Journal of Telecommunication, Electronic and Computer Engineering
ISSN: 2180 – 1843 e-ISSN: 2289-8131 Vol. 14 No. 4
jtec.utem.edu.my

A Comparative Study of Deep Learning Parameters for Arcus Senilis Classification

Nur Farahin AH[1], RA Ramlee[1,4], MZ Mas'ud[2] and M. A. Alias[3]
[1]*Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Durian Tunggal, Melaka, Malaysia.*
[2]*Faculty of Information and Communication Technology, UTeM, Durian Tunggal, Melaka, Malaysia.*
[3]*Faculty of Manufacturing Engineering, UTeM, Durian Tunggal, Melaka, Malaysia.*
[4]*Advanced Sensors Embedded Control System Research Group (ASECs), Durian Tunggal, Melaka, Malaysia.*
ridza@utem.edu.com

Article Info

Article history:
Received June 25th, 2022
Revised Aug 15th, 2022
Accepted Dec 26th, 2022

Index Terms:
Arcus Senilis
Convolutional Neural Network
Deep Learning
Hyperparameters
Non-Invasive

Abstract

Deep learning technique has recently yielded positive results that have increased productivity system for artificial intelligent task, especially the digital image processing and advance machine vision. The popularity of deep learning technique has a major impact on solving complex problems in many fields, particularly the medical science owing to its applications in medical imaging, disease diagnosis, and much more. However, successful application of deep learning depends upon the appropriate setting of the parameters to achieve better result. Therefore, this paper presents a comparative analysis of different base learning rate and batch size configurations for arcus senilis (AS) classification using deep learning techniques. In this analytical study, a dataset of 402 eye images comprising 158 normal and 244 abnormal eye images was employed. Two well-known ResNet-50, VGG-19 and pre-trained convolutional neural network (CNN) models have been trained and validated with 10-fold cross validation using the proposed dataset. Furthermore, base learning rate and batch size were adjusted accordingly to determine the optimal convergence of each model by observing the validation accuracy and error. Experimental result shows that the best combined system has achieved an overall accuracy of 99.78% with a base learning rate of 0.0001 and a batch size of 20 on CNN pre-trained model validation set. Moreover, CNN produces the best result on F1-score and standard deviation of 99.77% and 0.464 respectively. Thus, it can be concluded that CNN requires a considerably smaller number of parameters and reasonable computing time to achieve state-of-the-art performances. This study shows that CNN has the tendency to consistently improve inaccuracy with growing number of epochs, with no signs of overfitting and performance.

I. INTRODUCTION

Technological advancements like deep learning technique often led to health research improvements, allowing greater precision and clarity in data collection and interpretation. Essentially, it is a group of machine learning approaches that exploit many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification [1]. Despite of this success, the application of deep neural networks remains a black art, often frequently call for years of learning to effectively select the best hyperparameters, regularization, and network architecture, all of which are closely related [2]. Additionally, training the convolutional neural network (CNN) to classify the images accurately, necessitates the hyperparameters, such as learning rate and batch size to be adjusted with extensive trial and error that take a long period of time: These hyperparameters will have a big impact on the performance on the network's performance as it approaches convergence. As stated by Smith [3] and Bengio [4], it is widely known that a training algorithm will slowly converge if the learning rate is too low, and will diverge if the learning rate is too high [4], [5]. Moreover, Kandel l et al. stated that one of the most crucial hyperparameters is the batch size referred to the number of images used to train a single forward and backward pass[6]. Currently, the process of determining the hyperparameters, including the design of the network architecture, requires expertise and extensive trial and error, which is based more on chance rather than science. In this study, we propose a comparative study for finding and examine the optimal settings for base learning rate and batch size hyperparameters for AS classification using deep learning. ResNet-50, VGG-19 and pre-trained convolutional neural network (CNN) models were used in this study. This experimental study aims to provide an optimal value of the learning rates and batch size to be considered before it is fused into CNN model.

The rest of the paper is organized as follows: the related work is presented in section II. Methodology is presented in section III. Section IV discusses the obtained results are proposed in section III and the whole paper is concluded in section V.

II. RELATED WORK

Banowati et al. [7] developed an application that can detect AS using the camera of an Android smartphone. They used pre-trained Inception-v3 architectures with 10-fold cross-validation fused with epoch, batch size, and learning rate values of 800, 100 and 0.01 respectively. The proposed model has yielded 97.45% classification accuracy. However, these methods require longer time to train and test dataset in order to achieve better accuracy. Moreover, Kocejko et al. [8] used CNN, VGG16, Resnet and Inception architecture with 0.0001 initial learning rate using Adam optimizer to automatically detect the presence of AS. The performance of the models was evaluated on a set of images taken by volunteers using a custom mobile application. They achieved 88% accuracy for detecting AS in a real-world scenario and an F1 score of 86%. In addition to the AS study using deep learning approach, they are widely appreciated in a variety of medical applications, including brain hemorrhage [9], tuberculosis [10], COVID-19 [11] and pneumonia detection [12].

Awwal et al. [9] investigated the advantages of using SVM as a classifier instead of a three-layer neural network on brain computer tomography hemorrhage. The employed gradient descent optimizer with a batch size of 60 and a learning rate of 0.001 to train the model for 100 epochs. The validation classification accuracy of 90.65% is achieved. Furthermore, a study by Munadi et al. [10] proposed a technique in detecting the tuberculosis disease via enhancement of pre-processing approaches using Resnet18 architecture. To achieved the goals, batch size of 6 and a learning rate of 0.01 were fed into the pre-trained Resnet18 model and they achieved a satisfactory classification accuracy and area under curve scores of 89.92% and 94.8% respectively. Other than that, there was a study on well-known corona virus 2019 (COVID-19) disease via deep learning approach by Ahmad Musha et al. on 2022 [12]. This study presented a computer-aided deep learning model named COVID-CXDNetV2 to detect COVID-19 and pneumonia from the X-ray images in real-time. The network architecture was trained utilizing the Adam optimizer with 0.001 learning rate, 32 batch sizes, with a maximum number of epochs 100. The model has obtained an overall classification accuracy of 97.9% with a loss of 0.052 for multiclass classification (COVID-19, pneumonia, and normal) and 99.8% accuracy, 99.52% sensitivity, 100% specificity with a loss of 0.001 for binary classification (COVID-19 and normal).

Drawn from the above studies, it can be concluded that deep learning technique requires a considerably smaller number of hyper-parameters and reasonable computing time to achieve state-of-the-art performances.

III. METHODOLOGY

A. Datasets

Dataset is a collection of different types of data stored in a digital format. The datasets are mainly made up from images, text, audio, video, numerical data points and much more to solve various complex artificial intelligence tasks, such as image recognition or classification and video recognition. In this study, a dataset of 402 eye images was used, comprising 158 images for normal and 244 abnormal eyes referred to AS images. Moreover, the abnormal eye images were obtained from various public medical websites while the normal eye images were obtained from UBIRIS [26], which is a public database of eye images.

B. Image Pre-processing

The collected eye images were then processed in several stages consisting of image pre-processing and augmentation. The pre-processing phase was initially improved with CLAHE and unsharp masking algorithm in the preprocessing steps. CLAHE has been used primarily for the enhancement of low-contrast medical images [13], [14]. CLAHE is an improvement of HE, a method that applies local contrast enhancement to the local regions of the image [15], [16]. It is a technique commonly used by researchers for contrast enhancement of images, especially in the pre-processing phase, as in [17]–[19]. Moreover, the unsharp masking (UM) is a technique of adjusting the contrast of images to make the image look sharper. After the image quality were enhanced, the image augmentation was used to overcome the limitation of dataset.

C. Deep Learning Architecture

The proposed method was developed using MATLAB R2020b programme. After that, the training data was divided into training and validation data with a ratio of 70% to 30%, respectively. In this analytical study, pre-trained CNN (proposed technique), ResNet-50, and VGG-19 models were employed. These models were developed using a proposed dataset with 10-fold cross-validation with Adam optimizer on normal and CA images.

D. Model Evaluation Measure

To determine the best convergence for each CNN, we investigated the training validation accuracy and error improvement. Training will automatically stop if validation accuracy and errors do not improve. In this study, all models, namely the CNN (the proposed method), Resnet-50, and VGG-19 were evaluated using confusion matrix to measure the classification accuracy as in (1) that is, the total number of correct predictions divided by the total number of predictions. Table 1 shows a sample of confusion matrix with hyperparameter based on a batch size= 10 and a learning rate= 0.0 that have four possibilities: First, True positive (TP) rate, which is the number of images that classify or predict CA images accurately. Second, False positive (FP) rate, also known as type I error, which represents the number of images misclassified as CA images, in which the images should be classified as normal eye images. Third, False negative (FN) rate, also known as type II error, which is the number of images misclassified as normal eye images, in which the images should be classified as CA images. Finally, True negative (TN) rate, which represents the number of accurately classified normal images.

$$\text{Accuracy} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{1}$$

Table 1
A sample of confusion matrix with hyperparameter; batch size = 10, learning rate = 0.01

Model	TP	TN	FN	FP
CNN	45	42	0	5
Resnet50	3	26	42	41
VGG19	0	47	45	0

IV. RESULT AND DISCUSSION

Table 1 and 2 below shows the results of the proposed deep learning model with 10-fold cross training and validation using Adam optimizer on different batch size and learning rate on training and validation using the proposed architecture. In Table 1, it shows that CNN and Resnet50 outperformed VGG19 on each batch size and learning rate. VGG19 used less learning rate on each batch size to perform accurately but needed longer time to train the dataset, as shown in Table 4. Moreover, Resnet50 also performed poorly on learning rate of 0.01 on each batch size, while CNN performed better on a consistent basis for each batch size by learning rate.

Table 2 shows that all model gained a better accuracy if the learning rate were below than 0.01 on each batch size. CNN once again gained better accuracy consistently on every learning rate, while VGG19 needed less learning rate of 0.00001 to gain 100% classification accuracy. Apparently, it took a longer time to complete the validation task.

From the observation on the training and validation performance of the model, as shown in Table 1 and 2, we can conclude that the best result was demonstrated by the proposed CNN model that produced an excellent average classification accuracy of 99.78% with a good values standard deviation of 0.4638 on learning rate of 0.0001 and batch size of 20. Even though the performance of CNN model produced a good result, the Resnet50 model actually produced satisfactory result on the average classification accuracy of 100%. However, it still could not outperform CNN model as the execution time of Resnet50 (851.7 s) was higher than the CNN model (707.6 s).

Table 2
Training Performance of Deep Learning Architecture

BS	LR	CNN		Resnet-50		VGG-19	
		Acc. (%)	Std. Dev.	Acc. (%)	Std. Dev.	Acc. (%)	Std. Dev.
20	0.01	99.34	0.7891	72.02	18.1496	49.56	1.0627
	0.001	100	0	99.35	1.8864	72.93	18.7295
	0.0001	**100**	**0**	**100**	**0**	97.82	2.9873
	0.00001	98.52	0.4780	100	0	99.45	0.9348
30	0.01	99.43	0.6567	79.12	14.9199	50.66	0.9276
	0.001	100	0	99.35	1.5932	84.23	13.8946
	0.0001	100	0	**100**	**0**	99.23	1.0436
	0.00001	98.21	0.2961	100	0	99.34	1.0627

Table 3
Validation Performance of Deep Learning Architecture

BS	LR	CNN		Resnet-50		VGG-19	
		Acc. (%)	Std. Dev.	Acc. (%)	Std. Dev.	Acc. (%)	Std. Dev.
20	0.01	96.76	1.8963	67.27	16.9912	49.36	1.5457
	0.001	99.56	0.7691	98.58	3.7455	75.25	17.1358
	0.0001	99.78	0.4638	**100**	**0**	99.07	2.1889
	0.00001	98.68	0.4638	99.01	0.3479	100	0
30	0.01	96.84	2.3018	78.04	14.2647	50.96	1.3492
	0.001	99.02	1.6464	98.91	1.5322	82.51	13.0963
	0.0001	99.78	0.4638	99.56	0.5680	99.95	0.1581
	0.00001	98.9	1.4980	99.12	0.4638	100	0

Table 4
Validation Execution Time (s).

Batch size: 20				
Models	0.01	0.001	0.0001	0.00001
CNN	734.1	758.8	**707.6**	744
Resnet50	854.4	867.5	851.7	877.5
VGG19	1719.3	1726.4	1831.3	1821.6

V. CONCLUSION

This study is a comparative analysis of different base learning rate and batch size configurations for arcus senilis (AS) classification using deep learning techniques. For this purpose, 158 normal eye images and 244 AS eye images from UBIRIS and online medical images respectively were analyzed. Experimental result showed that the best combined system demonstrated an average classification accuracy of 99.78% with base learning rate of 0.0001 and batch size of 20 on CNN pre-trained model validation set. The performance of Resnet50 also can be considered as the best result for AS classification but it took a longer execution time compared to the CNN model. Overall, the model works well based on the hyperparameter tuning, but further research is needed to optimize the time execution and include larger databases to determine the impact on the classification process. The technique can also be used to diagnose eye problems caused by many types of eye diseases such as cataracts, diabetic retinopathy and pterygium.

Table 5
The selection of hyperparameters values using on AS classification

Parameters	Value
Model	CNN
Learning rate	0.0001
Batch size	20
Number of epochs	10
Optimizer	Adam
Dropout	0.5
Activation function of last layer	Softmax

ACKNOWLEDGMENT

We would like to thank to Universiti Teknikal Malaysia Melaka (UTeM) and Advanced Sensors & Embedded Control System Research Group (ASECs) for supporting us in accomplishing this study.

REFERENCES

[1] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4. Now Publishers Inc, pp. 197–387, 2013. doi: 10.1561/2000000039.

[2] L. N. Smith, “A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay,” Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.09820

[3] L. N. Smith, “A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay,” Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.09820

[4] Y. Bengio, “Practical Recommendations for Gradient-Based Training of Deep Architectures. In: Montavon, G., Orr, G.B., Müller, KR. (eds),” in *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science.*, vol. 7700, Springer, Berlin, Heidelberg, 2012, pp. 437–478. doi: https://doi.org/10.1007/978-3-642-35289-8_26.

[5] L. N. Smith, “Cyclical Learning Rates for Training Neural Networks,” Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.01186

[6] I. Kandel and M. Castelli, “The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset,” *ICT Express*, vol. 6, no. 4, pp. 312–315, Dec. 2020, doi: 10.1016/j.icte.2020.04.010.

[7] C. Banowati, A. Novianty, and C. Setianingsih, “Cholesterol Level Detection Based on Iris Recognition Using Convolutional Neural Network Method,” *2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies, CSUDET 2019*, pp. 116–121, 2019, doi: 10.1109/CSUDET47057.2019.9214690.

[8] T. Kocejko, J. Ruminski, M. Mazur-Milecka, M. Romanowska-Kocejko, K. Chlebus, and K. H. Jo, “Using convolutional neural networks for corneal arcus detection towards familial hypercholesterolemia screening,” *Journal of King Saud University - Computer and Information Sciences*, no. xxxx, 2021, doi: 10.1016/j.jksuci.2021.09.001.

[9] A. M. Dawud, K. Yurtkan, and H. Oztoprak, “Application of deep learning in neuroradiology: Brain haemorrhage classification using transfer learning,” *Comput Intell Neurosci*, vol. 2019, 2019, doi: 10.1155/2019/4629859.

[10] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, “Image Enhancement for Tuberculosis Detection Using Deep Learning,” *IEEE Access*, vol. 8, pp. 217897–217907, 2020, doi: 10.1109/ACCESS.2020.3041867.

[11] I. D. Apostolopoulos and T. A. Mpesiana, “Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks,” *Phys Eng Sci Med*, vol. 43, no. 2, pp. 635–640, Jun. 2020, doi: 10.1007/s13246-020-00865-4.

[12] A. Musha, A. al Mamun, A. Tahabilder, Md. J. Hossen, B. Hossen, and S. Ranjbari, “A deep learning approach for COVID-19 and pneumonia detection from chest X-ray images,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, p. 3655, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3655-3664.

[13] J. Ma, X. Fan, S. X. Yang, X. Zhang, and X. Zhu, “Contrast Limited Adaptive Histogram Equalization-Based Fusion in YIQ and HSI Color Spaces for Underwater Image Enhancement,” *Intern J Pattern Recognit Artif Intell*, vol. 32, no. 7, 2018, doi: 10.1142/S0218001418540186.

[14] S. M. Pizer *et al.*, “Adaptive Histogram Equalization and Its Variations,” 1987.

[15] G. F. C. Campos, S. M. Mastelini, G. J. Aguiar, R. G. Mantovani, L. F. de Melo, and S. Barbon, “Machine learning hyperparameter selection for Contrast Limited Adaptive Histogram Equalization,” *EURASIP J Image Video Process*, vol. 2019, no. 1, Dec. 2019, doi: 10.1186/s13640-019-0445-4.

[16] U. Kuran and E. Can Kuran, “Parameter selection for CLAHE using multi-objective cuckoo search algorithm for image contrast enhancement,” *Intelligent Systems with Applications*, vol. 12, p. 51, 2021, doi: 10.1016/j.iswa.2021.20.

[17] N. Kharel, A. Alsadoon, P. W. C. Prasad, and A. Elchouemi, “Early diagnosis of breast cancer using contrast limited adaptive histogram equalization (CLAHE) and Morphology methods,” in *2017 8th International Conference on Information and Communication Systems, ICICS 2017*, May 2017, pp. 120–124. doi: 10.1109/IACS.2017.7921957.

[18] A. M. Reza, “Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement,” 2004.

[19] B. K. Umri, M. Wafa Akhyari, and K. Kusrini, “Detection of COVID-19 in Chest X-ray Image using CLAHE and Convolutional Neural Network,” Oct. 2020. doi: 10.1109/ICORIS50180.2020.9320806.