Journal of Physics: Conference Series **1567** (2020) 022076 doi:10.1088/1742-6596/1567/2/022076

Deep convolutional neural networks for the detection of macular diseases from optical coherence tomography images

I Kharisudin*, M F Az-zahra, E R Winarti, and S B Waluya

Department of Mathematics FMIPA Universitas Negeri Semarang, Indonesia

*Corresponding author: iqbalkharisudin@mail.unnes.ac.id

Abstract. The purpose of this research is to design a system to recognize CNV image, DME, Drusen, and Normal, which were produced by Optical Coherence Tomography (OCT). A system would provide a training model, evaluation, and accuracy value. We used the Convolution Neural Network method with default parameter 50 epoch, one stride, 83484 train data images, and learning rate value 0.001 with the help of Python 3.7 software. The examination using epoch variation, stride, number of train data, and learning rate value resulted in different accuracy values. According to epoch variation, the best accuracy was 50 epoch with an accuracy value of 0.99 and loss validation of 0.2034. The best accuracy of stride value variation was one stride with an accuracy value of 0.99 and loss validation of 0.2267. The best accuracy of train data variation was 83484 images with an accuracy value of 0.99 and a loss validation value of 0.2524. The learning rate variation value with the best accuracy was 0.0001, with an accuracy value of 0.992 and validation loss value of 0.2524. According to the result of the research, it was obtained that convolution neural network architecture gain the best model with accuracy value 0.992 according to variation parameter 50 epoch, one stride, 83484 train data images, and learning rate 0.0001.

1. Introduction

Nowadays, there are many statistical computing applications exist. One of them is in the field of pattern recognition technology using neural networks and has been applied in some science fields. Some research about artificial neural networks (ANN) and pattern recognition are [1][2][3][4]. Mathematically, the digital image can be written in the form of a matrix as follows.

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0 & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix}$$

Where $f(x,y)$is the intensity value which is received by the sensor in every pixel point (x,y), and $0 < f(x,y) < \infty$.

Artificial neural network consists of many simple processing units that have a natural tendency to safe information, experiences and are reliable alike human brain [5][6]. Another definition about ANN is a set (N,V,w) with N set neuron, V set $\{(i,j)|i,j \in \mathbb{N}\}$ with each element called connection between neuron i and neuron j, and function $w: V \to \mathbb{R}$ defines value of $w(i,j)$ which is connection value between neuron i and neuron j [7].

Pattern recognition can be applied in several fields; one of them is in the medical area (see [8-10]). One of the data which is produced comes from the ophthalmology field. Ophthalmology is closely related to eyes, which eyes are one of the human's vital senses. Eyes are senses that functioned to see various surrounding objects [11].

Some degenerative diseases that attack eyes are Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen. Those three diseases attack macula in the retina, which can cause blindness. Degenerative disease is a disease that causes damage or destruction to body tissue or organs [12]. Optical Coherence Tomography (OCT) is an imaging technique that uses coherent light to capture high-resolution images of biological tissue [13]. OCT gives a considerable contribution to the new development of the ophthalmology field [14].

Figure 1. Optical Coherence Tomography (https://insightmedical.ca/product/zeiss-cirrus-4000-hd-oct/)

Convolutional Neural Network (CNN) is one of deep learning methods that is used to detect and recognize the digital image. This research uses deep learning methods, CNN method, and TensorFlow as the library on phyton, and will produce a system to acknowledge digital image. CNN architecture is showed in Figure 2.

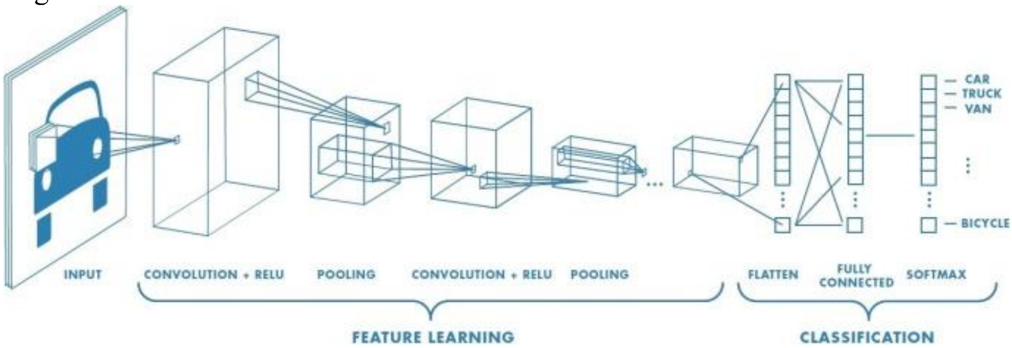

Figure 2. Convolutional Neural Network architecture

The first stage in CNN architecture is convolution. It was then continued by using ReLu (Rectifier Linear Unit) and the pooling process. This process is repeated until enough feature map is obtained, then keep to fully connected neural network so that output class can be derived. Some researches related to neural networks and CNN are [15-16]. According to the problems above, by using pattern identification technology, there is an idea to create a system that can recognize image patterns produced by OCT.

2. Research Methods

The method used in this research is a literature review and computing simulation analysis by collecting data or information from the library and internet related to the problems. The data used are OCT images of disease on the retina that are taken from internet sites Mendeley (https://data.mendeley.com/datasets/rscbjbr9sj/2). There are four conditions of the retina. The amount of OCT image data used is summarized in Table 1.

Table 1. Research data

Image	Data Amount	
	Training	Testing
Normal	26.315	250
CNV	37.205	250
DME	11.348	250
Drusen	8616	250

3. Result and Discussion

The research discusses the classification of 4 OCT images of the retinal eye conditions. The classification of the four image conditions uses the Convolutional Neural Network (CNN) algorithm. The process is divided into 2; they are data training process and validation process.

3.1. Network Architecture

The formation of a network architecture Convolutional Neural Network (CNN) algorithm is an important part. The structure of the CNN network architecture can affect the results of the accuracy of the model.

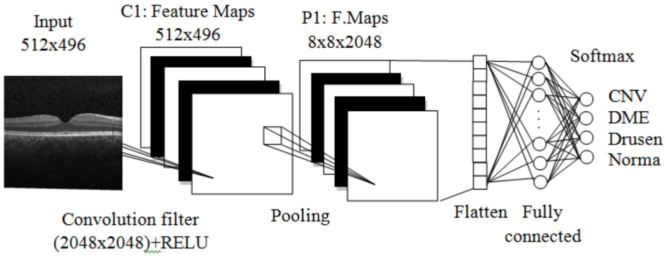

Figure 3. Network Architecture

Figure 3 is a network architecture of the training process in obtaining an optimal model. This study uses image input with an average size of 512 × 496. Convolution process is using kernel. The convolution process is a process of combining two different matrices to produce a new matrix value. After the convolution process, an activation function is added, which is RELU (Retrified Linear Unit). The purpose of adding the activation function is to change the negative value on the matrix to zero. The results in this process are the same size because using padding value 0. The pooling process is a process of reducing the matrix size. The pooling layer consists of a filter with a specific extent, which alternately shifts throughout the feature map area. This process uses max-pooling, which produces a new matrix and kernel. Flatten process or fully-connected layer is the process of changing the output pooling layer into a vector. This process uses dropout values before classifying. Dropout value is a neural network regulation technique that aims to select several neurons randomly and will not be used during the training process. The purpose of this process is to reduce overfitting during the training process. The last method is the process of activating the softmax function.

3.2. Data Training Result

After several processes in the Convolution Neural Network (CNN) algorithm, the results of training and validation are obtained. The parameters used are 50 epoch, one stride, 83484 OCT images, and a learning rate of 0.001. The following graph shows the accuracy and loss validation values of the training results. Based on Figure 4, the accuracy of the training model reaches to 99% with a validation loss value of 0.0693.

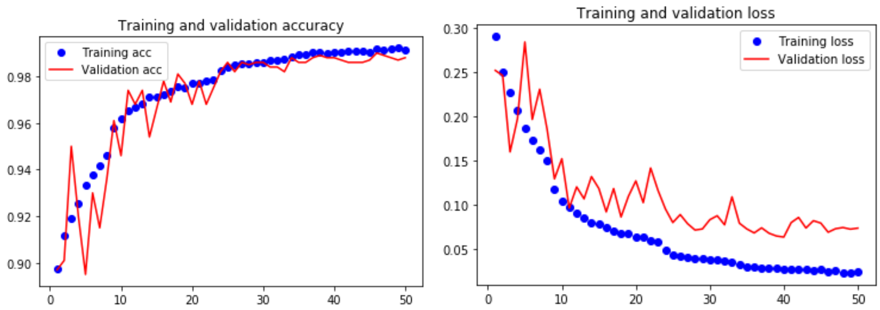

Figure 4. Training and validation accuracy graphic and Loss training and validation value graphic

3.2.1. Epoch Variation

Epoch is an iterative process on a data set that is repeated after the training process [17]. The default parameters used in this case are the value of stride 1, the number of data train 83484, and a learning rate value of 0.001.

Table 2. Accuracy value based on the effects of epoch amount

Number of Epoch	Accuracy validation		Loss validation		Time (second)
	Min	Maks	Min	Maks	
20	0.832	0.985	0.3706	0.0681	37367
50	0.897	0.99	0.2524	0.0693	63232
100	0.923	0.987	0.2034	0.0697	80317

Based on Table 2, the maximum value of accuracy obtained in the case of epoch 50 is 0.99. However, the minimum loss validation value that is obtained at epoch 100 is 0.2034. The conclusion on the accuracy value is parallel with the number of epoch that is too much or too little can affect the accuracy value.

3.2.2. Stride Variation

Stride is a shift size value that is used to shift the filter through image input [18]. If we use the value of stride 1, the convolution filter will shift 1 pixel horizontally then vertically. The default parameters used are epoch amount of 50, data train amount of 83484, and a learning rate value of 0.001.

Table 3. Accuracy value based on the effects of stride amount

Stride value	Accuracy validation		Loss validation		Time (second)
	Min	Maks	Min	Maks	
1	0.897	0.99	0.2524	0.0693	63232
2	0.885	0.989	0.2688	0.0557	91708
5	0.917	0.988	0.2267	0.0546	109059

Based on Table 3, the maximum value of accuracy obtained in the case of stride 1 is 0.99. However, the minimum value of validation loss is obtained on stride 5, which is 0.2267. The conclusion on the accuracy value is parallel with the smaller stride used, more detailed information will be achieved.

3.2.3. Data Training Variation
Data train is data that is tested to the program in order to recognize image input patterns that have similarities or are close to the actual image [18]. The default parameters used are epoch amount of 50, amount of stride 1, and learning rate values 0.001.

Table 4. Accuracy value based on the effects of data train amount

Number of data train	Accuracy validation		Loss validation		Time (second)
	Min	Max	Min	Max	
83484	0.897	0.99	0.2524	0.0693	63232
20000	0.855	0.989	0.3455	0.0557	91708
4000	0.831	0.917	0.5324	0.4725	4050

Based on Table 4, the maximum value of accuracy obtained in the case of 83848 data trains is 0.99. This is parallel with the value of the minimum validation loss of 0.2524. The conclusion on the accuracy value is that the more data train used, the higher the accuracy value will be.

3.2.4. Learning Rate Variation.
The value of the learning rate is one of the parameters that can be determined by researchers. Generally, image classification uses learning rate values from 0.1 to 0.0001 [18]. The default parameters used are epoch amount of 50, amount of stride 1, and data train amount of 83484 images.

Table 5. Accuracy value based on the effects of learning rate value

l_r value	Accuracy validation		Loss validation		Time (second)
	Min	Max	Min	Max	
0.01	0.778	0.988	0.5526	0.0774	64794
0.001	0.87	0.99	0.2524	0.0693	63232
0.0001	0.89	0.992	0.2684	0.033	65925

Based on Table 5, the maximum value of accuracy is obtained in the case of learning rate 0.0001 is 0.992. However, the minimum value of validation loss is obtained at a learning rate of 0.001 is 0.2524. The conclusion on the value of accuracy is the smaller the learning rate, the process of convergence of the loss value during the training process is faster than other learning rate values.

Based on the above experiments with four-parameter variations, the best model with a maximum accuracy value of 0.992 is obtained with parameter criteria number of epoch 50, amount of stride 1, amount of data train 83484, and a learning rate of 0.0001. Henceforth the model obtained will be used for the evaluation process.

3.3. Evaluation Result
The following section presents the results of an evaluation of 1000 OCT image data. Each condition class, either CNV, DME, Drusen, or Normal, consists of 250 images. We tested 1000 images using the best model obtained from the above experiments. The results of the confusion matrix is presented in Table 6 below.

Based on Table 6, the program can predict 250 CNV images correctly, 243 DME images correctly, 249 Drusen images correctly, and 250 Normal images correctly. By using model.features.47-0.99.hdf5, the program mispredicts 8 images. 7 DME images based on evaluation results predicted two as CNV, two as Drusen, and three as Normal. There is 1 Drusen image predicted as CNV. We calculate accuracy based on the confusion matrix Table 6 using the following formula.

Table 6. Confusion matrix on model "model.features.47-0.99.hdf5"

Matrix		Predict class			
		CNV	DME	Drusen	Normal
Actual Class	CNV	250	0	0	0
	DME	2	243	2	3
	Drusen	1	0	249	0
	Normal	0	0	0	250

$$\text{Accuracy} = \frac{\sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} \times 100\% = \frac{992}{1000} \times 100\% = 99.2\%.$$

According to these results, the accuracy value of the model is 99.2%. Although these results are relatively sound, the use of this model still needs competent expert judgment. This model can be used to help as an initial recommendation for the analysis of four macular eye conditions.

4. Conclusion

Based on the experimental results of the parameters epoch, stride, learning rate, and the number of data train, the following findings are obtained. The number of epoch can affect the accuracy value. According to the stride parameter, the small stride value will get a high accuracy value because the information obtained will be more detailed. In the learning rate parameter, the lower learning rate, the process of convergence of the loss value during the training process is faster than the relatively high learning rate. In the number of data train, the more data used, the higher the accuracy value will be. However, the data train parameter is the parameter that most influences accuracy value. This is because the more data that is used as a data train, the system will learn a lot of image patterns. This makes the system will more predict the image correctly.

References

[1] Hu Z Bodyanskiy Y V, Tyshchenko O K and Tkachov V M 2017 *Int. J. Intell. Syst. Appl.* **9**(6) 24
[2] Acharya U R Oh S L Hagiwara Y Tan J H and Adeli H 2018 *Comput. Biol. Med.* **100** 270
[3] Leibe B, Matas J, Sebe N, and Welling M Eds 2016 *Computer Vision ECVV 2016* (Springer)
[4] Coudray N Ocampo P S Sakellaropoulos T Narula N Snuderl M Fenyo D Moreira A L Razavian N and Tsirigos A 2018 *Nat. Med.* **24** 1559
[5] Ciaburro G and Venkateswaran B 2017 *Neural Networks with R* (Packt Publishing)
[6] Sewak M, Karim M R, and Pujari P 2018 *Practical Convolutional Neural Networks* (Birmingham: Packt Publishing,)
[7] Aghdam H H and Heravi E J 2017 *Guide to Convolutional Neural Networks* (Switzerland: Springer)
[8] Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E and Darrell T 2014 *Proc. 31st Int. Conf. Mach. Learn.* **32**(1) 647
[9] Krizhevsky B A Sutskever I and Hinton G E 2012 *Res. Highlight* **60** 842012
[10] Gupta D and Choubey S 2015 *Int. J. Emerg. Technol. Adv. Eng.* **4**(3) 598
[11] Michael F L and and Fernald R D 1997 *Brain. Behav. Evol.* **50**(4) 253
[12] Harahap J and Andayani L S 2018 *Talent. Conf. Ser. Trop. Med.* **1**(1) 142
[13] Kermany D. S. *et al.* 2018 *Cell* **172**(5) 1122
[14] Novita H D and Moestidjab 2018 *J. Oftalmol. Indones.* **6** 169
[15] Lee C S Baughman D M and Lee A Y 2017 *Kidney Int. Reports* **2**(4) 322
[16] Mudau T and Coulter D 2018 *IST-Africa Week Conf.* **1**
[17] Goodfellow I Bengio Y and Courville A 2016 *Deep Learning* (MIT Press)
[18] Khan S Rahmani H Shah S S A and Bennamoun M 2018 *A Guide to Convolutional Neural Networks for Computer Vision* (Morgan & Claypool)