**, 398–412 (2015).
21. Korolev, I. O. *et al.* Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification. *PloS One* **11**, e0138866 (2016).
22. Ye, J. *et al.* Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. *BMC Neurol* **12**, 46 (2012).
23. Gaser, C., Franke, K., Kloppel, S., Koutsouleris, N. & Sauer, H. BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease. *PLoS One* **8**, e67346 (2013).
24. Zhang, D. *et al.* Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage* **59**, 895–907 (2012).
25. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105 (2012).
26. Liu, S. *et al.* Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Transactions on Biomed. Eng.* **62**, 1132–1140 (2015).
27. Liu, S. *et al.* Early diagnosis of Alzheimer's disease with deep learning. In *Biomedical Imaging* (*ISBI*), *2014 IEEE 11th International Symposium on*, 1015–1018 (IEEE, 2014).
28. Suk, H.-I. *et al.* Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* **101**, 569–582 (2014).
29. Zhang, W., Zelinsky, G. & Samaras, D. Real-time accurate object detection using multiple resolutions. In *Computer Vision*, *2007. ICCV 2007. IEEE 11th International Conference on*, 1–8 (IEEE, 2007).
30. Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. journal computer vision* **60**, 91–110 (2004).
31. Tang, Y. & Mohamed, A.-R. Multiresolution Deep Belief Networks. In *AISTATS*, 1203–1211 (2012).

32. Dale, A. M., S., M. & Fischl, B. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**(2), 195–207 (1999).
33. Raamana, P. R. *et al*. Thickness network features for prognostic applications in dementia. *Neurobiol. aging* **36**, S91–S102 (2015).
34. Beg, F., Miller, M., Trouvé, A. & Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. journal computer vision* **61**(2), 139–157 (2005).
35. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
36. Bengio, Y. *et al*. Greedy layer-wise training of deep networks. *Adv. neural information processing systems* **19**, 153 (2007).
37. Kingma, D. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
38. Bengio, Y. Practical recommendations for gradient-based training of deep architectures. In *Neural networks*: *Tricks of the trade*, 437–478 (Springer, 2012).
39. Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
40. Liu, K., Chen, K., Yao, L. & Guo, X. Prediction of Mild Cognitive Impairment Conversion Using a Combination of Independent Component Analysis and the Cox Model. *Front. human neuroscience* **11** (2017).
41. Cheng, B., Liu, M., Zhang, D., Munsell, B. C. & Shen, D. Domain transfer learning for MCI conversion prediction. *IEEE Transactions on Biomed. Eng.* **62**, 1805–1817 (2015).
42. Zhu, X. *et al*. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Med. image analysis* (2017).
43. Xu, L., Wu, X., Chen, K. & Yao, L. Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment. *Comput. methods programs biomedicine* **122**, 182–190 (2015).
44. Zhang, D. & Shen, D. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS One* **7**, e33182 (2012).
45. An, L. *et al*. A Hierarchical Feature and Sample Selection Framework and Its Application for Alzheimer's Disease Diagnosis. *Sci. Reports* **7** (2017).
46. Huang, M. *et al*. Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. *Sci. reports* **7** (2017).
47. Abadi, M. *et al*. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. http://tensorflow.org/Software availablefromtensor flow.org (2015).

Acknowledgements

This work was supported by National Science Engineering Research Council (NSERC), Canadian Institutes of Health Research (CIHR), Michael Smith Foundation for Health Research (MSFHR), Brain Canada, Genome BC and the Pacific Alzheimer Research Foundation (PARF). 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.

Author Contributions

Donghuan Lu and Gavin Weiguang Ding built the deep neural network. Donghuan Lu and Karteek Popuri processed the neuroimage data. Donghuan Lu, Karteek Popuri and Mirza Faisal Beg designed the experiments. Donghuan Lu, Rakesh Balachandar and Mirza Faisal Beg interpreted the results. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

© The Author(s) 2018

Consortia

Alzheimer's Disease Neuroimaging Initiative

Michael Weiner[2], Paul Aisen[3], Ronald Petersen[4], Cliford Jack[5], William Jagust[6], John Trojanowki[7], Arthur Toga[8], Laurel Beckett[9], Robert Green[10], Andrew Saykin[11], John Morris[12], Leslie Shaw[12], Jefrey Kaye[13], Joseph Quinn[13], Lisa Silbert[13], Betty Lind[13], Raina Carter[13], Sara Dolen[13], Lon Schneider[8], Sonia Pawluczyk[8], Mauricio Beccera[8], Liberty Teodoro[8], Bryan Spann[8], James Brewer[14], Helen Vanderswag[14], Adam Fleisher[14,53], Judith Heidebrink[15], Joanne Lord[15], Sara Mason[5], Colleen Albers[5], David Knopman[5], Kris Johnson[5], Rachelle Doody[16], Javier Villanueva-Meyer[16], Munir Chowdhury[16], Susan Rountree[16], Mimi Dang[16], Yaakov Stern[17], Lawrence Honig[17], Karen Bell[17], Beau Ances[12], Maria Carroll[12], Mary Creech[2], Erin Franklin[12], Mark Mintun[12], Stacy Schneider[12], Angela Oliver[12], Daniel Marson[18], Randall Grifth[18], David Clark[18], David Geldmacher[18], John Brockington[18], Erik Roberson[18], Marissa Natelson Love[18], Hillel Grossman[19], Efe Mitsis[19], Raj Shah[20], Leyla deToledo-Morrell[20], Ranjan Duara[21], Daniel Varon[21], Maria Greig[21], Peggy Roberts[21], Marilyn Albert[22], Chiadi Onyike[22], Daniel D'Agostino[22], Stephanie Kielb[22], James Galvin[23], Brittany Cerbone[23], Christina Michel[23], Dana Pogorelec[23], Henry Rusinek[23], Mony de Leon[23], Lidia Glodzik[23], Susan De Santi[23], P. Doraiswamy[24], Jefrey Petrella[24], Salvador Borges-Neto[24], Terence Wong[24], Edward Coleman[24], Charles Smith[25], Greg Jicha[25], Peter Hardy[25], Partha Sinha[25], Elizabeth Oates[25], Gary Conrad[25], Anton Porsteinsson[26], Bonnie Goldstein[26], Kim Martin[26], Kelly Makino[26], M. Ismail[26], Connie Brand[26], Ruth Mulnard[27], Gaby Thai[27], Catherine Mc-Adams-Ortiz[27], Kyle Womack[28], Dana Mathews[28], Mary Quiceno[28], Allan Levey[29], James Lah[29], Janet Cellar[29], Jefrey Burns[30], Russell Swerdlow[30], William Brooks[30], Liana Apostolova[31], Kathleen Tingus[31], Ellen Woo[31], Daniel Silverman[31], Po Lu[31], George Bartzokis[31], Neill Graf-Radford[32], Francine Parftt[32], Tracy Kendall[32], Heather Johnson[32], Martin Farlow[11], Ann Marie Hake[11], Brandy Matthews[11], Jared Brosch[11], Scott Herring[11], Cynthia Hunt[11], Christopher Dyck[33], Richard Carson[33], Martha MacAvoy[33], Pradeep Varma[33], Howard Chertkow[34], Howard Bergman[34], Chris Hosein[34], Sandra Black[35], Bojana Stefanovic[35], Curtis Caldwell[35], Ging-Yuek Robin Hsiung[36], Howard Feldman[36], Benita Mudge[36], Michele Assaly[36], Elizabeth Finger[37], Stephen Pasternack[37], Irina Rachisky[37], Dick Trost[37], Andrew Kertesz[37], Charles Bernick[38], Donna Munic[38], Marek-Marsel Mesulam[39], Kristine Lipowski[39], Sandra Weintraub[39], Borna Bonakdarpour[39], Diana Kerwin[39], Chuang-Kuo Wu[39], Nancy Johnson[39], Carl Sadowsky[40], Teresa Villena[40], Raymond Scott Turner[41], Kathleen Johnson[41], Brigid Reynolds[41], Reisa Sperling[42], Keith Johnson[42], Gad Marshall[42], Jerome Yesavage[43], Joy Taylor[43], Barton Lane[43], Allyson Rosen[43], Jared Tinklenberg[43], Marwan Sabbagh[44], Christine Belden[44], Sandra Jacobson[44], Sherye Sirrel[44], Neil Kowall[45], Ronald Killiany[45], Andrew Budson[45], Alexander Norbash[45], Patricia Lynn Johnson[45], Thomas Obisesan[46], Saba Wolday[46], Joanne Allard[46], Alan Lerner[47], Paula Ogrocki[47], Curtis Tatsuoka[47], Parianne Fatica[47], Evan Fletcher[48], Pauline Maillard[48], John Olichney[48], Charles DeCarli[48], Owen Carmichael[48], Smita Kittur[49], Michael Borrie[50], T.-Y. Lee[50], Rob Bartha[50], Sterling Johnson[51], Sanjay Asthana[51], Cynthia Carlsson[51], Steven Potkin[52], Adrian Preda[52], Dana Nguyen[52], Pierre Tariot[53], Anna Burke[53], Nadira Trncic[53], Stephanie Reeder[53], Vernice Bates[54], Horacio Capote[54], Michelle Rainka[54], Douglas Scharre[55], Maria Kataki[55], Anahita Adeli[55], Earl Zimmerman[56], Dzintra Celmins[56], Alice Brown[56], Godfrey Pearlson[57], Karen Blank[57], Karen Anderson[57], Laura Flashman[58], Marc Seltzer[58], Mary Hynes[58], Robert Santulli[58], Kaycee Sink[59], Leslie Gordineer[59], Jef Williamson[59], Pradeep Garg[59], Franklin Watkins[59], Brian Ott[60], Henry Querfurth[60], Geofrey Tremont[60], Stephen Salloway[61], Paul Malloy[61], Stephen Correia[61], Howard Rosen[62], Bruce Miller[62], David Perry[62], Jacobo Mintzer[63], Kenneth Spicer[63], David Bachman[63], Nunzio Pomara[64], Raymundo Hernando[65], Antero Sarrael[64], Norman Relkin[65], Gloria Chaing[65], Michael Lin[65], Lisa Ravdin[65], Amanda Smith[66], Balebail Ashok Raj[66] & Kristin Fargher[66]

[2]Magnetic Resonance Unit at the VA Medical Center and Radiology, Medicine, Psychiatry and Neurology, University of California, San Francisco, USA. [3]San Diego School of Medicine, University of California, California, USA. [4]Mayo Clinic, Minnesota, USA. [5]Mayo Clinic, Rochester, USA. [6]University of California, Berkeley, USA. [7]University of Pennsylvania, Pennsylvania, USA. [8]University of Southern California, California, USA. [9]University of California, Davis, California, USA. [10]MPH Brigham and Women's Hospital/Harvard Medical School, Massachusetts, USA. [11]Indiana University, Indiana, USA. [12]Washington University St. Louis, Missouri, USA. [13]Oregon Health and Science University, Oregon, USA. [14]University of California–San Diego, California, USA. [15]University of Michigan, Michigan, USA. [16]Baylor College of Medicine, Houston, State of Texas, USA. [17]Columbia University Medical Center, South Carolina,

USA. [18]University of Alabama, Birmingham, Alabama, USA. [19]Mount Sinai School of Medicine, New York, USA. [20]Rush University Medical Center, Rush University, Illinois, USA. [21]Wien Center, Florida, USA. [22]Johns Hopkins University, Maryland, USA. [23]New York University, NY, USA. [24]Duke University Medical Center, North Carolina, USA. [25]University of Kentucky, Kentucky, USA. [26]University of Rochester Medical Center, NY, USA. [27]University of California, Irvine, California, USA. [28]University of Texas Southwestern Medical School, Texas, USA. [29]Emory University, Georgia, USA. [30]University of Kansas, Medical Center, Kansas, USA. [31]University of California, Los Angeles, California, USA. [32]Mayo Clinic, Jacksonville, USA. [33]Yale University School of Medicine, Connecticut, USA. [34]McGill University, Montreal-Jewish General Hospital, Montreal, Canada. [35]Sunnybrook Health Sciences, Ontario, USA. [36]U.B.C. Clinic for AD & Related Disorders, Vancouver, BC, Canada. [37]Cognitive Neurology - St. Joseph's, Ontario, USA. [38]Cleveland Clinic Lou Ruvo Center for Brain Health, Ohio, USA. [39]Northwestern University, San Francisco, USA. [40]Premiere Research Inst (Palm Beach Neurology), west Palm Beach, USA. [41]Georgetown University Medical Center, Washington DC, USA. [42]Brigham and Women's Hospital, Massachusetts, USA. [43]Stanford University, California, USA. [44]Banner Sun Health Research Institute, Sun City, AZ, 85351, USA. [45]Boston University, Massachusetts, USA. [46]Howard University, Washington DC, USA. [47]Case Western Reserve University, Ohio, USA. [48]University of California, Davis – Sacramento, California, USA. [49]Neurological Care of CNY, Liverpool, NY, 13088, USA. [50]Parkwood Hospital, Pennsylvania, USA. [51]University of Wisconsin, Wisconsin, USA. [52]University of California, Irvine – BIC, California, USA. [53]Banner Alzheimer's Institute, Phoenix, 85006, Arizona, USA. [54]Dent Neurologic Institute, NY, USA. [55]Ohio State University, Ohio, USA. [56]Albany Medical College, NY, USA. [57]Hartford Hospital, Olin Neuropsychiatry Research Center, Connecticut, USA. [58]Dartmouth-Hitchcock Medical Center, New Hampshire, USA. [59]Wake Forest University Health Sciences, North Carolina, USA. [60]Rhode Island Hospital, state of Rhode Island, Providence, RI, 02903, USA. [61]Butler Hospital, Providence, Rhode Island, USA. [62]University of California, San Francisco, USA. [63]Medical University South Carolina, Charleston, SC, 29425, USA. [64]Nathan Kline Institute, Orangeburg, New York, USA. [65]Cornell University, Ithaca, New York, USA. [66]USF Health Byrd Alzheimer's Institute, University of South Florida, Tampa, FL, 33613, USA.