Auto-Detection of Alzheimer's Disease Using Deep Convolutional Neural Networks

Lulu Yue, Xiaoliang Gong, Kaibo Chen, Mingze Mao, Jie Li
Department of Computer Science and Technology, Tongji University
Shanghai，200065,P.R. China

Asoke K. Nandi , Maozhen Li
Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, UK

Alzheimer's Disease, Neuroimaging Initiative

Abstract—**Alzheimer's disease(AD) is a kind of progressive neurodegenerative disease. One who is diagnosed as an Alzheimer's disease patient may has many symptoms, such as deterioration of memory and language. Once those symptoms was noticed, they usually can survive 4 to 20 years. So far, Alzheimer's disease has become the sixth leading cause of death, and it has become a worldwide health and social challenge. Traditional methods of diagnosing AD and mild cognitive impairment(MCI), mostly depend on capturing features from variable modalities of brain image data. It is a big challenge to pick out the MCI from normal controller (NC) and AD, especially for those who are lacking experience. In this article, we employ deep convolutional neural network (DCNN) to extract the most useful features of the structural magnetic resonance imaging (MRI). Firstly, the structural MRIs are pre-processed in a strict pipeline. Then, instead of parcellating regions of interest, we re-slice each volume, and put the resliced images into a DCNN directly. Finally, four stages of Alzheimer's are identified, and the average accuracy is 94.5% for NC versus LMCI, 96.9% for NC versus AD, 97.2% for LMCI and AD, 97.81% for EMCI versus AD, 94.8% for LMCI versus EMCI. The results show that the DCNN outperforms existing methods.**

Keywords-Deep Learning; Alzheimer's Disease; MRI; Early Diagnose

I. INTRODUCTION

Until 2017, the world's population of aged older than 60 years is above 962 million, and the number will be twice more by 2050. People suffering from Alzheimer's disease is also increasing at a rapidly speed. AD is a kind of progressive neurodegenerative disease [1]. Suffering from this kind of disease with language deterioration and memory loss may disrupts the one's daily life. People with Alzheimer's gradually get worse over a number of years. Memory loss is mild in the early stage of Alzheimer's, as time goes on, in their late stage, individuals gradually lose their ability to communicate with others and even couldn't take good care of themselves. According to preliminary statistics， Alzheimer's now has become the sixth leading cause of death[1]. According to the survey, people who is suffering from dementia all over the world is about 47 million. The number is already more than the total population of Spain and is increasing rapidly. Dementia has also severely affected the world economy. The global consumption of dementia is estimated to be about $818 billion, and the number will be trillions of billions by 2018. [2]. It is no doubt that Alzheimer's disease has become a worldwide health and social challenge.

AD is characterized by memory impairment, language dysfunction, and impairment of recognition, so that AD patients always cannot manage themselves with neurofibrillary tangles [3]. Though some treatments may temporarily improve the symptoms of AD, there is still no powerful evidence that can tell the reason or can stop its progression [4]. As the condition getting worse, patients become more and more dependent on the help of others. This also put a heavy burden on the caregiver; including social, psychological, physical and economic factors. It is found that, although AD is incurable, in its earlier stage, mild cognitive impairment patients may maintain the diagnosis of MCI even after many years.

Nuclear magnetic resonance as a noninvasive imaging technique, is getting more and more popular among researchers who are trying to identify AD from the normal controllers. Multi-model magnetic resonance may provide a deep view into the brain mechanism, such as structural and functional MRIs, diffusion MRIs, Diffusion Tensor Imaging (DTI) and Positron Emission Computed Tomography (PET). Inherent information extracted from these images give us an opportunity to understand aspects of the brain. Machine learning proved to be efficient in classifying AD form health controllers. But the accuracy of distinguishing Alzheimer's from its early stage is still not so satisfactory.

The rest of the paper is organized as follows. Section 2 gives a review on related work. Section 3 details the design of the DCNN. Section 4 evaluates the performance of DCNN and analyzes the results and Section 5 concludes the paper.

II. RELATED WORK

With the rapid developments of machine learning in recent years, it has been subsequently used in many fields. Many publications have proved that machine learning could significantly improve the accuracy of clinical diagnosis of AD. Due to the fact that the original neuroimage data can be high dimensional, researchers had made many efforts to reduce the dimension. Feature selection plays a rather important role here

978-1-5386-8097-1/18/$31.00 ©2018 IEEE

to identify AD/MCI from health controllers. After segmenting each volume into gray matter (GM), White matter (WM) and cerebrospinal fluid (CSF). The gray matter was used to separate AD patients from NCs as GM has the strongest relationship with the diagnosis of AD and MCI. Then the hippocampal volume, ventricular volume, whole brain volume or cortical thickness are usually taken as features to identify AD from NC, because the hippocampus volume of the AD patients is smaller than the healthy controllers; on the contrary, the ventricle of AD patients may be larger [4]. The GM are parcellated into regions of interests considering the fact the voxel dimension of gray matter is also extremely high. Above of all, when precise results are required, we should not only take multi-modal clinical training data, experienced experts are also indispensable. Therefore, single model with high precision automatic diagnosis system is very helpful.

Sparse representation-based classifier (denoted as SRC) was proposed to fix the problem of high dimension [5]. In [6], the authors proposed a local patch-based subspace ensemble framework, which constructs multiple separate classifiers to avoid the difficulties of selecting an optimal subset of discriminative features for the single classifier, yet at the same time, could capture the local spatial consistency. The connectivity between each pair of regions was also taken into consideration as a feature to increase the accuracy of the identification of AD/MCI from NC. A whole brain hierarchical network was built by regions of the whole brain which is divided within each subject based on Automated Anatomical Labeling (AAL) atlas [7]. However, they took both the texture features and the spatial-correlations as features which led to a challenging high-dimension problem. Gerardin et al. used spherical harmonics coefficients, a parametric boundary description approach to extract hippocampal shape features [9]. Different from the conventional computer-aided diagnostic systems which consider the simple low-level features, deep learning exploits the latent high-level features to improve the precision of the diagnosis. Suk et al. extracted latent features with both Stack Auto-Encoder and Deep Boltzmann Machine (DBM) respectively [10].

Although the methods presented in the existing articles had proven the effectiveness in AD/MCI classification. The tissue segment and manually-defined features, such as ventricular volume, cortical thickness, whole brain volume and hippocampal volume [11], require a lot of prior knowledge of the major. Besides, that is a real waste of time. Trying to avoid all these problems, we implement a work flow shown in Figure 1. Firstly, skull stripping and segmentation were both done by a MATLAB toolbox named VBM. We selected the GM data and resliced it just as the number each volume was scanned. Due to the excellent spatial resolution of structural MRI, we decide to use a deep convolutional neural network to extract the spatial features. Besides, we also segmented the gray matter into 90 regions of interest based on the Automated Anatomical Labeling(AAL) atlas as the control group.

III. DCNN DESIGN

The pipeline of data processing framework is shown in Figure 1. Firstly, the structural MR images have been registered to Montreal Neurological Institute (MNI) space. Then each volume is segmented into GM, WM and CSF. Many articles explicitly indicated that the gray matter has the strongest correlation with AD/MCI diagnosis.

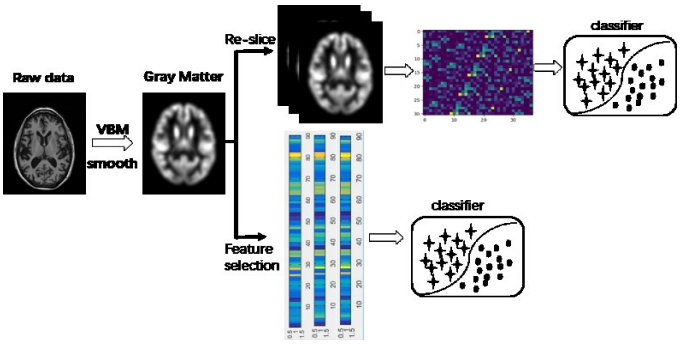


Figure 1. Data Processing Framework.

A. Description of the data source

In this study, all the subjects were selected from the Alzheimer's Disease neuroimaging (ADNI) (http://adni.loni.usc.edu/). ADNI researchers collected and integrated multimode data. Participants of the study are mostly from the North American. Different stages of Alzheimer's disease are available from this site.

In order to prove the effectiveness of the proposed framework, we selected data from different phases of ADNI. At last, we got 785 normal controllers, 542 MCI patients and 336 AD patients from ADNI1 and ADNI-GO as the first group of subjects. In the second group, 1106 normal controllers, 1583 early mild cognitive impairment (EMCI), 1304 late mild cognitive impairment (LMCI), and 336 NC were picked. The age of all the subjects range from 60 to 85 years old. There is no noticeable difference in this three groups in terms of gender and age. The demographic information of dataset1 is shown in Table 1 and the demographic information of dataset2 is shown in Table 2. The second group group/dataset contains the most difficult to diagnose MCI subjects; these were further divided into EMCI and LMC, which can better demonstrate the effectiveness of classification of our framework.

TABLE I. DEMOGRAPHIC AND CLINICAL INFORMATION OF THE SUBJECTS (DATASET1)

SUBJECTS TYPE	*Number*	*Age*	*Gender (F /M)*	*MMSE*
NC	785	74.63±3.69	416/369	29.07±1.32
MCI	542	76.53±5.35	193/349	26.56±2.63
AD	336	78.86±5.34	156/180	23.84±2.10

a. MMSE stands for the Mini-mental State Examination.

TABLE II. DEMOGRAPHIC AND CLINICAL INFORMATION OF THE SUBJECTS (DATASET2)

SUBJECTS TYPE	*Number*	*Age*	*Gender (F /M)*	*MMSE*
NC	1106	74.63±3.69	554/552	29.10±1.25
EMCI	1583	76.86±4.97	570/1013	28.37±1.48
LMCI	1304	76.53±5.35	639/665	27.19±2.23
AD	366	78.58±5.38	138/228	21.84±4.10

a. MMSE stands for the Mini-mental State Examination.

B. Data preprocessing

When downloading data from the ADNI database, we chose the NII format data. In order to obtain the data with the same parameters to avoid many other problems in processing procedure, the pre-processed data were preferred here. The steps of pre-processing include gradient unwarping and a non-parametric non-uniformed bias correction (N3) algorithm [5] [11], both can be done with Grinder Pipeline. The structural MR images were obtained from 3.0 tesla SIEMENS scanners. Flip angle is 9.0 degree; slice thickness is about 1.2mm and echo time (TE) is almost 3.0 ms. Repletion time (TR) is 2300.0 ms with pixel size=1.0 mm × 1.0 mm; the matrix size of each volume is 240×256×176. The demographic information of each group of subjects and their Mini Mental State Examination (MMSE) score are detailed in Table1 and Table 2. For more information about the parameters of the images, one can search the website of ADNI (http://adni.loni.usc.edu/).

Then we performed brain extraction with a toolbox named FSL-BET [3] to remove all the non-brain tissues from the images. The T1 images had been pre-processed with the standard procedure just as shown in Figure1. As almost all the literatures had shown that gray matter (GM) made much more contribution to AD/MCI diagnosis than WM or CSF [13], we segmented each brain into GM, WM and CSF with a SPM toolbox named Voxel-based morphometry (VBM) [5] which ran in the MATLAB package. VBM is a neuroimaging analysis technique that use statistical methods of statistical parameter mapping to study local differences in brain anatomy. It is composed of several steps. Firstly, after checking the quality of the image, the image was segmented in to GM, WM and CSF [14]. The outputs will be used for achieving more accurate inter-subject alignment using DARTEL [6]. Then the GM was eventually warped to MNI space. The obtained GM images were reshaped to 121×145×121. The Gaussian FWHM was set to 8mm here for smoothing the pre-processed data. The smoothed images were resliced into 121×145 and labeled for the final classification. Each class of subjects were divided into two parts. We took 10% randomly from the total dataset for testing and the training data accounts for 90%. After that, we had two processes for these GM images. We re-sliced each volume and reshape each slice into a 121×145 matrix saved as jpg format. As shown in Figure 1, the re-sliced images were then fed into the CNN platform. Furthermore, we also parcellated each volume into 126 regions among which 36 regions belonging to cerebella was ignored. The Automated Anatomical Labeling (AAL) atlas is used as template [15]. We took the GM tissue volume of each region as features to put into other work flow using an SVM classifier as comparision.

C. Computational Framwork

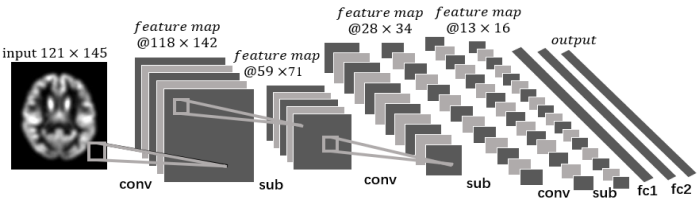


Figure 2. Convolutional neuro network framework adopt for struct-MRI

Convolutional neural networks were inspired by the workings of the visual cortex. CNN is a class of deep, feed-forward neural network which has excellent performance in large-scale image processing. Like the traditional neural network, CNN is composed of the input layer, convolutional layer, pooling layer, fully connected layer and the output layer. Different layers were connected with tied weights. The convolutional layer and pooling layer mainly used for dimensionality reduction and feature extraction. Due to each layer of CNN shares the tied weights, it is easier for it to train with much fewer parameters. The general feed-forward back propagation training is improved in the following way. In each convolution layer, the outputs of its previous layer are convolved with a learnable kernels. Then the feature map was formed by the activation function as the outputs. Generally, the formula can be described as

$$y_j^{\ell} = f\left(\sum_{i \in N_j} y_i^{\ell-1} * w_{ij}^{\ell} + b_j^{\ell}\right),$$

where N_j represents the number of the input maps and f is the activation function.

The down sampling layer validly reduced the dimensionality of the inputs. Even the number of output maps and inputmaps are the same, the output maps will be much smaller. More formally,

$$y_j^{\ell} = f\left(\beta_i^{\ell} down(y_j^{\ell-1}) + b_j^{\ell}\right),$$

where f is an activation function and $down(\cdot)$ represents the function of the sub-sampling.

The backpropagation technique here uses a feedforward structure to propagate errors in the neural network in order to adapt the weights. Backpropagation is a method of achieving gradient descent in neural networks. The output layer error is defined as

$$\delta_j^{(\ell)} = a_j^{(\ell)} - y_j,$$

where hidden layer error signal is written as

$$\delta^{(i)} = (\theta^{(i)})^T \delta^{(i+1)} * \Delta a^{(i)}$$

where $\theta^{(i)}$ represents weights of layer i. The $\delta^{(i)}$ represents the back-propagated error signal, which is used to update the activation values in layer i and $\Delta a^{(i)}$ represents the gradients of the activation function in layer i.

The CNN we implemented in this article is shown in Figure 2 which included three convolutional and sub-sampling layers. After each pooling layer, we set two norm layers after the first

two polling layers with two fully connected layers behind the last pooling layer. In order to achieve the best result, we have tried several activation functions. Such as Sigmoid function, Tanh function, Rectified Linear Unit (Relu) activation function and scaled exponential linear units (Selu) activation function. Finally, the Relu activation function was adopted. Feature maps of each convolutional layer are shown in Figure 3.

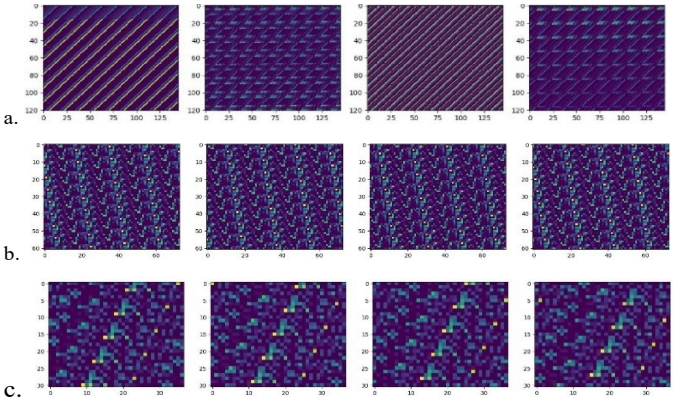


Figure 3. Feature map of each layer. a.displays the extracted features of the first convolutional, b.diplays the extracted features of the second concolutional, and c. displays the extracted features of the third convolutional.

IV. RESULTS AND DISCUSSION

We repeated each experiment 10 times because the accuracy of the classification differed when randomly changing the testing data. Each task was repeated 10 times in order to check the robustness of the performance. For the first dataset, we considered 3 classes of tasks including AD versus NC, AD versus MCI and MCI versus NC. The fluctuations of the mean accuracy to each task are shown in Figure 5. It can be seen from Figure 5 that the accuracy of classifying NC with AD is 95.45%, the accuracy of classifying NC with MCI is 95.39% and the accuracy of classifying NC with AD is 93.97%. The training loss trade can be seen from Figure 4. The performance comparison with other algorithms is displayed in Table 3. As can be seen from Table 3, the convolutional neural network framework we implemented here offers to much more accuracy.

TABLE III. DEMOGRAPHIC AND CLINICAL INFORMATION OF THE SUBJECTS (DATASET1)

methods	*NC vs AD*	*NC vs MCI*	*AD vs MCI*
SVM	81.04±6.28	76.98±5.48	70.15±4.31
SAE	91.65±1.37	87.59±2.73	77.27±1.76
CNN	95.441±0.004	95.388±0.003	93.888±0.044

a. The inputs of the competed methods are 90 regions of interest, while the inputs of CNN are the re-sliced volumes which were saved as .jpg format.

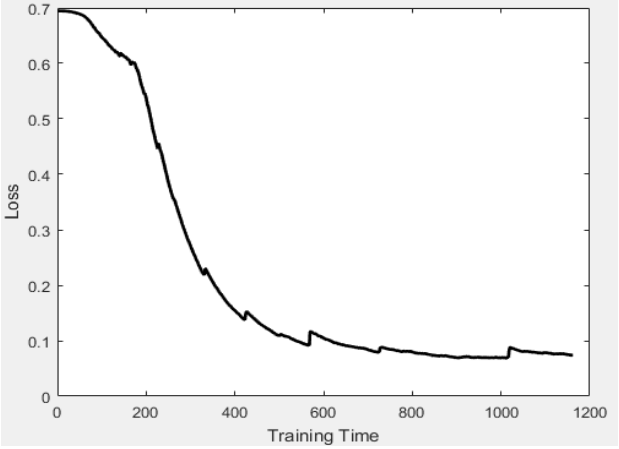
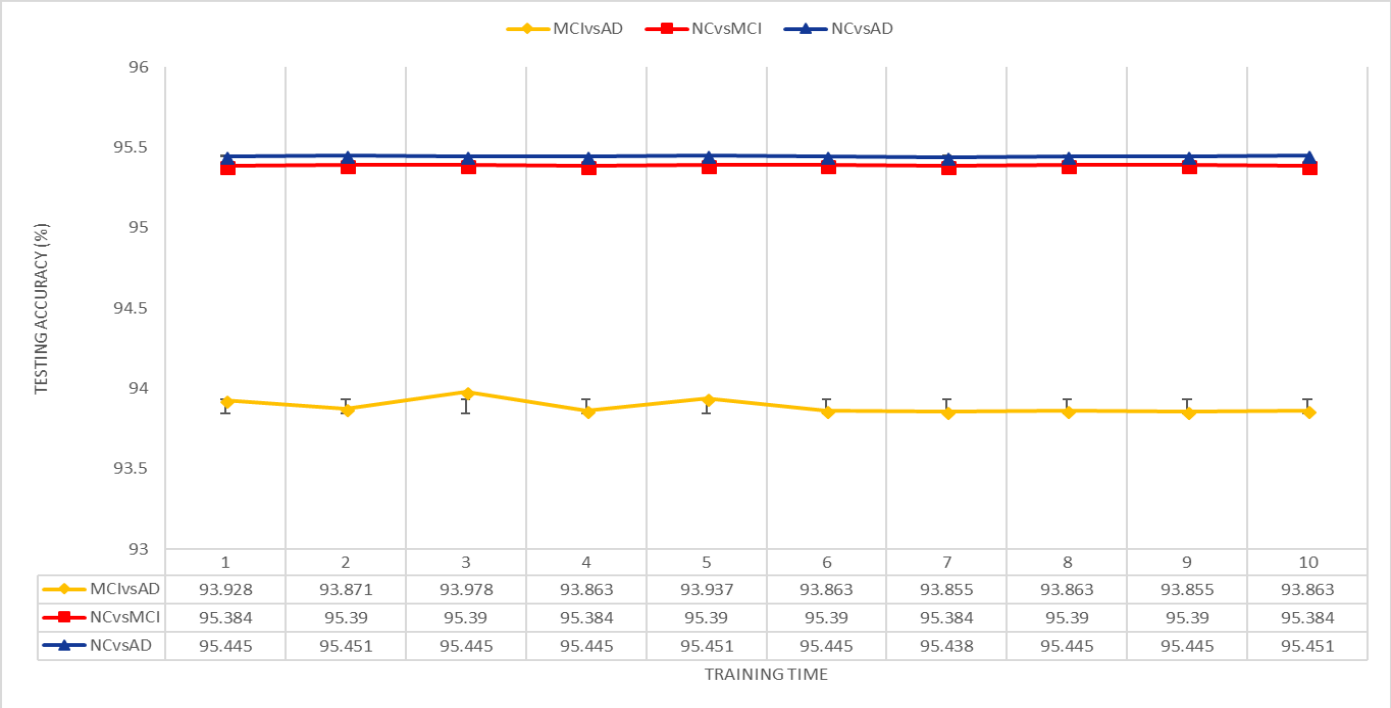


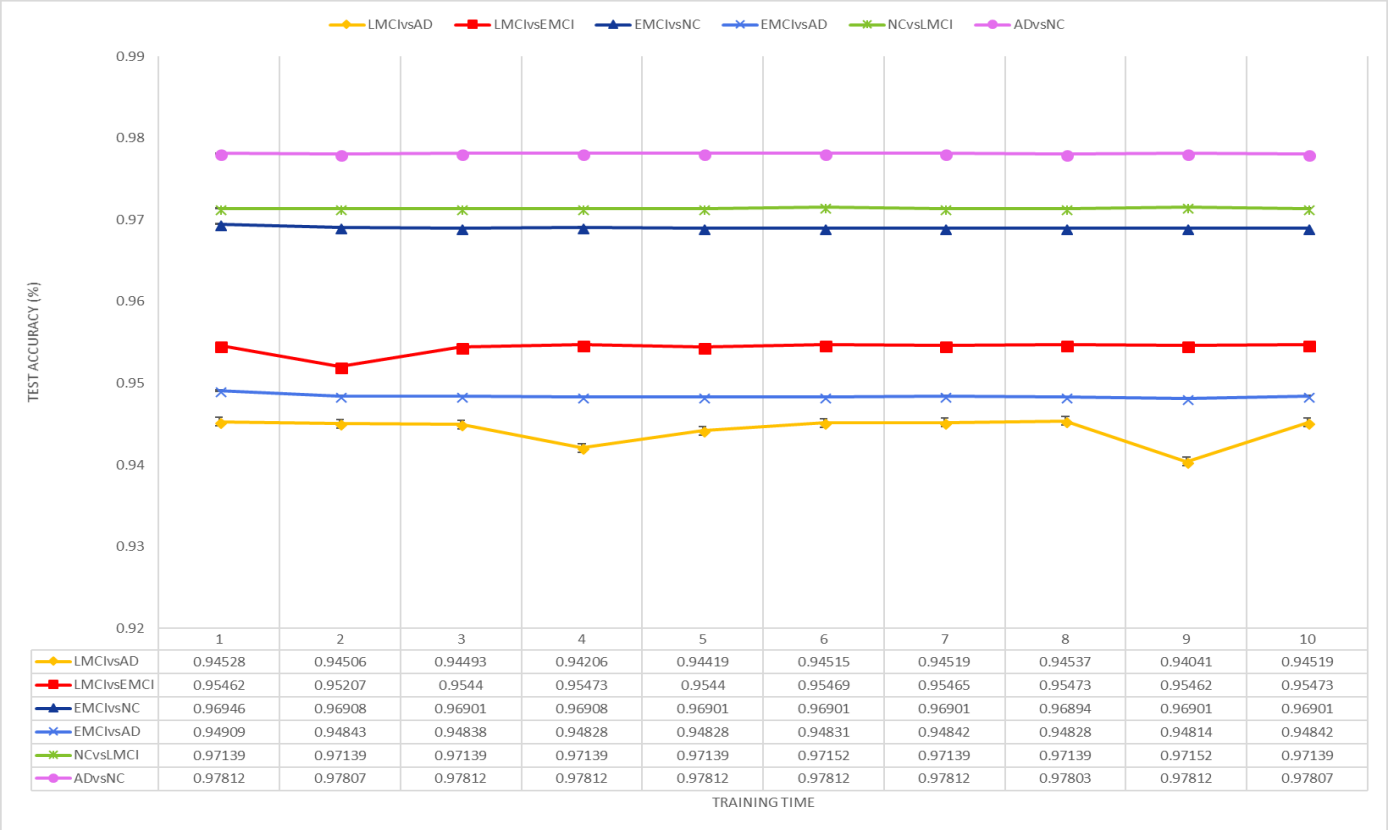
Figure 4. Training loss fluctuation



	1	2	3	4	5	6	7	8	9	10
MCIvsAD	93.928	93.871	93.978	93.863	93.937	93.863	93.855	93.863	93.855	93.863
NCvsMCI	95.384	95.39	95.39	95.384	95.39	95.39	95.384	95.39	95.39	95.384
NCvsAD	95.445	95.451	95.445	95.445	95.451	95.445	95.438	95.445	95.445	95.451

Figure 5. Test accuracy of each group,MCI vs AD means the accuracy of classification between MCI patients and AD pationts. NC vs MCI stands for the accuracy of classification between normal controllers and MCI patients. NC vs AD means the accuracy of classification between normal controllers and AD patients.

In order to demonstrate the effectiveness of the proposed framework, we tested the other group dataset2 which classified the MCI into EMCI and LMCI. If we can identify the patients in the earlier stage of MCI, that may be much more helpful to their health and daily life. The results can be seen from the Figure 6. It is clear that even the LMCI and EMCI group got an excellent performance of about 94.85%. The performance of classifying normal controllers and LMCI is about 94.53%. In the NC and AD group, the accuracy we got is about 96.91%. The best accuracy of the LMCI and AD group was about 97.16%. In discriminating the EMCI and AD group the accuracy can be 97.81%. The excellent result can be attributed to the number of the AD patients is much smaller than the groups of EMCI, LMCI and normal controllers.



	1	2	3	4	5	6	7	8	9	10
LMCIvsAD	0.94528	0.94506	0.94493	0.94206	0.94419	0.94515	0.94519	0.94537	0.94041	0.94519
LMCIvsEMCI	0.95462	0.95207	0.9544	0.95473	0.9544	0.95469	0.95465	0.95473	0.95462	0.95473
EMCIvsNC	0.96946	0.96908	0.96901	0.96908	0.96901	0.96901	0.96901	0.96894	0.96901	0.96901
EMCIvsAD	0.94909	0.94843	0.94838	0.94828	0.94828	0.94831	0.94842	0.94828	0.94814	0.94842
NCvsLMCI	0.97139	0.97139	0.97139	0.97139	0.97139	0.97152	0.97139	0.97139	0.97152	0.97139
ADvsNC	0.97812	0.97807	0.97812	0.97812	0.97812	0.97812	0.97812	0.97803	0.97812	0.97807

Figure 6. NC vs AD means the accuracy of classification between normal controllers and AD patients. NC vs LMCI stands for the accuracy of classification between normal controllers and LMCI patients. NC vs EMCI stands for the accuracy of classification between normal controllers and EMCI patients. LMCI vs AD means the accuracy of classification between LMCI patients and AD pationts. EMCI vs LMCI stands for the accuracy of classification between EMCI and LMCI patients. EMCI vs AD means the accuracy of classification between EMCI patients and AD pationts.

V. CONCLUSION

In this paper, we have presented DCNN for diagnosing the diseases of AD/MCI. Thanks to the latent feature extracted by the DCNN, a high level of accuracy has been achieved. A further work will research whether the subject has the risk of turning to the next stage.

VI. ACKNOWLEDGMENT

This work was supported by a grant from the Natural Science Foundation Program of Shanghai (16JC1401300,17ZR1431600, 18ZR1442700), the Shanghai Sailing Program(16YF1415300), and the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, “A nonparametric method for automatic correction of intensity nonuniformity in MRI data,” IEEE Trans. Med. Imag., vol. 17, no. 1, pp. 87–97, Feb. 1998.

[2] Jin Liu, Jianxin Wang, Bin Hu, Fang-Xiang Wu and Yi Pan, “Alzheimer’s Disease Classification Based on Individual Hierarchical Networks Constructed With 3-D Texture Features,” IEEE Trans on NanoBioscience, vol. 16, pp. 428-435,september 2017.

[3] DaoqiangZhang,DinggangShena and The Alzheimer's Disease Neuroimaging Initiative, “Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, ” NeuroImage,Volume 59, Issue 2, Pages 895-907,16 January 2012.

[4] Manhua Liu, Daoqiang Zhang, Dinggang Shen and The Alzheimer's Disease Neuroimaging Initiative, “Ensemble sparse classification of Alzheimer's disease,” NeuroImage, vol. 60, 1106–1116, January 2012.

[5] Jin Liu, Min Li, Wei Lan, Fang-Xiang Wu, Yi Pan, and Jianxin Wang .“Classification of Alzheimer’s Disease Using Whole Brain Hierarchical Network,” IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 14, NO. 8, August 2015.

[6] Jun Zhang, Yue Gao, Yaozong Gao, Brent C. Munsell and Dinggang Shen “Detecting Anatomical Landmarks for Fast Alzheimer’s Disease Diagnosis,” IEEE transections on medical imaging, vol. 35, NO. 12, December 2016.

[7] E. Gerardin et al., “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.

[8] Heung-Il Suk, Dinggang Shen and the Alzheimer’s Disease Neuroimaging Initiative, “Deep Learning in Diagnosis of Brain Disorders,” Business Media Dordrech, Springer Netherlands, 203-213, 2015

[9] Friston, K.J., Holmes, a. P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189–210 (1995).

[10] A. Chincarini, F. Sensi, L. Rei, G. Gemme, S. Squarcia, R. Longo,F. Brun, S. Tangaro, R. Bellotti, N. Amoroso et al., "Integrating longitudinal information in hippocampal volume measurements for the early detection of Alzheimer's disease," NeuroImage, vol. 125, pp. 834–847, 2016.

[11] Li, F. , Tran, L., Han Thung, K. et al , "A Robust Deep Model for Improved Classification of AD/MCI Patients," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 5, pp. 1610-1619, 2015.

[12] Alzheimer's disease from healthy controls. A European Alzheimer's Disease Consortium (EADC) study," NeuroImage: Clinical, vol. 7, pp. 34-42, 2015.