Neuroinformatics
https://doi.org/10.1007/s12021-018-9370-4

ORIGINAL ARTICLE

Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis

Manhua Liu [1,2] · Danni Cheng [1] · Kundong Wang [1] · Yaping Wang [3] · the Alzheimer's Disease Neuroimaging Initiative

Abstract

Accurate and early diagnosis of Alzheimer's disease (AD) plays important role for patient care and development of future treatment. Structural and functional neuroimages, such as magnetic resonance images (MRI) and positron emission tomography (PET), are providing powerful imaging modalities to help understand the anatomical and functional neural changes related to AD. In recent years, machine learning methods have been widely studied on analysis of multi-modality neuroimages for quantitative evaluation and computer-aided-diagnosis (CAD) of AD. Most existing methods extract the hand-craft imaging features after image preprocessing such as registration and segmentation, and then train a classifier to distinguish AD subjects from other groups. This paper proposes to construct cascaded convolutional neural networks (CNNs) to learn the multi-level and multimodal features of MRI and PET brain images for AD classification. First, multiple deep 3D-CNNs are constructed on different local image patches to transform the local brain image into more compact high-level features. Then, an upper high-level 2D-CNN followed by softmax layer is cascaded to ensemble the high-level features learned from the multi-modality and generate the latent multimodal correlation features of the corresponding image patches for classification task. Finally, these learned features are combined by a fully connected layer followed by softmax layer for AD classification. The proposed method can automatically learn the generic multi-level and multimodal features from multiple imaging modalities for classification, which are robust to the scale and rotation variations to some extent. No image segmentation and rigid registration are required in pre-processing the brain images. Our method is evaluated on the baseline MRI and PET images of 397 subjects including 93 AD patients, 204 mild cognitive impairment (MCI, 76 pMCI +128 sMCI) and 100 normal controls (NC) from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Experimental results show that the proposed method achieves an accuracy of 93.26% for classification of AD vs. NC and 82.95% for classification pMCI vs. NC, demonstrating the promising classification performance.

Keywords Alzheimer's disease diagnosis · Multi-modality brain images · Convolutional neural networks (CNNs) · Cascaded CNNs · Image classification

Alzheimer's Disease Neuroimaging Initiative (ADNI) is a Group/Institutional Author.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.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.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

✉ Manhua Liu
mhliu@sjtu.edu.cn

✉ Yaping Wang
ieypwang@zzu.edu.cn

[1] Department of Instrument Science and Engineering, School of EIEE, Shanghai Jiao Tong University, Shanghai 200240, China

[2] Shanghai Engineering Research Center for Intelligent Diagnosis and Treatment Instrument, Shanghai Jiao Tong University, Shanghai 200240, China

[3] School of Information Engineering, Zhengzhou University, Zhengzhou, China

Published online: 23 March 2018

Introduction

Alzheimer's disease (AD) is an irreversible brain degenerative disorder with progressive impairment of the memory and cognitive functions. Mild cognitive impairment (MCI) is a transitional state from normal control (NC) to dementia and it is often considered as a clinical precursor of AD when it is associated with memory loss (Krizhevsky et al. 2012; Minati et al. 2009). Currently there is no cure for AD, but it is of great interest for developing treatments to delay its progression, especially if AD can be diagnosed at an early stage where those treatments would have the most impact. Thus, accurate and early diagnosis of AD/MCI is important for patient care and future treatment. But it is still a challenging problem for accurate and early diagnosis of AD/MCI in clinic. Multi-modality neuroimages such as magnetic resonance images (MRI) and positron emission tomography (PET) are providing powerful imaging information to help understand the anatomical and functional neural changes related to AD. In recent years, extensive studies have been done to find the biomarkers and develop computer-aided system using the pattern recognition methods for AD diagnosis with the different types of neuroimaging modalities (Cheng et al. 2015; Suk et al. 2014, 2015; Suk and Shen 2013; Zhang et al. 2011).

MRI is a non-invasive medical imaging modality used for imaging the internal body structures. It uses a powerful magnetic field and radio frequency pulses to produce detailed images of organs, soft tissues, bone and virtually all other internal body structures. Currently, MRI is the most sensitive imaging test of the brain in routine clinical practice. MRI scans are specially used in brain imaging where it provides information about the morphology of the white matter, gray matter and cerebrospinal fluid (CSF). Structural MRIs are often used to non-invasively capture the regional brain atrophy and help understand the brain anatomical changes. Thus, they are recognized as an important biomarker for AD progression and are widely studied with pattern recognition methods for AD diagnosis (Hinrichs et al. 2009; Hosseini-Asl et al. 2016; Kloppel et al. 2008, epub). The raw brain images are too large and noisy to be directly used for classification. For morphological analysis of brain images, multiple anatomical regions, i.e., regions of interest (ROIs), were produced by grouping voxels through the warping of a labeled atlas and the regional measurements are computed as the features for image classification (Liu et al. 2015; Zhang et al. 2011). To capture the rich image information, voxel-wise features were extracted after registering all brain images to associate each voxel with a vector of scalar measurements for AD diagnosis (Ishii et al. 2005; Kloppel et al. 2008, epub). The brain volume is segmented to gray matter (GM), white matter (WM), and CSF parts, and the voxel-wise tissue density maps are computed for classification. Lerch et al., (Lerch et al. 2008) proposed to extract the cortical thickness features by calculating the distances between corresponding points at the WM and GM surfaces. Gerardin et al., (Gerardin et al. 2009) segmented and spatially aligned the hippocampus regions and modeled their shape with a series of spherical harmonics to quantify the hippocampus shape for AD classification.

In addition to MRI, Positrons Emission Tomography (PET) is a functional medical imaging modality which can help physicians to diagnose AD. A positron-emitting radionuclide (tracer) with a biologically active molecule, such as (18)F-fluorodeoxy-glucose ((18)FDG), is introduced in the body. Concentrations of this tracer are imaged using a camera and indicate tissue metabolic activity by virtue of the regional glucose uptake (Silveira and Marques 2010). There are some studies on prediction of the standard-dose PET image from low-dose PET and multimodal MR images (Wang et al. 2016; Yan et al. 2017). Fluorodeoxiglucose positron emission tomography (FDG-PET) provides a powerful functional imaging biomarker to help understand the neural changes for AD diagnosis. In recent years, various pattern recognition methods have been investigated in analysis of PET brain images to identify the patterns related to AD and decode the disease states for computer-aided-diagnosis (CAD). A region based method was proposed to extract features for classification of AD on PET images (Silveira and Marques 2010). In this method, brain images are mapped into 116 anatomical regions of interest (ROIs) and the first four moments and the entropy of the histograms of these regions are computed as the regional features. Receiver Operating Characteristics curves are then used to rank the discriminability of ROIs to distinguish PET brain images and the features from top 21 regions are input to both support vector machine (SVM) and random forest classifiers for AD classification. In (Lu et al. 2015), 286 features were extracted from 116 cerebral anatomical volumes of interest (VOIs) based on the automated anatomical labeling (AAL) cortical parcellation map, and a semi-supervised method was proposed to integrate the labeled and unlabeled data by random manifold learning with affinity regularization for AD classification. To capture the rich image information, the voxel-wise intensity features were extracted after preprocessing PET images, including co-registration to their baseline PET scan, reorientation into a standard space, voxel intensity normalization and smoothing with a 8 mm FWHM Gaussian filter for AD classification. In (Silveira and Marques 2010), a boosting method was proposed for classification of PET images based on a mixture of simple classifiers, which performs feature selection concurrently with classification to solve high dimensional problem. A favorite class ensemble of classifiers was proposed with each base classifier using a different feature subset which is optimized for a given class (Cabral et al. 2013).

Multi-modality images including MRI and PET are providing powerful imaging modalities to help understand the anatomical and neural changes related to AD (Alberdi et al. 2016;

Li et al. 2014; Liu et al. 2015; Zhang et al. 2011). The changes of multi-modality biomarkers may provide complementary information for the diagnosis and prognosis of AD and recent studies show that the combination of multimodal features can improve the classification performance. Zhang et al., (Zhang et al. 2011) proposed a multi-kernel SVM to ensemble the multimodal features such as tissue volumes extracted from 93 ROIs, intensity values of PET images and CSF biomarkers for disease classification. Recently, deep learning networks were also used to extract the latent features from measurements of ROIs with different image modalities for AD classification (Liu et al. 2015; Suk et al. 2015). Liu et al., (Liu et al. 2015) extracted a set of latent features from 83 ROIs of MRI and PET scans and trained a multi-layered neural network consisting of several auto-encoders to combine multimodal features for classification. Suk et al., (Suk et al. 2015) used a stacked Autoencoder to learn the latent high-level features separately from the multimodal ROI features as those in (Zhang et al. 2011) and a multi-kernel SVM was used to combine these features to improve the classification performance. The voxel-wise features of MRI and PET neuroimages including the GM density map o and the PET intensities are combined with a sparse regression classifier, and a deep learning based framework was proposed for estimating missing PET images for multimodal classification of AD (Li et al. 2014).

Previous studies have shown that it is not only important but also challenging to extract the representative biomarkers for image classification. Although promising results have been reported by investigating various pattern recognition methods for multimodal neuroimage analysis, there are still some limitations in the above feature extraction methods. The ROI-based feature extraction can significantly reduce the feature dimension and provide robust representations, but some minute abnormal changes may be ignored. In addition, the ROIs are generated by prior hypotheses and the abnormal brain regions relevant to AD might not well fit to the predefined ROIs, thus limiting the representation power of extracted features. The voxel-wise features, such as gray matter density map, can alleviate this problem, but they are of huge dimensionality, far more features than training subjects, which may lead to low classification performance due to the 'curse of dimensionality'. The cortical thickness and hippocampus shape features neglect the correlated variations of the whole brain structure affected by AD in other ROIs, e.g., the ventricle's volume. In addition, extraction of these handcrafted features highly depends on image preprocessing steps such as segmentation and registration, which often require the domain expert knowledge.

Recently, deep learning methods have gained a good reputation especially to extract informative features for computer vision and medical image analysis (Shen et al. 2017). Usually, the features learned via deep learning have better representations of the data than the handcrafted features. Instead of extracting features based on the expert's knowledge about the target domain, deep learning can discover the discriminant representations inherent in data by incorporating the feature extraction into the task learning process. Thus, it can be used by nonexperts for their researches and/or applications, especially in medical image analysis (Shen et al. 2017). In addition, deep learning can construct multi-layer neural networks to transform image data to task outputs (e.g., disease/normal) while learning hierarchical features from data such that high-level features can be derived from low-level features. Thus, complex patterns can be discovered with deep learning. Convolutional neural networks (CNNs) have been explored to learn the generic features of neuroimages for AD diagnosis (Adrien 2015; Hosseini-Asl et al. 2016). Hosseini-Asl et al., (Hosseini-Asl et al. 2016) proposed a deep 3D-CNNs method to learn discriminant features and predict AD using the structural MRI scans. In this method, the deep 3D-CNNs were built upon a 3D-CAES (convolutional Autoencoders) pre-trained with the rigidly registered training images to capture anatomical shape variations, followed by fully connected network for classification. Adrien et al., (Adrien 2015) proposed a deep learning based classification algorithm for AD diagnosis using both structural and functional MRI. In this method, the CNN model was built with one convolutional layer trained with sparse Autoencoder, which was explored to extract the imaging features for AD classification. The above methods can learn generic features capturing AD biomarkers via convolutional network. However, all of them require the convolutional filters pretrained on Autoencoder with carefully preprocessed data to extract features and then classify them for task-specific target. In addition, the above methods focused on the AD diagnosis from MRI. In the case of multimodal neuroimages, further investigation is still needed to determine their ability for AD diagnosis. The two-stream CNNs have been used to learn different types of features which are combined for image classification (Lin et al. 2015; Weinzaepfel et al. 2015). In (Weinzaepfel et al. 2015), a spatial-CNN was trained to capture the static appearance of the actor and the scene using the RGB image, while a motion-CNN took the optical flow as input to capture the motion pattern. These two-stream CNN features are concatenated to discriminate the actions against background regions for action localization in realistic videos. In (Kloppel et al. 2008, epub), a bilinear model has been proposed to multiply the outputs of two CNNs using outer product at each location of the image and pooled to obtain a bilinear vector for image classification.

Motivated by the success of CNN in image classification, this paper proposes a novel multi-modality classification algorithm based on cascaded CNNs model to learn and combine the multi-level and multi-modality features of MRI and PET images for AD diagnosis. First, a number of local 3D patches are uniformly extracted from the whole brain image, and a deep 3D CNN is built to hierarchically and gradually transform each image patch into more compact discriminative features. Second, an upper high-level 2D CNN is cascaded to ensemble the features learned by deep 3D CNNs from multimodality and generate the high-level features. Finally, these learned features are combined by fully connected layers followed by softmax prediction layer for image classification in AD diagnosis. For training the whole deep learning networks, the multiple local 3D-CNNs for different patches are individually trained to generate the compact features, and the upper convolution and fully connected layers are fine-tuned to combine the features learned by multiple 3D-CNNs and multiple image modality for image classification. The proposed method can be used to learn the generic features from the imaging data and combine the multi-level and multimodality features for classification. Our experimental results on ADNI database demonstrate the effectiveness of the proposed method for AD diagnosis. Comparing to the methods based on two-stream CNNs (Lin et al. 2015; Weinzaepfel et al. 2015), our proposed method uses two-stream CNNs to capture the individual features of MRI and PET for each local patch. Different from these two methods that use the concatenation and bilinear model to combine the features of two-stream CNNs, our proposed method will apply the cascaded CNNs to integrate the multi-level and multi-modality features for classification.

The rest of this paper is organized as follows. Section 2 presents the materials of multimodal imaging data used in this paper and the proposed multimodal classification algorithm based on cascaded CNNs. Experiments and comparisons are provided in Section 3. Section 4 concludes this paper.

Material and Method

In this section, we will present the multimodal brain image sets that used in this work and the proposed classification algorithm in detail. The MRI and PET images are powerful imaging modalities which are often used as biomarkers to help physicians for AD diagnosis. It is still challenging to make use of the high-dimensional and multi-modality image data. In this work, a cascaded deep CNNs model has been proposed to hierarchically learn and combine the multi-level and multimodality features for AD diagnosis using MRI and PET images. There are 3 main advantages to apply the cascaded deep CNNs to our task. First, the deep architecture of CNNs can gradually extract the features from the low-, mid- to high-levels with a high volume of training images. Second, they can explicitly make use of the spatial structure of brain images and learn local spatial filters useful for the classification task. Finally, cascading multiple CNNs can construct a hierarchy of more complex multi-modality features representing larger spatial regions, finally providing a global label. The features are robust to some variations such as scales and rotation to some extent. The proposed algorithm can integrate the feature extraction and classification processes by deep convolutional learning and data-driven methods without the domain expert knowledge. Figure 1 shows the flowchart of the proposed classification algorithm based on cascaded CNNs, which consists of 3 main steps: image preprocessing, feature learning by building multi-CNNs and final multimodal classification by cascading CNNs, as detailed below.

Data Sets

All data sets used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, which are publicly available in the website (www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of the ADNI was to test whether serial magnetic resonance imaging (MRI), Positron Emission Tomography (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). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The principal investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California, San Francisco. ADNI was the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. The study subjects were recruited from over 50 sites across the U.S. and Canada and gave written informed consent at the time of enrollment for imaging and genetic sample collection and completed questionnaires approved by each participating sites Institutional Review Board (IRB).

In this work, we used the T1-weighted magnetic resonance (MR) imaging data and the 18-Fluoro-DeoxyGlucose PET (FDG-PET) imaging data from the baseline visit for evaluation. These imaging data are acquired from 397 ADNI participants including 93 AD, 204 MCI (76 MCI converters (pMCI) and 128 MCI non-converters (sMCI)), 100 NC. Table 1 presents the demographic details of the studied subjects in this paper, where CDR denotes the Clinical Dementia Rating (CDR).

In ADNI, MRI and PET image acquisition had been done according to the ADNI acquisition protocol in (Jack Jr. et al. 2008). We used the structural MR images acquired from 1.5 T scanners. The T1-weighted MR images were acquired

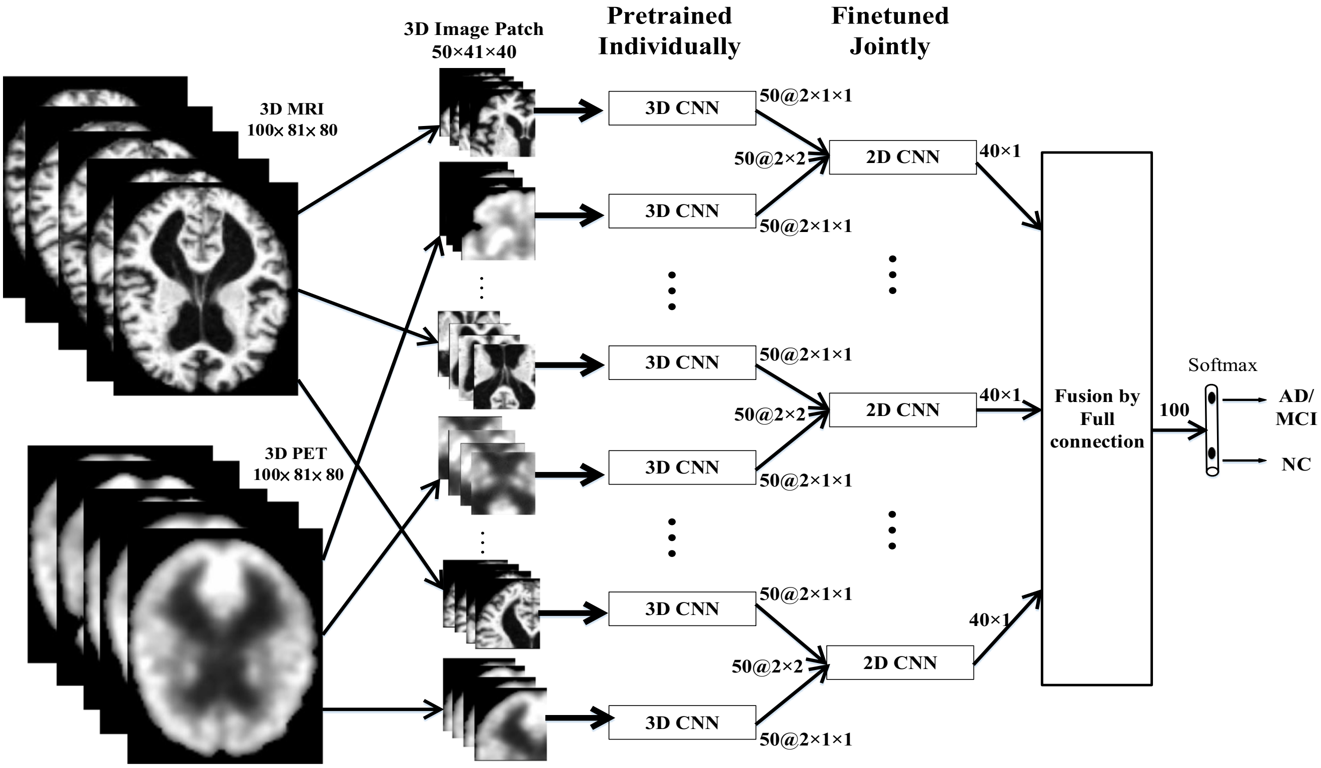


Fig. 1 The flowchart of the proposed classification algorithm, where "3D CNN" denotes the deep 3D CNN models for 3D local patches, "2D CNN" demotes the cascaded CNN model for fusion of multi-modality, "Fusion by full connection" denotes a fully connected layer for fusion of the learned high-level multimodality features, and "softmax" denotes the softmax prediction layer

sagittally using volumetric 3D MPRAGE with 1.25 × 1.25 mm^2 in-plane spatial resolution and 1.2 mm thick sagittal slices. The FDG-PET images were acquired 30–60 min post-injection, averaged, spatially aligned, interpolated to a standard voxel size, normalized in intensity, and smoothed to a common resolution of 8 mm full width at half maximum. More detailed information about MRI and FDG-PET acquisition procedures is available at the ADNI website.

The MRI and PET images were pre-processed to make the images from different systems more similar. All T1-weighted MR images were preprocessed by applying the typical procedures of Anterior Commissure (AC)–Posterior Commissure (PC) correction, skull-stripping, cerebellum removal and affine registration. Specifically, nonparametric nonuniform intensity normalization, N3 (Sled et al. 1998) was first applied to correct non-uniform tissue intensities, followed by skull stripping and cerebellum removal (Wang et al. 2011). After that, we manually checked the skull-stripped images to ensure the clean and dura removal. Affine registration was performed to register the MR images to a template with FSL (FMRIB Software Library) 5.0, which can be freely downloaded from the website https://fsl.fmrib.ox.ac.uk/. Regarding FDG-PET images, they were affine-registered to the respective MR images with the intensity normalization and conversion to a uniform isotropic resolution of 8 mm FWHM as in (Silveira and Marques 2010). No image segmentation and rigid registration are required in pre-processing the brain images. For consistency, all images were resampled to size 256 × 256 × 256 and resolution $1 \times 1 \times 1$ mm^3. The images are further down-sampled to 128 × 128 × 128 voxels. The voxels outside the brain are removed from the image analysis and the MRI and PET images finally used are of size 100 × 81 × 80 voxels. The voxel intensities of each MRI and PET image are used

Table 1 Demographic characteristics of the studied subjects from ADNI database. The values are denoted as mean ± standard deviation

Diagnosis	Number	Age	Gender (M/F)	MMSE	Education	CDR
AD	93	75.49 ± 7.4	36/57	23.45 ± 2.1	14.66 ± 3.2	0.8 ± 0.25
MCI	204	74.97 ± 7.2	68/136	27.18 ± 1.7	15.75 ± 2.9	0.5 ± 0.03
NC	100	75.93 ± 4.8	39/61	28.93 ± 1.1	15.83 ± 3.2	0 ± 0

for classification. The whole brain images are simply divided into 3 × 3 × 3 parts to extract 27 patches of size 50 × 41 × 40 voxels. Each patch has half overlaps with its neighbor in every direction.

Feature Learning with 3D CNNs

Different from the conventional methods which explicitly extract the handcrafted features of brain images, such as tissue density map, cortical thickness, hippocampus shape and volumes, deep CNNs are used to learn the generic features from the multi-modality brain images. CNNs are a special kind of multi-layer neural networks, which are trained with the back-propagation algorithm. CNN has been widely used in several domains such as image classification and object detection (He et al. 2015; Krizhevsky et al. 2012; Lécun et al. 1998). Convolutional neural networks are designed to recognize visual patterns directly from the images with minor preprocessing. Most of the mature CNN architectures are designed for 2D image recognition. To adapt to 3D brain image, the volumetric data is split along the third dimension into 2D image slices for training 2D CNN. However, this scheme is inefficient to encode the spatial information of 3D image due to the absence of kernel sharing across the third dimension. To efficiently encode the richer spatial information of 3D brain images, the 3D convolution kernel is used in this work. The deep 3D CNN is built by alternatively stacking convolutional and sub-sampling layers to hierarchically learn the multi-level features of multi-modality brain images, which is followed by the fully connected and softmax layers for image classification, as shown in Fig. 2.

A typical convolutional layer convolves the input image with the learned kernel filters, followed by adding a bias term and applying a non-linear activation function, and finally produce a feature map by each filter. More formally, the 3D convolutional operation is defined as:

$$u^l_{kj}(x,y,z) = \sum_{\delta_x}\sum_{\delta_y}\sum_{\delta_z} F^{l-1}_k\left(x+\delta_x, y+\delta_y, z+\delta_z\right) \times W^l_{kj}\left(\delta_x,\delta_y,\delta_z\right) \tag{1}$$

where x, y and z denote the pixel positions for a given 3D image, $W^l_{kj}\left(\delta_x,\delta_y,\delta_z\right)$ is the *jth* 3D kernel weight connecting the *kth* feature map of the *l-1* layer with the *jth* feature map of the l layer, F^{l-1}_k is the kth feature map of the previous *l-1* layer, and δ_x, δ_y, δ_z are the kernel sizes corresponding to the x, y and z coordinates. The output $u^l_{kj}(\mathrm{x,y,z})$ is the convolutional response of 3D kernel filter. After convolution, *tanh* is adopted as the activation function following each convolution layer:

$$F^l_j(x,y,z) = \tanh\left(\mathrm{b}^l_j + \sum_k u^l_{kj}(x,y,z)\right) \tag{2}$$

where b^l_j is the bias term for the jth feature map of the l layer. The jth 3D feature map of l layer $F^l_j(\mathrm{x,y,z})$ is obtained by summation of the response maps of different convolution kernels. By using 3D kernel to capture the spatial correlations, the CNNs can take full advantages of the volumetric contextual information.

After each convolutional layer, there usually has a pooling layer. There are several ways for pooling, such as taking the average value or the maximum, or a learned linear combination of the neurons in the cube. In our work, max pooling, i.e., the maximum of the pooling cube, is used to obtain more compact and efficient features as in (Hosseini-Asl et al. 2016). Max pooling reduces the feature map along the spatial dimensions by replacing each cube with their maximum. It can keep the most influential features for distinguishing images. Through max pooling, the features become more compact from low to high layers, which can achieve the robustness to some variations.

The third type of layer is the fully connected layer. After alternatively stacking several convolutional and max pooling layers, the high-level reasoning in the deep CNN is done by fully connected layers. All 3D feature maps are flattened into a 1D vector as the inputs of fully connected layer. A fully connected layer consists of a number of output neurons, which generate the learned linear combination of all neurons from the previous layer and passed through a nonlinearity. The inputs and outputs of fully connected layers are 1D vector and not spatially located anymore.

Finally, a softmax classification layer is appended to the last fully connected layer and is fine-tuned by back-propagation with negative log-likelihood to predict class probability. The softmax function is a derivation of logistic function that highlights the largest values in a vector while

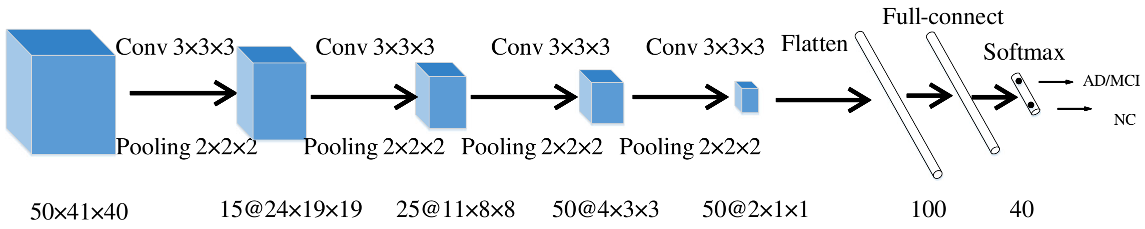


Fig. 2 The architecture of deep 3D CNNs denoted with the sizes of input, convolution, max pooling and output layers and the numbers and sizes of generated feature maps

suppressing those significantly below the maximum. The outputs of softmax layer, ranging from 0 to 1, can be interpreted as class prediction probabilities, and the sum of its outputs is equal to 1.

For brain image analysis, one direct way is to take the whole image as the input and construct a deep 3D CNNs of large depth for feature learning and classification. However, this may require training a large number of parameters which easily results in overfittings and needs large training dataset. Acquisition of large dataset is challenging for multi-modality brain images. In addition, training a deep CNNs model for the whole brain image not only requires high computation and memory costs especially for the high-resolution images but also is aimless for analyzing the local features related to AD. To avoid these problems, this paper proposes to uniformly partition the whole brain image into many local patches with some overlap and construct a deep 3D CNNs for each patch to learn the local features. Therefore, the learned features are more suitable for extracting the subtle local patterns of the high-dimensional brain images. The global classification will be performed by ensemble of these multiple deep 3D CNNs trained for the local patches.

The same network architecture is used to build all deep 3D CNNs, as illustrated in Fig. 2. In our implementation, each deep CNN is built by stacking 4 convolutional layers, 3 max pooling layers, a fully connected layers and a softmax layer. The sizes of all convolutional filters are set to 3 × 3 × 3 and the numbers of filters are set to 15, 25, 50, 50 for 4 convolution layers, respectively. Max pooling is applied for each 2 × 2 × 2 region. *Tanh* function is adopted as the activation function in these layers. The 3D convolutional kernels are randomly initialized in the Gaussian distribution. The other trainable parameters of the networks are tuned using the standard backpropagation with stochastic gradient descent by minimizing the loss of cross entropy. In addition, the dropout strategy is employed to reduce the co-adaption of intermediate features and overfitting problem, and improve the generalization capability.

Multi-Modality Cascaded CNNs for AD Diagnosis

To combine the multi-modality brain images, one direct method is to concatenate the learned features by deep CNNs from all local patches of multi-modality images and design a classifier to make the final classification. However, this method cannot make use of the correlation information between MRI and PET brain images. Thus, we propose to build the cascaded CNNs to combine the learned multimodal features of the corresponding local patches as shown in Fig. 3. The learned features of the local patches from the multi-modality images at the same position will be combined by cascading high-level 2D CNNs to further learn the features associated to both modalities. Finally, the learned high-level features containing both the intrinsic properties of each modality and the correlations between different modalities are combined with a fully connected layer followed by a softmax layer to predict the final global classification. While the lower layers of the predictive 3D-CNNs extract discriminative image features, the upper fully connected and softmax layers have to facilitate the task-specific classification with these features. Thus, training of the proposed multimodal classification consists of pretraining individual 3D CNNs, and final task-specific finetuning for 2D CNN networks and the ensemble classification.

Initially, a deep 3D CNN is pre-trained for each local patch in each modality by directly mapping the outputs of the fully connected layer to the probabilistic scores of all class labels with softmax function. Then, the initial-trained parameters of 3D CNNs are used to fix the first 3 convolution and pooling layers of the 3D CNNs, while the parameters of the multimodal 2D CNNs are fine-tuned jointly with the upper fully connected and softmax prediction layers. Finally, in the ensemble learning process, the initial-trained parameters of 3D CNNs and 2D CNNs are fixed, while the parameters of the upper fully connected layers and softmax prediction layer are finetuned jointly to combine the features for the task-specific classification. The training iteration ends when the validation error rate stops decreasing. The jointly fine training is included to

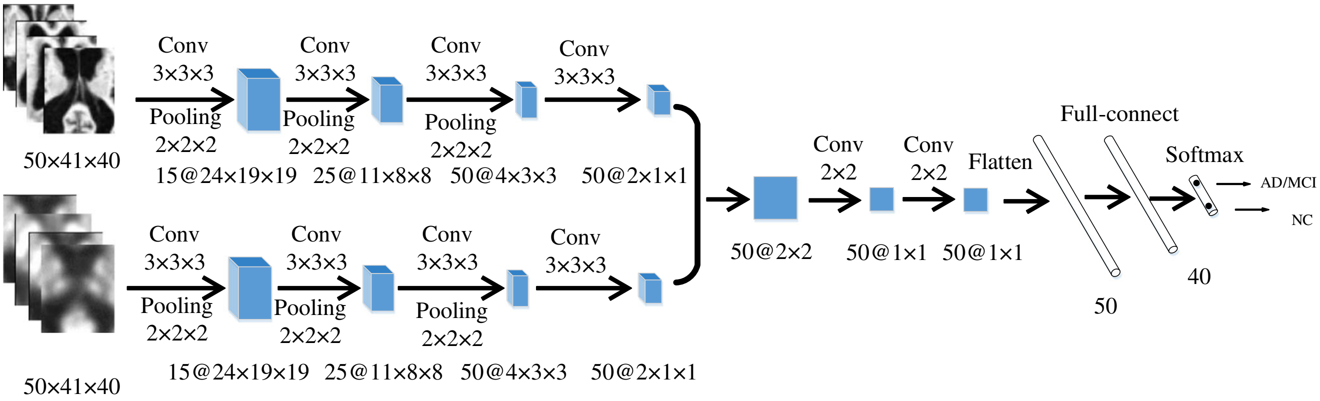


Fig. 3 The network architecture of multi-modality cascaded CNNs

adjust the parameters to handle the heterogeneity and produce a more reliable estimate from the multimodal images. However, training deep CNNs is challenged by the risk of over-fitting as current datasets for AD diagnosis are relatively small compared to other computer vision tasks such as face recognition. A common practice is to initialize the weights with the pre-trained models on some large dataset. To mitigate the problem, we adopt dropout layer for regularization (Srivastava et al. 2014). Data augmentation by shift and downsample is also employed to cover the diversity and variability of training samples.

Experimental Results

Experiments

In this section, experiments are performed to test the proposed multi-modality classification algorithm on prediction of the disease status AD, pMCI and sMCI from normal controls (NC). There are MRI and PET multi-modality images of 397 subjects including 93 AD, 76 pMCI, 128 sMCI and 100 NC subjects from ADNI database for our experiments. The proposed algorithm is tested on classifications of AD vs. NC, pMCI vs. NC and sMCI vs. NC. Ten-fold cross-validation is used to avoid random factors affecting the results. Each time, one fold of the image set is used for testing, another one fold used for validation while the left eight folds were used for training. The validation part is used for early stopping the training process to obtain the model weights with the optimized performance. To increase training data, augmentation is conducted by subsampling the brain image of 256 × 256 × 256 voxels in 8 shift ways to generate additional images of 128 × 128 × 128 for the training set. Augmentation is not conducted on the validation and test sets.

The proposed classification algorithm is implemented with the Keras library in Python based on Theano. The experiments are conducted on PC with GPU NVIDIA GTX1080. In the low level, 27 deep CNNs are independently trained to extract the local features with the output of the prediction scores for disease classification. The Adadelta gradient descent algorithm (Zeiler 2012) is used to train the local deep CNNs. To avoid overfitting problem, dropout, L1 and L2 regulation are adopted in our network (Srivastava et al. 2014). The batch size is set to 64, and the model begins to converge after 15~30 epochs. The transfer learning is also used in our experiments to alleviate the problems of limited training data. We initially train the local 3D CNN models for classification of AD vs. NC. The trained CNN model on AD vs. NC is used to initialize the parameters of the 3D CNN model of pMCI vs. NC classification to reduce the training time and improve the classification performance. Similarly, the trained CNN model on pMCI vs. NC is also used to initialize the parameters of the 3D CNN model for sMCI vs. NC classification. To evaluate the classification performance, we demonstrate the receiver operating characteristic (ROC) curves and compute the classification accuracy (ACC), the sensitivity (SEN), the specificity (SPE) and the area under ROC (AUC) for comparison in the experiments.

Table 2 Comparison of classification performances on AD vs. NC

AD vs. NC	ACC%	SEN%	SPE%	AUC%
MRI	84.97	82.65	87.37	90.63
PET	88.08	90.70	85.98	94.51
Multi-modality	93.26	92.55	93.94	95.68

Table 3 Comparison of classification performances on pMCI vs. NC

pMCI vs. NC	ACC%	SEN%	SPE%	AUC%
MRI	77.84	76.81	78.50	82.72
PET	78.41	77.94	78.70	85.96
Multi-modality	82.95	81.08	84.31	88.43

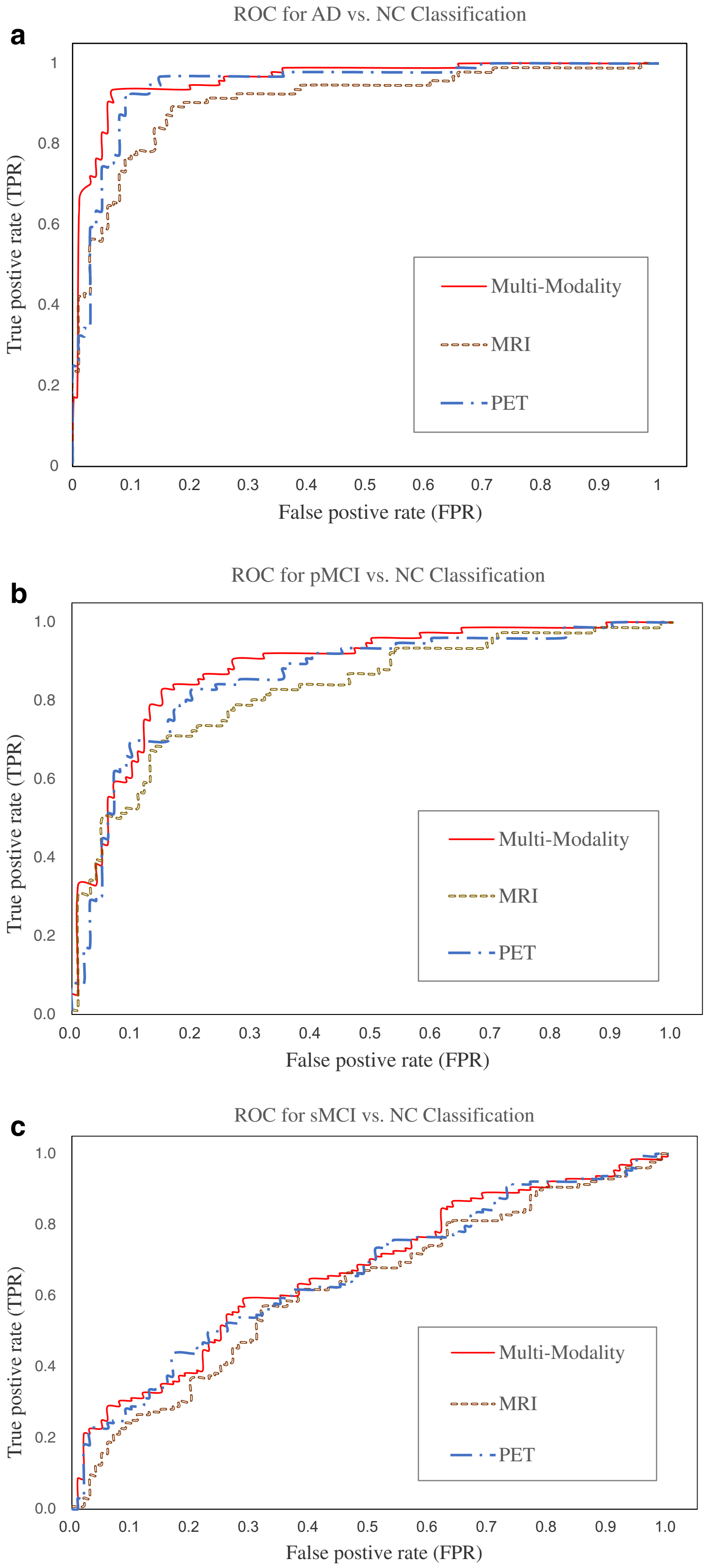
Classification Results on MRI, PET and Multi-Modality

The first experiment is to test the proposed classification algorithm in prediction of AD, pMCI and sMCI from NC, based on MRI and PET biomarkers and multi-modality of 397 subjects in ADNI. We compare the results of the proposed method based on the different modalities. For the single modality such as MRI or PET, the features learned by 27 3D CNNs are combined by fully connected layers for final classification. Tables 2, 3, and 4 show the comparisons of their classification performances for AD vs. NC, pMCI vs. NC and sMCI vs. NC, respectively. Figure 4 (a), (b) and (c) also compare the ROC (receiver operating characteristic) curves of different modalities for classification of AD vs. NC, pMCI vs. NC and sMCI vs. NC, respectively. From these results, the multi-modality performs better than each individual modality. The results also show that the performances of PET by CNN are better than those of MRI, which are different from some existing methods (Li et al. 2014; Suk et al. 2014; Zhang et al. 2011). This may be due to two important factors. One is that MRI can capture structural information of brain regions, which contains some variations for different subjects, while PET imaging consists

Table 4 Comparison of classification performances on sMCI vs. NC

sMCI vs. NC	ACC%	SEN%	SPE%	AUC%
MRI	60.09	65.29	54.21	62.38
PET	63.35	63.84	65.59	66.62
Multi-modality	64.04	63.07	67.31	67.05

Fig. 4 ROC curves of MRI, PET and Multi-modality for classifications of (**a**) AD vs. NC; (**b**) pMCI vs. NC; (**c**) sMCI vs. NC



of injecting radiotracer that contains a positron emitter to patients, detecting the emitted radiation by a scanner and computing a digital image that represents the distribution of radiotracer in the body. There are few structural information in

Table 5 Comparison of classification performances on different multimodal combination methods

Methods	AD vs. NC (%)				pMCI vs. NC (%)				sMCI vs. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Averaging	90.16	93.02	87.85	94.81	78.98	74.68	82.47	87.55	61.40	61.63	60.71	64.34
Full Connection	91.19	91.30	91.09	95.02	78.98	78.26	79.44	85.71	60.53	62.03	57.14	63.12
Concatenate (Weinzaepfel et al. 2015)	89.64	90.11	89.22	94.87	80.68	77.63	83.00	87.45	59.21	59.16	59.46	65.43
Bilinear (Lin et al. 2015)	90.16	90.22	90.10	94.73	81.82	78.95	84.00	87.78	62.28	61.80	64.00	65.48
Cascaded CNNs	93.26	92.55	93.94	95.68	82.95	81.08	84.31	88.43	64.04	63.07	67.31	67.05

PET images. The structural variations of MRI will bring the difficulties in consistent extraction of discriminant features and degrade the classification performance. Another important factor is that no rigid registration is performed for both the MRI and PET brain images used in this work, which will cause more local variations for MRI than for PET.

Comparison of Different Combination Methods

The second experiment is to compare the proposed multimodal classification algorithm by the cascaded CNNs to other multimodal combination methods. One combination method is to directly average the prediction scores of the low-level 3D CNNs from all multimodal image patches. Another combination method is to apply the fully connected layer to combine the learned features of low-level 3D CNNs from all multimodal image patches, followed by a softmax layer for classification. In addition, we also compare our method with the concatenating method published in (Weinzaepfel et al. 2015) and the bilinear published in (Lin et al. 2015). We implement these two methods with our best efforts. To implement the method in (Weinzaepfel et al. 2015), we concatenate the features of all multimodal 3D CNNs into a feature vector as the input to a softmax classification layer. We also implement the bilinear model in (Lin et al. 2015) by multiplying the outputs of two multimodal 3D CNNs using outer product at each local patch to obtain the bilinear vector. The vector is then passed through a softmax layer to obtain classification predictions. Table 5 shows the comparison of their classification results. From the results, we can see that the proposed cascaded CNNs method performs better than other combination methods. The cascaded CNNs can capture the high-level multimodal correlation features which can further improve the classification performance.

Comparison with Existing Methods

In this section, we will compare our proposed method with those published in the literature. First, we compare our proposed method based on deep 3D CNNs to the method based on 3D convolutional Autoencoder (Hosseini-Asl et al. 2016), which was proposed to learn the imaging features of structural MRI scans with stacked 3D Convolutional Autoencoder (CAE) for prediction of AD. Different from our method, the 3D convolutional Autoencoder method extracts the features of a 3D brain image based on reconstructing the input. Thus, training of the Autoencoder uses back-propagation and constraints on the properties of feature space to reduce the reconstruction error. For fair comparison, we downloaded the source codes released by the authors of (Hosseini-Asl et al. 2016) in the website and implemented it with our best effort by using each single modality and multi-modality of our data set. The same training and test sets are used in the experiments. Table 6 shows the comparison of classification results by our proposed 3D CNN method and the 3D CAE method (Hosseini-Asl et al. 2016) using the MRI, PET and multimodality images for classification of AD and NC. We can see that our proposed method performs better than the 3D CAE method.

Furthermore, we compare the proposed multimodal classification method to other multimodal classification methods published in the literature (Li et al. 2014; Liu et al. 2015;

Table 6 Comparison of classification results by our method and the 3D CAE method for AD vs. NC classification

Method	Modality	ACC(%)	SEN(%)	SPE(%)	AUC(%)
Auto-Encoder (Hosseini-Asl et al. 2016),	MRI	81.87	81.00	82.80	87.09
	PET	84.97	84.95	85.00	91.34
	Muti-modality	87.56	81.72	93.00	93.90
Proposed CNN Method	MRI	84.97	82.65	87.37	90.63
	PET	88.08	90.70	85.98	94.51
	Muti-modality	93.26	92.55	93.94	95.68

Zhang et al. 2011). These methods are based on the handcrafted ROI or voxel-wise features of multi-modality. A multi-kernel SVM was proposed to combine the multi-modality features of multiple anatomical regions of interest, which were generated by grouping voxels through the warping of a labeled atlas, to improve the classification performance (Zhang et al. 2011). Liu et al., (Liu et al. 2015) extracted a set of latent MRI and PET features from 83 anatomical regions of interest and trained a multi-layered neural network consisting of several Autoencoder to combine multimodal features for classification. Instead of extracting ROI features, the voxel-wise features of multi-modality images including the GM density map of MRI and the intensity values of PET are combined with a sparse regression classifier, and a deep learning based framework was proposed for estimating missing PET imaging data for multimodal classification (Li et al. 2014). However, it is difficult to implement these published methods on the same settings for fair comparison. Thus, the results of these methods reported in the literature are directly used for comparison. Since these methods combines the pMCI and sMCI into MCI for classification, to compare with them, the pMCI and sMCI are also combined into one class for evaluation. Table 7 compares our results with the reported results of these methods that also used the multi-modality data of MRI and PET from ADNI. It is worth noting that the differences of the reported results may be due to the use of different feature extraction and classification methods for MRI and PET images, and also the use of different ADNI subjects. All these variations make the results comparison complicated. In addition, the differences in the size of test samples, the use of cross-validation to separate the training and testing sets can also make the fair comparison difficult to achieve. Compared to these methods, our proposed method requires less image preprocessing steps for feature extraction. No segmentation and rigid registration are required in our method, which can reduce the computation costs.

Discussion

Different from the traditional methods based on the handcrafted features, the proposed method built the cascaded deep CNNs to learn the multi-level and multimodal features for classification of brain images. Each 3D CNN layer combines the low-layer feature maps to generate higher-level features which can achieve more robustness to some variations of translation and rotation etc. in images. No segmentation and rigid registration are required in pre-processing the brain images. However, there are still some limitations in the proposed method. First, the parameters of the deep CNN model, such as the number of layers, the size and number of kernels in each layer, may not be optimally determined. Second, only MRI and PET modalities are used in this work, more information such as CSF can be included to further improve performance. Third, it is not easy to visualize the learned features by the proposed method for interpretation of the brain and neurodegenerative disease (i.e., AD or MCI) in the clinical application. The learned features have no sufficient clinical information to find the related ROIs for clinical understanding of the brain abnormalities.

However, there are some suggestions to address the above limitations. For setting the convolutional kernel size of deep CNN, large kernel size is effective to capture the large patterns but it may ignore the small ones. Thus, the size of all 3D kernels is set to $3 \times 3 \times 3$, but multiple convolutional layers are used to hierarchically capture the large patterns. The other optimal parameters can be obtained by cross validation in our experiments. More auxiliary information such as CSF and clinical information may be considered to improve the performance if they have high correlations to AD. Since AD has important relations to certain pathological patterns, only a subset of image regions is closely related to AD. For interpretation of diseases, we aim to identify the impact of local regions based on evaluation of classification prediction after individually excluding these areas from the whole image as in (Zeiler and Fergus 2014). To achieve this, we systematically exclude different local areas of brain images with a 3D grey box and monitor the classifier outputs. If the excluded patch covers the important area related to AD, the prediction probability of the correct class drops significantly. In our experiments, the important and network attention areas are identified for MRI and PET images with the AD subjects from a test subset as follows. First, the top 3 local patches are selected with the better classification performance for each modality.

Table 7 Comparison of the multimodal classification performances (%) reported in the literature

Method	Subjects	AD vs. NC (%)				MCI vs. NC (%)			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Liu et al. 2015	85 AD +168 MCI+ 77 NC	91.40	92.32	90.42	–	82.10	60.00	92.23	–
Li et al. 2014	93 AD +204 MCI 101 NC	–	–	–	89.82	–	–	–	70.14
Zhang et al. 2011	51 AD +52 HC	90.60	90.50	90.70	–	–	–	–	–
Our method	93 AD +204 MCI 100 NC	93.26	92.55	93.94	95.68	74.34	70.08	84.91	80.23

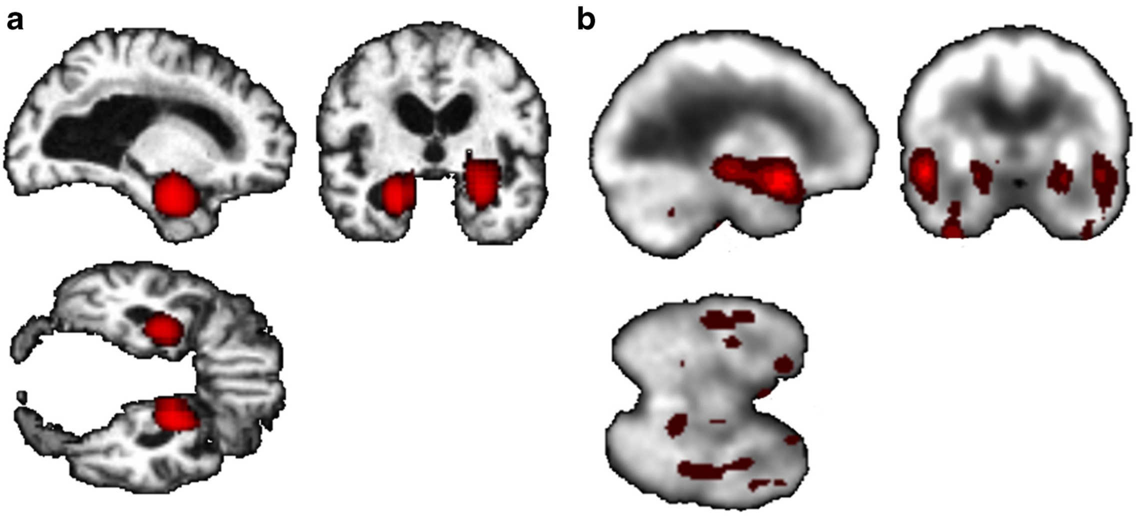


Fig. 5 The network attention areas generated by systematically excluding local patches of brain images and measuring the drop of correct class prediction probability for (**a**) MRI and (**b**) PET images

Then, the selected image patches are systematically obstructed with 15 × 15 × 15 grey box and a heatmap is generated by measuring the drop of the correct class prediction probability. This heatmap shows the importance and attention of the corresponding areas in prediction of disease status. Finally, the network attention areas are generated by compiling the prediction masks of heatmaps. The areas with the highest attention for MRI and PET images are demonstrated in Fig. 5 (a) and (b), respectively. To facilitate the interpretation, we also map the network attention areas to a template with 93 manually labeled ROIs for reference (Kabani et al. 1998). The top ROIs covered by these attention areas are shown in Fig. 6, which seem to be consistent with the ones that are most affected by AD, mainly hippocampus, Amygdata, and Parahippocampal gyrus etc. (Liu et al. 2015; Zhang et al. 2011).

Conclusion

Multimodal neuroimages can provide complementary information for the diagnosis and prognosis of AD. In this paper, we have presented a multimodal classification algorithm

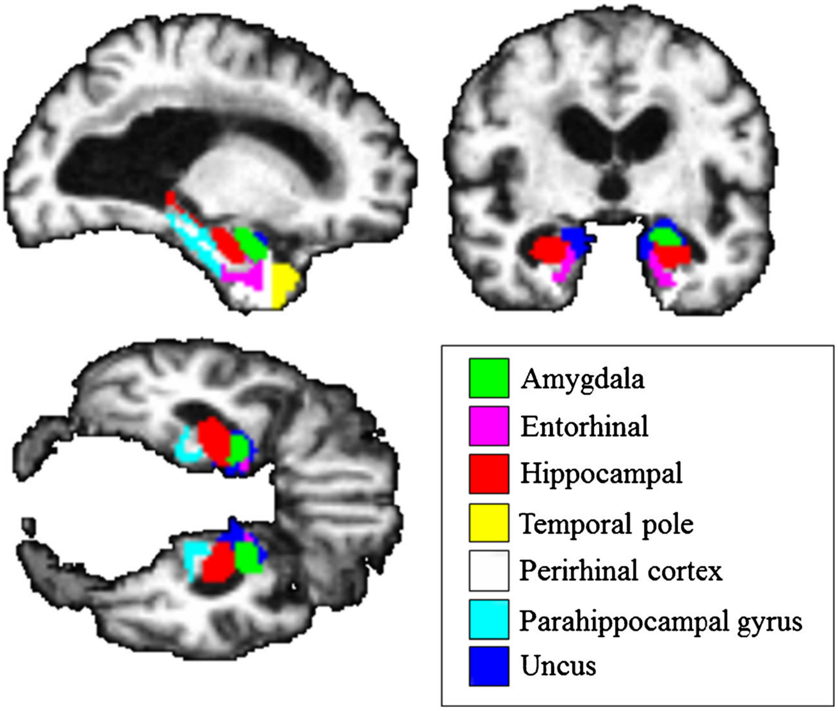


Fig. 6 The top affected brain regions of interests (ROIs) covered by the generated network attention areas

based on the cascaded CNNs to predict AD and MCI from normal controls using MRI and PET images. Multiple deep 3D-CNNs are built on different local image patches to learn the discriminative features of MRI and PET images. Then, a set of upper high-level CNNs are cascaded to combine the features learned from local CNNs and learn the latent multimodal features for image classification. The proposed algorithm can gradually and automatically learn the generic multilevel and multimodal features from multiple imaging modalities for disease classification. The proposed method requires no image segmentation and rigid registration in preprocessing the brain images, which can save computation costs. Our experimental results and comparison on ADNI database demonstrate the performance improvement of the proposed method for AD diagnosis.

Information Sharing Statement

The dataset used in preparation of this paper was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI, RRID:SCR_003007) database which is freely available at the website www.loni.ucla.edu. The ADNI investigators contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this paper. A complete list of ADNI investigators can be found at:http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. In this paper, the proposed method is implemented with the Keras library in Python based on Theano. The CNN code used in this paper can be obtained at https://github.com/Tristacheng/mutimodality_cnn/.

Acknowledgments This work was supported in part by National Natural Science Foundation of China (NSFC) under grants (No. 61375112, 61773263, U1504606), The National Key Research and Development Program of China (No.2016YFC0100903) and SMC Excellent Young Faculty program of SJTU. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (NIH Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co.,Medpace, Inc.,Merck and Co., Inc., Novartis AG, Pfizer Inc., F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles.

References

Adrien, P.A.G.M. (2015). Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. arXiv: 1502.02506 [cs.CV].

Alberdi, A., Aztiria, A., & Basarab, A. (2016). On the early diagnosis of Alzheimer's disease from multimodal signals: A survey. *Artificial Intelligence in Medicine, 71*, 1–29.

Cabral, C., Silveira, M., Neuroimaging, A.S.D. (2013). Classification of Alzheimer's disease from FDG-PET images using favourite class ensembles. 2013 35th Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc), pp. 2477–2480.

Cheng, B., Liu, M., Suk, H. I., Shen, D., & Zhang, D. (2015). Multimodal manifold-regularized transfer learning for MCI conversion prediction. *Brain Imaging and Behavior, 9*, 913–926.

Gerardin, E., Chetelat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H. S., Niethammer, M., Dubois, B., Lehericy, S., Garnero, L., Eustache, F., Colliot, O., & Initi, A.s. D. N. (2009). Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage, 47*, 1476–1486.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. pp. 770–778.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M. K., & Johnson, S. C. (2009). Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *NeuroImage, 48*, 138–149.

Hosseini-Asl, E., Keynton, R., & El-Baz, A. (2016). Alzheimer's disease diagnostics by adaptation of 3D convolutional network. 2016 I.E. International Conference on Image Processing (ICIP), pp 126–130.

Ishii, K., Kawachi, T., Sasaki, H., Kono, A. K., Fukuda, T., Kojima, Y., & Mori, E. (2005). Voxel-based morphometric comparison between early- and late-onset mild Alzheimer's disease and assessment of diagnostic performance of z score images. *AJNR American Journal of Neuroradiology, 26*(2), 333–340.

Jack Jr., C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L Whitwell, J., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C. S., Krueger, G., Ward, H. A., Metzger, G. J., Scott, K. T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J. P., Fleisher, A. S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., & Weiner, M. W. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: JMRI, 27*, 685–691.

Kabani, N., MacDonald, D., Holmes, C. J., & Evans, A. (1998). A 3D atlas of the human brain. *NeuroImage, 7*, S717.

Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr, C.R., Ashburner, J., & Frackowiak, R.S.J. (2008). Automatic classification of MR scans in Alzheimer's disease Brain 131(Pt 3):681–689.

Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems, pp. 1097–1105.

Lécun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc IEEE, 86*, 2278–2324.

Lerch, J. P., Pruessner, J., Zijdenbos, A. P., Collins, D. L., Teipel, S. J., Hampel, H., & Evans, A. C. (2008). Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiology of Aging, 29*, 23–30.

Li, R., Zhang, W., Suk, H.I., Wang, L., Li, J., Shen, D., Ji, S., (2014). Deep learning based imaging data completion for improved brain disease diagnosis. Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention 17, 305–312.

Lin, T.Y., Roychowdhury, A., & Maji, S. (2015). Bilinear CNN models for fine-grained visual recognition. IEEE International Conference on Computer Vision, Santiago, Chile, pp 1449–1457.

Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., & Fulham, M. J. (2015). Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Transactions on Biomedical Engineering, 62*, 1132–1140.

Lu, S., Xia, Y., Cai, T.W., & Feng, D.D. (2015). Semi-supervised manifold learning with affinity regularization for Alzheimer's disease identification using positron emission tomography imaging. 2015 37th Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc), pp. 2251–2254.

Minati, L., Edginton, T., Bruzzone, M. G., & Giaccone, G. (2009). Reviews: Current concepts in Alzheimer's disease: A multidisciplinary review. *American Journal of Alzheimers Disease & Other Dementias, 24*, 95–121.

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering, 19*, 221.

Silveira M, Marques, J. (2010). Boosting Alzheimer disease diagnosis using PET images. 20th IEEE international conference on pattern recognition (ICPR), pp. 2556–2559.

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.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.

Suk, H.I., Shen, D., 2013. Deep learning-based feature representation for AD/MCI classification. Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention 16, 583–590.

Suk, H. I., Lee, S. W., & Shen, D. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage, 101*, 569–582.

Suk, H. I., Lee, S. W., & Shen, D. (2015). Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Structure and Function, 220*, 841–859.

Wang, Y., Nie, J., Yap, P. T., Shi, F., Guo, L., & Shen, D. (2011). Robust deformable-surface-based skull-stripping for large-scale studies. *Medical Image Computing and Computer-Assisted Intervention – MICCAI, 14*(3), 635–642.

Wang, Y., Zhang, P., An, L., Ma, G., Kang, J., Shi, F., Wu, X., Zhou, J., Lalush, D. S., & Lin, W. (2016). Predicting standard-dose PET image from low-dose PET and multimodal MR images using mapping-based sparse representation. *Physics in Medicine and Biology, 61*(2), 791–812.

Weinzaepfel, P., Harchaoui, Z., & Schmid, C. (2015). Learning to track for spatio-temporal action localization. pp. 3164–3172.

Yan, W., Ma, G., Le, A., Feng, S., Pei, Z., Xi, W., Zhou, J., & Shen, D. (2017). Semi-supervised tripled dictionary learning for standard-dose PET image prediction using low-dose PET and multimodal MRI. *IEEE Transactions on Biomedical Engineering, 64*, 569–579.

Zeiler, M.D. (2012). ADADELTA: An adaptive learning rate method. Computer Science.

Zeiler, M. D., & Fergus, R. (2014). *Visualizing and understanding convolutional networks*. Basel: Springer International Publishing.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage, 55*, 856–867.