Magnetic Resonance Imaging 78 (2021) 119–126

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

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

Original Contribution

A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification

Jie Zhang [a,b], Bowen Zheng [a], Ang Gao [a], Xin Feng [b,c], Dong Liang [a,d], Xiaojing Long [a,d,*]

[a] Reasearch Center for Medical AI, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[b] Computer Science and Engineering, Chongqing University of Technology, China
[c] Big Data and Machine Learning Lab, China
[d] Key Laboratory for Magnetic Resonance and Multimodality Imaging of Guangdong Province, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Convolutional neural network
Attention mechanism
Early detection
Structural MRI

ABSTRACT

Purpose: Alzheimer's disease (AD) is a progressive and irreversible neurodegenerative disease. In recent years, machine learning methods have been widely used on analysis of neuroimage for quantitative evaluation and computer-aided diagnosis of AD or prediction on the conversion from mild cognitive impairment (MCI) to AD. In this study, we aimed to develop a new deep learning method to detect or predict AD in an efficient way.
Materials and methods: We proposed a densely connected convolution neural network with connection-wise attention mechanism to learn the multi-level features of brain MR images for AD classification. We used the densely connected neural network to extract multi-scale features from pre-processed images, and connection-wise attention mechanism was applied to combine connections among features from different layers to hierarchically transform the MR images into more compact high-level features. Furthermore, we extended the convolution operation to 3D to capture the spatial information of MRI. The features extracted from each 3D convolution layer were integrated with features from all preceding layers with different attention, and were finally used for classification. Our method was evaluated on the baseline MRI of 968 subjects from ADNI database to discriminate (1) AD versus healthy subjects, (2) MCI converters versus healthy subjects, and (3) MCI converters versus non-converters.
Results: The proposed method achieved 97.35% accuracy for distinguishing AD patients from healthy control, 87.82% for MCI converters against healthy control, and 78.79% for MCI converters against non-converters. Compared with some neural networks and methods reported in recent studies, the classification performance of our proposed algorithm was among the top ranks and improved in discriminating MCI subjects who were in high risks of conversion to AD.
Conclusions: Deep learning techniques provide a powerful tool to explore minute but intricate characteristics in MR images which may facilitate early diagnosis and prediction of AD.

1. Introduction

Alzheimer's disease (AD) is a neurological degenerative disease. Clinically, it is characterized by memory impairment, visual spatial impairment and personality behavioral changes. Nowadays, at least 50 million people are believed to be living with Alzheimer's disease or other dementias [1]. Moreover, this number will continue to grow in the next two to three decades [2,3]. Mild cognitive impairment (MCI) is generally considered to be a transitional state from normal control (NC) to AD, and is often considered as a precursor of AD when it is associated with memory loss and poor judgment. Although the etiology of AD has not been completely known and there is no effective treatment for AD, so far, early detection of AD is of importance for prevention and treatment of the disease. However, early diagnosis is still challenging, especially the precise distinction between stable MCI subjects (who does not develop to AD) and MCI converters (who gradually develops to AD). Magnetic resonance image (MRI) provides a non-invasive and powerful tool to facilitate our understanding and evaluation of anatomical and functional brain changes related to AD. They are playing important roles in the routine clinical practice and also recognized as important

* Corresponding author at: Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Blvd., Shenzhen, Guangdong 518055, China.

E-mail address: xj.long@siat.ac.cn (X. Long).

https://doi.org/10.1016/j.mri.2021.02.001
Received 28 September 2020; Received in revised form 18 January 2021; Accepted 3 February 2021
Available online 13 February 2021

biomarkers for AD progression [4–6]. For decades, researchers have devoted their efforts to develop various approaches for early diagnosis of AD on individual basis, including both ordinary machine learning methods and deep learning networks.

Many computer-aided systems have been developed using various machine learning methods to decode disease states from MR images [7–10]. These algorithms were trained to produce a desired output from a set of training data, such as features obtained from voxel intensity, tissue density or shape descriptor. These methods can be roughly divided into two categories in terms of the brain coverage for feature extraction: the whole brain based [6,11] and the region of interest (ROI) based [8,12,13]. Kloppel et al. [6] mapped the gray matter segment of the whole brain to a high dimensional space, where voxels were treated as coordinates and the value of each voxel was taken as intensity value. The subjects were then classified by using linear support vector machine (SVM). Long et al. [11] quantified the deformation vectors on the whole brain gray matter as image dissimilarity, then applied SVM for classification. As the whole-brain methods may be computational expensive due to the high-dimensional features, methods focusing on regional features either chose one or more brain areas that were regarded as being relevant to AD or selected ROIs that were adapted to the specific cohorts by algorithm. The 3D volume and shape characteristics of hippocampus, parahippocampal gyrus, and entorhinal cortex were commonly used features in ROI-based methods. Silveira et al. [12] separated the brain images into 116 anatomical ROIs, and the boosting classification was adopted for classification. Zhang et al. [8] proposed a multi-kernel SVM to ensemble the multi-modal features such as tissue volumes extracted from 93 ROIs. Gutman et al. [13] presented the first SVM classification study using the feature space of shape invariants of hippocampal surface. The shape invariants were based on rotationally invariant properties of spherical harmonics (SPH). The ROI-based methods were shown to be effective, but errors or feature variance may be induced by region delineation and thus affect classification results.

Recently, the popular deep learning technologies have made great success in the field of computer vision. It is shown that deep neural networks could discover discriminative and complex patterns from sufficiently labelled data. From AlexNet to DenseNet, the model can further extract those intricate but useful information. The DenseNet which connects each layer to every other layer in a feed-forward fashion, alleviates the gradient vanishing problem, strengthens feature propagation, encourages feature reuse, which improves the performance of classification. Researchers also extended the usage of deep learning technologies in AD detection. At the early stage, deep learning methods were mainly used for region segmentation or feature extraction followed by traditional machine learning algorithms such as SVM and bboosting. Liu et al. [14] proposed a method to learn deep convolutional features using both unsupervised and supervised learning for AD and MCI classification based on 2D MRI images. Suk et al. [10] proposed to use a stack sparse autoencoder for feature extraction along with SVM for classification. Li et al. [15] developed a deep learning framework based on three-dimensional convolutional neural network (CNN). The 3D CNN features were then combined with MRI gray matter density map and PET intensity values for multi-modal AD discrimination. Recently, deep neural networks have been applied to the throughout classification process. Payan et al. [16] constructed a 3D CNN model for both feature extraction and the following classification, where the 3D convolution layers were trained by sparse automatic encoder. Sarraf et al. [17] used the classic architecture LeNet-5, while Hosseini-Asl et al. [5] proposed to build a 3D convolutional autoencoder named 3D-CAES where the model was pre-trained to capture anatomical shape variations in structural brain MRI scans.

Among recent advanced deep learning techniques, attention mechanism (AM) has shown to be a powerful tool and been an essential component of neural architectures in a large number of applications in natural language processing [18,19], image classification [20,21] and segmentation [22]. The basic idea of attention mechanism comes from visual attention in human vision system, which illustrates that human vision always focuses on selective parts of the whole visual screen, and learning process could be similar to selectively learn weights of interest. Xu et al. [18] first introduced visual attention into deep learning model for image captioning, where soft attention and hard attention were both proposed. Attention mechanism has also been applied in the field of medical imaging. Ypsilantis et al. [23] explored where to look in chest X-rays by using a recurrent attention model (RAM) based on recurrent neural network (RNN) and reinforcement learning. The model was shown to obtain 90.6% and 91.0% accuracy for the recognition of medical devices and enlarged hearts, respectively. Schlemper et al. [24] applied an attention-gated network to real-time automated scan plane detection for fetal ultrasound screening. They demonstrated that the network with attention modules led to better performance than general networks.

Inspired by the success of deep learning methods and attention mechanisms in medical imaging areas, we proposed an improved convolution neural network for AD diagnosis and prediction with MR images. We first proposed the densely connected convolution neural network with connection-wise attention mechanism (short for 'CAM-CNN') to learn the multi-level features of MR brain images. To capture the 3D features from MR images, we used a 3D convolution on the classic DenseNet structure. A connection-wise attention mechanism was applied to integrate feature maps from different layers to hierarchically transform the MR image into more compact high-level features. Furthermore, we also extended the convolution operation to 3D to capture the spatial information of MRI. The features extracted from each 3D convolution layer were interconnected with the features from all previous layers with different attention and were finally used for classification. The proposed model was verified on the data from the ADNI database for discrmination of AD and prediction of MCI conversion.

2. Material and methods

2.1. Data and processing

Data in this paper were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and the Food and Drug Administration (FDA). It aims to investigate the role of using serial magnetic resonance imaging (MRI), Positron Emission Tomography (PET), and other biological markers, together with clinical and neuropsychological assessment in the diagnose of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). In ADNI, the T1-weighted MR images were acquired sagittally using the volumetric 3D MPRAGE with 1.25×1.25 mm^2 in-plane spatial resolution and 1.2 mm thick sagittal slices. Most of these images were obtained with 1.5 T scanners. More detailed information about MR acquisition procedures is available at the ADNI website.

In this work, we used the T1-weighted MRI data from the baseline visits of 968 participants including 280 AD patients, 162 subjects who were diagnosed as MCI and had converted to AD within 18 months (short for 'cMCI'), 251 subjects with MCI who had not developed to AD within 5 years (short for 'ncMCI'), and 275 normal controls (NC) who remained cognitively normal within 3 years for evaluation. The demographic and cognitive examination details of each group were shown in Table 1.

A routine pre-processing procedure was applied to each image using the FSL software (https://fsl.fmrib.ox.ac.uk). Specifically, a non-parametric and non-uniform bias correction algorithm were used to correct the intensity inhomogeneity followed by an intensity normalization step. The BET module was used to remove the skull. All processed images were aligned to the standard space using FLIRT. Finally, all the images were resampled to the size of $182\times 218 \times 182$.

Table 1
Demographic and cognitive examination scores of all subjects.

Diagnostic type	Number	Age	Gender (M/F)	MMSE
AD	280	76.13 ± 6.14	132/148	23.54 ± 2.08
cMCI	162	75.13 ± 5.23	86/76	26.94 ± 1.22
ncMCI	251	77.61 ± 5.92	115/136	27.54 ± 1.32
NC	275	76.16 ± 6.29	144/131	29.16 ± 0.82

AD: Alzheimer's disease; cMCI: MCI converters; ncMCI: MCI non-converters; NC: Normal control.

2.2. Image patch generation

Image biomarkers of AD may be spatially associated with widespread areas in the brain. Such information may be ignored if the whole brain was considered as multiple 2D slices in 2D-CNN. Inspired by the great success of densely connected neural network in image classification, the proposed method extended the 2D convolution layers of DenseNet to 3D convolution layers to capture the unitary features of AD from MRI. However, the 3D model has brought new problems that may affect the experimental computation. Firstly, considering the brain in a 3D scope increased the complexity to acquire training parameters and was time-consuming. Secondly, backgrounds with trivial information would be unfavorable for network training and convergence.

We proposed a method that removed the blank areas of the brain image. After the skull-strip and spatial normalization step in pre-processing, the binarized masks of all brains were obtained. The maximum value of the brain size across all slices of all individuals was extracted, and the region was expanded for 5 voxels in all directions to determine the effective area. Then we removed the peripheral background outside the effective brain area and resampled the image to 160 × 180 × 160. Then the new volume was segmented into 64 patches through the following steps: First, we set the origin of image at the superior-posterior cornor of the image. A 3D window with the size of 96 × 120 × 96 was defined and placed at the origin. Then the window slid along each axis (*x, y, z*) of the image with a step size of 20, and finally generated a total of 64 patches as the input for training. Removing the marginal trifling information increased the efficiency for network training, while the patch which had smaller size compared with original brain images reduced the amount of calculation. Fig. 1 showed the patch generation procedure.

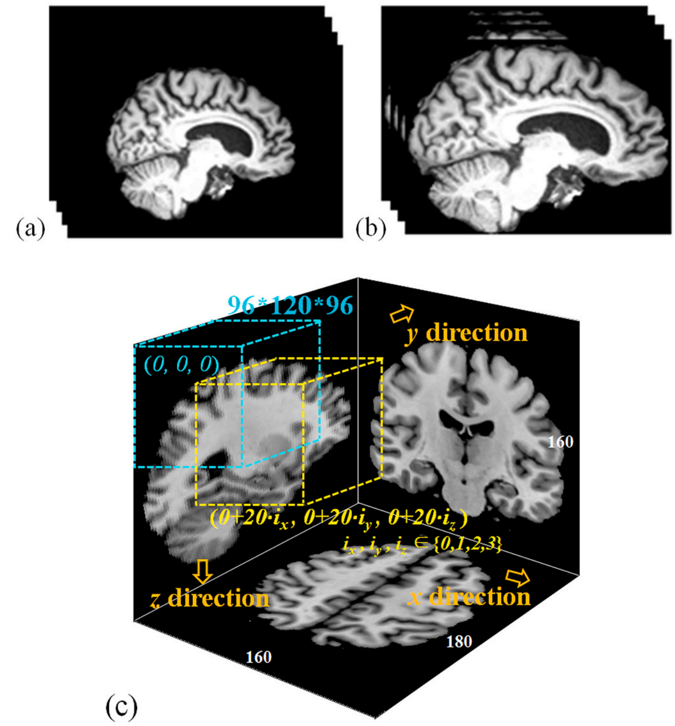

Fig. 1. The procedure of image patch generation. (a) Original images with large uninformative background. (b) Images removed marginal background. (c) Brain patch generation. The defined 3D window with the size of 96 × 120 × 96 slid from the up-left of the pre-processed image along x, y, z direction, at a step of 20. Finally, 64 patches were generated for each subject.

2.3. A convolution neural network based on connection-wise attention model

We proposed a 3D connection-wise-attention-model-based densely connected convolution neural network (CAM-CNN) to learn the multi-level features of brain MR images.

2.3.1. Dense connections in DenseNet

He et al. [25] proposed the ResNet (residual network) to combat the vanishing gradient problem during training of deep convolutional networks. ResNet eases the training of networks that are substantially deeper than those used previously and add a skip connection that bypasses the on-linear transformations with an identity function. An advantage of ResNet is that the gradient can flow directly through the identity function from an earlier layer to subsequent layers. Fig. 2(a) illustrated the layout of the ResNet schematically. To further improve the information flow between layers, Huang et al. [26] proposed a different connectivity pattern named densely connected convolutional networks (DenseNet). Instead of drawing representational power from extremely deep or wide architectures, DenseNet exploits the potential of the network through feature reuse, yielding condensed models that are easy to train and have high parameter efficiency. The feature maps that connect the information of all previous layers has been demonstrated to increase variation in the input of subsequent layers and improve network efficiency. This constitutes a major difference between DenseNet and ResNet. Fig. 2(b) illustrated the layout of the resulting DenseNet schematically. Consider a single image x_0 that is passed through a convolutional network. The network comprises *L* layers, each of which implements a non-linear transformation $H_l(\cdot)$, where *l* indexes the layer. Consequently, the *l*th layer receives the feature maps of all preceding layers, x_0, x_1, $\cdots x_l$ as input.

$$x_l = H_l[(x_{l-1}, x_{l-2}, x_{l-3}, \cdots, x_0)] \tag{1}$$

Visual attention was applied to provide an effective construction approach of this network. The connection-wise attention model proposed by this work allowed the network to place the visual attention on the feature maps of each layer (as shown in Fig. 3). Instead of concatenating the features of all preceding layers to the subsequent layer as done in the DenseNet, the proposed CAM-CNN algorithm introduced a model that integrated the feature maps of all previous layers by a weighted summation, where the weighting parameters were self-learned during network training. This has made the network more simple and efficient by paying attention on the most contributory information. The CAM-CNN assigned a weighting coefficient W to the *i*th layer of the network as shown in Formula 2, where W_i denoted an attention vector consisting of *i-1* elements. Formula 3 illustrated the layout of the *l*th layer with connection-wise attention, where H_l (.) was a non-linear transformation and $x_j(1 \le j \le l-1)$ represented feature maps from the *j*th layer.

$$W_i = \left[w_{i-1,i}, w_{i-2,i}, \cdots, w_{2,i}, w_{1,i}\right] \tag{2}$$

$$x_l = H_l\left(w_{l-1,l}x_{l-1}\right) + w_{l-2,l}x_{l-2}\cdots + w_{1,l}x_1 \tag{3}$$

The network consisted of 4 layer types. The first type was the input layer which the image patches were fed into the network. The second type was convolutional layer which convolved the learned filters with the input images and produced feature maps for each filter. The CAM-

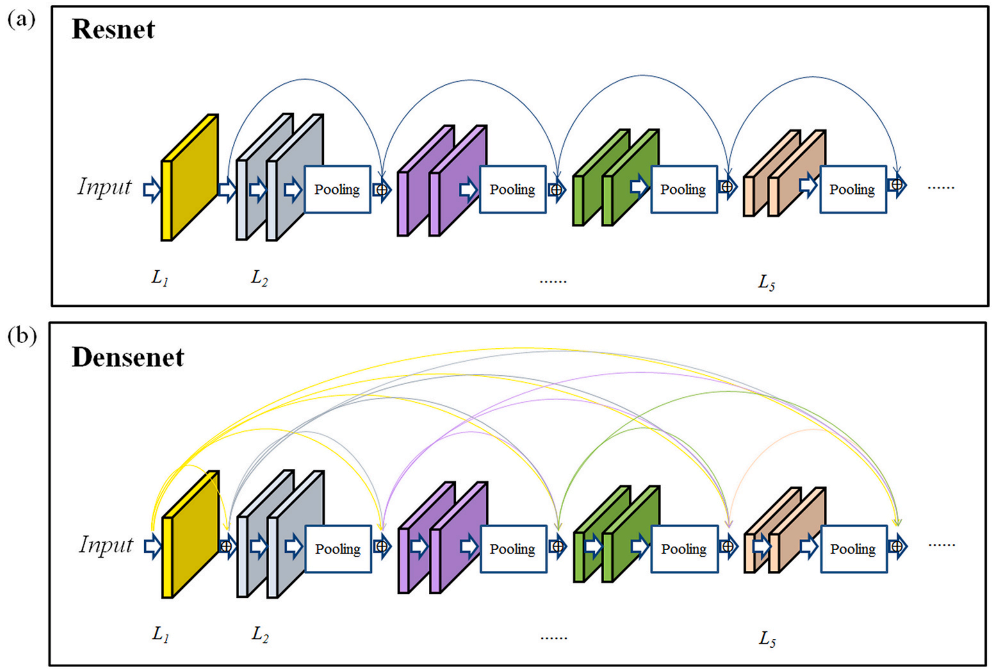

Fig. 2. The layout of the ResNet (a) and DenseNet (b). (a) The ResNet builds skip connections to jump over layers. (b) The DenseNet constructs an architecture that connects each layer to every subsequent layer.

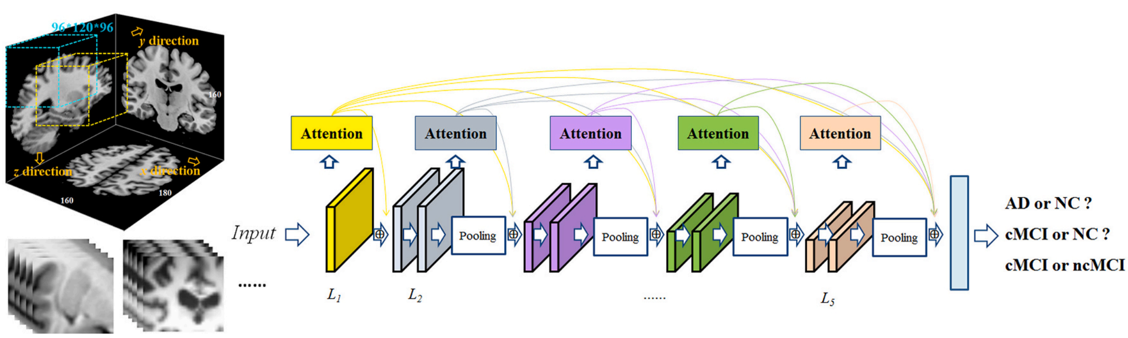

Fig. 3. The layout of the proposed CAM-CNN. The attention mechanism was introduced to the dense connections between layers, which allowed the network to place attention on the feature maps that were more contributive for classification.

CNN was built with 9 convolutional layers, and each two convolutional layers were followed by a max pooling layers. The sizes of the convolution filters were 3×3×3, and the filter numbers were set to 35. The third type was the pooling layer, where max pooling was used in this work. Max pooling reduced the feature map along the spatial dimensions by replacing each cube with their maximum value and kept the most influential features for distinguishing images. In this work, max pooling was applied for each 2×2×2 region, and Tanh was adopted as the activation function in these layers due to its good performance for CNNs. The fourth type of layer was the fully connected layer which consisted of a number of input and output neurons that generated the learned linear combination of all neurons from the previous layer and passed through a nonlinearity. The inputs and outputs of the fully connected layers were a 1D vector and not spatially located anymore. The designed model had two characteristics: First, the CAM-CNN referred to the ResNet connection method that the identity function and the output of H_l were integrated by summation which reduced the number of parameters during training. Second, the feature maps of the current layer were obtained through the maps of all preceding layers. Each layer of the network was assigned a self-learned weighting coefficient, allowing the CAM-CNN to concentrate on more contributive features.

2.4. Experiments

The original MR images were first registered to the standard space using affine transformation and resampled to the size of 160 × 180 × 160. Then 64 brain patches were generated from each subject for training and testing, which finally resulted in 61,954 feature patches including 17,920 in AD group, 10,368 in cMCI, 16,064 in ncMCI, and 17,600 in NC group.

In our experiments, each group was randomly split into training set (70% of subjects), validation set (15% of subjects), and testing set (15% of subjects). The proposed classification method was implemented with the Keras library in Python3.6 based on tensorflow, and then performed on a PC with GPU NVIDIA GTX1080 in the environment of Ubuntu14.04-x64. The network parameters were randomly initialized at the beginning, and stochastic gradient decent (SGD) optimizer was adopted with the initial learning rate of 0.001. The momentum was set to 0.9, and the batch size was set to 64. To avoid overfitting, the dropout as well as L1 and L2 regularization were used in our network. We has applied the proposed algorithm to classify (1) mild AD patients versus normal controls (AD vs. NC), (2) MCI converters versus normal controls (cMCI vs. NC), and (3) MCI converters versus non-converters (cMCI vs.

ncMCI).

The proposed algorithm was compared with three classic networks: the basic CNN, the ResNet, and the DenseNet. The basic CNN network consisted of 9 3D convolutional layers that used a kernel with a size of $3\times3\times3$, step size of $1\times1\times1$, and channel number of 35. Each two convolutional layers were followed by a pooling layer, so there were a total of 4 pooling layers, all of which used maximum pooling with a size of $2\times2\times2$. Two fully connected layers were constructed at the end of the network, and the dropout technique was used as a regular term to prevent overfitting and improve the robustness of the model. The ResNet introduced four residual modules to the basic CNN, while the DenseNet changed the design of information flow on the basis of the ResNet to ensure the interconnection between all layers. The general parameter settings of the three networks were the same with those of the proposed method.

3. Results

3.1. Classification results of the proposed method and comparison with basic convolutional neural networks

The proposed method presented good classification performance with accuracy of 97.35%, 87.82%, and 78.79% for discriminating mild AD, MCI converters (cMCI) and stable MCI subjects (ncMCI) against normal controls, respectively. Compared with the basic CNN, ResNet, and DenseNet on the same datasets, our method demonstrated higher accuracy in all group separation.

The proposed CAM-CNN method presented better performance than the popular networks such as ResNet and DenseNet. For AD vs. NC, the three indicators of classification performance including sensitivity (SEN), specificity (SPE), and accuracy (ACC) using the basic CNN was lower than 90%. ResNet improved the accuracy to 93.43%, which was 8.36% higher than basic CNN. By strengthening feature propagation and reuse, DenseNet led to a raise of the accuracy to 94.96%. The proposed CAM-CNN model achieved the highest accuracy of 97.35%. The area under curve (AUC) of the four networks were all higher than 90%, where AUC of CAM-CNN almost attained 100%. For cMCI vs. NC, the accuracy using basic CNN was lower than 80%. DenseNet and CAM-CNN performed well with accuracy of 87.53% and 87.82% respectively, and the AUC of the two models have exceeded 90%. For cMCI vs. ncMCI, the classification accuracy of the four networks were between 70% and 80%. The basic CNN produced accuracy of 72.19%, sensitivity of 72.50% and specificity of 71.86%. The accuracy of ResNet and DenseNet were 73.24% and 76.02%. The proposed CAM-CNN method again showed the highest accuracy of 78.79% for cMCI vs. ncMCI classification, with AUC of 86.79%. The results were summarized in Table 2 and Fig. 4.

3.2. Comparison with other existing methods

We compared the classification results of our model with those reported in previous studies also using the ADNI database (as shown in Table 3).

First, we compared our proposed model to traditional machine learning methods. Wolz et al. [27] and Cho et al. [28] proposed ROI-based methods to extract brain features, then used machine learning methods such as SVM or linear discriminant analyses (LDA) to perform the classification. Wolz et al. [27] obtained the accuracy of 85%, 82%, and 69% for a single modality dataset of structural MRI, but none of the accuracy was greater than 90%. Similar results were shown in the work of Cho et al. [28]. Furthermore, we compared our model to some existing deep learning methods. Liu et al. [29] designed an architecture which contained stacked auto-encoders (SAE) and a softmax output layer, where they showed 87.86% accuracy for AD vs. NC classification and 76.92% for cMCI vs. NC. Later they enhanced the algorithm by applying a zero-mask strategy for data fusion to extract complementary information from images [30]. The algorithm improved the accuracy for AD vs. NC classification to 91.4%. Suk et al. [31] proposed a novel framework that combined two conceptually different methods of sparse regression and deep learning method, and obtained the accuracy of 91.02%, 73.02%, 74.82% for AD vs. NC, cMCI vs. NC, cMCI vs. ncMCI classification. They also introduced PET data along with MRI data and constructed a deep network with a restricted boltzmann machine, which increased the accuracy to 95.35% and 75.92% for classifying AD vs. NC and cMCI vs. ncMCI respectively [32]. Cheng et al. [33] proposed to construct multiple deep 3D convolutional neural networks to learn various features from local brain images which were combined to make the final classification for AD diagnosis. They achieved 86.36% accuracy for classifying AD vs. NC. Liu et al. [34] constructed cascaded convolutional neural networks to learn the multi-level features of MRI for AD classification. The method yielded 92.75% classification accuracy for AD vs. NC, and 76.90% for cMCI vs. ncMCI. Basais et al. [35] adopted the CNN model on a large amount of samples and obtained the accuracy of 99.2%, 87.1%, 75.1% for AD vs. NC, cMCI vs. NC, and cMCI vs. ncMCI classification. The proposed CAM-CNN method presented a good classification performance with accuracy of 97.35%, 87.82%, and 78.79% to distinguish AD vs. NC, cMCI vs. NC, and cMCI vs. ncMCI respectively, where it demonstrated the highest capacity to differentiate MCI converters from normal controls and MCI non-converters.

4. Discussion

Effective and accurate AD diagnosis is critical for early intervention and management of the disease. Therefore researchers have devoted their efforts to develop computer-aided systems that aim to diagnose AD in an early stage. Different with traditional methods based on hand-crafted features, we proposed a densely connected convolutional network with connection-wise attention mechanism (CAM-CNN) that facilitated detection and prediction of individuals with AD or MCI using structural MR brain scans.

CNN model has provided a tool for image-assisted diagnosis and currently have attracted much attention in disease classification. The increasing depth and complexity of the network improved the classification performance, but at the same time it brought new problems, such

Table 2
Classification results of the proposed network compared with some general networks on AD vs. NC, cMCI vs. NC and cMCI vs. ncMCI.

	AD vs. NC				cMCI vs. NC				cMCI vs. ncMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
	%	%	%	%	%	%	%	%	%	%	%	%
Basic CNN	85.07	82.59	87.56	93.21	79.32	83.14	75.45	85.11	72.19	72.50	71.86	78.29
ResNet	93.43	94.27	92.59	95.00	82.99	81.39	84.59	86.36	73.24	73.05	73.43	81.80
DenseNet	94.96	94.50	95.43	96.24	87.53	76.19	**88.91**	90.81	76.02	73.59	78.44	85.08
CAM-CNN	**97.35**	**97.10**	**97.95**	**99.70**	**87.82**	**87.56**	88.84	**92.85**	**78.79**	**75.16**	**82.42**	**86.79**

The bold numbers denote the maximum value of each column, that is the highest rate of each index.
ACC:Accuracy; AD: Alzheimer's disease; AUC: Area under curve; cMCI: MCI converter; ncMCI: MCI non-converter; NC:Normal control; SEN: Sensitivity; SPE: Specificity.

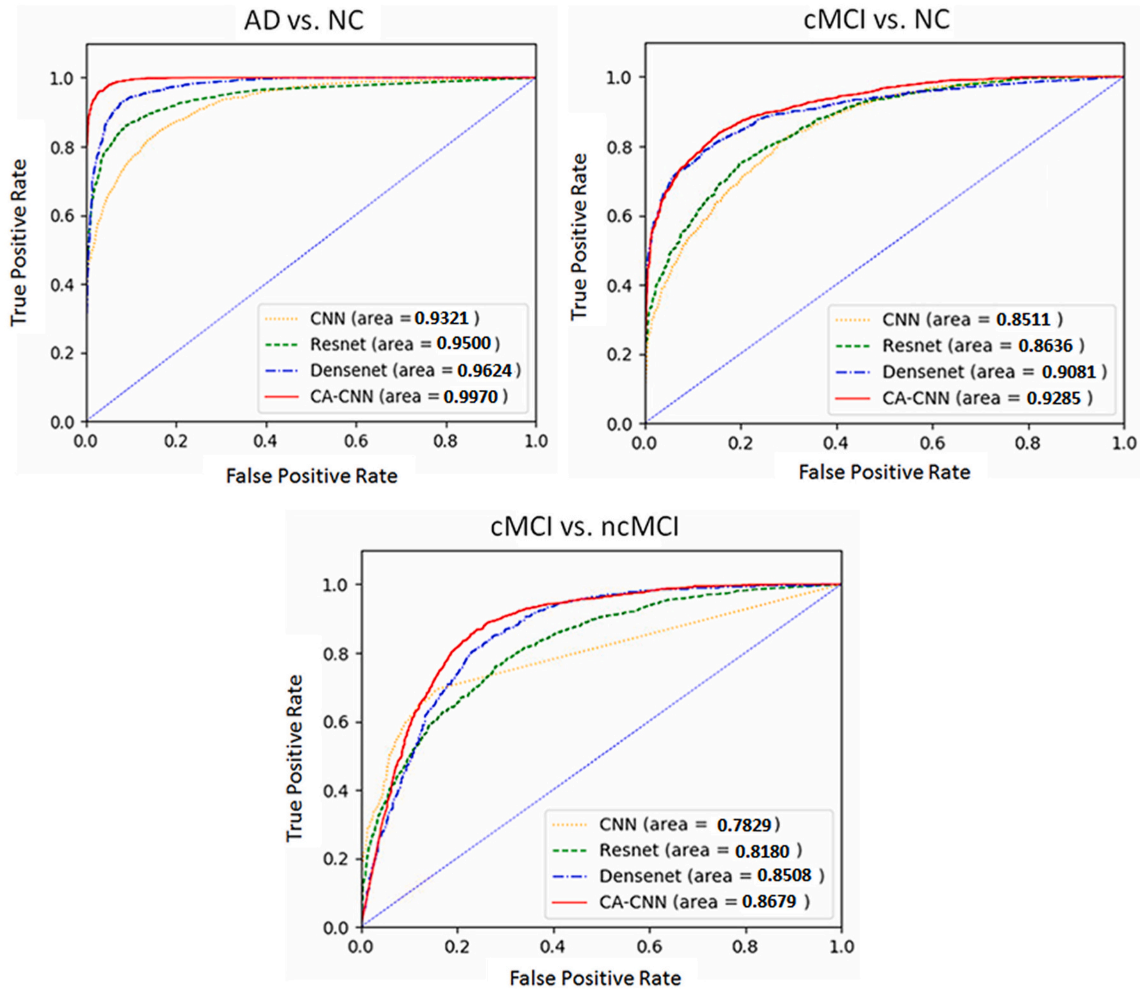

Fig. 4. ROC curves of classification by different networks. The proposed network demonstrated higher accuracy and AUC compared with the basic CNN, ResNet, and DenseNet models.
AD: Alzheimer's disease; AUC: Area under curve; cMCI: MCI converter; ncMCI: MCI non-converter; NC: Normal control; ROC: Receiver Operating Characteristic.

Table 3
Classification results of the proposed method andsome published methods on AD vs. NC, cMCI vs. NC, and cMCI vs. ncMCI.

Algorithms	Data modality	Number of samples	AD vs NC			cMCI vs NC			cMCI vs ncMCI		
			SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)
Ours	MRI	AD-280, cMCI-162, ncMCI-251, NC-275	97.10	97.95	97.35	87.56	88.84	**87.82**	**75.16**	82.42	**78.79**
Wolz et al., 2011	MRI	AD-198, cMCI-238, ncMCI-167, NC-231	89	93	85	84	86	82	68	67	69
Cho et al., 2012	MRI	AD-128, cMCI-72, ncMCI-131, NC-160	82	93	–	66	**89**	–	63	76	–
Liu et al., 2014	MRI + PET	AD-65, cMCI-67, ncMCI-102, NC-77	88.57	87.22	87.86	74.29	78.13	76.92	–	–	–
Suk et al., 2014	MRI	AD-93, cMCI-76, ncMCI-128, NC-101	91.54	94.56	92.38	99.58	53.79	84.24	36.70	**90.98**	72.42
Liu et al., 2015	MRI	AD-180, cMCI-160, ncMCI-214, NC-204	92.32	90.42	91.40	60.00	92.32	82.10	–	–	–
Suk et al., 2017	MRI	AD-186, cMCI-167, ncMCI-226, NC-226	92.72	89.94	91.02	77.6	68.22	73.02	70.93	78.82	74.82
Cheng et al., 2017	MRI	AD-199, cMCI-0, ncMCI-0, NC-229	85.93	87.15	86.36	–	–	–	–	–	–
Liu et al., 2018	MRI	AD-93, cMCI-76, ncMCI-128, NC-100	93.48	91.30	92.75	–	–	–	42.11	82.43	76.90
Basaia et al., 2019	MRI	AD-294, cMCI-253, ncMCI-510, NC-352	**98.9**	**99.5**	**99.2**	**87.8**	86.5	87.1	74.8	75.3	75.1

The bold numbers denote the maximum value of each column, that is the highest rate of each index.
ACC: Accuracy; AD: Alzheimer's disease; cMCI: MCI converter; ncMCI: MCI non-converter; NC:Normal control; SEN: Sensitivity; SPE: Specificity.

as gradient vanishing or explosion. Feature information in basic CNN architecture was a one-way and one-time flow, which meant that feature maps from different layers did not interact much. In our experiments, the basic CNN method presented classification accuracy of 85.07%, 79.32%, and 72.19% for mild AD, MCI converters and stable MCI subjects versus normal controls, not as good as other networks. The emergence of ResNet was a landmark in the development of deep learning methods, which brought in a residual unit that linked the current layer to previous layer. The residual unit was introduced to solve the degradation problem. The skip-connection allowed ResNet to become deeper and show better performance than basic CNN. The DenseNet has borrowed the idea of ResNet but proposing a brand-new structure. In this network, there were direct connections between any of the two layers, that is, the input of each layer of the network was the union of the outputs of all previous layers, and the feature map generated by the current layer would also be directly passed to all the following layers as input. Such an architecture has alleviated the problem of gradient vanishing, encouraged feature reuse, and greatly reduced the amount of parameters. The classification accuracy of DenseNet thus has been improved to 94.96%, 87.53%, and 76.02% respectively for classification of AD vs. NC, cMCI vs. NC, and cMCI vs. ncMCI.

This paper proposed an improved densely connected network with connection-wise attention mechanism named CAM-CNN. Existing attention mechanisms are mainly divided into two categories in neural networks, the first is the spatial attention mechanism, and the second is the channel attention mechanism. Spatial attention is to train to find areas that need attention for image information [36,37]. Channel attention allows the network to focus on different filters, thereby improving the accuracy of network classification [38,39]. Different from the above two attention models, this paper for the first time proposed a new attention mechanism that placed the focus point on the feature maps between different network layers. We know that the DenseNet conducts feature fusion of different layers. However, the information extracted from all preceding layers and their contribution for classification performance may be different. It could cause information redundancy and increase the time and amount of calculation for network training if all information were equally considered in the model. Therefore, the proposed algorithm introduced a dense connection attention mechanism to the improved network. The dense connection attention module assigned a weight coefficient to each layer in the network, which was a weighting parameter obtained through network training. By applying different weight parameters, the network automatically adjusted the contribution of different information of different layers to increase the efficiency for classification. Thus the proposed method has demonstrated higher classification accuracy than ResNet and DenseNet in the experiments.

Discriminating MCI converters who were in the high risk of developing AD is of particular importance for clinical control and management of the disease. However, the image characteristics of MCI converters compared to those of healthy elderly or MCI non-converters are less notable. With the proposed method, we obtained the accuracy of 78.79% for cMCI vs. ncMCI, 87.82% for cMCI vs. NC which have shown improvement on the performance compared with some popular network models and indicated the potential of CAM-CNN model to detect subjects in prodromal dementia.

There were limitations of the proposed method. Although we have reduced the size of input data by generating MR patches instead of using the whole brain image, it was still computational expensive to train parameters. Moreover, in the training process, we adjusted the parameters of the deep CNN model, including the number of layers, the size and number of kernels in each layer, nevertheless network convergence was still challenging. Reducing the size of patch to 1/3 of that used in this work help solve the non-convergence problem as well that the training time was shortened. In practice, network parameters and settings should be considered for a trade off to achieve the best classification efficiency and effect.

5. Conclusions

In this study, we developed a network to diagnose and predict AD conversion based on structural MR data, by combining CNN with an attention model. We used the densely connected neural network to extract multi-scale features from the pre-processed data, and a connection-wise attention mechanism to combine connections between different features from different layers to hierarchically transform the MR data into more compact high-level feature maps. We achieved the classification accuracy of 97.35%, 87.82% and 78.79% for distinguishing AD vs. NC, cMCI vs. NC, and cMCI vs. ncMCI. Compared with some general neural networks and previous methods on the same dataset, the proposed algorithm has demonstrated top-ranked classification accuracy rates for detection of AD and prediction of MCI conversion.

Funding

This work was supported by Shenzhen Overseas High-level Talent Innovation and Entrepreneurship Special Fund (Peacock Plan) (KQTD20180413181834876), the Guangdong grant 'Key technologies for treatment of brain disorders' (2018B030332001) and Guangdong Provincial Key Laboratory for Magnetic Resonance and Multimodality Imaging (2014B030301013).

Acknowledgements

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, Alzheimers Association; Alzheimers 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 Immunothe-rapy Research and Development, LLC.; Johnson and Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and 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 Alzheimers 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.

References

[1] Patterson C. World Alzheimer report 2018. The state of the art of dementia research: New frontiers. London: Alzheimer's Disease International; 2018.
[2] Zhan L, Zhou J, Wang Y, Jin Y, Jahanshad N, Prasad G, et al. Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease. Front Aging Neurosci 2015;7:48.
[3] Zhu X, Suk HI, Zhu Y, Thung KH, Wu G, Shen D. Multi-view classification for identification of Alzheimer's disease. International workshop on machine learning in medical imaging. Cham: Springer; 2015. p. 255–62.
[4] Hinrichs C, Singh V, Mukherjee L, Xu G, Chung MK, Johnson SC. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage 2009;48(1):138–49.
[5] Hosseini-Asl E, Keynto R, El-Baz A. Alzheimer's Disease diagnostics by adaptation of 3D convolutional network. In: 2016 IEEE international conference on image processing (ICIP). Phoenix, AZ, USA: IEEE; 2016. p. 126–30.
[6] Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008;131(3):681–9.
[7] Lerch JP, Pruessner J, Zijdenbos AP, Collins DL, Teipel SJ, Hampel H, et al. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol Aging 2008;29(1):23–30.
[8] Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 2011;55(3):856–67.
[9] Liu S, Liu S, Cai W, Che H, Fulham MJ. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans Biomed Eng 2015;62(4):1132–40.
[10] Suk HI, Lee SW, Shen D. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct Funct 2015;220(2):841–59.
[11] Long X, Wyatt C. An automatic unsupervised classification of MR images in Alzheimer's disease. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE; 2010. p. 2910–7.

[12] Silveira M, Marques JS. Boosting Alzheimer disease diagnosis using PET images. In: 20th international conference on pattern recognition. IEEE; 2010. p. 2556–9.
[13] Gutman B, Wang Y, Morra J, Toga AW, Thompson PM. Disease classification with hippocampal shape invariants. Hippocampus 2009;19(6):572–8.
[14] Liu F, Shen C. Learning deep convolutional features for MRI based Alzheimer's disease classification. In: IEEE conference on computer vision and pattern recognition (CVPR); 2014.
[15] Li R, Zhang W, Suk HI, Wang L, Ji S. Deep learning based imaging data completion for improved brain disease diagnosis. In: International conference on medical image computing & computer-assisted intervention. Cham: Springer; 2014. p. 305–12.
[16] Payan A, Montana G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. arXiv preprint arXiv:150202506. 2015.
[17] Sarraf S, Tofighi G. Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. In: 2016 future technologies conference (FTC). IEEE; 2016. p. 816–20.
[18] Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhutdinov R, et al. Show, attend and tell: neural image caption generation with visual attention. In: International conference on machine learning; 2015. p. 2048–57.
[19] Nam H, Ha JW, Kim J. Dual attention networks for multimodal reasoning and matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 299–307.
[20] Wang X, Gao L, Song J, Shen H. Beyond frame-level CNN: saliency-aware 3D CNN with LSTM for video action recognition. IEEE Signal Process Lett 2016;24(4): 510–4.
[21] Jetley S, Lord NA, Lee N, PHS Torr. Learn to pay attention. arXiv preprint arXiv: 180402391. 2018.
[22] Ren M, Zemel RS. End-to-end instance segmentation and counting with recurrent attention. arXiv preprint arXiv:160509410. 2017.
[23] Ypsilantis PP, Montana G. Learning what to look in chest X-rays with a recurrent visual attention model. arXiv preprint arXiv:170106452. 2017.
[24] Schlemper J, Oktay O, Chen L, Matthew J, Knight C, Kainz B, et al. Attention-gated networks for improving ultrasound scan plane detection. arXiv preprint arXiv: 180405338. 2018.
[25] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE conference on computer vision & pattern recognition; 2016. p. 770–8.
[26] Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–8.
[27] Wolz R, Julkunen V, Koikkalainen J, Niskanen E, Zhang DP, Rueckert D, et al. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS One 2011;6(10).
[28] Cho Y, Seong J-K, Jeong Y, Shin SY. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage 2012;59(3):2217–30.
[29] Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D. Early diagnosis of Alzheimer's disease with deep learning. In: IEEE 11th international symposium on biomedical imaging (ISBI). Beijing, China: IEEE; 2014. ISBN: 978-1-4673-1961-4).
[30] Liu S, Liu S, Cai W, Che H, Pujol S, Kikinis R, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans Biomed Eng 2015;62(4):1132–40.
[31] Suk HI, Lee SW, Shen D. Deep ensemble learning of sparse regression models for brain disease diagnosis. Med Image Anal 2017;37:101–13.
[32] Suk HI, Lee SW, Shen D. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage 2014;101:569–82.
[33] Cheng D, Liu M, Fu J, Wang Y. Classification of MR brain images by combination of multi-CNNs for AD diagnosis. In: 9th international conference on digital image processing (ICDIP 2017); Hong Kong, China; 2017.
[34] Liu M, Cheng D, Wang K, Wang Y. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. Neuroinformatics 2018;16(3–4): 295–308.
[35] Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. Neuroimage Clin 2019;21:101645.
[36] Xu H, Saenko K. Ask, attend and answer: exploring questionguided spatial attention for visual question answering. In: European Conference on computer vision; 2016.
[37] Yang Z, He X, Gao J, Deng L, Smola A. Stacked attention networks for image question answering. In: IEEE conference on computer vision and pattern recognition; 2016.
[38] Bastidas AA, Tang H. Channel attention networks. In: CVPR workshops; 2019.
[39] Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: efficient channel attention for deep convolutional neural networks. In: IEEE conference on computer vision and pattern recognition; 2020.