

## Alzheimer's disease diagnosis via 5-layer Convolutional Neural Network and Data Augmentation

Shuangshuang Gao\*

<sup>1</sup>School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, P R China

### Abstract

**OBJECTIVES:** Alzheimer's disease (AD) is a progressive neurodegenerative disease with insidious onset and one of the biggest challenges in geriatrics. Because the cause of the disease is unknown and there is currently no cure, AD Early diagnosis is particularly important.

**METHODS:** In this paper, we built a 5-layer convolutional neural network based on deep learning technology. We used six data augmentation methods to increase the size of the training set. Batch normalization and dropout techniques are also used, which are respectively associated with the convolutional layer and the fully connected layer, Form convolution batch normalization (CB) and dropout fully connected (DOFC) block respectively.

**RESULTS:** Our 5-layer CNN has shown excellent results on the training set, a sensitivity of 94.80%, a specificity of 93.98%, a precision of 94.04% and an accuracy of 94.39%, and has good performance compared with several other state-of-the-art methods.

**CONCLUSION:** In terms of classification performance, our method performs better than 8 state-of-the-art approaches and the performance of human observers. Therefore, this proposed method is effective in the detection of Alzheimer's disease.

**Keywords:** Alzheimer's disease, Convolutional neural network, Data augmentation, Batch normalization, Dropout.

Received on 07 August 2021, accepted on 29 August 2021, published on 16 September 2021

Copyright © 2021 Shuangshuang Gao *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.16-9-2021.170957

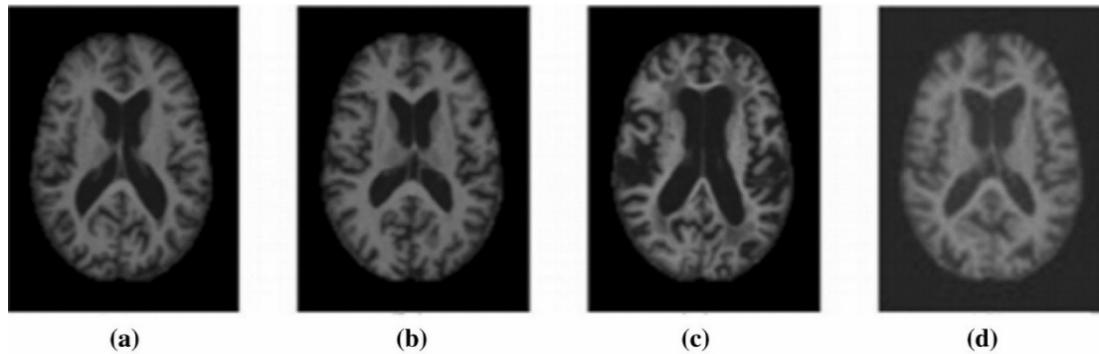
### 1. Introduction

Alzheimer's disease (AD) is a clinical syndrome characterized by progressive deterioration of cognitive and memory abilities. The clinical manifestations are generally memory loss, language impairment, cognitive decline, etc., mostly in old age. But it is not a patent for the elderly. Alzheimer's disease has slowly begun 20 years before the onset of the patient's disease. According to statistics, a new patient with Alzheimer's disease will be added every 3 seconds in the world. The incidence of AD is inextricably linked with age. The prevalence rate of people over 65 years old is 5%, and people over 80 years old. The prevalence rate has risen to 10%, and by the age of 90, the prevalence rate is as high as 50%. That is to say, more than half of the elderly in the centenarian period have Alzheimer's disease.

Some people describe Alzheimer's disease like this, he is like the eraser of the human brain, cruelly erasing the memory in the brain piece by piece, and slowly, it will also lose the basic ability to survive. It is mainly caused by damage to cells in a specific area of the brain, causing atrophy of the brain area. Alzheimer's disease is generally divided into three stages, as shown in Fig. 1. From the initial stage to the final death, the average time is about 4 to 10 years. The first stage is mild dementia, usually from the first year to the third year. The most basic symptom is memory loss, the judgment ability is weakened, and the patient cannot analyze and think about things, anxiety increases and emotional indifference. The second stage is moderate dementia, usually from the second year to the tenth year, and memory has been severely damaged; independent actions cannot be performed, help is needed from others, and one's negative emotions cannot be controlled. The third stage is severe dementia, which usually occurs in the eighth to twelfth year. The memory has been

\* Corresponding author. Email: shuangshuanggao@home.hpu.edu.cn

completely lost; the life cannot be taken care of by oneself, and the caregiver is completely needed, and eventually he will die.



**Fig. 1** The three main stages of AD. (a) Original Image (b) Mild (c) Moderate (d) Severe

The incidence of Alzheimer's disease is so high, but so far, there is still no effective cure, and only drugs and post-care can be used to slow down the condition. But it can also be prevented in advance. There are six main aspects to prevent Alzheimer's disease: diet, sleep, exercise, social interaction, stress reduction, and brain use. For the treatment of Alzheimer's disease, we can use medications for symptomatic treatment, such as antidepressants, anti-anxiety drugs, antipsychotics, etc., and can also be combined with cognitive rehabilitation therapy, vocational training, cognitive training, etc. Slow down the condition. Since Alzheimer's disease cannot be treated, it is particularly important for the early diagnosis of Alzheimer's disease.

At present, relevant scholars have proposed many methods based on computer vision and artificial intelligence for medical image classification, which can generally be divided into traditional machine learning methods and deep learning methods. Traditional machine learning medical image classification methods mainly include artificial neural network (ANN), genetic algorithm (GA), support vector machine (SVM), etc. Macedo Firmino et al. [1] have achieved certain results by extracting two-dimensional textures and using the SVM classifier to classify lung cancer with the possibility of malignant tumors, but the system does not have a good solution in the case of severe lesions. Michael C. Lee et al. [2] used a two-step supervised learning method to combine GA and random subspace method (RSM). In the combined classifier ensemble algorithm, 125 lung nodules and 125 pulmonary nodules were diagnosed. Better than the performance of a single step, but the generality of this model is weak. Yu et al. [3] classify non-small cell lung cancer by fully automated microscopic pathological image features, but it does not have much effect on large cell carcinoma in lung cancer. Suzuki et al. [4] developed a large-scale training artificial neural network (MTANN) based on ANN's pattern recognition technology, which achieved a sensitivity of 80.3%. We can find that these methods have achieved good detection results for medical image classification, but they cannot capture medical image features well, nor can they extract information containing medical features, so that the overall classification effect is not very good.

With the continuous advancement of technology, deep learning methods have become the preferred method for

analyzing medical images, providing new ideas for solving the problems of traditional machine learning in medical image classification. Therefore, medical image research based on deep learning has received extensive attention from scholars and conducted in-depth research. Esteva et al. [5] used a data set containing 129,450 clinical images to train a deep convolutional neural network. Experimental results showed that CNN's ability to diagnose skin cancer is comparable to that of dermatologists. Ehsan et al. [6] proposed to use a deep 3D convolutional neural network (3D-CNN) to classify AD patients. This classifier is superior to other conventional classifiers in terms of accuracy and robustness. Bidart et al. [7] developed a fully convolutional neural network (FCN) to automatically locate breast cancer tissue slices, and then divide the nuclear pattern into benign epithelial (BE), lymphocyte (L) and malignant epithelial (ME) cells for three categories, the final classification accuracy rate is 94.6%. Through analysis, we can know that deep learning technology has achieved great success in the field of medical image classification. However, medical images contain far more feature information than natural images, and their adaptive capabilities are poor. Therefore, when constructing deep learning models, we must fully consider the characteristics of medical images to build the most suitable model, effectively positioning and Extract feature information. Savio and Grana [8] put forward a Deformation-based Morphometry (DBM) method. Yuan [9] proposed an eigenbrain (EB) method to identify AD. Zhang [10] presented to use displacement field (DF) and machine learning (ML) methods. Liu, et al. [11] extends to a three-dimensional displacement field (3D-DF) method to identify AD. Zhang, et al. [12] proposed a characteristic brain for diagnosing AD, extending the 2D characteristic to 3D to form a 3D eigenbrain (3D-EB), combined with the Pol-SVM classifier, and achieved 92.81% accuracy. Li [13] combined wavelet entropy (WE) and biogeography-based optimization (BBO). Du [14] combined Pseudo Zernike Moment (PZM) and Linear Regression Classification (LRC) for AD diagnosis. Gorriz [15] presented a predator-prey particle swarm optimization (PPPSO). Jiang and Chang [16] combined batch normalization and dropout technique into convolutional neural network. Zhou [17] presented a novel network with Convolutional Block Attention Module (CBAM).

In this paper, we have constructed a five-layer convolutional neural network (5L-CNN) model by analyzing the characteristics of Alzheimer's disease images, and building a 5L-CNN to automatically extract effective features from the image as accurately as possible. In order to solve the problem of small data sets of medical images, we use data augmentation (DA) methods to expand our data sets to achieve better training results.

The rest of this paper is organized as follows. In the second section, we describe the dataset used and introduce pre-processing steps. Then in the third section, the method we used is explained in detail. The experimental results and discussion are given in Section 4. Finally, the fifth section summarizes and discusses the full paper.

## 2. Dataset

### 2.1. Subjects

There are two sources of brain imaging data used:

- (i) One is downloaded from the "Open Access Series of Imaging Studies (OASIS)", this dataset contains cross-sections of 416 male and female subjects aged 18 to 96. For each subject, this includes obtaining 3 or 4 individual T1-weighted MRI scans in a single scan session. The subjects are all right-handed. Due to the incomplete clinical data of some subjects, we finally selected 126 subjects with complete records, of which 28 were AD patients and 98 were HC subjects.
- (ii) The other is from four local hospitals, Zhong-Da Hospital of Southeast University, Children's Hospital of Nanjing Medical University, First Affiliated Hospital of Nanjing Medical University, and Affiliated Nanjing Brain Hospital of Nanjing Medical University. We selected 70 AD patients to balance the data set. Our inclusion criteria: are patients who meet the diagnostic criteria of AD, have a Clinical Dementia Rating (CDR) score of 1 to 2, and are over 60 years and have a Mini-Mental State Examination (MMSE) score of less than 24. The scoring criteria are shown in Table 1 and Table 2. The exclusion criteria were those with contraindications to MR scanning, incomplete clinical data, patients with mental illness. The study has been approved by the medical ethics committee of the local hospital, and the patients and their families have been informed and signed the informed consent before the study.

Table 1 Diagnostic criteria for MMSE

| $\leq 17$  | $\leq 20$             | $\leq 22$          | $\leq 23$         |
|------------|-----------------------|--------------------|-------------------|
| Illiteracy | Primary school degree | High school degree | University degree |

Table 2 Diagnostic criteria for CDR

| CDR 0        | CDR 0.5            | CDR 1.0       | CDR 2.0           | CDR 3.0         |
|--------------|--------------------|---------------|-------------------|-----------------|
| Non-demented | Suspected dementia | Mild dementia | Moderate dementia | Severe dementia |

Table 3 Demographic status of subjects

| Characteristic       | OASIS        |              | Local Hospitals |
|----------------------|--------------|--------------|-----------------|
|                      | AD           | HC           | AD              |
| Subject #            | 28           | 98           | 70              |
| Age                  | 77.75 ± 6.99 | 75.91 ± 8.98 | 76.34 ± 7.81    |
| Gender (M/F)         | 9/19         | 26/72        | 24/46           |
| Education            | 2.57 ± 1.31  | 3.26 ± 1.31  | 2.63 ± 1.42     |
| Socioeconomic status | 2.87 ± 1.29  | 2.51 ± 1.09  | 2.89 ± 1.16     |
| MMSE                 | 21.67 ± 3.75 | 28.95 ± 1.20 | 21.12 ± 4.62    |
| CDR                  | 1            | 0            | 1               |

The 70 AD data obtained by the local hospital was scanned on patients with the MR Prisma 3.0 magnetic resonance instrument produced by Siemens, Germany was used to examine the patients. Before the scan, explain the situation to the patients in detail, all patients are lying on their backs, and earplugs are used to reduce the impact of noise from the scanning equipment. The imaging parameters of the scan are: TE = 2.48 ms, TR = 1900 ms, TI = 900 ms, FA = 9°, FOV = 256 mm × 256 mm, matrix size = 256 × 256, slice thickness = 1 mm. Through the MP-RAGE sequence, a total of 176 sagittal slices covering the whole brain were obtained.

OASIS is combined with all the data of the local hospital to form our dataset. There is a total of 196 image data, 98 ADs and 98 HCs. The detailed information is listed in Table 3.

### 2.2. Preprocessing

The purpose of preprocessing the data is to output the region of interest (ROI), because the data will inevitably be interfered by the outside world and may lose characteristic information [18]. Through the preprocessing, the image quality can be enhanced and the image details can be highlighted to achieve better Experimental results. Fig. 3 shows two samples of preprocessed data, of which (a) shows the AD sample, (b) shows its healthy control sample. Fig. 2 shows a schematic diagram of preprocessing. First use the brain extraction tool to extract brain features, then FLIRT and FNIRT are used for spatial normalization, and the Gaussian kernel is smoothed. We choose the hippocampus as the slice selection criterion, because compared with HC, the hippocampus of AD patients is relatively small, slice

selection is implemented in the MNI space of  $Z = -22$  mm. Due to the different sources of our data, the gray level distribution of the image may be uneven, so we perform histogram stretching to standardize the scanned image. We first set the original data set as  $M_1$ , which consists of 98 ADs and 98 HCs

$$M_1 = \{m_1(1), m_1(2), \dots, m_1(i), \dots, m_1(196)\} \quad (1)$$

For the  $i$ -th image  $m_1(i)$ , it becomes a new image  $m_2(i)$  through histogram stretching

$$m_2(i|(a, b)) = \frac{m_1(i|(a, b)) - \sigma_{min}(i)}{\sigma_{max}(i) - \sigma_{min}(i)} \quad (2)$$

where  $\sigma_{min}(i)$  and  $\sigma_{max}(i)$  represent the minimum 0% and maximum 100% intensity values of the original image, but under normal circumstances, 5% and 95% will be used instead of 0% and 100%, because pixel with minimum and maximum values are more susceptible to noise. Finally, we get the image  $M_2$  after the histogram is stretched.

$$M_2 = HS(M_1) = \{M_1(1), M_1(2), \dots, M_1(i), \dots, M_1(196)\} \quad (3)$$

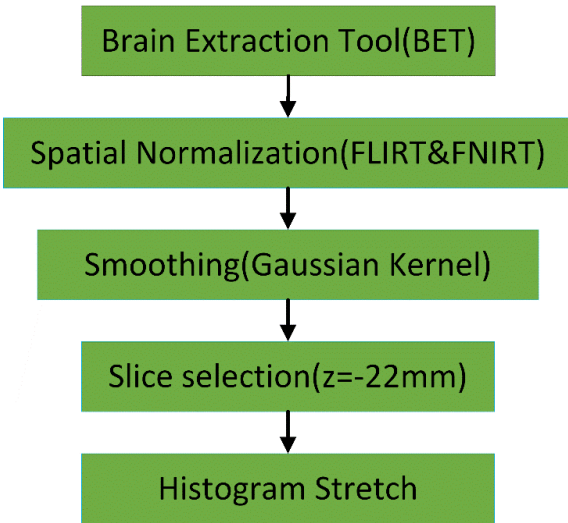
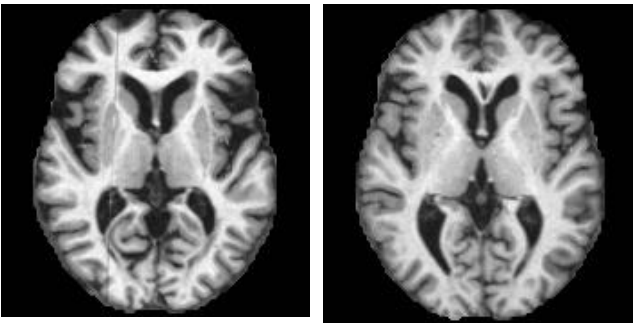


Fig. 2 Preprocessing flow chart



(a) AD

(b) HC

Fig. 3 Preprocessed images

### 3. Methodology

In this section, we will elaborate on the convolutional network and data augmentation methods we used.

#### 3.1. Convolutional Neural Network

Convolutional Neural Network (CNN) usually consists of two stages [19, 20], feature extraction and classification. Feature extraction consists of convolutional layers (CL) and pooling layers (PL) alternately. As shown in Fig. 4, the classification consists of fully connected layers (FCL) and softmax layers

##### 3.1.1. Convolutional Layer

The most basic operation in CNN is convolution, and CL performs 2D convolution along the width and height directions [21]. Assuming there is an input matrix, the kernel  $F_K$  and the output, its size is defined as [22, 23]

$$S(\text{Input}) = X \times Y \times C \quad (4)$$

$$S(\text{Filter } F_K) = X_F \times Y_F \times C(k = 1, 2, \dots, K) \quad (5)$$

$$S(\text{Output}) = X_O \times Y_O \times K \quad (6)$$

Where  $X$  and  $Y$  represent the length and height of the matrix [24],  $C$  represents the number of input channels, the subscripts  $F$  and  $O$  represent the filter and output respectively, and  $K$  represents the number of filters [25]. The output of each filter is stacked together to form a stacked output, we can get the output length  $X_O$  and height  $Y_O$

$$X_O = 1 + \frac{X - X_F + 2 \times D}{A} \quad (7)$$

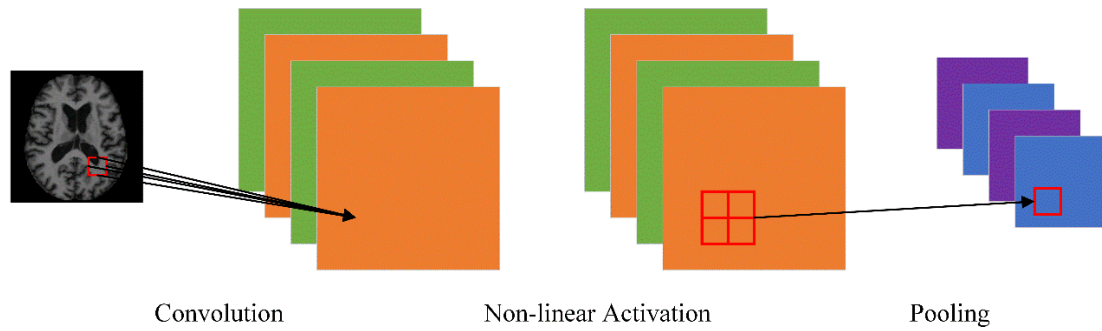
$$Y_O = 1 + \frac{Y - Y_F + 2 \times D}{A} \quad (8)$$

where  $A$  represents stride size, and  $D$  the margin. Therefore, the activation diagram is  $X_O \times Y_O \times K$ , as shown in Fig. 5.

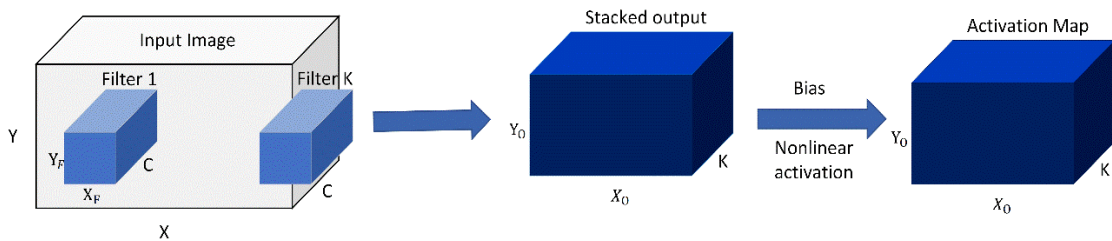
##### 3.1.2. Activation function

The stacked output adds a biased, and then through a nonlinear activation function, the output activation map can be obtained. The activation function of the traditional shallow neural network uses the sigmoid function [26], which is defined as

$$f_{sigmoid}(z) = \frac{1}{1 + \exp(-z)} \quad (9)$$



**Fig. 4.** Alternating convolution-pooling structure



**Fig. 5** Illustration of convolution layer

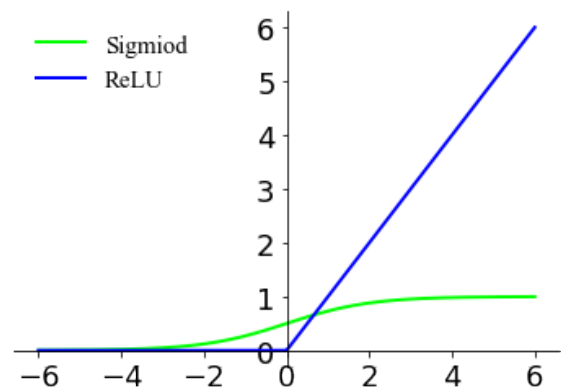
The sigmoid function can transform the continuous real value of the input into the output between 0 and 1. In the deep neural network [27, 28], the gradient will explode and disappear when the gradient is passed back. The probability of gradient explosion is very small, but the probability of gradient disappears is relatively large, and its analytical formula contains exponentiation, which is relatively time-consuming to solve by computer. For large-scale deep networks, this will greatly increase the training time. In order to solve this problem, the rectified linear unit (ReLU) has become popular. It not only solves the gradient vanishing problem, but also has a very fast calculation speed. It only needs to judge whether the input is greater than 0, and the convergence speed is much faster than sigmoid. Therefore, we choose the ReLU activation function. It is defined as

$$f_{relu}(z) = \max(0, z) \quad (10)$$

However, the output of ReLU is not zero-centered, and the Dead ReLU Problem means that certain neurons will never be activated [29], resulting in the corresponding parameters not being updated. Therefore, a leaky ReLU (LReLU) has been proposed, which has all the advantages of ReLU, and will not have zero-centered and dead ReLU problems. LReLU is a deformation made on the basis of ReLU. It is defined as

$$f_{trelu}(z) = \begin{cases} z & z > 0 \\ 0.01z & z \leq 0 \end{cases} \quad (11)$$

However, there is no accurate conclusion on the effect of LReLU. Some experiments performed well, while some experiments were not. The difference between sigmoid and ReLU activation functions is shown in Fig. 6.



**Fig. 6** Curves of sigmoid and ReLU activation functions

### 3.1.3. Pooling layer

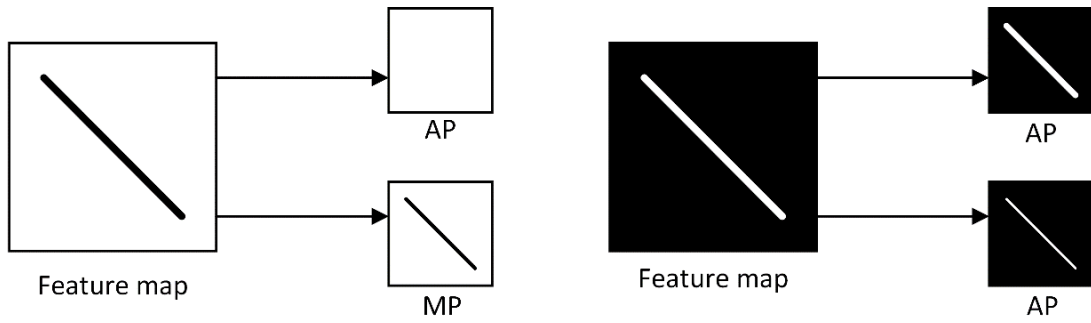


Fig. 7 Comparison of AP and MP

Because the activation map in CNN contains many features of the image, all of them are usually very large, which may lead to excessive computational cost and overfitting. Therefore, we need to use a pooling layer (PL) to reduce the size of the activation map, perform feature dimensionality reduction [30], reduce overfitting, and improve the fault tolerance of the model. In addition, PL can also achieve invariance, that is, rotation invariance, scale invariance and translation invariance.

The common pooling operations are average pooling (AP) and maximum pooling (MP). We first define a spatial neighborhood. AP is to calculate the average value in the corrected feature map in the window as the pooled value of the region, and MP is to select the maximum value in the corrected feature map in the window as the value after zone pooling. Fig. 7 shows two examples of pooling operations. A feature map is divided into 4 regions, that is, the pooling window has a size of  $2 \times 2$  and a step size of 2.

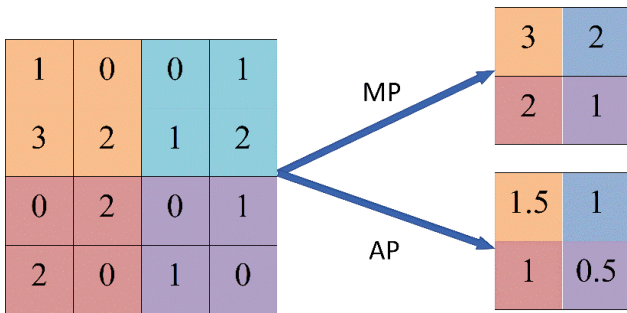


Fig. 8 Pooling graph

Both pooling operations have their own pros and cons. Average pooling can preserve the background well, but it is easy to make the picture blurred. Maximum pooling can well retain texture features and reduce the impact of useless information, but may encounter a situation where the output value fluctuates greatly. The comparison diagram of the two operations is shown in Fig. 8. Observing the left image, you can see that the foreground brightness is lower than the background brightness, the maximum pooling fails at this time, but in practice, the brightness of most foreground objects is greater than the background, so the maximum pooling may be used more.

### 3.2. Batch Normalization

From machine learning perspective, if the input distribution of a layer changes, then its parameters need to be relearned, which may make the network performance worse. This phenomenon is internal covariate shift (ICS) [31]. In order to solve this problem, it is necessary to keep the input distribution of each layer of neural network consistent during the training process [32, 33], so we have led to Batch Normalization (BN) to normalize the layer's inputs  $A$  at every mini batch.

For a layer with a mini batch of size  $m$ , we have training set  $A$  defined as

$$A = \{x_1, x_2, \dots, x_m\} \tag{12}$$

To ensure that the normalized output  $B$  can be evenly distributed

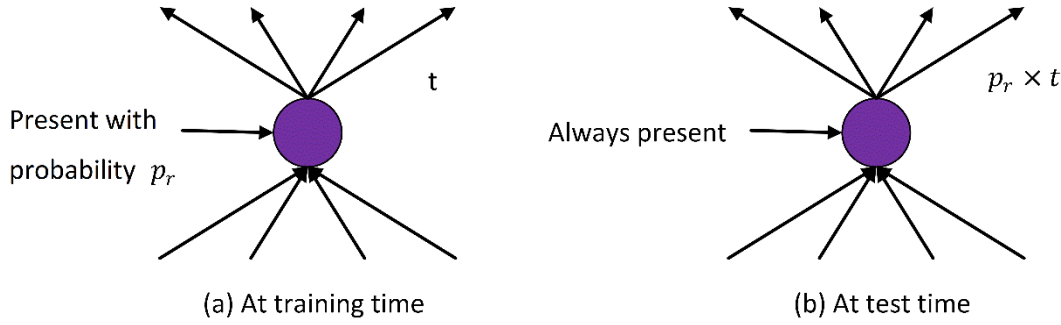
$$B = \{y_1, y_2, \dots, y_m\} \tag{13}$$

Its mean and variance are expressed as

$$\mu_A = \frac{1}{m} (\sum_{i=1}^m x_i) \tag{14}$$

$$\sigma_A^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_A)^2 \tag{15}$$

The standard normalization of input  $A$  will make it concentrate near 0. If the sigmoid activation function is used, this value interval is just close to the linear transformation interval, which will weaken the nature of the nonlinear transformation of the neural network [34]. In order to make the normalization operation not have a negative impact on the network's representation ability, an additional scaling and translation transformation can be used to change the value range



**Fig. 9** Dropout neuron at training and test stages

$$\bar{A}_i = \gamma \times \frac{(x_i - \mu_A)}{\sqrt{(\sigma_A^2 + \varphi)}} + \beta \quad (16)$$

$\gamma$  and  $\beta$  represent the parameter vector of scaling and translation respectively, and they are also two learnable parameters. This is the whole process of batch normalization

$$A \xrightarrow{BN} B \quad (17)$$

$$A = \{x_i\}, (i = 1, 2, \dots, m) \quad (18)$$

$$B = \{y_i\}, (i = 1, 2, \dots, m) \quad (19)$$

The normalization remains in the current layer, and the transformed output  $y_i \in B$  is passed to the next layer [35, 36]. In this paper, we propose to construct CB with one convolutional layer and one batch normalization layer.

### 3.3. Dropout Layer

In previous studies, the convolutional layer was used for feature learning, and the fully connected layer (FCL) was used as the classifier. FCL needs to be multiplied by a weight matrix, and then the result is added to a bias vector [37]. Because it is connected to all neurons in the previous layer, it has a large number of parameters, which may cause overfitting. Many solutions have been proposed to solve overfitting, among which dropout is simple and the effect is very good. In this paper, we add dropout layer (DOL) before each FCL to form a dropout fully connected (DOFC) block to prevent overfitting and improve the generalization ability of the model [38].

In the learning and training process of the dropout method, the neural network training unit is randomly removed from the network according to a certain probability [39, 40], and the selection of dropout neurons is random, Fig. 9 illustrates the performance of dropout neuron in training and test stages. The selections of dropout are random with a retention probability  $P_r$  (default is 0.5).

$$\text{Training: } T = \begin{cases} t & \text{with } P_r \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

where  $T$  means the weights of dropped out neurons, and  $t$  means before dropout. During the test, the entire neural network did not use dropout, but the weight of the neural network will be reduced due to  $P_r$ .

$$\text{Test: } T = P_r \times t \quad (21)$$

### 3.4. Data Augmentation

In the field of medical imaging, lack of dataset is a common phenomenon. Since our dataset has only 196 images, the corresponding training cannot be performed, and the model cannot achieve the expected effect [41]. In order to solve this problem, we use the data augmentation (DA) method [42] to expand the data set. First, the entire data set  $E$  is randomly divided into 98 training sets  $C$  and 98 test sets  $D$ , as shown in Table 4. The training set of 98 images is significantly smaller for the convergence of the convolutional neural network, and we use DA to increase it in this training set.

Table 4 The division of training set and test set

|          | HC | AD | Total |
|----------|----|----|-------|
| Training | 49 | 49 | 98    |
| Test     | 49 | 49 | 98    |
| Total    |    |    | 196   |

In 98 training set, we use six types of data augmentation methods (DA-6) for each picture  $c(k) \in C, k = 1, 2, \dots, 98$ , rotation, scaling, random translation, Gamma correction, noise injection and random affine.

(1) Rotation. The rotation angle is set between  $-30^\circ$  and  $30^\circ$ , in  $2^\circ$  increments each time, skipping  $\theta = 0$ .

$$c^{rot}(k) = \text{rotate}[c(k)] = [c_1^{rot}(k, \theta_1), \dots, c_{30}^{rot}(k, \theta_{30})] \quad (22)$$

where

$$\theta_1 = -30^\circ, \theta_2 = -28^\circ, \dots, \theta_{15} = -2^\circ, \theta_{16} = 2^\circ, \dots, \theta_{30} = 30^\circ \quad (23)$$

(2) Scaling. The scaling factor  $\vartheta$  is 0.7 to 1.3, the step size is 0.02, skip  $\vartheta = 1$ .

$$c^{scale}(k) = scale[c(k)] = [c_1^{scale}(k, \vartheta_1), \dots, c_{30}^{scale}(k, \vartheta_{30})] \quad (24)$$

where

$$\vartheta_1 = 0.7, \vartheta_2 = 0.72, \dots, \vartheta_{15} = 0.98, \vartheta_{16} = 1.02, \dots, \vartheta_{30} = 1.3 \quad (25)$$

(3) Random translation (RT). Random horizontal displacement  $\tau_x$  and vertical displacement  $\tau_y$  for each image are 30 times, the displacement range is  $[-10, 10]$ , and it obeys the random distribution  $\cup$ .

$$c^{RT}(k) = PT[c(k)] = [c_1^{RT}(k, \tau_x, \tau_y), \dots, c_{30}^{RT}(k, \tau_x, \tau_y)] \quad (26)$$

where

$$\tau_x \sim \cup [-10, 10], \tau_y \sim \cup [-10, 10] \quad (27)$$

(4) Gamma correction (GC). The gamma-value  $\alpha$  varied from 0.7 to 1.3 with step of 0.02, skipping the value of 1.

$$c^{GC}(k) = GC[c(k)] = [c_1^{GC}(k, \alpha_1), \dots, c_{30}^{GC}(k, \alpha_{30})] \quad (28)$$

where

$$\vartheta_1 = 0.7, \vartheta_2 = 0.72, \dots, \vartheta_{15} = 0.98, \vartheta_{16} = 1.02, \dots, \vartheta_{30} = 1.3 \quad (29)$$

(5) Noise injection (NI). Inject noise into each image to generate thirty new images, using zero-mean Gaussian noise with a variance of 0.01.

$$c^{NI}(k) = NI[c(k)] = [c_1^{NI}(k), c_2^{NI}(k), \dots, c_{30}^{NI}(k)] \quad (30)$$

(6) Random affine (RA). Random affine transformation is performed on each image, and the maximum affine factor  $\gamma$  is 0.2.

$$c^{RA}(k) = RA[c(k)] = [c_1^{RA}(k), c_2^{RA}(k), \dots, c_{30}^{RA}(k)] \quad (31)$$

Combining the above 6 DA methods, 180 new pictures are made for each original picture, so our training data set has a total of 17,738 pictures. The results are shown in Table 5.

Table 5 Training dataset data augmentation result

|                           | HC   | AD   | Total |
|---------------------------|------|------|-------|
| Original Training Dataset | 49   | 49   | 98    |
| Rotation                  | 1470 | 1470 | 2940  |
| Scaling                   | 1470 | 1470 | 2940  |
| Random translation        | 1470 | 1470 | 2940  |
| Gamma correction          | 1470 | 1470 | 2940  |
| Noise injection           | 1470 | 1470 | 2940  |
| Random affine             | 1470 | 1470 | 2940  |
| Training Dataset after DA | 8869 | 8869 | 17738 |

### 3.5. Measurement

In order to verify the performance of our model, we will evaluate it on 98 test sets. We select sensitivity, specificity [43, 44], precision, accuracy, MCC, F1, and FMI as performance test indicators. We use the confusion matrix visualization tool to obtain these indicators. Our ideal confusion matrix (CM) is

$$CM_{ideal} = \begin{bmatrix} TP & 0 \\ 0 & TN \end{bmatrix} \quad (32)$$

We assume that the confusion matrix is

$$CM_{realistic} = \begin{bmatrix} \text{True Positive}(TP) & \text{False Negative}(FN) \\ \text{False Positive}(FP) & \text{True Negative}(TN) \end{bmatrix} \quad (33)$$

Where the positive category is AD and the negative category is HC. We obtain other performance indicators for model performance verification.

(1) Sensitivity. It represents the proportion of all positive samples that the model judges to be truly positive, and it measures the classifier's ability to recognize positive samples.

$$Sensitivity = \frac{TP}{TP+FN} \quad (34)$$

(2) Specificity. Represents the proportion of negative samples that the model judges to be true negatives, and measures the classifier's ability to recognize negative samples.

$$Specificity = \frac{TN}{FP+TN} \quad (35)$$

(3) Precision. It is defined as the proportion of true positives among all the results that are judged to be positive.



The larger the value, the stronger the ability to identify true positive samples.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (36)$$

(4) Accuracy. The correct rate is our most common evaluation indicator, the number of samples that the model predicts correctly is divided by the number of all samples. Generally speaking, the higher the correct rate, the better the classifier.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad (37)$$

(5) MCC. MCC is a more balanced indicator that considers true positives, true negatives, false positives, and false negatives. It is usually used to measure classification performance in the case of dichotomy. It is still applicable even when the content of samples in the two categories is very different.

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \quad (38)$$

(6) F1-score. It is a new single indicator that combines accuracy and sensitivity. When the accuracy and

sensitivity are both high, a higher F1-score can be obtained. The maximum value is 1 and the minimum value is 0.

$$\text{F1} = \frac{2TP}{2TP+FP+FN} \quad (39)$$

(7) FMI. FMI is defined as the geometric mean of precision rate and recall rate. It is an external indicator of clustering performance measurement. The larger the value, the better the clustering result.

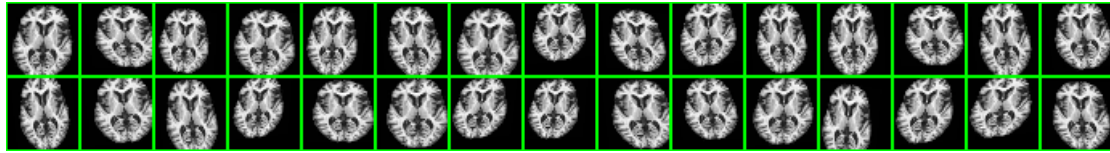
$$\text{FMI} = \frac{TP}{\sqrt{(TP+FP)(TP+FN)}} \quad (40)$$

## 4. Experiment Result and Discussions

### 4.1. Augmentation result

Take Fig. 3(a) as an example. The data augmentation result is shown in Fig. 10. There were 30 new images for each DA type, so an original image generates a total of 180 new images. Fig. 10 a-f show the results of rotation, scaling, random translation, Gamma correction, noise injection and random affine respectively.





(f) Random affine

Fig. 10 Data augmentation results

### 4.2. CNN

The details of our proposed 5-layer CNN architecture are shown in Table 6, which includes three convolutional layers as the feature extraction stage and two fully connected layers as the classification stage. The padding value of each layer is the same. In the convolution stage, we named each block "Convolution Batch Normalization (CB)". In the fully connected phase, each block is named "Dropout fully connected (DOFC)". Here the parameter  $n a \times a / b$  represents a filter of size  $b \times b$  and pooling size of  $c$ . The last column of the table shows the size of the AM, as shown in Fig. 11.

Table 6 Details of each layer in proposed 5L-CNN

| Layer   | Parameters                           | AM                        |
|---------|--------------------------------------|---------------------------|
| Input   | $n a \times a / b$                   | $128 \times 128 \times 1$ |
| CB-1    | $32 \ 3 \times 3 / 2$                | $64 \times 64 \times 32$  |
| CB-2    | $64 \ 3 \times 3 / 2$                | $32 \times 32 \times 64$  |
| CB-3    | $128 \ 3 \times 3 / 2$               | $16 \times 16 \times 128$ |
| Flatten |                                      | $1 \times 32768$          |
| DOFC-4  | $W(50 \times 32768); B(50 \times 1)$ | $1 \times 50$             |
| DOFC-5  | $W(2 \times 50); B(2 \times 1)$      | $1 \times 2$              |

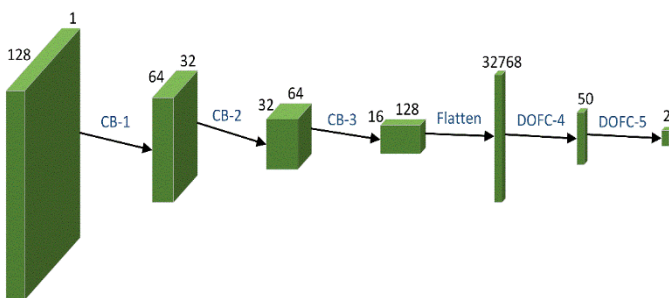


Fig. 11 Structure of proposed 5L-CNN

### 4.3. Statistical Analysis

In our experiment, using the original 98 pictures, and 17,640 pictures generated by data enhancement, a total of 17,738 pictures were used to train our model. In order to evaluate the performance of our 5L-CNN model, we obtain

the sensitivity, specificity, precision, accuracy, F1, MCC, and FMI indicators through the confusion matrix. We perform ten runs of 5L-CNN, and the results of the ten runs are shown in Table 7, and we use Fig. 12 to show the results more intuitively. The data shows: the average values of sensitivity, specificity, precision, accuracy, F1, MCC, and FMI are  $94.80 \pm 1.48\%$ ,  $93.98 \pm 1.63\%$ ,  $94.04 \pm 1.56\%$ ,  $94.39 \pm 1.32\%$ ,  $94.41 \pm 1.31\%$ ,  $88.79 \pm 2.63\%$ ,  $94.41 \pm 1.31\%$ . The overall performance of our method has good results, and it can be clearly seen that the accuracy of 94.39% has been achieved, which also shows the superiority of our 5L-CNN.

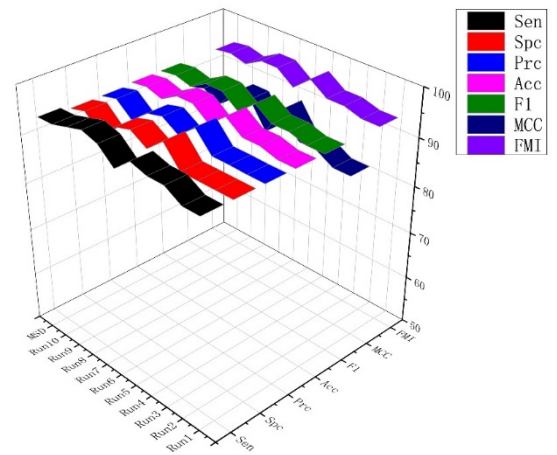


Fig. 12 Performance results of 5L-CNN

### 4.4. Comparison with State-of-the-art Approaches

In order to further prove the effectiveness of our 5L-CNN method, compare it with several state-of-the-art approaches: DBM, EB, DF-ML, 3D-DF, 3D-EB, WE-BBO, PZM-LRC, PPPSO. Test set the detailed results of the ten runs are shown in Table 8. Fig. 13 is a graphical comparison of several models. It can be clearly observed from the results that our proposed 5L-CNN is superior to the other eight methods. From the point of view of accuracy, DBM is  $90.36 \pm 1.31\%$ , EB is  $90.36 \pm 0.81\%$ , DF-ML is  $91.89 \pm 0.70\%$ , 3D-DF is  $92.56 \pm 1.03\%$ , 3D-EB of  $92.04 \pm 0.77\%$ , WE-BBO is  $91.58 \pm 0.60\%$ , PZM-LRC is  $93.06 \pm 1.30\%$ , PPPSO of  $92.45 \pm 0.67\%$ , and our proposed 5L-CNN achieves an accuracy of  $94.39 \pm 1.32\%$ . In terms of sensitivity, the result of 5L-CNN is  $94.80 \pm 1.48\%$ , which is also better than several other methods, such as  $88.27 \pm 3.27\%$  DBM or  $91.63 \pm$

1.91% 3D-EB. It can be seen that we use data augmentation methods to expand our dataset so that the model is better trained, and by using batch normalization and dropout layers to improve the stability and generalization ability of the model. Therefore, our method is superior to other methods. In the future, our method may be applied to other fields.

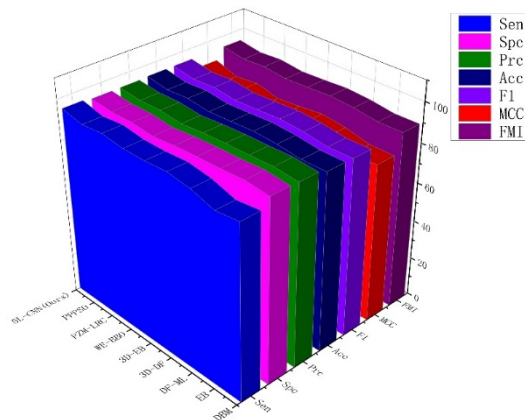


Fig. 13 Performance comparison

## 5. Conclusion

In this research on AD diagnosis, we proposed 5L-CNN. We use DA-6 to solve the problem of small-scale dataset, add batch normalization and dropout layer to the network, each convolutional layer is associated with BNT to form a CB block, and use DOFC to replace the traditional fully connected layer. The results show that our 5L-CNN has good performance: the average values of sensitivity, specificity, precision, accuracy, F1, MCC, and FMI are  $94.80 \pm 1.48\%$ ,  $93.98 \pm 1.63\%$ ,  $94.04 \pm 1.56\%$ ,  $94.39 \pm 1.32\%$ ,  $94.41 \pm 1.31\%$ ,  $88.79 \pm 2.63\%$ ,  $94.41 \pm 1.31\%$ , its results are better than eight state-of-the-art algorithms.

Our method also has shortcomings. Although the dataset is expanded by data augmentation technology, it is still relatively small. Some new network technologies have not been tested. In future research, we can use more advanced data augmentation techniques and try transfer learning techniques or pretraining models.

## References

- [1] M. Firmino, G. Angelo, H. Morais, M. R. Dantas, and R. Valentim, "Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy," *Biomedical Engineering Online*, vol. 15, Jan 6 2016, Art no. 2, doi: 10.1186/s12938-015-0120-7.
- [2] M. C. Lee *et al.*, "Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction," *Artificial Intelligence in Medicine*, vol. 50, no. 1, pp. 43-53, Sep 2010, doi: 10.1016/j.artmed.2010.04.011.
- [3] K.-H. Yu *et al.*, "Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features," *Nature Communications*, vol. 7, Aug 2016, Art no. 12474, doi: 10.1038/ncomms12474.
- [4] K. Suzuki, S. G. Armato, F. Li, S. Sone, and K. Doi, "Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography," *Medical Physics*, vol. 30, no. 7, pp. 1602-1617, Jul 2003, doi: 10.1118/1.1580485.
- [5] A. Esteva, B. Kuprel, and R. A. Novoa, "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Oncologie*, vol. 19, no. 11-12, pp. 407-408, Nov 2017. [Online]. Available: <Go to ISI>://WOS:000419735600012.
- [6] E. H. Asl *et al.*, "Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network," *Frontiers in Bioscience-Landmark*, vol. 23, pp. 584-596, Jan 1 2018, doi: 10.2741/4606.
- [7] M. J. Gangeh, R. Bidart, M. Peikari, A. L. Martel, and S. Nofech-Mozes, "Localization and classification of cell nuclei in post-neoadjuvant breast cancer surgical specimen using fully convolutional networks," in *Digital Pathology*, 2018.
- [8] A. Savio and M. Grana, "Deformation based feature selection for Computer Aided Diagnosis of Alzheimer's Disease," (in English), *Expert Systems with Applications*, Article vol. 40, no. 5, pp. 1619-1628, Apr 2013, doi: 10.1016/j.eswa.2012.09.009.
- [9] T. F. Yuan, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," (in English), *Front. Comput. Neurosci.*, Article vol. 9, Jun 2015, Art no. 66, doi: 10.3389/fncom.2015.00066.
- [10] Y. Zhang, "Detection of Alzheimer's disease by displacement field and machine learning," *PeerJ*, vol. 3, 2015, Art no. e1251, doi: 10.7717/peerj.1251.
- [11] G. Liu, P. Phillips, and T.-F. Yuan, "Detection of Alzheimer's Disease by Three-Dimensional Displacement Field Estimation in Structural Magnetic Resonance Imaging," *Journal of Alzheimer's Disease*, vol. 50, no. 1, pp. 233-248, 2016.
- [12] Y. Zhang *et al.*, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," *Frontiers in Computational Neuroscience*, vol. 9, Jun 2 2015, Art no. 66, doi: 10.3389/fncom.2015.00066.
- [13] Y.-J. Li, "Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10393-10417, 2018, doi: 10.1007/s11042-016-4222-4.
- [14] S. Du, "Alzheimer's Disease Detection by Pseudo Zernike Moment and Linear Regression Classification," *CNS & Neurological Disorders - Drug Targets*, vol. 16, no. 1, pp. 11-15, 2017.
- [15] J. M. Gorriz, "Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization," *Journal of Alzheimer's Disease*, vol. 65, no. 3, pp. 855-869, 2018, doi: 10.3233/JAD-170069.
- [16] X. Jiang and L. Chang, "Classification of Alzheimer's Disease via Eight-Layer Convolutional Neural Network with Batch Normalization and Dropout Techniques," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 5, pp. 1040-1048, 2020.

- [17] Q. Zhou, "ADVIAN: Alzheimer's Disease VGG-Inspired Attention Network Based on Convolutional Block Attention Module and Multiple Way Data Augmentation," *Front. Aging Neurosci.*, vol. 13, 2021, Art no. 687456.
- [18] B. Peng, Y.-X. Liang, J. Yang, and K. So, "Image processing methods to elucidate spatial characteristics of retinal microglia after optic nerve transection," *Scientific Reports*, vol. 6, 2016, Art no. 21816, doi: 10.1038/srep21816.
- [19] G. Bhadane, A. A. Jadhav, V. S. Bhong, S. A. Inamdar, and D. P. Narsale, "Misfire Detection of Automotive Engines with Convolutional Neural Network," in *Techno-Societal 2018: Proceedings of the 2nd International Conference on Advanced Technologies for Societal Applications - Vol 1*, India, P. M. Pawar, B. P. Ronge, R. Balasubramaniam, A. S. Vibhute, and S. S. Apte, Eds., 2020: IEEE, pp. 121-131, doi: 10.1007/978-3-030-16848-3\_12. [Online]. Available: <Go to ISI>://WOS:000625869400012
- [20] D. Santos *et al.*, "Real-Time Traffic Sign Detection and Recognition using CNN," *IEEE Latin America Transactions*, vol. 18, no. 3, pp. 522-529, Mar 2020, doi: 10.1109/ta.2020.9082723.
- [21] S.-H. Wang, "Covid-19 Classification by FGCNet with Deep Feature Fusion from Graph Convolutional Network and Convolutional Neural Network," *Information Fusion*, vol. 67, pp. 208-229, 2020/10/09/2021, doi: 10.1016/j.inffus.2020.10.004.
- [22] T. Ameen, L. Chen, Z. X. Xu, D. D. Lyu, and H. Y. Shi, "A Convolutional Neural Network and Matrix Factorization-Based Travel Location Recommendation Method Using Community-Contributed Geotagged Photos," *Isprs International Journal of Geo-Information*, vol. 9, no. 8, Aug 2020, Art no. 464, doi: 10.3390/ijgi9080464.
- [23] S. Y. Lee and S. K. Lee, "Deep convolutional neural network with new training method and transfer learning for structural fault classification of vehicle instrument panel structure," *Journal of Mechanical Science and Technology*, vol. 34, no. 11, pp. 4489-4498, Nov 2020, doi: 10.1007/s12206-020-1009-3.
- [24] G. Zhao, "Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units," *Journal of Real-Time Image Processing*, vol. 15, no. 3, pp. 631-642, 2018, doi: 10.1007/s11554-017-0717-0.
- [25] D. S. Guttery, "Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network," *Information Processing and Management*, vol. 58, 2, 2021, Art no. 102439.
- [26] S. Xie, "Alcoholism Identification Based on an AlexNet Transfer Learning Model," (in English), *Frontiers in Psychiatry*, Original Research vol. 10, 2019-April-11 2019, Art no. 205, doi: 10.3389/fpsy.2019.00205.
- [27] K. Maeda, K. Horii, T. Ogawa, and M. Haseyama, "Multi-Task Convolutional Neural Network Leading to High Performance and Interpretability via Attribute Estimation," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E103A, no. 12, pp. 1609-1612, Dec 2020, doi: 10.1587/transfun.2020SML0006.
- [28] A. Dutta, D. Sil, A. Chandra, and S. Palit, "CNNbased musical instrument identification using time-frequency localized features," *Internet Technology Letters*, 2020, Art no. e191, doi: 10.1002/itl2.191.
- [29] M. C. Kim and J. H. Lee, "Color reproduction in virtual lip makeup using a convolutional neural network," *Color Research and Application*, vol. 45, no. 6, pp. 1190-1201, Dec 2020, doi: 10.1002/col.22549.
- [30] C. Tang, "Cerebral Micro-Bleeding Detection Based on Densely Connected Neural Network," (in English), *Frontiers in Neuroscience*, Original Research vol. 13, 2019-May-17 2019, Art no. 422, doi: 10.3389/fnins.2019.00422.
- [31] M. Chen, Y. Li, and L. Han, "Detection of dendritic spines using wavelet-based conditional symmetric analysis and regularized morphological shared-weight neural networks," *Computational and Mathematical Methods in Medicine*, 2015, Art no. 454076, doi: 10.1155/2015/454076.
- [32] M. Ahmad, I. Ahmed, F. A. Khan, F. Qayum, and H. Aljuaid, "Convolutional neural network-based person tracking using overhead views," *International Journal of Distributed Sensor Networks*, vol. 16, no. 6, Jun 2020, Art no. 1550147720934738, doi: 10.1177/1550147720934738.
- [33] S. Memon, M. Ahmed, S. Narejo, U. A. Baig, B. S. Chowdry, and M. R. Anjum, "SELF-DRIVING CAR USING LIDAR SENSING AND IMAGE PROCESSING," *International Journal of Grid and Distributed Computing*, vol. 13, no. 2, pp. 77-88, Dec 2020, doi: 10.33832/ijgdc.2020.13.2.06.
- [34] Y.-D. Zhang, D. R. Nayak, X. Zhang, and S.-H. Wang, "Diagnosis of secondary pulmonary tuberculosis by an eight-layer improved convolutional neural network with stochastic pooling and hyperparameter optimization," *Journal of Ambient Intelligence and Humanized Computing*, doi: 10.1007/s12652-020-02612-9.
- [35] Y. N. Zolotukhin, K. Y. Kotov, A. A. Nesterov, and E. D. Semenyuk, "Object Tracking in the Video Stream by Means of a Convolutional Neural Network," *Optoelectronics Instrumentation and Data Processing*, vol. 56, no. 6, pp. 642-648, Nov 2020, doi: 10.3103/s8756699020060163.
- [36] S. Katiyar and S. K. Borgohain, "Comparative Evaluation of CNN Architectures for Image Caption Generation," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 12, pp. 793-801, Dec 2020. [Online]. Available: <Go to ISI>://WOS:000606823500092.
- [37] S. C. Satapathy, "A five-layer deep convolutional neural network with stochastic pooling for chest CT-based COVID-19 diagnosis," *Machine Vision and Applications*, vol. 32, 2021, Art no. 14.
- [38] S.-H. Wang, "DenseNet-201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 2s, p. Article 60, 2020, doi: 10.1145/3341095.
- [39] J. H. Jo and C. B. Moon, "Learning Collision Situation to Convolutional Neural Network Using Collision Grid Map Based on Probability Scheme," *Applied Sciences-Basel*, vol. 10, no. 2, Jan 2020, Art no. 617, doi: 10.3390/app10020617.
- [40] A. Noor, Y. Q. Zhao, R. Khan, L. W. Wu, and F. Y. O. Abdalla, "Median filters combined with denoising convolutional neural network for Gaussian and impulse noises," *Multimedia Tools and Applications*, vol. 79,

- no. 25-26, pp. 18553-18568, Jul 2020, doi: 10.1007/s11042-020-08657-4.
- [41] J. Yang, "Pathological brain detection in MRI scanning via Hu moment invariants and machine learning," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 29, no. 2, pp. 299-312, 2017, doi: 10.1080/0952813X.2015.1132274.
- [42] V. V. Govindaraj, "High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 9, pp. 2012-2021, 2019.
- [43] V. Srivastava and R. K. Purwar, "Classification of CT Scan Images of Lungs Using Deep Convolutional Neural Network with External Shape-Based Features," *Journal of Digital Imaging*, vol. 33, no. 1, pp. 252-261, Feb 2020, doi: 10.1007/s10278-019-00245-9.
- [44] A. Sezer and H. B. Sezer, "Convolutional neural network based diagnosis of bone pathologies of proximal humerus," *Neurocomputing*, vol. 392, pp. 124-131, Jun 2020. [Online]. Available: <Go to ISI>://WOS:000531900100012.