An Explainable 3D Residual Self-Attention Deep Neural Network FOR Joint Atrophy Localization and Alzheimer's Disease Diagnosis using Structural MRI

Xin Zhang, Liangxiu Han, Wenyong Zhu, Liang Sun, Daoqiang Zhang

***Abstract*—Computer-aided early diagnosis of Alzheimer's disease (AD) and its prodromal form mild cognitive impairment (MCI) based on structure Magnetic Resonance Imaging (sMRI) has provided a cost-effective and objective way for early prevention and treatment of disease progression, leading to improved patient care. In this work, we have proposed a novel computer-aided approach for early diagnosis of AD by introducing an explainable 3D Residual Attention Deep Neural Network (3D ResAttNet) for end-to-end learning from sMRI scans. Different from the existing approaches, the novelty of our approach is three-fold: 1) A Residual Self-Attention Deep Neural Network has been proposed to capture local, global and spatial information of MR images to improve diagnostic performance; 2) An explainable method using Gradient-based Localization Class Activation mapping (Grad-CAM) has been introduced to improve the explainable of the proposed method; 3) This work has provided a full end-to-end learning solution for automated disease diagnosis. Our proposed 3D ResAttNet method has been evaluated on a large cohort of subjects from real dataset for two changeling classification tasks (i.e. Alzheimer's disease (AD) vs. Normal cohort (NC) and progressive MCI (pMCI) vs. stable MCI (sMCI)). The experimental results show that the proposed approach outperforms the state-of-the-art models with significant performance improvement. The accuracy for AD vs. NC and sMCI vs. pMCI task are 97.1% and 84.1% respectively. The explainable mechanism in our approach regions is able to identify and highlight the contribution of the important brain parts (hippocampus, lateral ventricle and most parts of the cortex) for transparent decisions.**

***Index Terms*—Deep learning, 3D CNN, MRI brain scans, Model Explanation/Explainable Artificial Intelligence**

I. Introduction

ALZHEIMER'S disease (AD) is the most common cause of dementia among the old people, which is irreversible and progressive neurodegenerative brain disease. It contributes to 60-70% of dementia cases and affects over 30 million individuals[1]. With the exponentially growing aging population across the globe, the prevalent increased cases of Alzheimer's disease (AD) has presented unprecedented pressures on public healthcare service. There is currently no cure for AD. However, the progress could be slowed through medicine and optimised physical cognition and activity. Therefore accurate and timely diagnosis of Alzheimer's disease (AD) and its early form mild cognitive impairment (MCI) is essential for optimal management and improved patient care [2]. Clinically, Structural Magnetic Resonance Imaging (sMRI) has been used for AD diagnosis. The structural MRI measurement is considered as a marker of AD progression, which can help detect the structural abnormalities and track the evolution of brain atrophy[3], [4], [5]. However, the disease identification process is mainly performed manually by specialists, which is time-consuming and expensive.

To solve this problem, much effort has been devoted to developing computer-aided diagnostic systems for automated discrimination of progression of AD (e.g. mild cognitive impairment (MCI) including progressive MCI (pMCI) and stable MCI (sMCI) and normal cohort (NC)) from sMRI scans based on voxel-wise global features, or predetermined regional features or combination of both [6], [7], [8], [9]. The volumetric or voxel-based approaches extract global features for detecting the structure changes and identifying voxel-wise disease associated microstructures for AD diagnosis [10], [11], [12], [13]. The Tensor-based morphology (TBM) diagnostic approach is a voxel-wise optimisation approach, which can recognize local structural changes through mapping orders of local tissue volume loss or income over time to understand the neurodegenerative or neurodevelopment processes for AD diagnosis[14]. In [15], the gray matter voxels was selected as features and was used to trained an machine learning model to classify AD and NC.

Since some specific brain regions e.g. hippocampal region of interest (ROI) are strongly correlated to the disease, several existing works focused on some predetermined ROIs guided by prior biological knowledge and extracted regional features for AD diagnosis [16], [17], [18], [19], [20]. For instance, Magnin [17] and Zhang [21] applied SVM to learn regional features for AD diagnosis by splitting the brain into some non-overlapping areas.

Recently, deep neural networks have shown successful for various computer vision tasks [22], [23]. A few deep learning methods have been proposed for AD diagnosis with sMRI scans and achieved better performance than the machine learning-based methods. These methods focused on learning either regional features from prior Knowledge regions (e.g., hippocampus [24], [19], cortical [25]), global features [7] or

X. Zhang and L. Han is with Department of Computing and Mathematics, Manchester Metropolitan University, Manchester M1 5GD U.K.

D. Zhang, W. Zhu, L. Sun, T. Zhang and Z. Li are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Corresponding authors: L. Han and D. Zhang (e-mail: l.han@mmu.ac.uk; dqzhang@nuaa.edu.cn)

combination of both [26]. Lian et al. proposed a hybrid deep learning approach using CNNs to learn combined features at multi-scale [26]. Hosseini-Asl et al. predicted AD with a 3D CNN based on the pretrained 3D convolutional autoencoder model to capture anatomical shape variations from sMRI [7].

Despite the existing encouraging work, it suffers several limitations. Firstly, extracting global features using voxel-based approaches involve processed high-dimensional 3D data, which is computationally intensive. Secondly, regional based features focusing on certain brain regions of interest (e.g. The cortical thickness and hippocampus shape) may neglect possible pathological locations in the brain and fail to get global structural information for accurate AD diagnosis. Moreover, these methods require domain-specific prior-knowledge and multi-stage training. Thus, it is hard to provide an end-to-end solution for automatic disease diagnosis. Thirdly, the existing methods [26], [7] used combined features or global features to improve disease diagnostic performance based on deep learning approaches. However, the use of hybrid loss functions for each layer with the same shared weight may lead to difficulty in training and reproduction. Finally, most of existing deep learning-based approaches for AD diagnosis lack transparency in terms of model explanation due to the nature of black-box.

To overcome the aforementioned limitations, this work proposes a novel computer-aided approach for early diagnosis of AD from sMRI by developing an explainable 3D Residual Attention Deep Neural Network (3D ResAttNet) for end-to-end learning from sMRI. Different from the existing approaches, our contributions lie in:

1) A Residual Attention Deep Neural Network has been designed and implemented, allowing for capturing local, global and spatial information to improve diagnostic performance;
2) An explainable Gradient-based Localization Class Activation mapping (Grad-CAM) has been introduced, enabling visual explanation and interpretation of model predictions;
3) The proposed work has provided a full end-to-end solution for automated disease diagnosis.

The rest of this paper is organised as follows: Section 2 presents related work; Section 3 details the proposed method; In Section 4, the experimental evaluation is described; Section 5 concludes the work and highlights the future work.

II. Related work

A. Computer-aided AD diagnosis

Computer-aided diagnosis of AD treatment has a long history. The aim of the computer-aided diagnosis is to extract useful features for automatic classification. According to the feature extracted method, it can be broadly divided into three categories: 1) Global feature-based approaches (Voxel-based approaches; 2) Regional feature-based approaches; 3) Combination of both global and regional based approaches.

The early works on AD diagnosis mainly focused on the extracted global features from the whole MR image. The volumetric-based approach using voxel intensity features has been widely used for AD classification. Ashburner [12] introduced voxel-based morphometry (VBM) method, which used voxel-wise comparison on the smoothed gray-matter images. It showed the difference bwtween white and gray voxels in local concentrations compared with the normal cohort(NC) brains. Based on this voxel-wise features, Klöppel et al. trained a support vector machine (SVM) model to diagnosis AD with sMRI image [15]. Hinrichs et al. [27] also employed the gray-matter density to extract voxel-wise feature, then a linear programming boosting method was trained to classify AD with sMRI image. However some limitations include 1) computationally intensive and over-fitting due to high dimensionality of feature vectors extracted from image with the relatively small number of images for model training; 2) neglecting the regional information that has been proven important to AD diagnosis;

The second category is the methods that use regional features. The majority of the works in this category mainly relied on prior knowledge to ROI. Several works in the literature extracted features from the predetermined ROI based on biological prior knowledge the shrinkage of cerebral cortices and hippocampi, the enlargement of ventricles, and the change of regional glucose uptake [18], [19], [20]. Magnin [17] and Zhang [21] extracted regional features by splitting the whole brain into smaller regions to train the machine learning classifiers for AD diagnosis. The work in [16] used Gauss-Laguerre Harmonic Functions (GL-CHFs) and SURF [28] descriptors to extract local features on sMRI scans in hippocampus and posterior cingulate cortex (PCC) structures of the brain. Fan [29] partitioned the sMRI images into an adaptive set of brain areas based on watershed algorithm, and then extracted the regional volumetric features to train SVM-based AD classification. However, these aforementioned methods are based on empirically regions, which might neglect possible pathological locations in the whole brain. Moreover, the features extracted from ROIs may not be able to reflect the subtle changes involved in the brain [30].

In order to address these limitations, a hierarchical method was introduced by combination of global and regional features. Lian et al. divided sMRI images into small 3D patches and extracted features, and then combined the features hierarchically [26]. Suk et al. also proposed a systematic method for a joint feature representation from the paired patches of sMRI images [30] using the patch-based approach. These patch-based methods have been proven to efficiently deal with the problem of the high feature size and also the sensitivity to small changes. However, these models always require multi-stage training, which are not an end-to-end solution.

After determining the methods for feature extraction, a classifier need to be selected for automatic classification. The early work on computer-aided classification used several machine learning methods applied to sMRI. Support Vector Machine (SVM), Random forest (RF) and boosting were widely used to build classification models that is helpful for clinical diagnosis automation [5]. However, most of these methods rely on manual/semi-automatic feature selection which are inadequate to present the whole structural information.

Recently, deep learning has achieved a great success in

computer vision field, which has also become a popular and useful method for medical image analysis including Alzheimer's disease (AD) diagnosis based on MRI image. The convolutional neural network (CNN) [31] [32] has been proven to be suitable for grid-like data such as RGB images and MRI images. Billones et al. proposed a modified 16-layered VGG network to AD classification with sMRI images [25]. The method selected 20 central slices of sMRI image and reached high accuracy level on classification tasks using 900 sMRI images from the ADNI database. Residual network is the most widely used DCNN architecture which won the Imagenet classification competition [22]. It aims to alleviate the issue of disappearing/exploding gradients when the network is larger. In ResNet Block, a shortcut connection is added to link the input with the output, thus the Resnet learns the residual of input. Li et al.[33] proposed a deep network with residual blocks to classify sMRI images by using 1776 sMRI images of ADNI database.

B. Explainable deep learning

Due to the nature of blackbox, one challenge facing in the deep learning models is their explainable capability [34]. For the AD diagnosis task, most of existing deep learning based approaches lack transparency with difficulty in reasoning and explaining why and how a model decision is reached. To explain the image classification result by DCNN models, several explainable methods for CNNs have recently been proposed.

Saliency map [32] is firstly used for interpreting CNNs based models, which can highlight and explain which part of image features that contribute the most to the activity of a specific layer in a network or the decision of the network as a whole. It computes the gradients of logits based on the back-propagation algorithm and visualizes the feature contributions based on the amount of gradient they receive. This Saliency map is suitable for visualization but not good for localization and segmentation due to the noisy results [35]. Some improved methods based on saliency map [36], [37] have been proposed. The most widely used is the guided back-propagation by preventing the backward flow of negative gradients on ReLU activation from the higher layer in the CNN architecture [38], and other optimised visualization method such as PatternNet and PatternAttribution [39], Layer-Wise Relevance Propagation (LRP) [40] and DeepTaylor [41].

Class Activation mapping (CAM) is another explainable method for CNNs. In CAM merthod, the top fully connected layers was replaced by convolutional layers to maintain the object positions and can find the spatial distribution of distinguished regions for predict category [42].

The CAM requires retraining the model since it changed the model architecture. However, to address this issue, Grad-CAM has been proposed as a generalisation of the CAM method [43], which keeps the origin classification architecture and calculate the weight by pooling the gradient. This method has been widely used to explain the DCNN classification models. However, since Grad-CAM extracts the spatial distribution from the last layer of feature map with low resolution, this results in smaller size than the input image size. In order to obtain more accurate location information at high resolution, some optimised CAM methods were proposed, such as Adversarial Complementary Learning for Weakly Supervised Object Localization (ACOL) [44], Self-produced Guidance for Weakly-supervised Object Localization (SPG) [45] and guided attention inference networks (GAIN) [46]. As far as we know, only a few works presented the explanations of deep learning based AD Diagnosis. Montavon et al. and Yange et al. [41], [47] tried to explain 3D-CNNs by using visual interpretation methods. These methods were able to show how the CNNs made the classification decision. But there is no attempt made yet to explain 3D data classification tasks for diagnosis of MCI.

III. METHOD

The aim of this work is to develop an end-to-end deep learning framework to automatically classify discriminative atrophy localization on sMRI image for AD diagnosis, which consists of two levels of classifications: Alzheimer's disease (AD) vs. Normal cohort (NC) and progressive MCI (pMCI) vs. stable MCI (sMCI).

A. 3D Explainable Residual Self-Attention Convolutional Neural Network (3D ResAttNet)

We proposed a 3D explainable residual attention network (3D ResAttNet), a deep convolutional neural network that adopts self-attention residual mechanism and explainable gradient-based localisation class activation mapping (Grad-CAM) for AD diagnosis. The high-level conceptual framework is shown in Fig. 1, which consists of several major building blocks including: 3D Conv block, Residual Self-attention block, and Explainable blocks. The rationale behind of this architecture design includes:

1) The residual mechanism is designed to allow for more efficient training with fewer parameters for performance enhancement when increasing the depth of the network. Existing methods [22] have shown that residual learning can alleviate the issue of disappearance/exploding gradients when the network becomes larger. In addition, the residual connection avoids losing global features to ensure the integrity of the original information [48].
2) The self-attention mechanism is added to learn long-range dependencies. Capturing long-range dependencies is of central importance in deep neural networks. Since the convolution operator has a local receptive field, the long distance dependencies can only be processed after several convolutional layers. This is computational inefficient and causes optimisation difficulties.
3) The gradient-based localization class activation mapping (Grad-CAM) is introduced to provide visual explanations of predictions of Alzheimer's disease.

B. 3D CNN

Deep convolutional neural networks provide an effective way to learn multi-level features with multi-layers of convolutional operations in an end-to-end fashion [23]. Essentially,

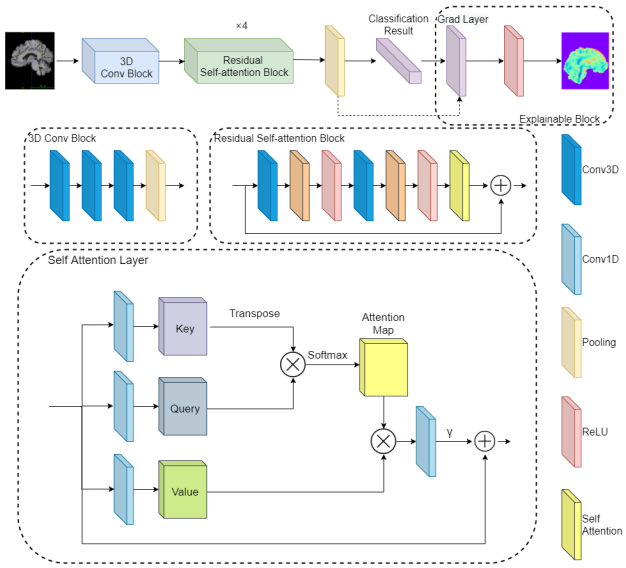


Fig. 1. The architecture of 3D residual attention deep Neural Network

the high-level features are obtained by composing low-level features and the levels of features can be enriched by the number of stacked layers (i.e. depth). We have used 3D CNNs. 3D convolutions apply a 3D filter to the dataset and the filter moves 3 directions x, y, z to calculate the low-level feature representations of the output shape as a 3-dimensional volume space. The stack of three $3 \times 3 \times 3$ convolutional layer is used to improve computing efficiency, compared to the widely used $7 \times 7 \times 7$ convolutional layer.

C. Residual Self-attention block (ResAttNet)

In this work, for the first time, we have combined self-attention with residual module to capture both global and local information based on 3D images to avoid information loss. Attention mechanism has been popular and a useful tool in recent year [49], [50], which can learn and focus on the important areas and suppress non-essential information through a weight matrix in whole image. On the other hand, this may cause global information loss. Therefore, we have added residual connection to address this issue. Residual network was originally designed to solve the issue of disappearance/exploding gradients when a network becomes deeper [22]. A residual connection was added between the origin input and the processed layer, which allow gradients to propagate more easily through the network.

1) Residual network layer:

A residual network can be formulated as follows:

$$y = x + r(x) \tag{1}$$

Where y is the output of the residual module, x is the input and $r(x)$ is the residual function. This module includes two Conv3D blocks consisting of $3 \times 3 \times 3$ 3D convolution layers, 3D batch normalization and rectified-linear-unit nonlinearity layer (ReLU).

2) Self-attention layer:

The self-attention layer was added to the end of original residual module $r(x)$. A self-attention function can be described as mapping a query, a key, and a value to the output, where those are all vectors. In our work, key and value are the features of the whole MRI image extracted by the 3D convolution network and query is the weight matrix that the network needs to learn. The key, query and value are denoted by $f(x)$, $g(x)$, $h(x)$ as follows:

$$Key: \ f(x) = W_f x \tag{2}$$

$$Query: \ g(x) = W_g x \tag{3}$$

$$Value: \ h(x) = W_h x \tag{4}$$

Here $x \in R^{C \times N}$ is the features from the previous layer. C is the number of channels and N is the number of feature locations of features from the previous layer. W_f, W_g and W_h are all $1 \times 1 \times 1$ convolution filters. The self-attention map $(a_{i,j})$ can be calculated as:

$$a_{i,j} = \frac{exp(f(x_i)^T g(x_j))}{\sum_{i=1}^{n} exp(f(x_i)^T g(x_j))} \tag{5}$$

where $a_{i,j}$ indicates the correlative degree of attention between each region i and all other regions. j is the index of an output position. The output of the attention layer is $o = (o_1, o_1 \ldots o_j, o_N) \in R^{C \times N}$, where

$$o_j = W_v(\sum_{i=1}^{N} a_{i,j} h(x_i)) \tag{6}$$

In order to keep the same number of channels as the original input and for memory efficiency, a $1 \times 1 \times 1$ convolution filter (W_v) was used to reduce the channel number of final outputs.

3) Residual Self-attention block (ResAttNet):

Therefore, the final output of the Residual Attention Block is given by:

$$y = x + r(x) + \gamma o(r(x)) \tag{7}$$

Where the $o(r(x))$ is the output of self-attention map, $r(x)$ is the output of original output of residual function and x is input feature, the γ is a learnable parameter. We set γ as 0 as default to allows the network to first rely on the cues in the local neighborhood. When γ increased, the model gradually learns to assign more weight to the non-local evidence.

D. The explainable 3D-CNNs

To understand inside the proposed deep model, the 3D Grad-CAM have been applied to explain the model decision.

We first calculated the gradient of the probabilities of disease areas with respect to the activation of unit k at location x, y, z in the last convolutional layer of the network. Then, the global average pooling of the gradients (a_k^c) was used to show the importance weights for unit k.

$$a_k^c = \frac{1}{Z} \sum_x \sum_y \sum_z \frac{\partial y(c)}{\partial A_{x,y,z}^k} \tag{8}$$

where Z is the number of voxels in the corresponding convolutional layer. Then, we combined the unit weights with

the activations, $A^k_{x,y,z}$, to get the heatmap of 3D gradient-weighted class activation mapping.

$$L^c_{3D-Grad-CAM} = ReLU\left(\sum_k a^c_k A^k_{x,y,z}\right) \quad (9)$$

IV. Experimental evaluation

A. Dataset description

Data used in this study were obtained from the ADNI cohort (http://adni.loni.usc.edu), consisting of baseline MRI scans of 1193 subjects from ADNI-1. The demographic information of subjects is presented in Table II.

TABLE I
DEMOGRAPHIC INFORMATION IN THE USED DATASET.GENDER REPORTS ARE MALE AND FEMALE. THE AGE, EDUCATION YEARS, AND MINI-MENTAL STATE EXAMINATION (MMSE) VALUES ARE REPORTED.

Category	Gentle (male/Female)	Age (Mean± Std)	Edu (Mean± Std)	MMSE (Mean± Std)	CDR (Mean± Std)
AD	202/187	75.95± 7.53	13.90± 4.97	23.28± 2.03	0.75± 0.25
pMCI	105/67	75.57± 7.13	15.69± 2.85	26.59± 1.71	0.50± 0.00
Smci	155/77	75.71± 7.87	15.56± 3.17	27.27± 1.78	0.49± 0.04
NC	202/198	76.02± 5.18	15.85± 3.51	29.10± 1.01	0.00± 0.00

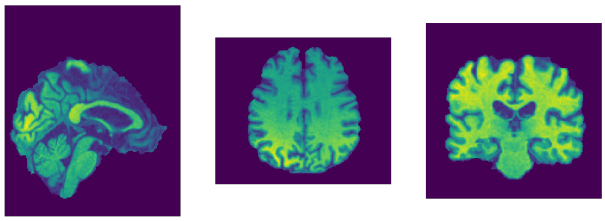
For this experiment, we have used a subset of ADNI structural MRI data that has been pre-processed with alignment and skull-stripping marked as "Spatially Normalized, Masked and N3 corrected T1 images". The subjects of MCI which converted to AD within 0.5 to 3 years from the baseline scan were defined as progressive MCI (pMCI), otherwise they are considered as stable MCI (sMCI). The whole dataset has 1193 scans of four classes: 389 of Alzheimer's Disease (AD) patients and 400 of Normal Cohort (NC), 232 sMCI and 172 pMCI.

As the original dataset is in Neuroimaging Informatics Technology Initiative (NIfTI) format, the preprocess is needed for spatial distortion correction caused by gradient nonlinearity and B1 field inhomogeneity, which is a standard pipeline process including anterior commissure (AC)-posterior commissure (PC) correction, intensity correction [51], skull stripping [52], and cerebellum removing. We have used MIPAV(Medical Image Processing, Analysis, and Visualization) application to complete AC-PC correction and use FSL(FMRIB Software Library v6.0) to complete skull stripping. A line align registration strategy (flirt instruction in FSL) is also executed to align every sMRI linearly with the Colin27 template [53] to delete global linear differences (including global translation, scale, and rotation differences), and also to re-sample all sMRIs to have identical spatial resolution. The sample images after the preprocessing are shown in Fig. 2.

B. Evaluation metrics

We have evaluated two binary classification tasks of AD classification (i.e., AD vs. NC) and MCI conversion prediction (i.e., pMCI vs. sMCI). The classification performance has been evaluated based on four commonly used standard metrics, including classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and Area under the curve (AUC). These metrics are defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$



a) Sagittal view b) Axial view c) Coronal view

Fig. 2. Sagittal View, Coronal View and Axial View of MRI image

$$SEN = \frac{TP}{TP + FN} \quad (11)$$

$$SPE = \frac{TN}{TN + FP} \quad (12)$$

where $TP = TruePositive$, $TN = TrueNegative$, $FP = FalsePositive$ and $FN = FalseNegative$. The AUC is calculated based on all possible pairs of SEN and $1 - SPE$ obtained by changing the thresholds performed on the classification scores yielded by the trained networks.

C. Experimental evaluation

We have evaluated our approach through comparing it with the state-of-the-art models for AD diagnosis. Two sets of evaluation have been conducted including 1) comparison study with state-of-the-art 3D convolutional neural networks; 2) Comparison study with other existing machine learning/deep learning methods for AD diagnosis.

1) Evaluation 1: Comparison study with state-of-the-art 3D convolutional neural networks:

We have performed comparison study with most commonly used 3D convolutional neural networks including 3D-VGGNet, 3D-ResNet under two conditions: with and without self-attention mechanism in 18 and 34 layers. The structures of 3D-VGG Block, 3D-ResNet Block and 3D-ResAttNet Block are shown in Fig. 3:

Each Conv3D layer consists of 3 consistent operations: 3D convolution, batch normalization 3D and RELU. The 14 layers and 34 layers network contain 8 and 14 3D Resnet block and 3D-ResAttNet block respectively. A $3 \times 3 \times 3$ 3D convolution is 3 times more expensive than 2D version in computational cost which will run out of GPU memory. In order to reduce the computational cost. We replace the $7 \times 7 \times 7$ convolution in the 3D Conv block with three conservative $3 \times 3 \times 3$ convolutions. The details configuration is shown in Fig. 4:.

To optimise model parameters, Adam, a stochastic optimisation algorithm, with a batch size of 8 samples, has been used for optimisation to train the proposed network [54]. We firstly set the learning rate (LR) as 1×10^{-4}. The LR was decreased to 1×10^{-6} with increased iterations. CrossEntropy has been selected as the loss function for this task [55]).

All the experiments have been implemented based on PyTorch and executed on a server with an Intel(R) Xeon(R) CPU E5-2650, NVIDIA TITAN and 64 GB memory.

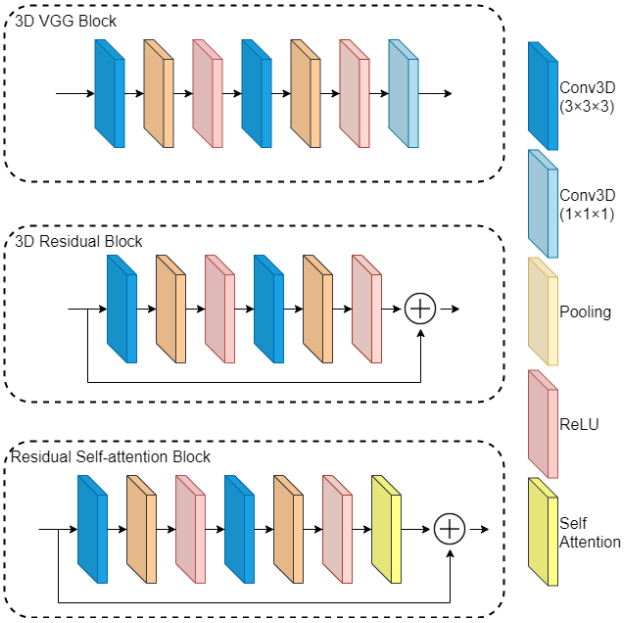


Fig. 3. The structure of 3D-VGG Block, 3D-ResNet Block and 3D-ResAttNet Block.

Layer name	Output size	3D-VGGNet	3D-ResNet18	3D-ResAttNet18	3D-ResNet34	3D-ResAttNet34
Input	181×217×181					
Input stem	91×109×91	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\3\times3\times3\end{pmatrix}$ Pooling layer				
Conv1	46×55×46	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}$ Pooling layer	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}\times2$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\Attention\end{pmatrix}\times2$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}\times3$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\Attention\end{pmatrix}\times3$
Conv2	23×28×23	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\1\times1\times1\end{pmatrix}$ Pooling layer	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}\times2$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\Attention\end{pmatrix}\times2$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}\times4$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\Attention\end{pmatrix}\times4$
Conv3	12×14×12	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\1\times1\times1\end{pmatrix}$ Pooling layer	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}\times2$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\Attention\end{pmatrix}\times2$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}\times6$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\Attention\end{pmatrix}\times6$
Conv4	6×7×6	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\1\times1\times1\end{pmatrix}$ Pooling layer	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}\times2$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\Attention\end{pmatrix}\times2$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\end{pmatrix}\times3$	$\begin{pmatrix}3\times3\times3\\3\times3\times3\\Attention\end{pmatrix}\times3$
Pooling	Max Pooling Global Average Pooling					
Total trainable parameters		13,673,172	33,817,992	33,822,088	64,127,624	64,131,720

Fig. 4. Overall architecture of 3D DCNN models, including 3D-VGGNet, 3D-ResNet 18, 3D-ResAttNet 18, , 3D-ResNet 34 and , 3D-ResAttNet 34.

2) Evaluation 2: Comparison study with other existing machine learning/deep learning methods for AD diagnosis:

For this evaluation, we have selected most recent and state-of-the-art machine learning methods reported in the literature for comparison using baseline sMRI data from ADNI for indirect comparison [10], [29], [15], [56], [57], [58], [59].

D. Result and discussion

1) Results from evaluation 1:

As introduced previously, we have added the attention mechanism in the Resnet block. In this group of experiments, we compared the models including 3D-VGGNet and 3D-ResNet models with and without attention layer. The results are presented in Table II. The classification performance of models with attention layer are significantly higher than models without it, especially on pMCI vs. sMCI classification. Our proposed model (3D-ResAttNet34) shows the best performance in all experiments. Fig. 7 shows the examples of classification results for two classification tasks: AD vs. NC and pMCI vs. sMCI tasks. Fig. 7 a) shows an example on AD vs. NC classification where the classification result using out proposed model 3D-ResAttNet34 with attention layer classify the image into a normal category (i.e. NC) while the result from 3D-ResNet34 classifies the image into disease category (i.e. AD). Similarly, Fig. 7 b) shows an example on pMCI vs. sMCI classification. When using our proposed method, the classification result is same as the ground truth.

TABLE II
RESULTS OF CLASSIFICATION FOR AD VS. NC AND PMCI VS. SMCI.

	AD vs. NC classification				pMCI vs. sMCI classification			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
3D-VGGNet	0.863	0.824	0.902	0.883	0.686	0.640	0.732	0.718
3D-ResNet18	0.943	0.917	0.968	0.888	0.792	0.756	0.828	0.758
3D-ResAttNet18	0.952	0.941	0.963	0.956	0.809	0.785	0.833	0.826
3D-ResNet34	0.965	0.966	0.964	0.975	0.824	0.795	0.853	0.845
3D-ResAttNet34	0.971	0.967	0.976	0.987	0.841	0.825	0.857	0.857

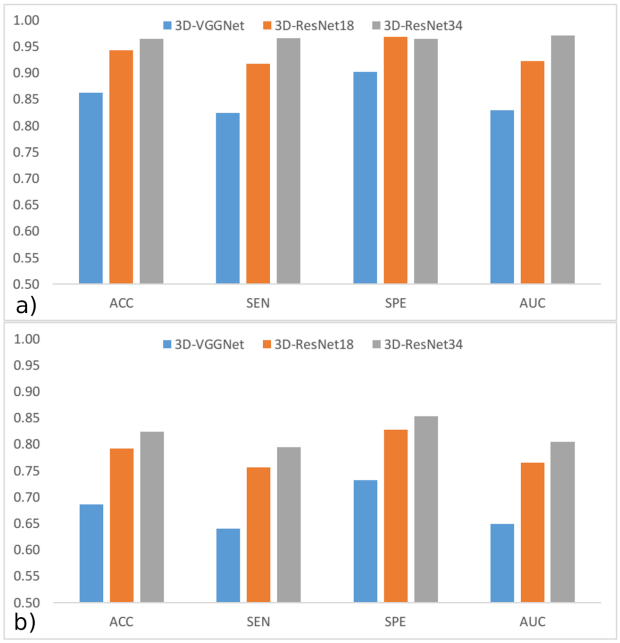


Fig. 5. Comparison between difference depth modes. a) and b) show the classification results of AD diagnosis and MCI conversion prediction, respectively.

For the explainable model evaluation, the 3D Grad-CAM has been used to explain the model for Alzheimer's Disease Diagnosis. We have only applied 3D Grad-CAM to 3D-ResAttNet34 with the best classification performance. The heat-map was created to show how the network learns the importance of the areas. As described earlier, 3D Grad-CAM can be used on an arbitrary layer. The output feature maps of each convolution block automatically extracted by the DCNN are shown in Fig. 8. As more convolutions are processed, the resolution of the feature map also gradually decreases. Some important areas and details information can be observed in the first two convolution layers. In the third and fourth convolution layers, the feature maps look like binary patterns where global feature maps can be extracted. The attention heatmap of 3D-ResAttNet34 result is presented in Fig. 9. In the first line, we have applied the activation mapping heat-map to last convolutional layer. The heatmap is blurry because the last convolutional layer of 3D-ResAttNet34 is only of size $6 \times 7 \times 6$. The heat map tends to show global information.

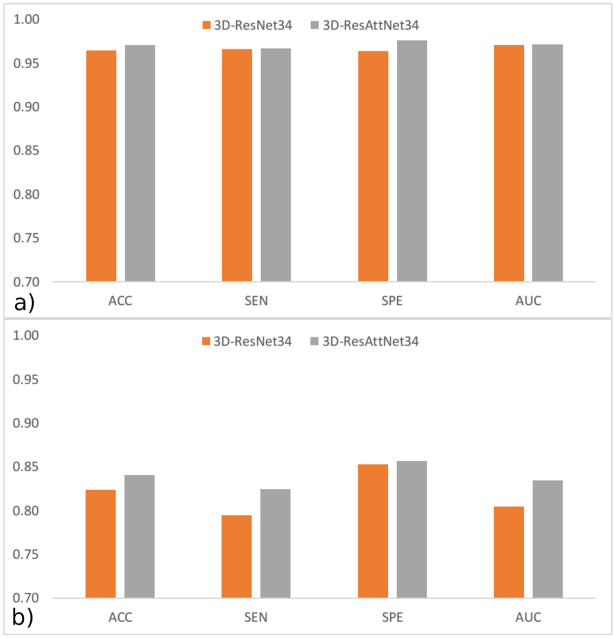


Fig. 6. Comparison between no attention layer (3D-ResNet34) and with attention layer (3D-ResAttNet34). (a) and (b) show the classification results of AD diagnosis and MCI conversion prediction, respectively.

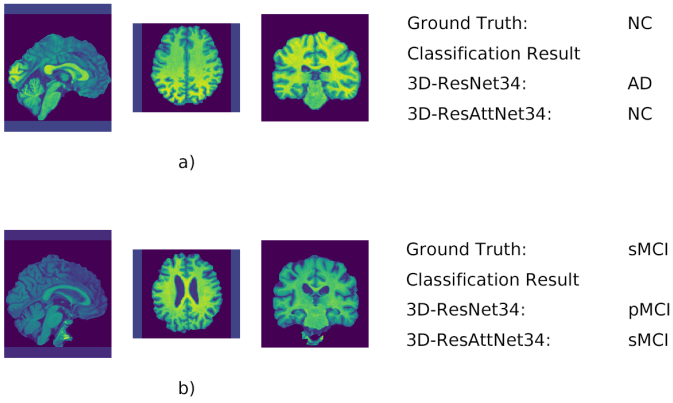


Fig. 7. Examples of classification results on a) the AD vs. ND classification task. The result shows the proposed method classified the image into the right category same as the ground truth (NC) while 3D ResNet34 classified into a wrong catagory (AD); b) the sMCI VS. pMCI task. The result shows the proposed method classified it into the right category same as the ground truth while 3D ResNet34 classified into a wrong category (pMCI)

To obtain a higher resolution and more detailed 3D class activation mapping heat-map, we have applied the 3D Grad-CAM to upper convolutional layers. The second line in Fig. 9 is obtained from 3D-ResAttNet34 with more convolutional layers removed. It is up sampled from a $46 \times 55 \times 46$ heat-map and thus provides more detail. It identifies and highlights the hippocampus, lateral ventricle and most parts of the cortex as important areas, which matches the human expert's approach [24], [60].

2) Results for evaluation 2:

The comparison results with 6 state-of-the-art machine learning based methods [61], [30], [62], [63], [64], [26] are shown in Table III. It shows that the proposed 3D-ResAttNet

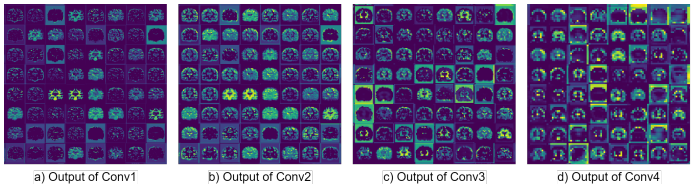


Fig. 8. Visualization results of selected convolutional layer feature maps. From left to right: first, second, third and fourth convolutional block.

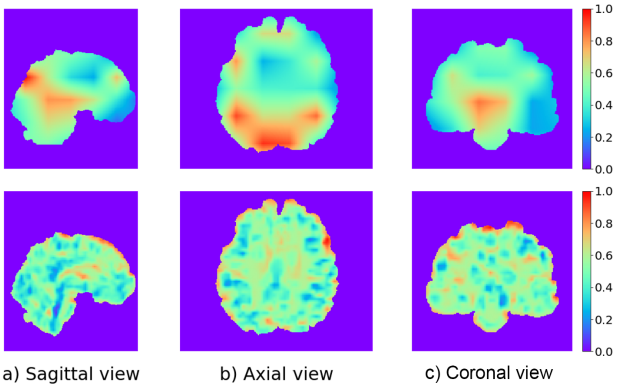


Fig. 9. Sagittal, Axial and coronal view of the brain MRI and the visual explanation heatmaps.

outperforms other approaches, especially on the challenging task of pMCI vs. sMCI classification. Compared with the traditional region- and voxel-level pattern analysis methods, our proposed method takes the whole MRI image as input and automatically extracts high dimensional and nonlinear features, which leads to better classification performance for AD diagnosis.

TABLE III
COMPARATIVE PERFORMANCE OF THE CLASSIFIER VS. FIVE COMPETITORS ON ADNI DATASET.

References	Modality	Method	AD vs. NC classification			pMCI vs. sMCI classification		
			ACC	SEN	SPE	ACC	SEN	SPE
(Liu et al., 2014)	MRI, PET	Stacked auto-encoder	0.88	0.89	0.87	0.77	0.74	0.78
(Suk et al., 2014)	MRI, PET	Deep Boltzmann machine	0.95	0.95	0.95	0.76	0.48	0.95
(Aderghal et al., 2017)	MRI	2D-CNN	0.91	0.94	0.89	0.66	0.66	0.65
(Liu et al., 2018)	MRI	Landmark detection + 3D CNN	0.91	0.88	0.94	0.77	0.42	0.82
(Shi et al., 2018)	MRI	Deep polynomial network	0.95	0.94	0.96	0.75	0.63	0.85
(Lian et al., 2018)	MRI	Hierarchical-CNN	0.903	0.824	0.965	0.809	0.526	0.854
Our 3D-ResAttNet	MRI	3D-CNN	0.971	0.967	0.976	0.841	0.825	0.857

V. CONCLUSION

Inspired by the attention mechanism and residual learning, we have proposed an end-to-end framework based on 3D Residual Self-Attention Network (3D ResAttNet) for better early diagnosis of AD diseases at two levels (i.e., AD vs. NC and pMCI vs. sMCI) from sMRI scans. The proposed method combines residual learning with self-attention mechanism, which can fully exploit both global and local information and avoid the information loss. Meanwhile, to understand inside our model and how our model reach decisions, we have also applied the 3D Grad-CAM method to identify and visualise those important areas contributing to our model decisions. The experimental evaluation has been performed by comparing it with the state-of-the-art AD diagnosis models on a large

cohort of subjects from real datasets. The experimental results show that the proposed approach outperforms the state-of-the-art models with significant performance improvement. The accuracies for AD vs. NC and sMCI vs. pMCI task are 97.1% and 84.1% respectively. The explainable mechanism in our approach regions is able to identify and highlight the contribution of the important brain parts (hippocampus, lateral ventricle and most parts of the cortex) for transparent decisions.

ACKNOWLEDGMENT

The work reported in this paper has formed part of the project by Royal Society - Academy of Medical Sciences Newton Advanced Fellowship (NAF\R1\180371).

REFERENCES

[1] WHO. Dementia. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia/

[2] O. B. Ahmed, M. Mizotin, J. Benois-Pineau, M. Allard, G. Catheline, C. B. Amar, and A. D. N. Initiative, "Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex," *Computerized Medical Imaging and Graphics*, vol. 44, pp. 13–25, 2015.

[3] C. Huang, L.-O. Wahlund, O. Almkvist, D. Elehu, L. Svensson, T. Jonsson, B. Winblad, and P. Julin, "Voxel- and VOI-based analysis of SPECT CBF in relation to clinical and psychological heterogeneity of mild cognitive impairment," *NeuroImage*, vol. 19, no. 3, pp. 1137–1144, Jul. 2003.

[4] G. B. Frisoni, N. C. Fox, C. R. Jack Jr, P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, p. 67, 2010.

[5] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811917302823

[6] H. Allioui, M. Sadgal, and A. Elfazziki, "Utilization of a convolutional method for Alzheimer disease diagnosis," *Machine Vision and Applications*, vol. 31, no. 4, p. 25, May 2020. [Online]. Available: http://link.springer.com/10.1007/s00138-020-01074-5

[7] E. Hosseini-Asl, G. Gimel'farb, and A. El-Baz, "Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network," *arXiv preprint arXiv:1607.00556*, 2016.

[8] A. Payan and G. Montana, "Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks," *arXiv preprint arXiv:1502.02506*, 2015.

[9] T. D. Vu, H.-J. Yang, V. Q. Nguyen, A.-R. Oh, and M.-S. Kim, "Multimodal learning using convolution neural network and Sparse Autoencoder," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2017, pp. 309–312.

[10] J. C. Baron, G. Chetelat, B. Desgranges, G. Perchey, B. Landeau, V. De La Sayette, and F. Eustache, "In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease," *Neuroimage*, vol. 14, no. 2, pp. 298–309, 2001.

[11] J. Ashburner and K. J. Friston, "Why voxel-based morphometry should be used," *Neuroimage*, vol. 14, no. 6, pp. 1238–1243, 2001.

[12] J. Ashburner and K. J, "Voxel-based morphometrythe methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.

[13] A. Mechelli, C. J. Price, K. J. Friston, and J. Ashburner, "Voxel-based morphometry of the human brain: methods and applications," *Current medical imaging reviews*, vol. 1, no. 2, pp. 105–113, 2005.

[14] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change," *IEEE transactions on Medical Imaging*, vol. 25, no. 5, pp. 626–639, 2006.

[15] S. Klppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimers disease," *Brain : a journal of neurology*, vol. 131, no. Pt 3, pp. 681–689, Mar. 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2579744/

[16] O. B. Ahmed, J. Benois-Pineau, M. Allard, C. B. Amar, G. Catheline, and A. D. N. Initiative, "Classification of Alzheimers disease subjects from MRI using hippocampal visual features," *Multimedia Tools and Applications*, vol. 74, no. 4, pp. 1249–1266, 2015.

[17] B. Magnin, L. Mesrob, S. Kinkingnhun, M. Plgrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehricy, and H. Benali, "Support vector machine-based classification of Alzheimers disease from whole-brain anatomical MRI," *Neuroradiology*, vol. 51, no. 2, pp. 73–83, 2009.

[18] E. Gerardin, G. Chtelat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S. Kim, M. Niethammer, B. Dubois, S. Lehricy, and L. Garnero, "Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging," *Neuroimage*, vol. 47, no. 4, pp. 1476–1486, 2009.

[19] B. Gutman, Y. Wang, J. Morra, A. W. Toga, and P. M. Thompson, "Disease classification with hippocampal shape invariants," *Hippocampus*, vol. 19, no. 6, pp. 572–578, 2009.

[20] V. Planche, A. Ruet, P. Coup, D. Lamargue-Hamel, M. Deloire, B. Pereira, J. V. Manjon, F. Munsch, N. Moscufo, and D. S. Meier, "Hippocampal microstructural damage correlates with memory impairment in clinically isolated syndrome suggestive of multiple sclerosis," *Multiple Sclerosis Journal*, vol. 23, no. 9, pp. 1214–1224, 2017.

[21] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, and A. D. N. Initiative, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] Y. Mu and F. H. Gage, "Adult hippocampal neurogenesis and its role in Alzheimer's disease," *Molecular Neurodegeneration*, vol. 6, p. 85, Dec. 2011.

[25] C. D. Billones, O. J. L. D. Demetria, D. E. D. Hostallero, and P. C. Naval, "Demnet: A convolutional neural network for the detection of Alzheimer's Disease and Mild Cognitive Impairment," in *2016 IEEE Region 10 Conference (TENCON)*. IEEE, 2016, pp. 3724–3727.

[26] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis using Structural MRI," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8585141/

[27] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, S. C. Johnson, A. D. N. Initiative *et al.*, "Spatially augmented lpboosting for ad classification with evaluations on the adni dataset," *Neuroimage*, vol. 48, no. 1, pp. 138–149, 2009.

[28] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[29] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "COMPARE: classification of morphological patterns using adaptive regional elements," *IEEE transactions on medical imaging*, vol. 26, no. 1, pp. 93–105, 2006.

[30] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical Feature Representation and Multimodal Fusion with Deep Learning for AD/MCI Diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4165842/

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[33] H. Li, M. Habes, and Y. Fan, "Deep ordinal ranking for multi-category diagnosis of alzheimer's disease using hippocampal MRI data," *arXiv preprint arXiv:1709.01599*, 2017.

[34] I. Sample, "Computer says no: why making ais fair, accountable and transparent is crucial," *The Guardian*, vol. 5, pp. 1–15, 2017.

[35] J. Adebayo, J. Gilmer, I. Goodfellow, and B. Kim, "Local explanation methods for deep neural networks lack sensitivity to parameter values," *arXiv preprint arXiv:1810.03307*, 2018.

[36] M. Alber, S. Lapuschkin, P. Seegerer, M. Hägele, K. T. Schütt, G. Montavon, W. Samek, K.-R. Müller, S. Dähne, and P.-J. Kindermans, "innvestigate neural networks," *Journal of Machine Learning Research*, vol. 20, no. 93, pp. 1–8, 2019.

[37] M. Bojarski, A. Choromanska, K. Choromanski, B. Firner, L. Jackel, U. Muller, and K. Zieba, "Visualbackprop: efficient visualization of cnns," *arXiv preprint arXiv:1611.05418*, 2016.

[38] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.
[39] P.-J. Kindermans, K. T. Schütt, M. Alber, K.-R. Müller, D. Erhan, B. Kim, and S. Dähne, "Learning how to explain neural networks: Patternnet and patternattribution," *arXiv preprint arXiv:1705.05598*, 2017.
[40] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PloS one*, vol. 10, no. 7, 2015.
[41] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, "Explaining nonlinear classification decisions with deep taylor decomposition," *Pattern Recognition*, vol. 65, pp. 211–222, 2017.
[42] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.
[43] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
[44] X. Zhang, Y. Wei, J. Feng, Y. Yang, and T. S. Huang, "Adversarial complementary learning for weakly supervised object localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1325–1334.
[45] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," *arXiv preprint arXiv:1805.08318*, 2018.
[46] K. Li, Z. Wu, K.-C. Peng, J. Ernst, and Y. Fu, "Guided attention inference network," *IEEE transactions on pattern analysis and machine intelligence*, 2019.
[47] C. Yang, A. Rangarajan, and S. Ranka, "Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification," *arXiv:1803.02544 [cs, stat]*, Jul. 2018, arXiv: 1803.02544. [Online]. Available: http://arxiv.org/abs/1803.02544
[48] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual Attention Network for Image Classification," *arXiv:1704.06904 [cs]*, Apr. 2017, arXiv: 1704.06904. [Online]. Available: http://arxiv.org/abs/1704.06904
[49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
[50] X. Zhang, Y. Wei, G. Kang, Y. Yang, and T. Huang, "Self-produced guidance for weakly-supervised object localization," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 597–613.
[51] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE transactions on medical imaging*, vol. 17, no. 1, pp. 87–97, Feb. 1998.
[52] Y. Wang, J. Nie, P.-T. Yap, F. Shi, L. Guo, and D. Shen, "Robust Deformable-Surface-Based Skull-Stripping for Large-Scale Studies," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2011*, ser. Lecture Notes in Computer Science, G. Fichtinger, A. Martel, and T. Peters, Eds. Berlin, Heidelberg: Springer, 2011, pp. 635–642.
[53] C. J. Holmes, R. Hoge, L. Collins, R. Woods, A. W. Toga, and A. C. Evans, "Enhancement of MR images using registration for signal averaging," *Journal of Computer Assisted Tomography*, vol. 22, no. 2, pp. 324–333, Apr. 1998.
[54] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Jan. 2017, arXiv: 1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980
[55] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of operations research*, vol. 134, no. 1, pp. 19–67, 2005.
[56] J. Koikkalainen, J. Ltjnen, L. Thurfjell, D. Rueckert, G. Waldemar, and H. Soininen, "Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease," *NeuroImage*, vol. 56, no. 3, pp. 1134–1144, Jun. 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3554792/
[57] M. Liu, D. Zhang, and D. Shen, "Relationship Induced Multi-Template Learning for Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment," *IEEE transactions on medical imaging*, vol. 35, no. 6, pp. 1463–1474, 2016.
[58] C. Mller, Y. A. L. Pijnenburg, W. M. van der Flier, A. Versteeg, B. Tijms, J. C. de Munck, A. Hafkemeijer, S. A. R. B. Rombouts, J. van der Grond, J. van Swieten, E. Dopper, P. Scheltens, F. Barkhof, H. Vrenken, and A. M. Wink, "Alzheimer Disease and Behavioral Variant Frontotemporal Dementia: Automatic Classification Based on Cortical Atrophy for Single-Subject Diagnosis," *Radiology*, vol. 279, no. 3, pp. 838–848, Jun. 2016.
[59] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, I. Castiglioni, and Alzheimer's Disease Neuroimaging Initiative, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach," *Frontiers in Neuroscience*, vol. 9, p. 307, 2015.
[60] B. R. Ott, R. A. Cohen, A. Gongvatana, O. C. Okonkwo, C. E. Johanson, E. G. Stopa, J. E. Donahue, G. D. Silverberg, and Alzheimer's Disease Neuroimaging Initiative, "Brain ventricular volume and cerebrospinal fluid biomarkers of Alzheimer's disease," *Journal of Alzheimer's disease: JAD*, vol. 20, no. 2, pp. 647–657, 2010.
[61] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2014, pp. 1015–1018, iSSN: 1945-8452.
[62] K. Aderghal, M. Boissenin, J. Benois-Pineau, G. Catheline, and K. Afdel, "Classification of sMRI for AD Diagnosis with Convolutional Neuronal Networks: A Pilot 2-D+$$\epsilon $$Study on ADNI," in *MultiMedia Modeling*, ser. Lecture Notes in Computer Science, L. Amsaleg, G. . Gumundsson, C. Gurrin, B. . Jnsson, and S. Satoh, Eds. Cham: Springer International Publishing, 2017, pp. 690–701.
[63] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Medical image analysis*, vol. 43, pp. 157–168, 2018.
[64] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 173–183, 2018.

Xin Zhang Xin Zhang is associate researcher in Manchester Metropolitan University (MMU), he received the B.S degree from The PLA Academy of Communication and Commanding, China, in 2009 and Ph.D. degree in Cartography and Geographic Information System from Beijing Normal University(BNU), China, in 2014. His current research interests include remote sensing image processing and deep learning.

Liangxiu Han received the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2002. She is currently a Professor of computer science with the School of Computing, Mathematics, and Digital Technology, Manchester Metropolitan University. Her research areas mainly lie in the development of novel big data analytics and development of novel intelligent architectures that facilitates big data analytics (e.g., parallel and distributed computing, Cloud/Service-oriented computing/data intensive computing) as well as applications in different domains using various large datasets (biomedical images, environmental sensor, network traffic data, web documents, etc.). She is currently a Principal Investigator or Co-PI on a number of research projects in the research areas mentioned above.

Wenyong Zhu Wenyong Zhu received the B.S degree from Nanjing University of Aeronautics and Astronautics (NUAA), China, in 2019. And he is studying for a master's degree in NUAA. His current research interests include machine learning and medical image classification.

Sun Liang Liang Sun received the B.S degree from Shandong University of Science and Technology, China, in 2014, and Ph.D. degree in Computer Science and Technology from Nanjing University of Aeronautics and Astronautics (NUAA), China, in 2020. His current research interests include machine learning and medical image segmentation.

Daoqiang Zhang received the B.Sc. and Ph.D. degrees in computer science from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1999 and 2004, respectively. He is currently a Professor in the Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics. His current research interests include machine learning, pattern recognition, and biomedical image analysis. In these areas, he has authored or coauthored more than 100 technical papers in the refereed international journals and conference proceedings.