

# Diagnosis of Alzheimer's Disease using Structural MRI and Convolution Neural Network

Shuyang Bian<sup>1,a</sup>

<sup>1</sup>High School Affiliated to Fudan University, Shanghai, China

**Abstract.** Background: Alzheimer's disease (AD) is a prevalent, neurological disease without effective treatment. However, if diagnosed early, the progression of the disease could be delayed through medication. Currently, one method to effectively diagnose AD early is to use Alternate Covering Neural Network (ACNN) network to discern various non-invasive Magnetic Resonance Imaging (MRI) images. This research aims to create an approach better than the current one and thus increase the accuracy of classifying MRI images, thereby diagnosing AD earlier and more perfectly. Methods: Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health Grant U19 AG024904) database provided 3013 different sets of 3D MRI images labeled as cognitively normal (CN), mild cognitive impairment (MCI), and AD. A newly-proposed, modified Residual Network (ResNet) and an ACNN network were then constructed. Their common goal was to learn how to classify these labeled MRI images. After training, the two models got unique parameters for using the updated network to diagnose new images. Finally, inference, or testing the diagnostic accuracy of the two models, were performed based on another 469 different 3D MRI image sets. The accuracy of classification for two separate models were compared. Results: Compared with the ACNN network with a weighted classification accuracy of 80.17%, the newly proposed ResNet network enhances the weighted accuracy to 85.07% and showed statistical significance ( $p < 0.001$ ). Through analyzing the occurrence of false-positive cases by two models, a Receiver Operating Characteristic (ROC) curve was drawn. The area under the curve of the ROC confirms this enhancement as the area under the curve of ROC is greater than that of the ACNN model in two of the three cases (MCI  $0.9293 > 0.9196$ ; AD  $0.9389 > 0.9146$ ). Conclusions: The research proposed a new deep learning convolutional network to classify 3D structural MRI images. The new ResNet is better in that it showed increased accuracy with statistical significance and had fewer false-positive results compare with the traditional ACNN network, thereby promising to help doctors diagnose AD more quickly and more accurately.

## 1 Introduction

### 1.1 Patient demography

It has been over 110 years since Doctor Alzheimer first reported a 51-year-old female case suffering from Alzheimer's disease (AD) [1]. According to the figures by Alzheimer's Disease International (ADI) in 2017, there have been close to 50 million cases of AD across the globe, with a rate of new cases increasing at 9.9 million cases per year or 1 case per 3 seconds. Now, about 58% of the people with dementia live in low to middle-income countries, while the fastest growth in older adults is in China, which makes AD a severe threat to our country [2]. With no cure for this prevalent disease, it is worthy of our awareness to pay attention to this disease that affects a whole family's life and brings tremendous trouble to society. It is estimated that the global cost of dementia will cost the world a trillion US dollars, whereas AD was the leading result for dementia.

### 1.2 Alzheimer's Disease: Medically relevant facts

As a neurodegenerative disease, AD is correlated with the aging of a human being and correlated with dementia. Clinically, AD is characteristic of defects in cognition. Patients suffer memory malfunction, aphasia, impaired in visual, and changes in personality. To this day, the cause of the disease is not yet known.

The mechanism of AD may result from multiple factors, including biological, social, and psychological reasons. Epidemic research shows that family history contributes to AD as most patients share the same disease with their family members. Recent gene study reveals that excessive amyloid protein, whose gene contributing to pathological conditions was on the 21st pair of chromosomes, was a phenotype, thereby hereditary factors may be helpful. Other factors are also underlying [3]. Diseases such as thyroid diseases, immunological diseases, epilepsy are previously considered as possible

<sup>a</sup> Corresponding author: [simonbian02@outlook.com](mailto:simonbian02@outlook.com)

factors in the disease. Chemicals such as aluminum in tap water will speed up the process of aging, making it detrimental to the normal function of a brain. Neurological conditions, such as depression, schizophrenia, and other functional mental disorders, were also reckoned as factors [4]. Cerebral trauma is also reported to become a dangerous factor indicating AD. AD is generally found in women.

Contemporary classification of AD patients is not so strict as to either classify a patient into AD or not. Instead, using imaging techniques. Clinical diagnosis of AD patients may AD, Mild Cognitive Impairment (MCI), and the control normal (CN) group.

### 1.3 Current Supplementary drugs for AD patients [5]

Acetylcholinesterase inhibitor (AChEI) or anti-cholinesterase can inhibit acetylcholinesterase enzyme from breaking down acetylcholine, a stimulator for neurons, allowing lengthened duration and level of excitement from neuron [6].

Flunarizine(hydrochloride), nimodipine, and other calcium antagonists can increase blood circulation in the brain and prevent loss of oxygen or blood in the brain, which may otherwise be detrimental to neurons.

Anti-amyloid drugs were previously thought to be beneficial in decreasing the speed of progressive regression of neurons. However, recent clinical trials have not been optimistic [7].

Nerve growth factor can improve the conditions of aging animals and prevent degeneration necrosis of central cholinergic neurons [8].

Most of the drugs listed above have not been recognized as an actual effective treatment for AD. However, most of them may reduce the symptoms of it, making them excellent supplementary drugs for enhancing the conditions of AD patients.

### 1.4 Current Diagnostic Method

Even though early diagnosis and intervention are critical to the healthcare of the patients, statistics show that nearly 75% of people with AD are not diagnosed [2]. With care for the patients dragging by and there being little caring for such patients, the life quality of the AD patient suffered severely, leading to the need for diagnosing AD.

Several methods have been proposed to diagnose AD while each have their advantages and drawbacks.

EEG of AD is characteristic of decreased  $\alpha$  wave, increased  $\theta$  wave, and reduced average frequency [9]. This method is of use, however, only if a doctor has a patient's full record of EEG, which many may not possibly have. EEG samples are not challenging to obtain if and only if a patient is available for laboratory work, thereby making the diagnostic process much laboring.

Previous methods for diagnosing AD also include psychology tests, social cognition scales, which mostly consist of questionnaires. Tests and scales may be biased since the cognition level may correspond with people's understanding of a question, people's education level, and

factors that do not correlate directly with AD conditions, making it awkward pick up confounding variables.

Blood tests for testing for AD were applied previously, which is non-invasive and collects various signaling proteins, such as IL-3, -11,  $-1\alpha$  the concentration of macrophage colony-stimulating factor [10].

Functional MRI (fMRI). The basic principle of fMRI was using hemoglobin is diamagnetism[11], while deoxyhemoglobin is paramagnetic, whereas both cells are in the bloodstream. Typical features of patients with AD include brain atrophy, narrow gyrus, increased sulci, and superficial atrophy of the frontal-parietal and frontal lobes.

Positron emission tomography (PET), which may reveal the brain glucose metabolism level, providing evidence for early malfunction in cognition.

Projection X-ray. X-ray imaging requires interactions of x-ray photons with the object. Therefore, the wavelength should be in a specific energy band to ensure that a machine can see through a body.

### 1.5 Drawback of neuroimages in the past

Experience in applying MRI images or PET images implied that the high-intensity magnetic field used in MRI and high dosage of radioactive rays used in PET are detrimental to human cells.

What is worse, in the past, experienced doctors may have to read a neurological image many times before all of them have come to a consensus, which not only takes time but also a great deal of sources for providing patients with useful treatments. Furthermore, the consultation of experts may result in bias due to the physical condition including lightness, shades, and structures of a certain neurological photograph. Nowadays, with the help of a computer to facilitate the diagnosis of AD, workload for doctors will require fewer technologies and mean quicker diagnosis, thereby creating an environment for all AD patients to maximally enhance their life quality.

### 1.6 "AI + MRI" make MRI more user-friendly

Artificial intelligence has achieved state-of-the-art results from its application in diverse areas in natural image processing, including multiple tasks such as image classification, object detection, image segmentation, and various visual tasks [12]. Applying deep learning algorithms onto medical images has brought unprecedentedly successful results. Since the price of caring for AD patients would rise drastically, the need to have a machine-aided model for early and effective AD diagnosis becomes vital.

Predicting whether a patient had AD based on MRI images will no longer rely on the expert's knowledge. Big data collected from local hospitals, when used appropriately with some fundamental computer science knowledge, will enable medical workers to train the computer to assist themselves with the help of the data they collected. By using a convolutional network that will learn under supervision, the AD patients will, for sure, be better and more quickly diagnosed to enhance their living.

## 2 Methods

The method is practically based on images from 3D MRI, using CNN as a model, and produce trinary classification results as AD, MCI, and CN groups. To enhance the processing speed, the MRI is not a specific 3D image but a combination of multiple 2D images.

### 2.1 General supervised learning problem [13]

$$f^* = \arg \min_{f \in \mathcal{F}} \sum_{i=1}^N \mathcal{L}(f_{\theta}(x_i), y_i)$$

Objective function (loss)

Hypothesis space                      Training data

**Figure 1.** The mathematical or theoretical formula for approaching a supervised learning problem, also known as the loss function

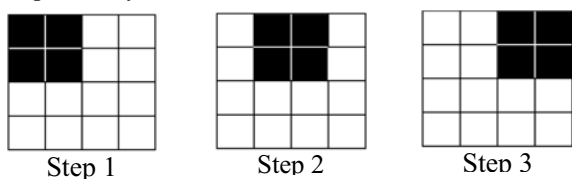
From the function above (Figure 1), it is shown that with the given hypothesis space and sufficient training data, a computer will try every possible combination to produce a loss function, to bes predict the outcome.

### 2.2 Definition of Convolution Layer in the use of Classification problems

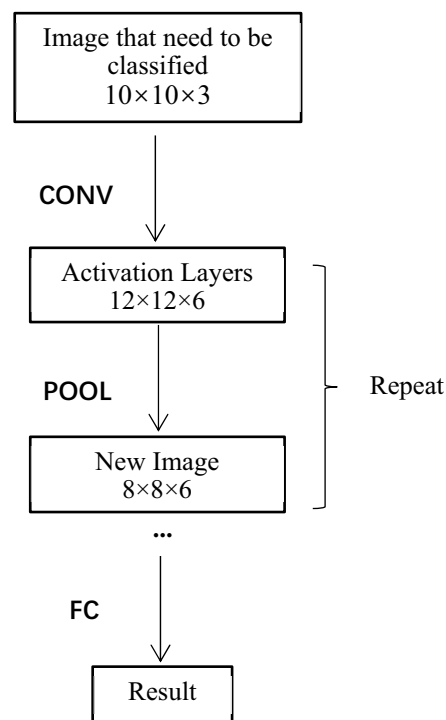
A convolution network is several stacks of a permutation of the Convolution layer (CONV), pooling layer (POOL), and Fully Connected Layers (FC) (Figure 2).

The convolution layer often preserves the spatial structures, thereby making the feature of a picture still somewhat similar. In order to convolve the data, we usually need a filter, which means that it slides over the image spatially, which means all spatial locations (Figure 3). For instance, a 3D image with a width of  $W_1$ , a height of  $H_1$ , and a dense of  $D_1$ , given  $N$  filters, with its spatial extent as  $SE$ , its stride as  $S$ , and the amount of padding as  $P$ , typically produces a volume size of  $W_2 \times H_2 \times D_2$  where  $W_2 = \frac{W_1 - SE + 2P}{S} + 1$ ,  $H_2 = \frac{H_1 - SE + 2P}{S} + 1$  and  $D_2 = N$ . The different filters enable us to get several separate activation layers. As using filters often reduces the size of the image, we could use padding to retain the basic size of the pictures because shrinking too fast is not good, does not work well.

After that, a typical network requires pooling, which can make representations smaller and more manageable. It also enables us to operate over each activation map independently.



**Figure 2** The mechanism of pooling methods



**Figure 3** A typical convolution network, where there are various combinations of activation layers

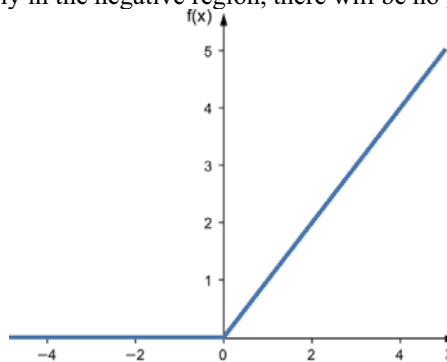
Generally, after numerous stacks of CONV and POOL, a graph of the same dimensional but different sizes will be output. The final stage for a convolution network is fully connected layers, where a probability of each class will be computed to determine to which type a figure belongs [14].

### 2.3 Various Activation Layers

There are various types of activation layers. The generic ones are the Rectified Linear Unit (ReLU) and Sigmoid or logistic regression. [13,14]

#### 2.3.1 ReLU

The function of a ReLU activation layer can be written as  $g(y) = \max(0; y)$ . It can be represented in the graph Figure 4. An advantage of using ReLU as the activation layer is that it can provide unbounded output and is efficient to mplement. However, if ReLU is applied strongly in the negative region, there will be no gradient.



**Figure 4.** A visual display of ReLU function

### 2.3.2 Sigmoid

The function of a ReLU activation layer can be written as  $g(y) = \frac{1}{1+e^{-y}}$ . It is shown in the graph Figure 5. An advantage of using Sigmoid is that the result bounded between (0; 1). Therefore, a comparison of the probability could reveal the likelihood of classifying a certain into a category.

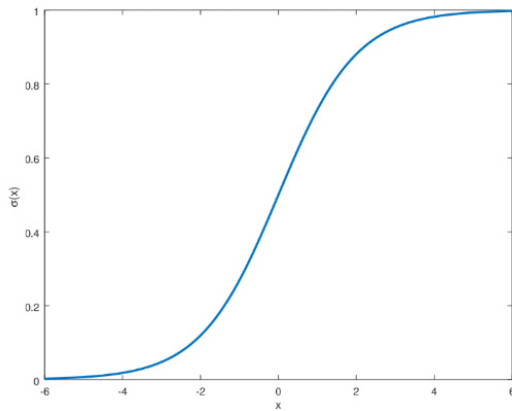


Figure 5. A visual display of Sigmoid function

### 2.4 Deep Residual learning

To increase the speed of running the model, based on previous literature analysis, we use the ResNet [15], which is an intensely deep network using residual connections (Figure 6).

A typical deep learning method is to obtain input as X and learns to adjust the parameters so that it could make the prediction, f(X), has the least loss with the actual value Y, so that  $\lim_{n \rightarrow \infty} (f(x) - y) \xrightarrow{\text{yields}} 0$ . However, in the deep residual learning method, the machine does not learn f(X). Instead, it learns  $H(X) - X$ , which yields,  $X \rightarrow (H(X) - X) + X$ .

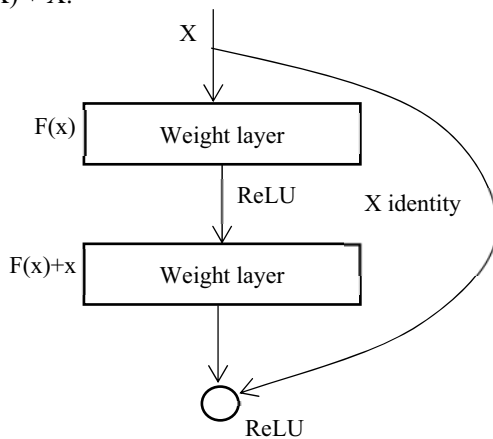


Figure 6 One Unit of ResNet Model, which display its mechanism

### 2.5 Training details on ResNet Model and Alternate Covering Neural Network (ACNN) Model

#### 2.5.1 Our ResNet Model (Figure 7)

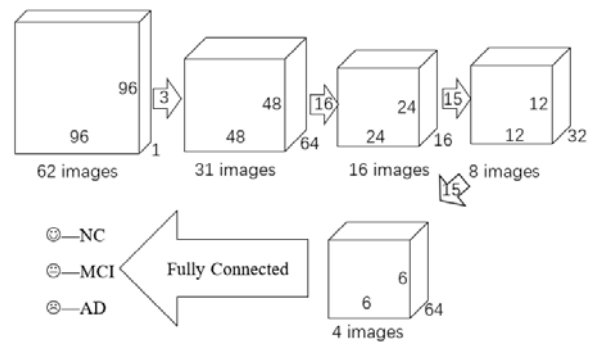


Figure 7 the proposed ResNet model. The number on the arrow means the number of activation layers.

Figure 7 shows a typical ResNet model combined with a convolutional network. The scale of the picture has become smaller, making it easier for the machine to run the fully connected layer to reduce the amount of calculating. All parameters are learned by calculating the least of the loss, thereby reducing cross-entropy. The training period is divided into two parts: the fitting of parameters on training data using ResNet Model and the adjustment of parameters on validation data. Neural networks are implemented with Keras and trained on an NVIDIA TITIAN V GPU using Adadelata optimizer with a learning rate of 1.0 and a batch size of 32.

#### 2.5.2 ACNN Model

A currently in use ACNN model was replicated to establish the effectiveness of our model and was used to compare with the newly proposed model. The basic structure of ACNN model that we use to compare with the ResNet Network is shown in Figure 8.

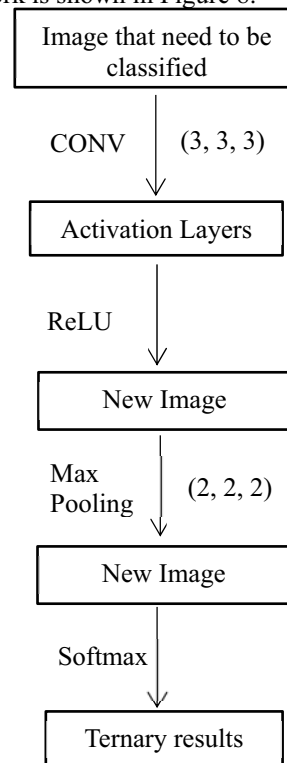


Figure 8 the structure of the ACNN model

A second training was conducted on the ACNN Model, using the same GPU, optimizer, learning rate, and batch size of those in the ResNet network implementation period.

## 2.6 Data Acquisition

Image samples are as in Figure 8. During this research, we evaluate the accuracy of the proposed model on the Alzheimer's Disease Neuroimaging Initiative (ADNI, National Institutes of Health Grant U19 AG024904) database on the website <http://adni.loni.usc.edu/>. The image was composed of 3013 scans from 321 subjects (130758 96×96 2D images for training, 26970 for validation, 29078 for the test) [16,17], which are about an equal balance of cognitively normal, mild cognitive impairment, and Alzheimer's disease. The dataset is also composed of demographic data, including clinical diagnosis, age, and sex, accompanied by a 3-D image[18], whereas one sample image is shown in the Figure 9.

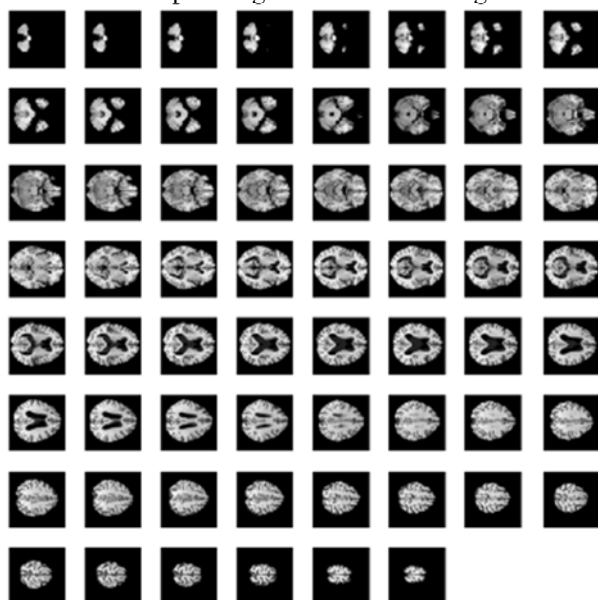


Figure 9 an example of one stack of 2D images that constitute the final 3D image

## 3 Results

### 3.1 Data on accuracy of classification collected on two models: ACNN and ResNet

The model was run and was tested on 469 3D image sets. Each of them has the dimension of (62, 96, 96, 1), meaning that they are 62 2D images in each set, with a width and height of 96 units.

#### 3.1.1 Accuracy on the ResNet model

The classification of the ResNet model can be summarized in the Table 1.

The total accuracy on the data is 0.8507, which is calculated by first summing up the positive values in AD and MCI cases and the negative value in the CN cases, then divide the number by the total 469.

Subgroup analysis on AD, MCI and CN were performed, and the respective value of accuracy is 0.85, 0.79 and 0.91, which is calculated by dividing the positive value by the Total in the AD and MCI part while dividing the negative value by the Total in the CN part.

Table 1. Classification report based on the comparison between prediction and demographic data based on ResNet Model

Actual/Predict	CN	MCI	AD	precision
CN	151	5	3	0.90
MCI	9	103	45	0.78
AD	7	24	122	0.82

#### 3.1.2 Accuracy on the ACNN model

Specifically, the classification of the ResNet model can be summarized in the Table 2.

The total accuracy on the data is 0.8017, which is calculated by first summing up the positive values in AD and MCI cases and the negative value in the CN cases, then divide the number by the Total 469.

Subgroup analysis on AD, MCI and CN were performed, and the respective value of accuracy is 0.82, 0.78 and 0.90, which is calculated by dividing the positive value by the Total in the AD and MCI part while dividing the negative value by the Total in the CN part.

Table 2. Classification report based on the comparison between prediction and demographic data based on the ACNN Model

Actual/Predict	CN	MCI	AD	precision
CN	144	12	3	0.91
MCI	10	128	19	0.79
AD	4	22	127	0.85

### 3.2 Statistical analysis of the comparison results

Let  $p_A$  and  $p_B$  represent the probability of the accuracy of the correct classification, which would be classified into AD, MCI, or control. The hypotheses to be checked are  $H_0: p_A > p_B$  versus  $H_a: p_A < p_B$ . The appropriate procedure is a two-proportion z-test for comparison. The testing data were randomly sampled from the population of all the MRI images. The second condition is the case that the sample sizes are large relative to the proportions involved. This condition is satisfied because all sample counts are larger than standard thresholds such as 5 and 10. In both cases, there are 376 out of 496 in the ACNN model, while there are 399 out of 496 cases in the ResNet model. Calculations show that  $p < 0.001$ . Because the p-value is very small, the null hypothesis is rejected and that the data provide convincing statistical evidence that ResNet is more accurate than the ACNN model.

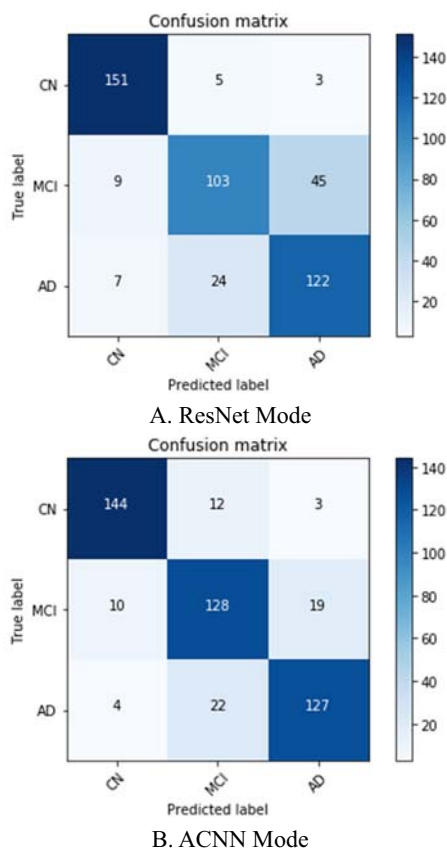
### 3.3 Visualizing the data

With Matplotlib being the tool, a curve for multi-class

classification accuracy was drawn below.

To better visualize the data, two plots comparing ACNN and ResNet Model are established as Figure 10, are drawn with the help of Matplotlib package in Python.

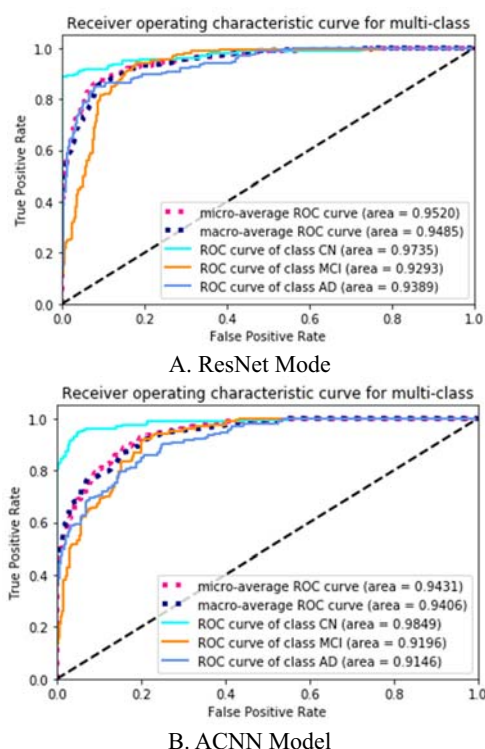
### 3.3.1 The Confusion Matrix



**Figure 10** The comparison of the confusion matrix between machine learning approaches.

### 3.3.2 The receiver operating characteristic (ROC) curve

When a machine classifies wrongly, especially when the machine considers a non-AD patient as an AD patient, a false positive result will occur. Contrarily, if a patient is classified correctly, we consider the case as a real positive. The ROC curve shown below demonstrates the change of the occurrence of false-positive cases and can reveal a characteristic of models. A plot of the area under the curve (AUC) is indicated using the Matplotlib module in Python to compare the quality of ROC. AUC with higher value shows a better fit of the model. It should also be noted that ROC is targeting binary classification problems. We have encountered a trinary classification problem, so it's necessary to find a method to make equalizations for these curves. The macro-average is the arithmetic mean, while the micro-average is the weighted average. The comparison of the ROC curve between two models are shown in Figure 11.



**Figure 11** The comparison of the ROC curve between two models with the figures drawn with the Matplotlib module in Python.

## 4 Discussion

As AD is a disease that requires extensive care and medication, it is, therefore, of high significance that early diagnosis is made. Not only will the patient get precautionary methods to avoid being troubled with the disease, but also the patient's family and the community can stay prepared in case of emergencies happened. With the growing population of elderly people, the quick yet accurate diagnosis is essential to improve the overall quality for all elderly persons under the current social situation that doctors are few in comparison to the demand of patients. By applying the residual network and convolutional networks, we could balance between the need for accuracy and the speed of recognition.

A model on ResNet is created to address this problem. The dataset is from Alzheimer's Disease Neuroimaging Initiative, taken from the North America population and may not represent the Chinese population who suffer from this disease. Nevertheless, considering the rife of the disease, much inference concluded from this experiment is still promising to be used in the future.

Coding ACNN model and ResNet model in Python (see attached files) and applying them onto the figures we obtain from the database, we get a model and get it tested on new, labeled 3D images. Two confusion matrixes have hence been shown (Figure 9). Observing more closely at the dark squares which stand for those classified wrongly, it should be noticed that the confusion matrix suggests an error about mistaking MCI for AD could be easily made. This might be a consequence of the magnification of human classification error. Suppose a human who

mistakenly labels an MCI case for that of AD or vice versa; the machine tends to copy the bias between the two circumstances, resulting in a not so precise classification.

Another way to compare the ACNN model and ResNet model is to consider the false positive rate. Observing the ROC curve is facilitated by calculating the area under the curve. In Figure 10, the AUC for MCI and AD all calculate to be more significant for ResNet (with  $0.9293 > 0.9196$  in the case of MCI and  $0.9389 > 0.9146$  in the case of AD), indicating possible increase incorrectness when using ResNet. Comparing two figures drawn from the area under macro-average and micro-average ROC curve, the area under both curve with the model ResNet is larger in size (with  $0.9520 > 0.9431$  and  $0.9485 > 0.9406$  in micro-average ROC and macro-average ROC respectively in Figure 10), which suggests a slightly advantageous model over ACNN.

The results provide a promising prospect of using ResNet for quickly and accurately diagnosing AD patients. Yet, there may be a drawback. Owing to the lack of high-quality equipment in most areas in China, 3D images can be challenging to obtain. We hope that the development of facilities and the reduction of the cost of using the machines can provide a more satisfactory answer.

## 5 Conclusion

This paper establishes effective predictive models for classifying the AD, MCI patients based on convolutional, ResNet Network, and machine learning methods. The parameters of the past ACNN models were used to simulate and make a comparison with the new model. The test results show, with statistical significance, that the new network, a version of the ResNet network, can predict the AD patients with higher accuracy compared with previous ones. The proposed algorithm can present a more accurate model that has reasonable practicability and convenience to perform on the future diagnosis.

## References

1. Ji Y. A 200-year history of Alzheimer's disease. *Chin J Contemp Neurol Neurosurg*. 2014; 14(3):156.
2. Alzheimer's Disease International. Dementia statistics. 2015 <https://www.alz.co.uk/research/statistics>
3. Goedert M, Spillantini MG, Jakes R, Rutherford D, Crowther RA. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron*. 1989; 3(4):519-26.
4. Perry EK, Perry RH, Blessed G, Tomlinson BE. Changes in brain cholinesterases in senile dementia of Alzheimer type. *Neuropathol Appl Neurobiol*. 1978; 4(4):273-7.
5. Sun X, Jin L, Ling P. Review of drugs for Alzheimer's disease. *Drug Discov Ther*. 2012; 6(6):285-90.
6. Nordberg A, Svensson AL. Cholinesterase inhibitors in the treatment of Alzheimer's disease: a comparison of tolerability and pharmacology. *Drug Saf*. 1998; 19(6):465-80.
7. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther*. 2014; 6(4):37.
8. Hefti F, Weiner WJ. Nerve growth factor and Alzheimer's disease. *Ann Neurol*. 1986; 20(3):275-81.
9. Jones L, Holmans PA, Hamshere ML, et al. Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *PLoS One*. 2010; 5(11):e13950.
10. Jafarian-Tehrani M, Sternberg EM. Animal models of neuroimmune interactions in inflammatory diseases. *J Neuroimmunol*. 1999; 100(1-2):13-20.
11. Prince JL, Links JM. Medical imaging signals and systems. Upper Saddle River: Pearson Prentice Hall; 2006.
12. Simard PY, Steinkraus D, Platt JC. Best practices for convolutional neural networks applied to visual document analysis. Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.
13. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. *Frontiers in Artificial Intelligence and Applications*. 2007; 160:3-24.
14. Hosseini-Asl E, Keynton R, El-Baz A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. 2016 IEEE International Conference on Image Processing (ICIP). 2016: 126-130.
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:770-778.
16. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010; 74(3): 201-9.
17. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008; 27(4): 685-91.
18. [https://github.com/jbrown81/ADNI\\_kaggle/blob/master/adni\\_kaggle.ipynb](https://github.com/jbrown81/ADNI_kaggle/blob/master/adni_kaggle.ipynb)