Neurocomputing 361 (2019) 185–195

Contents lists available at ScienceDirect

Neurocomputing

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

Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease

Fan Zhang[a,b,*], Zhenzhen Li[a], Boyan Zhang[c], Haishun Du[a,b], Binjie Wang[d], Xinhong Zhang[b,e,**]

[a] School of Computer and Information Engineering, Henan University, Kaifeng 475001, China
[b] Institute of Image Processing and Pattern Recognition, Henan University, Kaifeng 475001, China
[c] School of Mechanical, Electrical and Information Engineering, Shandong University at Weihai, Weihai 264209, China
[d] Huaihe Hospital of Henan University, Kaifeng 475001, China
[e] School of Software, Henan University, Kaifeng 475001, China

ARTICLE INFO

Article history:
Received 28 August 2018
Revised 5 March 2019
Accepted 21 April 2019
Available online 16 July 2019

Communicated by Zechao Li

Keywords:
Alzheimers disease
Auxiliary diagnosis
Deep learning
Correlation analysis
Multi-modal

ABSTRACT

Alzheimer's disease (AD) is one of the most difficult to cure diseases. Alzheimer's disease seriously affects the normal lives of the elderly and their families. The mild cognitive impairment (MCI) is a transitional state between the normal aging and Alzheimer's disease, and MCI is most likely to converted to AD later. MCI is often misdiagnosed as the symptoms of normal aging, which results to miss the best opportunity of treatment. Therefore, the accurate diagnosis of MCI is essential for the early diagnosis and treatment of AD. This paper presents a deep learning model for the auxiliary diagnosis of AD, which simulates the clinician's diagnostic process. During the diagnosis of AD, clinician usually refers to the results of various neuroimaging, as well as the results of neuropsychological diagnosis. In this paper, the multi-modal medical images are trained by two independent convolutional neural networks. Then the consistency of the output of two convolutional neural networks is judged by the correlation analysis. Finally, the results of multi-modal neuroimaging diagnosis are combined with the results of clinical neuropsychological diagnosis. The proposed model provides a comprehensive analysis about patient's pathology and psychology at the same time, therefore, it improves the accuracy of auxiliary diagnosis. The diagnosis process is closer to the process of clinician's diagnosis and easy to implement. Experiments on the public database of ADNI (Alzheimer's disease neuroimaging initiative) show that the proposed method has better performance, and can achieve an excellent diagnostic efficiency in the auxiliary diagnosis of AD.

1. Introduction

Alzheimer's disease (AD) is one of the most difficult diseases to cure. Alzheimer's disease, commonly refers to the senile dementia, is a degenerative neurological disease that manifests as the progressive loss of cognition and memory. After cardiovascular disease, cancer and stroke, Alzheimer's disease is the fourth leading cause of death in the world. AD is one of the most financially costly diseases. Alzheimers disease has taken over from cancer to become the most feared disease. It kills more people than breast cancer and prostate cancer combined. Gradually, bodily functions of AD patients are lost, ultimately leading to death. At least 50 million people with Alzheimer's disease in 2018. According to statistics of World Health Organization (WHO), 4–8% of adults may suffer from AD at the age of 65. After the age of 85, the risk of AD will increase to 35% [1,2]. Currently, the pathogenesis of AD is still not fully understood. The academic community usually believes that AD is related to the neurofibrillary tangles (NFT) and the extracellular Amyloid-β (Aβ) deposition, which cause neurons and synapses loss or damage [3,4]. The mild cognitive impairment (MCI) is a state of early AD, which is a transitional state between the normal aging and Alzheimer's disease. MCI is often misdiagnosed as the symptoms of normal aging. However, 44% of MCI may eventually convert to AD within a few years [5]. Medication and psychotherapy can effectively slow down the deterioration of MCI and improve the lives quality of patients. Therefore, the accurate diagnosis of MCI is very important for the early diagnosis and treatment of AD. Currently, the research of Alzheimer's disease is one of the hottest topics in the medical research fields.

* Corresponding author at: School of Computer and Information Engineering, Henan University, Kaifeng 475001, China.
** Corresponding author at: Institute of Image Processing and Pattern Recognition, Henan University, Kaifeng 475001, China.
E-mail addresses: zhangfan@henu.edu.cn (F. Zhang), zxh@henu.edu.cn (X. Zhang).

https://doi.org/10.1016/j.neucom.2019.04.093
0925-2312/

At least 100 billion U.S. dollars are invested to the researches of diagnosis and treatment of Alzheimer's disease each year.

Currently, the clinical examination methods of Alzheimer's disease mainly include: the cerebrospinal fluid (CSF) examination, the electroencephalogram examination, the neuropsychological examination, the neuroimaging examination (including the molecular imaging detection), the genetic detection, *etc* [6]. The clinical neuropsychological examinations usually include the mini-mental state examination (MMSE), the Clinical Dementia Rating (CDR), the Weissler intelligence scale (WAS-RC), the activity of daily living (ADL), the Alzheimer's disease assessment scale-cognitive subscale (ADAS-Cog), *etc* [7–9]. The MMSE and CDR are the most commonly used methods in the clinical Alzheimer's psychology diagnosis. The MMSE and CDR can help clinician to determine the severity of dementia and easily to accept by patients and their families. The neuropsychological examination is just a clinical auxiliary diagnosis method. With the rapid development of neuroimaging technology, neuroimaging diagnosis becomes the most intuitive and the most reliable method for the diagnosis of Alzheimer's disease. Among of the neuroimaging methods, magnetic resonance imaging (MRI) is commonly used for the diagnosis of AD. It has high resolution for soft tissues of brain and can displays brain tissues in three dimensions, which can clearly distinguishes the gray matter and white matter of brain [10]. MRI can be divided into the structural magnetic resonance imaging (sMRI) and the functional magnetic resonance imaging (fMRI). The positron emission computed tomography (PET) is also a commonly used neuroimaging technology for the diagnosis of AD. It can show the distribution of lesions and the changes of glucose metabolism by imaging agents. The diffusion tensor imaging (DTI) and the diffusion kurtosis imaging (DKI) can reflect the structural properties of white matter cellulose in the brain, so they often used for the study of microstructural properties of white matter tracts [11]. Neuroimaging has a great impact on the understanding, diagnosis, and treatment of neurological diseases.

Neuroimaging diagnosis can be regarded as an image understanding problem. Li and Tang propose a novel weakly supervised deep matrix factorization (WDMF) algorithm, which uncovers the latent image representations and tag representations embedded in the latent subspace by collaboratively exploring the weakly-supervised tagging information, the visual structure and the semantic structure [12]. They also investigate the problem of learning knowledge from the massive community-contributed images with rich weakly-supervised context information, which can benefit multiple image understanding tasks simultaneously, the encouraging performance demonstrates its superiority [13].

Since the single-modal neuroimaging method only contains the part of information related to the brain abnormalities, it may be unsatisfactory for the classification of cognitively normal (CN), MCI and AD. The multi-modal neuroimaging method can provide more complementary information and can improve the accuracy of classification results theoretically [14,15]. The structural MRI images of AD patients reflect the changes of brain structure. PET is the functional imaging method that can acquire the functional features of brain to enhance the ability of finding lesions [16]. The fusion of MRI and PET is an effective multi-modal neuroimaging method, which can provide more accurate data for clinical diagnosis and treatment [17,18]. Tong et al. present a multi-modality classification framework to efficiently exploit the complementarity in the multi-modal data. Pairwise similarity is calculated for each modality individually using the features including regional MRI volumes, voxel-based FDG-PET signal intensities, CSF biomarker measures, and categorical genetic information. Similarities from multiple modalities are combined in a nonlinear graph fusion process, which generates a unified graph for final classification. The random forest method are used to calculate similarity matrices and perform classification [19]. Zhang et al. propose a multi-layer multi-view classification (ML-MVC) approach to explore the complex correlation between the features and class labels. The high-order complementarity among different views are captured by means of the underlying information with a low-rank tensor regularization [20].

Deep learning (DL) is a computational model which is composed of multiple processing layers to learn the representations of data with multiple levels of abstraction [21]. In spite of deep learning gets the remarkable success in the field of computer vision, but the classification and recognition of medical images is still an important challenge. In the past few years, the application of deep learning in medical image processing has developed rapidly. Suk et al. propose a deep learning based method to distinguish the cognitively normal (CN) and AD, CN and MCI, MCI and AD. The accuracy rates achieves 95.9% (CN vs. AD), 85.0%(CN vs. MCI), and 75.8% (MCI vs. AD), respectively [22]. The underlying features are extracted from PET and MRI images by a deep Boltzmann machine, and the support vector machine (SVM) method is used for the finally classification. However, their method only uses four-layer networks, which is difficult to extract the more abstract features of images.

Convolutional neural network (CNN) is a class of deep, feed-forward artificial neural networks. CNN is one of the most influential deep learning methods for the classification and recognition of images. CNN directly uses the two-dimensional images as the input data, and then automatically learns from the training data, which avoids the different calculation errors caused by the traditional manual extraction features. CNN can extract the higher-level features to capture the subtle lesion sites [23–25]. Currently, CNN has been used to identify AD and normal healthy human brains [26]. Sarraf and Tofighi use CNN to recognize the brains of AD and the brains of healthy human, resulted in the accuracy rate of 96.86% [27,28]. In their method, the sMRI and fMRI images are fused used to classify AD by LeNet-5 networks and Google-Net networks. Although the higher accuracy is achieved, but only healthy elderly people and Alzheimer's patients can be diagnosed, their method not be used for MCI.

Deep learning can also be combined with a resting-state brain network method to distinguish the CN, MCI and AD by using fMRI images. Ju et al. propose an early diagnosis method for Alzheimer's disease based on resting-state state brain network and deep learning [29]. The training model combined the fMRI images with the clinically relevant information to distinguish between normal aging, MCI and AD. Compared with the traditional method, the prediction accuracy rate is increased by 31.21%. Lv et al. design an intensive AlexNet network model for the diagnosis of CN, MCI and AD using MRI images [30]. AlexNet is the champion of ImageNet challenge in 2012, and it has a large impact in the field of machine learning, especially in the application of image processing [31]. AlexNet is a convolutional neural network, originally written with CUDA and running with GPU support. The intensive AlexNet achieves 100% sensitivity for the diagnosis of AD vs. CN. Shi et al. propose a Alzheimer's disease diagnosis algorithm using the multi-modal fusion method [32]. They use MRI and PET images at the same time, and each PET image is aligned with MRI image by the rigid registration algorithm. Then, 93 features are extracted from the same region of MRI and PET images respectively. Their method performs well in the diagnosis of dichotomous problems, but may meet problems in the multiple classification problems because of the manual intervention. The above mentioned studies have verified that deep learning can effectively diagnose MCI and AD, and they provide some research directions and ideas for the further research on auxiliary diagnosis of AD.

This paper presents a deep learning model for the auxiliary diagnosis of AD, which simulates the clinician's diagnostic process.

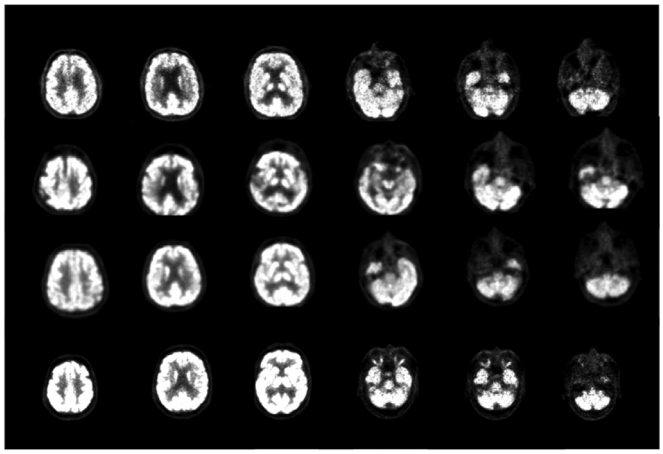


Fig. 1. PET images of AD patient.

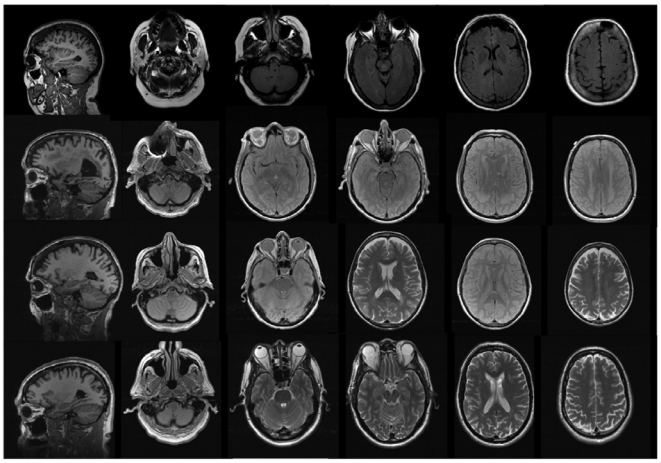


Fig. 2. MRI images of AD patient.

During the diagnosis of AD, clinician usually refers to the results of various of neuroimaging, as well as the results of neuropsychological diagnosis. In this paper, MRI, PET and the clinical neuropsychological diagnosis are combined. The multi-modal medical images are trained by deep learning method. Then the result of multi-modal neuroimaging diagnosis is combined with the result of clinical neuropsychological diagnosis. The proposed deep learning model provides a comprehensive analysis of patient's pathology and psychology, therefore, it improves the accuracy of auxiliary diagnosis.

2. Multi-modal deep learning model

Deep learning is a machine learning method that stems from the study of artificial neural networks (ANN) [33]. Deep learning technology learns features of data through the deep nonlinear network structures. It combines the low-level features to form the more abstract deep representations (attribute categories or features). Deep learning can realize the complex function approximation and learn the essential features of data sets. Deep learning architectures including the deep neural networks, the deep belief networks and the recurrent neural networks, *etc* [34]. Deep learning has been applied to many fields, such as the computer vision. Most of the modern deep learning models are based on the artificial neural network. A deep neural network (DNN) is an artificial neural network with multiple layers between the input and output layers. Convolutional neural network (CNN) is a class of feedforward artificial neural networks, which is commonly applied in image analysis [35,36].

PET images can reflect the changes in glucose metabolism of brain. In the normal human brain structures, such as the cerebral cortex, the subcortical structure and the cerebellum, the distribution of imaging agents is symmetrically in the PET images. In the brain of MCI patients, the distribution of imaging agents is abnormal. AD patients have a phenomenon of decreased glucose metabolism in the cingulate gyrus, the bilateral parietal lobe, the hippocampus, the temporal lobe and the frontal lobe. The left side of temporal and frontal lobe is obviously larger than the right side. Fig. 1 shows the PET images of a AD patient. Compared with PET image, MRI image can clearly show the boundaries of gray matter and white matter, and can provide the high resolution images for soft tissues. It can be observed from the MRI images that the atrophy degree of the hippocampus and the entorhinal cortex in the brain of MCI patients is more obvious than that of the normal aging peers. The atrophy degree of the cerebral cortex and the medial temporal lobe are more obvious in the brain of AD patients. Fig. 2 shows the MRI images of the same AD patient, including the coronal, sagittal, and transverse sections spin echo scan images. In this paper, the two different modal images are learned by two independent convolutional neural networks to evaluate the relationship between the brain function metabolism, the brain atrophy and the Alzheimer's disease.

Convolutional neural network (CNN) is currently a research hot spot of deep learning. Its weight-sharing network structure makes it more similar to the biological neural networks, which reduces the complexity of network models and reduces the number of weights. A convolutional neural network is a multi-layer artificial neural network, which mainly consists of the input layer, the convolutional layer, the down-sampling layer or a pooling layer, the fully connected layer and the output layer. Unlike the traditional neural network, the convolutional neural network is able to extract more abstract and generalized features from the original input data. The features are extracted through the convolutional layer and the down-sampling layer. The features are synthesized in the fully connected layer and finally output from the output layer. After many times trains and parameter adjustments, the best network model is finally obtained [37,38].

The convolutional layer is composed of several feature maps, each of which is composed of several neurons. Each neuron is locally connected with a feature map of the previous output layer by a weight matrix (the convolutional kernel). Compared with the full connected network structure, the number of neurons per convolutional layer is greatly reduced, which is helpful to reduce the computation complexity. The weight sharing means that all the neurons in a given convolutional layer respond to the same feature within their specific receptive field. The advantage of weight sharing is to reduce the training parameters in the network, effectively reduces the complexity of model and enhance the generalization ability. Each neuron in a neural network computes an output value according to the input values coming from the receptive field in the previous layer by applying activation function. The activation function is specified by a vector of weights and biases. Activation function of a neuron defines the output of neuron on a given input set. The nonlinear activation function such as Sigmoid function, Tanh function and ReLU function, *etc*. is used in CNN [39]. Different local features, such as edges and contours of image, can be extracted by convolution calculations, and deeper convolutional layers can extract more advanced features. According to these local features, comprehensive features can be obtained which are used to distinguish objects in image. The number of feature maps, the size of convolutional kernel, and the stride of convolution sliding will affect the features expression ability and learning ability,

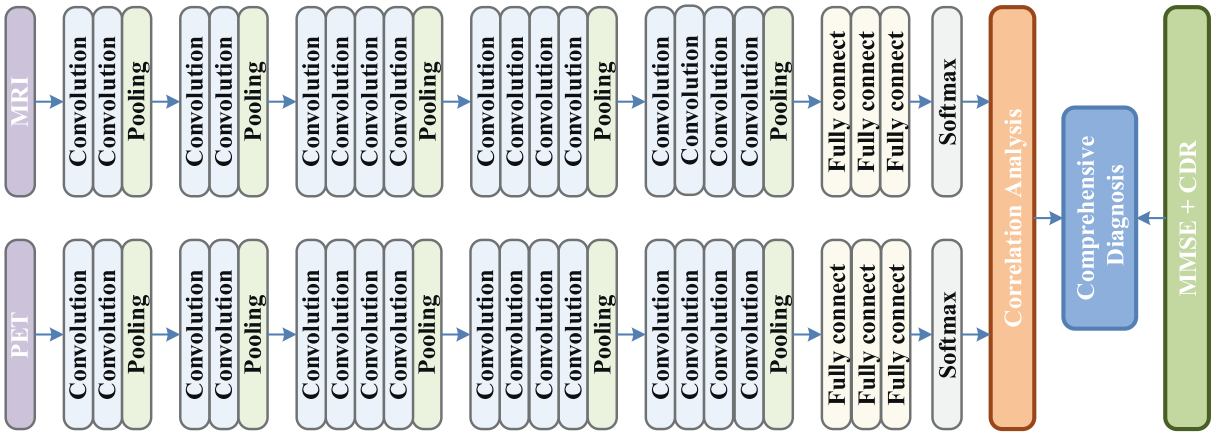


Fig. 3. The proposed auxiliary diagnosis model of Alzheimer's disease.

therefore, the network parameters should be selected appropriately according to the different applications.

Usually the convolutional layer and the down-sampling layer are alternately arranged. It is common to periodically insert a down-sampling layer between the successive convolutional layers in the CNN architectures. The down-sampling layer is also called the pooling layer. There are several non-linear functions to implement pooling. The down-sampling operation can reduce the number of neurons, and can ensure that the useful information in the image is preserved while reducing the feature dimension, and hence it can also control the over-fitting. The convolutional filter is slid from top to bottom and from left to right in a determined stride through a sliding window, and then the pixels in the block corresponding to the window are sampled and output. The pooling is further divided into the maximum pooling and the average pooling. The maximum pooling is equivalent to the sharpening of image, which can effectively compensate for the shortcomings of medical image features, such as the PET images. The average pooling can smooth the image and reduce the effects of noises.

After several convolutional layers and down-sampling layers, the high-level reasoning is done by the fully connected layers. Each neuron in the fully connected layer is connected to every neurons in the previous layer. The main purpose of the fully connected layer is to expand the multi-dimensional feature vector to a one-dimensional feature vector. The last fully connected layer will transfer the output value to the output layer, and the output layer will complete the task of identification and classification by the softmax logistic regression function [40]. In order to avoid the situation that the correct rate is high in the training set and the recognition effect is not good in practical applications, dropout technology [41] is usually used in the fully connected layer, that is, the output of some neurons will be set to 0 with a certain probability, which makes the neuron fail. Dropout can greatly reduce the complexity of mutual adaptation between neurons and can enhance the robustness of model.

This paper presents a deep learning model for the auxiliary diagnosis of Alzheimer's disease. Two independent convolutional neural networks are used to extract features from the MRI images and the PET images through a series of forward propagation convolution and down-sampling process. Then the consistency of the output of two convolutional neural networks is judged by the correlation analysis. If the results diagnosed by the two CNN models are similar, it is intuitive that the diagnosis by different modality is consistent for the same patient. Based on this idea, we calculate the correlation between the diagnosis result of PET images and the diagnosis result of MRI images, as the weight of the multi-modal neuroimaging diagnosis. Finally the neuroimaging diagnosis result and the clinical neuropsychological diagnosis result are integrated to make a comprehensive classification prediction. The advantage of this model is that it not only utilizes the method of deep learning, but also integrates the clinical neuropsychological diagnosis information. The diagnosis process is closer to the process of clinician's diagnosis. The auxiliary AD diagnosis model designed in this paper is shown in Fig. 3.

In this paper, two independent convolutional neural networks are used to learn the features of PET images and MRI images respectively. The structure of convolutional neural network is as follows:

(1) Convolutional layer

Assuming that the l layer is a convolutional layer, then the j feature map of the l layer is calculated as follows:

$$x_j^l = f\left(\sum_{i\in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right), \tag{1}$$

where x_i^{l-1} is the input feature map of the l layer, which is also the output feature map of the previous layer. M_j is the set of input feature map. k_{ij}^l represents the convolutional kernel corresponding to the partial input feature map. b_j^l represents the bias offset of the j feature map after convolution. $*$ represents the convolutional calculation, and $f(\cdot)$ represents the ReLU activation function. The essence of the above formula is to convolute the convolutional kernel with all the associated feature maps, then add an bias offset, and finally calculate the output value of the convolutional layer by the ReLU activation function.

(2) Down-sampling layer (pooling layer)

Assuming that the l layer is a down-sampling layer. In the down-sampling layer, the output feature map is in one-to-one correspondence with the input feature map. The calculation formula of the output feature map is as follows:

$$x_j^l = f\left[\beta_j^l \cdot down\left(x_j^{l-1}\right) + b_j^l\right], \tag{2}$$

where β_j^l and b_j^l represent the weight coefficients and bias offset parameters respectively, and different input feature maps correspond to different weight coefficients and bias offset parameters. $down(\cdot)$ represents the down-sampling function by dividing the input feature map into some non-overlapping regions with size $n \times n$ and then average pooling or maximum pooling the pixels in each region. The feature map is reduced to $1/n$ of the original size in each dimension. The down-sampling operation compresses the data, and improves the robustness and efficiency of computation.

(3) Fully connected layer

Assuming that the l layer is a fully connected layer. The fully connected layer is fully connected to the previous layer, that is, each neuron in the fully connected layer is connected to each neuron in the previous layer. The input two-dimensional feature map is expanded into a one-dimensional feature vector. The calculation formulas for the fully connected layer are as follows:

$$r^{l-1} = Bernoulli(p), \tag{3}$$

$$\tilde{x}^{l-1} = r^{l-1} * x^{l-1}, \tag{4}$$

$$\tilde{x}^{l} = f(w^{l}\tilde{x}^{l-1} + b^{l}), \tag{5}$$

where *Bernoulli* function randomly generates a vector r^{l-1} obeying the 0–1 distribution with a specified probability. The dimension of the vector is the same as x^{l-1}. The first two layers of the full connection layer use the dropout strategy to randomly block certain neurons according to a certain probability, which can effectively prevent the over-fitting phenomena in the deep networks. w^l and b^l are the weighting and offset parameters of the fully connected layer respectively. The calculation process of the fully connected layer is similar as the convolutional layer and the pooling layer. The output feature map after the dropout is weighted and added to the bias offset, the activation value is used as the input of the next layer by activation function. The local features extracted by the convolutional layer and the pooling layer are integrated into the global features.

(4) Output layer

The output of the final result uses the softmax function, and the calculation formulas are as follows:

$$x_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}, \tag{6}$$

$$y_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}, \tag{7}$$

where k is the total number of categories to be classified, and z_i is the i dimension component in the k dimension vector. x_i is the probability of the i class over k classes of the PET images, and y_i is the probability of the i class over k classes of the MRI images.

3. Correlation analysis

Correlation analysis is a statistical analysis method that studies the correlation between two or more random variables with the equal status. The correlation is a kind of non deterministic relation, and its analysis methods include the graph correlation analysis (line and scatter plot), the covariance and covariance matrix, the correlation coefficient, the regression analysis (simple linear regression and multiple regression), the information entropy analysis and the mutual information analysis. The correlation coefficient is a measure of the degree of linear correlation between the research variables. The correlation coefficient can be divided into the simple correlation coefficient, the multiple correlation coefficient, the canonical correlation coefficient and so on. The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and the standard deviation between the two variables. The output of convolution neural network can be understand as the result of neuroimaging diagnosis. After the deep learning training, the convolutional neural network classifies images (MRI or PET) into three categories: 0 represents cognitively normal (CN), 1 represents MCI and 2 represents AD (dementia). For example, if the output of a convolution neural network is 0, and the output of another convolution neural network is also 0, the two neuroimaging diagnoses are consistent and the result of neuroimaging diagnosis is reliable. If the output of another convolution neural network is 2, the diagnosis result of the deep learning model is unreliable. This paper calculates the correlation between the prediction of PET images and the prediction of MRI images based on Pearson correlation coefficient. The purpose of correlation analysis is to determine whether the results of two neuroimaging diagnoses are consistent. The calculation formula is as follows:

$$\gamma = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{8}$$

where $\bar{x}$ and $\bar{y}$ represent the mean of probabilities, and n is the number of categories for x and y. The γ value is between -1 and 1. The consistency of the prediction of two convolutional neural networks is judged by Pearson correlation coefficient. If the value of γ approaches 1, it means that the prediction of PET images by the convolutional neural network is consistent with the prediction of MRI images, and the diagnosis result of the multi-modal deep learning model is reliable. If the γ value approaches -1, the diagnosis result of the deep learning model is unreliable.

Based on the Pearson correlation coefficient, this paper presents the following formula to combine the neuroimaging diagnosis with the clinical neuropsychological diagnosis, and hence to obtain the comprehensive diagnosis result:

$$\begin{aligned}\text{CDscore} = {} & \lambda \times avg(\text{PETcnn} + \text{MRIcnn}) \\ & + (1-\lambda) \times avg(\text{MMSEscore} + \text{CDRscore}),\end{aligned} \tag{9}$$

where *avg* represents the mean function. PETcnn and MRIcnn represent the output results of two convolutional neural networks (neuroimaging diagnosis) respectively. MMSEscore and CDRscore represent the results of clinical neuropsychological examinations, this is, the score of the mini-mental state examination (MMSE), the clinical dementia rating (CDR) respectively. The value of CDscore is a comprehensive diagnosis result after the combination of the neuroimaging diagnosis and the clinical neuropsychological diagnosis. In the above formula, the influence of the neuroimaging diagnosis and the clinical neuropsychological diagnosis is controlled by the parameter λ. The calculation formula of λ is as follows:

$$\lambda = (1 + \gamma)/2, \tag{10}$$

where γ is the Pearson correlation coefficient between the prediction of PET images and the prediction of MRI images. When $\lambda = 1$, the algorithm only takes the neuroimaging diagnosis results. When $\lambda = 0$, the algorithm will focus on the diagnosis of clinical neuropsychology to prevent the possibility of other dementia types (for example, the dementia of Parkinson's disease). The value of λ determines the contribution of the different diagnosis results to the final comprehensive diagnosis result.

4. Experimental results and analysis

The experimental data of this paper mainly come from the database of Alzheimer's Disease Neuroimaging Initiative (ADNI), which is a large-scale clinical medical imaging database. ADNI makes all data and samples available for sharing with scientists worldwide [43,42]. ANDI is managed by the National Institute of Biomedicine Imaging and Bioengineering (NIBIB), the US Food and Drug Administration (FDA), the National Institutes of Aging (NIA), and other organizations. ADNI Database provides neuroimaging, biochemical, genetic biological markers, and other neuroscience data for research in all areas of AD. In addition to the data of ANDI, Huaihe hospital of Henan University also provides some clinical data for the experiments of this paper.

Table 1
The clinical information of the subjects.

	Number	Male/Female	Age	MMSE	CDR
CN	101	37/64	74.63 ± 4.8	23.25 ± 2.1	0 ± 0
MCI	200	68/132	74.97 ± 7.0	27.14 ± 1.7	0.5 ± 0.03
AD	91	34/57	75.48 ± 7.3	23.46 ± 2.1	0.8 ± 0.25

ADNI tracks and collects 2063 test subjects from 59 regions around the world. According to the Mini-Mental State Examination (MMSE) scores, subjects in ADNI can be divided into three categories: cognitively normal (CN) subjects, MCI subjects, and AD subjects. In the baseline ADNI-1 database, there are a total of 807 subjects, including 186 AD subjects, 226 CN subjects and 395 MCI subjects. The subjects in this paper are selected only to have both PET images (18F-FDG PET) and MRI images. The clinical information of selected subjects is shown in Table 1.

The original image archive of ADNI database is in the NIfTI format, and ADNI provides the preprocessing methods for MRI and PET images. All the pre-processed PET data is in the DICOM format, which simplifies the subsequent analysis process. The MRI images selected in this paper are derived from 1.5T magnetic resonance imaging device, and some of the images are derived from 3T magnetic resonance imaging devices. The MRI images are the pre-processed T_1 weighted MP-RAGE sequence images and the T_2 weighted sequence images. The image preprocessing methods will vary from device to device, and also depend on the system configuration. For example, for the images acquired from Philips MRI device, ADNI does not provide pre-processing method for the grad-warp correction and B_1 correction, so it is necessary to develop a special-purpose pre-processing method eliminating the gradient nonlinearity, the geometric distortion and the field inhomogeneity.

In this paper, MicroDicom software is used to filter the usable data (without missing information) in the ADNI database. We convert the image format from DICOM to PNG, which is more easily to process. The image size is adjusted to 224 × 224. Deep learning method requires a large amount of data, so the data set needs to be expanded. Different patients have different postures in the process of medical imaging, and it is impossible to maintain the same posture at all times. PET imaging needs to inject imaging agents, and the inspection time is longer. The FDG-PET imaging obtains 65 frames within 30 min to 60 min after the injection. The data set can be expanded by image rotating *etc*., which can make up for the number of images with different postures. Increasing the number of data samples can effectively prevent over-fitting and improve the generalization ability. In this paper, the images are processed by flipping, scaling and rotation to expand the data set.

In this paper, the data set is divided into the training set (90%), the test set (5%) and the verification set (5%), respectively. In order to make the training model have better generalization ability, division of data set is random. Considering the condition of data set and the performance of CNN models, this paper finally selects the 19-layer VGG (Visual Geometry Group) convolutional neural network model for the PET and MRI images recognition [44,45]. VGGNet-19 network model consists of 19 weight layers. VGGNet-19 has very deep convolutional architectures with smaller sizes of convolutional kernel (3 × 3), stride (1 × 1), and pooling window (2 × 2). The specific network structure of VGGNet-19 used in this paper is shown in Fig. 3.

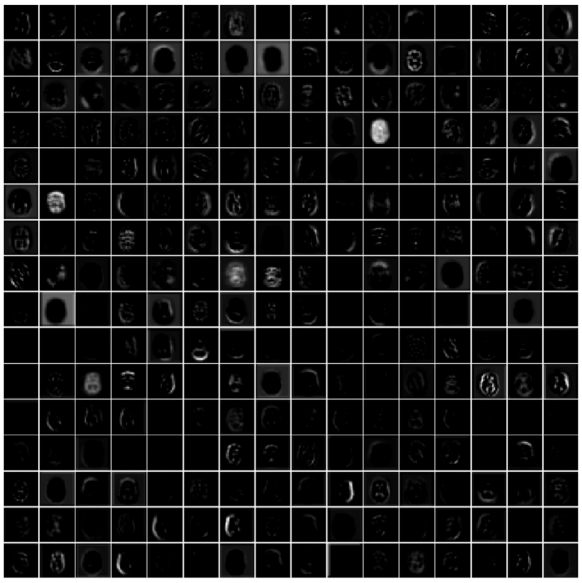
The input data source is the 224 × 224 × 3 images. After the calculation of five convolution networks (ConvA, ConvB, ConvC, ConvD, ConvE), the extracted feature vectors are input into three fully connected layers, and the classification result is finally output. Taking the ConvA network as an example, it consists of two convolutional layers and one maximum pooling layer. Each convolutional layer has 64 convolutional kernels (3 × 3), the stride is 1, and the padding is 1. The size of convolutional data is 224 × 224 × 6 4 (224 = (224 + 2 × 1 − 3)/1 + 1). The down-sampling layer uses the maximum pooling method. The maximum pooling size is 2 × 2. The maximum value neuron in the region is selected, and the data size of the next layer is 112 × 112 × 6 (112 = (224 − 2)/2 + 1). The calculation process of ConvB is similar as ConvA. The calculation processes of ConvC, ConvD, and ConvE are also similar as ConvA, except that they contain four convolutional layers. In the ConvE network, data are processed into a long vector and then input to the fully connected layer FC. There are three fully connected layers in total. The first fully connected layers contain 4096 neurons. The dropout strategy stops the output of some neurons according to a certain probability, which effectively prevents the over-fitting phenomenon due to the network depth. The dropout value is 0.5. The number of neurons in the last fully connected layer is consistent with the number of categories. In this paper, the VGGNet-19 network outputs three categories, which represent CN, MCI, and AD respectively.

The experimental environment of this paper is the Tensorflow deep learning framework based on Ubuntu 16.04 operating system. It is configured to accelerate GPU computing with NVIDIA CUDA 8.0. The software used for experiments is Matlab 2017a. In this paper, the convolutional neural network structure is used to extract the low-level and high-level features from a large amount of data, thus producing a high-precision prediction model. The convolutional neural network training process in this paper is as follows: firstly, the parameters of network model are initialized. The initialization weights of network are particularly important for the deep neural network. In order to avoid the phenomenon of stopping learning caused by the gradient instability, the Glorot uniform distribution initialization method is adopted. The parameter generation obeys [$-limit$, $limit$] uniform distribution. The limit is $sqrt(6/(num\_in + num\_out))$. $sqrt$ is the square root function. The num_in and num_out are the input and output units of weight tensor respectively, which can reduce the gradient dispersion problems. Each convolutional layer performs convolution calculation on the input batch data, and introduces nonlinear elements into the neural network by using the ReLU activation function, which improves the network sparse expression ability and solves the complex classification problem. Through the forward propagation, the output value of each layer is calculated, and then the error is propagated by layer-by-layer back propagation, and the weight parameters of each layer are adjusted according to the residual error. In this paper, the basic learning rate is set to 0.01, and the learning rate is reduced by 0.1 per 5000 iterations. The termination of training is determined by the number of iterations.

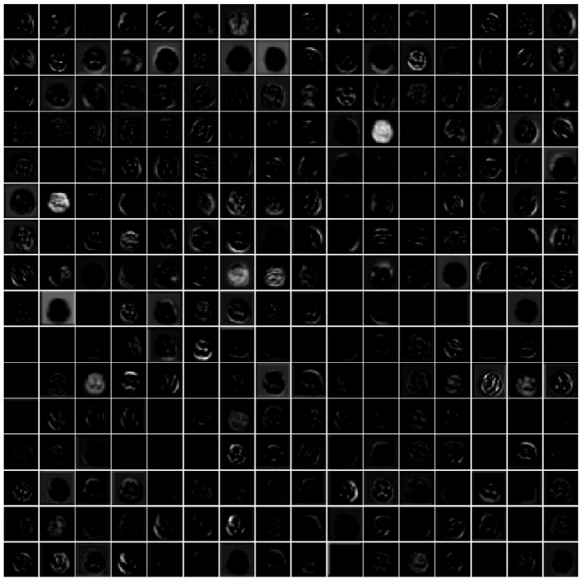
The human brain structure is complicated. It can be divided into three basic units: the forebrain, the midbrain and the hindbrain. The human brain can be segmented into 116 macro-regions according to the AAL (Automated Anatomical Labeling) atlas (90 cortical/subcortical regions and 26 cerebellar/vermis regions). VGGNet-19 reduces the size of convolution kernel and increases the number of convolution kernels, therefore, it is more suitable for the processing of brain images and can extract more abstract features. In VGGNet-19, ConvA extracts the lowest level features of input image, and each convolutional kernel extracts different features. ConvA can extract 64 local features from the input data, and the number of local features extracted by the subsequent convolutional networks are 128, 256 and 512, respectively.

Fig. 4 shows the partial features visualization of a PET image extracted by VGGNet. The partial features visualization of an MRI image extracted by VGGNet is shown in Fig. 5.

Fig. 7 is the accuracy curve of the MRI image classification model. The VGGNet-19 network has a larger depth and larger parameters amount. It can be seen from Fig. 7 that the VGGNet-19

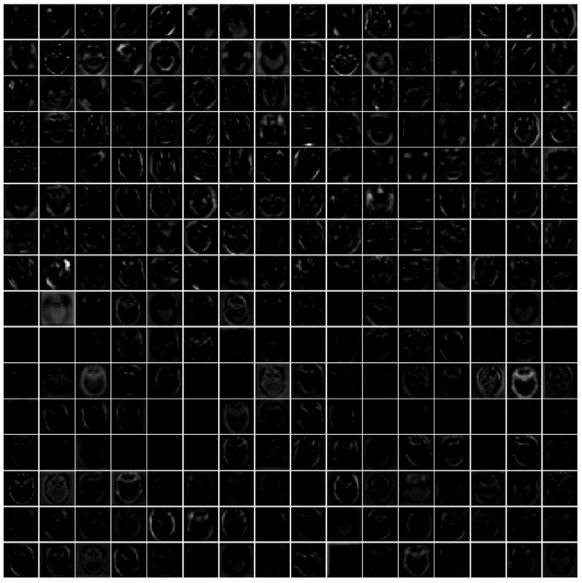


(a) MCI

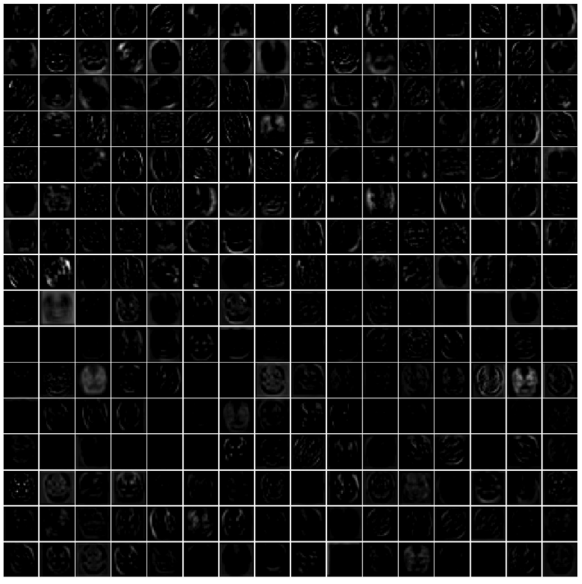


(b) AD

Fig. 4. Visualization of PET image local feature extracted by VGGNet.



(a) MCI



(b) AD

Fig. 5. Visualization of MRI image local feature extracted by VGGNet.

network converges after a fewer of iterations. The horizontal axis indicates the number of iterations.

The softmax layer outputs a probability distribution. Cross entropy indicates the distance between what the model believes the output distribution should be, and what the original distribution really is. Cross entropy loss function is used to guide the training process of the neural network. Fig. 6 shows the loss curve of PET image classification model. The number of test samples is 2000, and the batch size is set to 32. It can be seen from the figure that the loss value gradually decreases during the experiment, and finally it tends to a stable state (no longer decrease). After many times of parameters adjustments, the two independent CNN models can provide the neuroimaging diagnosis of PET and MRI images with high accuracy.

The clinical neuropsychological criteria are shown in Table 2. The MMSEscore and CDRscore used in Eq. (9) come from Table 2. Finally, the value of CDscore is calculated by Eq. (9), and the comprehensive diagnosis results are obtained, which combines the neuroimaging diagnosis and the clinical neuropsychological diagnosis.

Table 2
Clinical neuropsychological diagnostic criteria.

	CN	MCI	AD
MMSE	27	10–27	£10
MMSEscore	0	1	2
CDR	0–0.5	0.5–1	1–3
CDRscore	0	1	2

The classification accuracy is not the only evaluation criteria for the results of Alzheimer's disease diagnosis. Usually the sensitivity and the specificity also involved in the evaluation of Alzheimer's disease diagnosis. Sensitivity is also known as the true positive (TP) rate, which is the probability that a patient be diagnosed as positive. In this paper, sensitivity refers to the probability of having AD or MCI and being correctly diagnosed. Specificity is also known as the true negative (TN) rate, which is the probability that a person

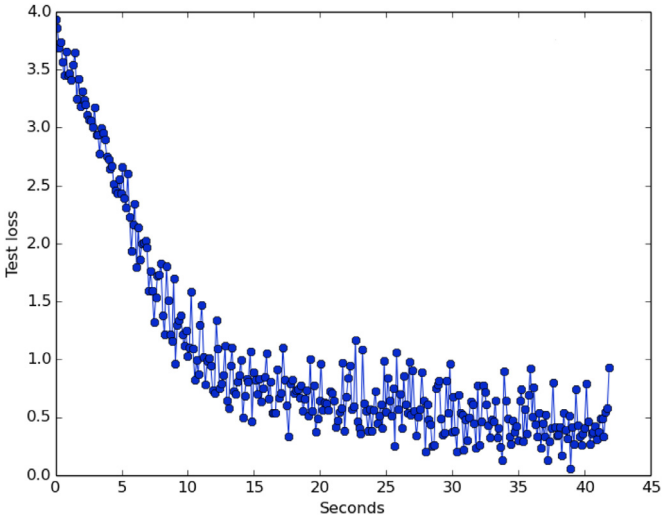


Fig. 6. Loss curve of the PET image classification model.

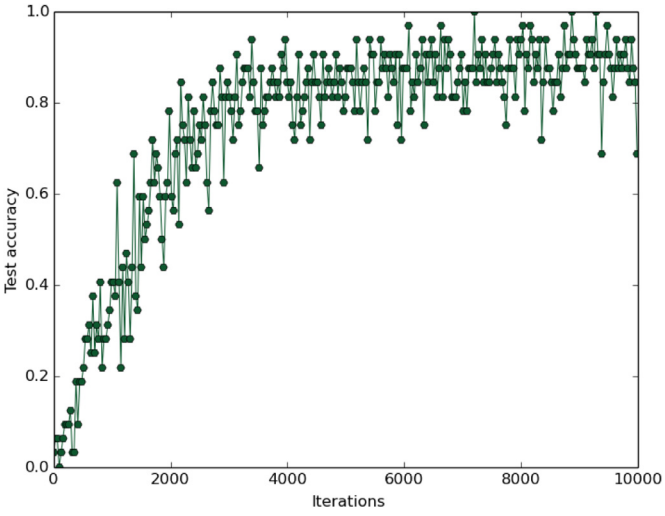


Fig. 7. Accuracy curve of the MRI image classification model.

who is actually not ill is diagnosed as negative. In this paper, specificity refers to the probability that the CN is correctly diagnosed. 1-Specificity is called the false positive (FP) rate, which is the probability that a person who is not ill but be diagnosed as positive. In the medical field, the higher sensitivity means the lower missed diagnosis rate. The lower 1-Specificity means the lower misdiagnosis rate. The calculation formulas of the sensitivity, the specificity and the accuracy are as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{11}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}, \tag{12}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{13}$$

where TP represents the true positive, that is, the number of AD (or MCI) samples predicted by the model as AD (or MCI). FN represents the false negative, which is the number of AD (or MCI) samples predicted as CN by the model. FP represents the false positive, which is the number of CN samples predicted as AD (or MCI) by the model. TN represents the true negative, which is the number of CN samples predicted as CN by the model.

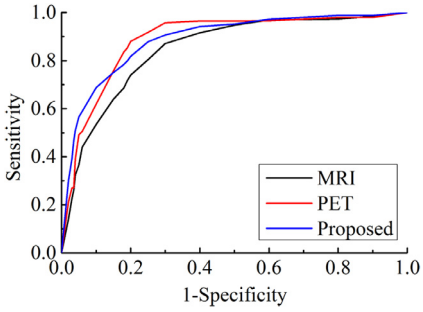


Fig. 8. AD vs. CN classification ROC curve.

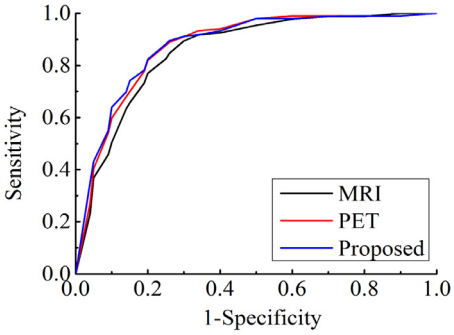


Fig. 9. AD vs. MCI classification ROC curve.

Ideally, we want the higher sensitivity and the higher specificity. However, in practice, we usually need to find a balance point between the sensitivity and the specificity. This process can be expressed by ROC (receiver operating characteristic) curve. ROC curve is a curve drawn according to the statistical data of two classification problem. In this paper, the ordinate of ROC curve indicates the sensitivity and the abscissa indicates the 1-Specificity, so the sensitivity and the specificity can be comprehensively measured by the ROC curve. The ideal indicator is that the sensitivity is 1, 1-Specificity is 0, that is, there are neither missed diagnosis nor misdiagnosis.

Figs. 8 and 9 are the experimental result classification ROC curve of AD vs. CN and AD vs. MCI respectively for the single-modal neuroimaging diagnosis and the multi-modal neuroimaging diagnosis. The area under the ROC curve reflects the accuracy of diagnosis. The single-modal neuroimaging diagnosis means that we only use MRI or PET images to train a VGGNet-19 deep convolution neural network, and the result of neuroimaging diagnosis is not be combined with the results of clinical psychological diagnosis.

The area under the ROC curve reflects the accuracy of diagnosis. It can be seen from the ROC curve that under the premise of using VGGNet-19 network to classify images, the comprehensive diagnosis of proposed method is more close to the results of the doctor's diagnosis compared with only one modal images is used to diagnose AD and CN. In other word, the multi-modal neuroimaging diagnosis is superior to the single modal neuroimaging diagnosis.

The area under ROC curve (AUC) is equivalent to the probability that a randomly chosen positive sample is ranked higher than a randomly chosen negative sample [51,52]. Assuming that the ROC curve is formed by sequential coordinates connection $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_m, y_m)\}$, the AUC can be estimated as follows,

$$\text{AUC} = \frac{1}{2}\sum_{i=1}^{m-1}(x_{i-1} - x_i)(y_i + y_{i+1}). \tag{14}$$

Table 3 shows a comparative analysis of experimental results. The various indicators of AD vs. CN, MCI vs. CN and AD vs. MCI classification are compared with the traditional methods and other

Table 3
The comparative analysis of the various indicators using the proposed method with the traditional methods and other deep learning methods.

Auxiliary diagnosis	Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
AD vs. CN	Liu et al. [46]	89.48	92.44	90.27	0.9697
AD vs. CN	Suk et al. [47]	94.65	95.22	95.35	0.9798
AD vs. CN	Intensive AlexNet [30]	**100.00**	93.57	96.14	0.9777
AD vs. CN	MM-SDPN (with SVM) [32]	97.13	**95.93**	**98.53**	0.9732
AD vs. CN	MM-SDPN (with LINEAR) [32]	96.93	95.02	98.37	0.9816
AD vs. CN	VGGNet-19 (MRI)	95.94	93.97	95.12	0.9789
AD vs. CN	VGGNet-19 (PET)	96.32	92.89	95.89	0.9828
AD vs. CN	The proposed method	96.58	95.39	98.47	**0.9861**
MCI vs. CN	Liu et al. [46]	**98.97**	52.59	83.90	0.8329
MCI vs. CN	Suk et al. [47]	95.37	65.87	85.67	0.8478
MCI vs. CN	Intensive AlexNet [30]	80.62	88.02	84.80	0.8081
MCI vs. CN	MM-SDPN (with SVM) [32]	87.24	**97.91**	67.04	0.8297
MCI vs. CN	MM-SDPN (with LINEAR) [32]	86.99	94.24	71.32	0.8808
MCI vs. CN	VGGNet-19 (MRI)	82.94	74.21	83.24	0.8467
MCI vs. CN	VGGNet-19 (PET)	84.54	75.36	84.62	0.8744
MCI vs. CN	The proposed method	90.11	91.82	**85.74**	**0.8815**
AD vs. MCI	3D CNN [48]	n/a	n/a	86.84	n/a
AD vs. MCI	Li et al. [49]	n/a	n/a	70.10	n/a
AD vs. MCI	Intensive AlexNet [30]	97.10	78.72	**89.66**	0.8691
AD vs. MCI	DLasso [50]	86.92	83.33	80.00	0.8026
AD vs. MCI	mfLasso [50]	84.62	83.26	84.00	0.8503
AD vs. MCI	VGGNet-19 (MRI)	86.76	74.65	82.41	0.8720
AD vs. MCI	VGGNet-19 (PET)	94.97	79.24	84.20	0.8792
AD vs. MCI	The proposed method	**97.43**	**84.31**	88.20	**0.8801**

state-of-the-art deep learning methods. The comparison methods include Liu et al. model [46], Suk et al. model [47], Intensive AlexNet model [30], MM-SDPN model [32], VGGNet-19 (MRI), VGGNet-19 (PET), 3D CNN model [48], Li et al. model [49], DLasso and mfLasso model [50]. Where the VGGNet-19 (MRI) and VGGNet-19 (PET) is the single-modal neuroimaging diagnosis that only MRI or PET images are used to train a VGGNet-19 deep convolution neural network, and the result of neuroimaging diagnosis is not be combined with the results of clinical psychological diagnosis.

From the comparison of experimental results in Table 3 we can see that the proposed model is superior to the traditional non-deep learning algorithms in the diagnosis sensitivity, specificity and accuracy. The experimental results of the proposed method are close to other deep learning methods. The proposed model incorporates the neuropsychological diagnosis information such as the MMSE and CDR. The MMSE is susceptible to the level of education. In the MMSE, MCI is similar to the normal aging elderly, so it is difficult to distinguish between MCI and CN, in other word, the MMSE is not very sensitive to MCI. In the clinical applications, the misdiagnosis is more unacceptable than the missed diagnosis, so a more specific diagnosis model is needed.

In the auxiliary diagnosis of AD vs. CN, and MCI vs. CN, the proposed method achieves the highest accuracy and the AUC value compare with other methods, while the sensitivity and specificity are also higher. In the auxiliary diagnosis of AD vs. MCI, the proposed method achieves the highest sensitivity, specificity and AUC value compare with other methods, and meanwhile the accuracy ranks the second. The sensitivity, specificity, accuracy and AUC value of proposed method achieve 97.39%, 84.27%, 88.25% and 0.8864, respectively. The experimental results show that the proposed multi-modal auxiliary diagnosis can achieve an excellent diagnostic efficiency.

5. Conclusions

This paper presents a deep learning model for the early auxiliary diagnosis of Alzheimer's disease, which simulates the diagnosis process of clinicians. During the diagnosis of AD, clinician usually refers to the results of various of neuroimaging, as well as the results of neuropsychological diagnosis. In this paper, the neuroimaging diagnosis and the clinical psychological diagnosis are combined. The multi-modal auxiliary diagnosis model are trained by the deep learning method. We designed two independent convolution neural networks for training of MRI and PET images. The two independent convolutional neural networks are used to extract MRI image features and PET image features through a series of calculations such as the convolution, the down-sampling and softmax. The consistency of the output of two convolutional neural networks is judged by correlation analysis. If the results of the two CNN models are similar, it is intuitive that the diagnosis for the same patient are consistent with the different modality diagnosis. Based on this idea, a new correlation calculation method is proposed. We calculate the Pearson correlation coefficient between the diagnosis result of PET images and the diagnosis result of MRI images. Then we combine the results of the multi-modal neuroimaging auxiliary diagnosis with the results of clinical psychological diagnosis, so the pathology and psychology of patients are analyzed in a comprehensive way. Finally, a comprehensive diagnosis result is got, which improves the accuracy of diagnosis. The advantage of the proposed model is that it combines the neuroimaging diagnosis with the clinical neuropsychological diagnosis. The diagnosis process is closer to the process of clinician's diagnosis and easy to implement. This paper expands the data set by image flipping, scaling and rotation. A large number of experiments on the open database of ADNI show that the diagnostic effect of the proposed method is superior to other auxiliary diagnostic models in many indicators. The experimental results show that the proposed multi-modal auxiliary diagnosis can achieve an excellent diagnostic efficiency.

Declarations of interest

None.

Acknowledgment

This research was supported by the Natural Science Foundation of China (Nos. 61771006 and U1504621), the Natural Science Foundation of Henan Province (No. 162300410032).

References

[1] K. Kumar, A. Kumar, R.M. Keegan, R. Deshmukh, Recent advances in the neurobiology and neuropharmacology of Alzheimers disease, Biomed. Pharmacother. 98 (2018) 297–307.

[2] M. Hecht, L.M. KrMer, V.A. Caf, M. Otto, D.R. Thal, Capillary cerebral amyloid angiopathy in Alzheimer's disease: association with allocortical/hippocampal microinfarcts and cognitive decline, Acta Neuropathol. 135 (5) (2018) 681–694.

[3] P. Theofilas, A.J. Ehrenberg, A. Nguy, J.M. Thackrey, S. Dunlop, M.B. Mejia, A.T. Alho, R.L. Paraizo, R.D. Rodriguez, C.K. Suemoto, Probing the correlation of neuronal loss, neurofibrillary tangles, and cell death markers across the Alzheimer's disease Braak stages: a quantitative study in humans, Neurobiol. Aging 61 (2018) 1–12.

[4] C. Wang, V. Saar, K.L. Leung, L. Chen, G. Wong, Human amyloid eptide and tau co-expression impairs behavior and causes specific gene expression changes in caenorhabditis elegans, Neurobiol. Dis. 109 (A) (2018) 88–101.

[5] Alzheimer's Association, 2018 Alzheimer's disease facts and figures, Alzheimers Dement. 14 (3) (2018) 367–429.

[6] Z.F. Dai, Applications, opportunities and challenges of molecular probes in the diagnosis and treatment of major diseases, Chin. Sci. Bull. 62 (1) (2017) 25–35.

[7] K.R. Chapman, H. Bingcanar, M.L. Alosco, E.G. Steinberg, B. Martin, C. Chaisson, N. Kowall, Y. Tripodis, R.A. Stern, Mini mental state examination and logical memory scores for entry into Alzheimers disease trials, Alzheimers Res. Therapy 8 (1) (2016) 9–20.

[8] M.E. Kowoll, C. Degen, S. Gladis, J. Schrder, Neuropsychological profiles and verbal abilities in lifelong bilinguals with mild cognitive impairment and Alzheimer's disease, J. Alzheimers Dis. 45 (4) (2015) 1257.

[9] P. Roberta, T.C. Stella, C. Giulia, F. Lucia, C. Carlo, C.G. Augusto, Neuropsychological correlates of cognitive, emotional-affective and auto-activation apathy in Alzheimer's disease, Neuropsychologia 18 (1) (2018) 12–21.

[10] G.A. Papakostas, A. Savio, M. Granā, V.G. Kaburlasos, A lattice computing approach to Alzheimer's disease computer assisted diagnosis based on MRI data, Neurocomputing 150 (PA) (2015) 37–42.

[11] Y. Chen, M. Sha, X. Zhao, J. Ma, H. Ni, W. Gao, D. Ming, Automated detection of pathologic white matter alterations in Alzheimer's disease using combined diffusivity and Kurtosis method, Psychiatry Res. 264 (2017) 35–45.

[12] Z. Li, J. Tang, Weakly supervised deep matrix factorization for social image understanding, IEEE Trans. Image Process. 26 (1) (2017) 276–288.

[13] Z. Li, J. Tang, M. Tao, Deep collaborative embedding for social image understanding, IEEE Trans. Pattern Anal. Mach. Intell. (2018). Early Access.

[14] R.J. Perrin, A.M. Fagan, D.M. Holtzman, Multimodal techniques for diagnosis and prognosis of Alzheimer's disease, Nature 461 (7266) (2009) 916–922.

[15] D.L. Bailey, B.J. Pichler, B. Gckel, H. Barthel, A.J. Beer, J. Bremerich, J. Czernin, A. Drzezga, C. Franzius, V. Goh, Combined PET/MRI: multi-modality multi–parametric imaging, Mol. Imaging Biol. 17 (5) (2015) 1–14.

[16] I. Riederer, K.P. Bohn, C. Preibisch, E. Wiedemann, C. Zimmer, P. Alexopoulos, S. Frster, Alzheimer disease and mild cognitive impairment: integrated pulsed arterial spin-labeling MRI and ^{18}F-FDG PET, Radiology 288 (1) (2018) 198–206.

[17] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, Multimodal classification of Alzheimer's disease and mild cognitive impairment, Neuroimage 55 (3) (2011) 856–867.

[18] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M.J. Fulham, Multimodal neuroimaging feature learning for multi-class diagnosis of Alzheimers disease, IEEE Trans. Biomed. Eng. 62 (4) (2015) 1132–1140.

[19] T. Tong, K. Gray, Q. Gao, L. Chen, D. Rueckert, Multi-modal classification of Alzheimer's disease using nonlinear graph fusion, Pattern Recognit. 63 (2017) 171–181.

[20] C. Zhang, E. Adeli, T. Zhou, X. Chen, D. Shen, Multi-layer multi-view classification for Alzheimer, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018, pp. 4406–4413.

[21] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

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

[23] P. Moeskops, M.A. Viergever, A.M. Mendrik, L.S. de Vries, M.J. Benders, I. Isgum, Automatic segmentation of MR brain images with a convolutional neural network, IEEE Trans. Med. Imaging 35 (5) (2016) 1252–1261.

[24] L. Chang, X.M. Deng, M.Q. Zhou, Z.K. Wu, Y. Yuan, S. Yang, H. Wang, Convolutional neural networks in image understanding, Acta Autom. Sin. 42 (9) (2016) 1300–1312.

[25] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, S. Mougiakakou, Lung pattern classification for interstitial lung diseases using a deep convolutional neural network, IEEE Trans. Med. Imaging 35 (5) (2016) 1207–1216.

[26] C.D. Billones, O.J.L.D. Demetria, D.E.D. Hostallero, P.C. Naval, DemNet: a convolutional neural network for the detection of Alzheimer's disease and mild cognitive impairment, in: Proceedings of the 2017 IEEE Region 10 Conference, 2017, pp. 3724–3727.

[27] S. Sarraf, G. Tofighi, Classification of Alzheimer's disease structural MRI data by deep learning convolutional neural networks, IEEE Trans. Med. Imaging 35 (5) (2016) 1252–1261.

[28] S. Sarraf, G. Tofighi, Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data, in: Proceedings of the 2017 IEEE Future Technologies Conference (FTC), 2017, pp. 816–820.

[29] R. Ju, C. Hu, P. Zhou, Q. Li, Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning, IEEE/ACM Trans. Comput. Biol. Bioinform. 16 (1) (2019) 244–257.

[30] Lv Hong-meng, D. Zhao, X.b. Chi, Deep learning for early diagnosis of Alzheimer's disease based on intensive AlexNet, Comput. Sci. 44 (S1) (2017) 50–60.

[31] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the 2012 International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105. Lake Tahoe, USA

[32] J. Shi, X. Zheng, Y. Li, Q. Zhang, S. Ying, Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease, IEEE J. Biomed. Health Inf. 22 (1) (2017) 173–183.

[33] Y. Bengio, Learning deep architectures for ai, Found. Trends Mach. Learn. 2 (1) (2009) 1–127.

[34] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.

[35] S. Lawrence, C.L. Giles, A.C. Tsoi, A.D. Back, Face recognition: a convolutional neural-network approach, IEEE Trans. Neural Netw. 8 (1) (1997) 98–113.

[36] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing 234 (2016) 11–26.

[37] L. Xiang, Q. Wang, D. Nie, L. Zhang, X. Jin, Y. Qiao, D. Shen, Deep embedding convolutional neural network for synthesizing CT image from T1-weighted MR image, Med. Image Anal. 47 (2018) 31–44.

[38] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. Cook, A.D. Marvao, T. Dawes, D. O'Regan, Anatomically constrained neural networks (ACNN): application to cardiac image enhancement and segmentation, IEEE Trans. Med. Imaging 37 (2) (2018) 384–395.

[39] T.N. Sainath, B. Kingsbury, G. Saon, H. Soltau, A.R. Mohamed, G. Dahl, B. Ramabhadran, Deep convolutional neural networks for large-scale speech tasks, Neural Netw. 64 (2015) 39–48.

[40] C.S. Chin, J.T. Si, A.S. Clare, M. Ma, Intelligent image recognition system for marine fouling using softmax transfer learning and deep convolutional neural networks, Complexity 2017 (12) (2017) 1–9.

[41] A. Poernomo, D.K. Kang, Biased dropout and crossmap dropout: learning towards effective dropout regularization in convolutional neural network, Neural Netw. 104 (2018) 60–67.

[42] C.R. Jack Jr., M. Bernstein, N. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. Britson, J. Whitwell, C. Ward, The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, J. Magn. Reson. Imaging 27 (4) (2008) 685–691.

[43] M. Liu, J. Zhang, P.T. Yap, D. Shen, View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data, Med. Image Anal. 36 (2017) 123–134.

[44] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the 2015 International Conference on Learning Representations, 2015, pp. 1–14. San Diego, USA

[45] Y. Tang, X. Wu, Scene text detection and segmentation based on cascaded convolution neural networks, IEEE Trans. Image Process. 26 (3) (2017) 1509–1520.

[46] M. Liu, D. Zhang, D. Shen, Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis, Hum. Brain Map. 35 (4) (2014) 1305–1319.

[47] H.I. Suk, S.W. Lee, D. Shen, A hybrid of deep network and hidden Markov model for MCI identification with resting-state fMRI, in: Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015, pp. 573–580.

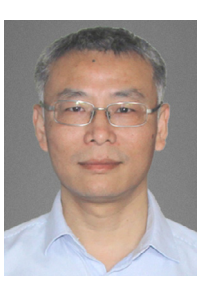
[48] A. Payan, G. Montana, Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, in: Proceedings of the 2015 International Conference on Pattern Recognition Applications and Methods, Lisbon, Portugal, 2015, pp. 1–9.

[49] F. Li, L. Tran, K.H. Thung, S. Ji, D. Shen, J. Li, A robust deep model for improved classification of AD/MCI patients, IEEE J. Biomed. Health Inf. 19 (5) (2015) 1610–1616.

[50] X. Wang, Y. Ren, W. Zhang, Multi-task fused Lasso method for constructing dynamic functional brain network of resting-state fMRI, J. Image Graph. 22 (7) (2017) 978–987.

[51] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.

[52] W. Yu, J.K. Kim, T. Park, Estimation of area under the ROC curve under nonignorable verication bias, Stat. Sin. 28 (4) (2018) 1–25.



Fan Zhang is a professor at the School of Computer and Information Engineering at Henan University, China, and he is also the director of the Institute of Image Processing and Pattern Recognition, Henan University, China, and the associate director of the Open Laboratory of Intelligent Technology and Systems, Henan Province, China. He is a lecturer for undergraduate and graduate programs, including operating system, digital image processing, and so on. He received his B.S. degree from North China University, China, and received his M.S. degree and Ph.D. degree from Jiangsu University, China and Beijing University of Technology, China respectively. His research focuses on digital image processing, pattern recognition, information visualization and deep learning.

Zhenzhen Li is a candidate of M.S. degree in Computer Science and Technology at School of Computer and Information Engineering, Henan University, China. His research interests include pattern recognition and computer vision.

Boyan Zhang is a senior undergraduate student at School of Mechanical, Electrical and Information Engineering, Shandong University at Weihai, China. His major is computer science. His research interests are digital image processing and deep learning.

Haishun Du received his Ph.D. degree from Southeast University, China, in 2007. Now, he is a professor at the School of Computer and Information Engineering, Henan University, China. His research interests include pattern recognition, computer vision, and image processing.

Binjie Wang is an attending physician at Huaihe Hospital of Henan University, China. He received his Medical Imaging B.S. degree from ZhengZhou University, China, in 2003, and received his Basic Medicine M.S. degree from Henan University, China, in 2013. Now he works in radiology department at Huaihe Hospital of Henan University, China. He has more than 10 years' clinical experience in medical imaging diagnosis, especially in the central nervous system disease.

Xinhong Zhang is an associate professor at the School of Software at Henan University, China. She received her B.S. degree and M.S. degree from Henan University, China. Her research focuses on digital image processing and pattern recognition.