Comparison and Combination of Leaky ReLU and ReLU Activation Function and Three Optimizers on Deep CNN for COVID-19 Detection

Almuzhidul Mujhid[a] , Sugiyarto Surono[a, 1], Nursyiva Irsalinda[a], Aris Thobirin[a]
[a] *Mathematics Department, University of Ahmad Dahlan, Indonesia*

Abstract. COVID-19 detection is an interesting field of study in the medical world and the commonly used method is classification. In determining the best detection model, several classification architectures, such as SVM, KNN, and CNN were utilized. The CNN is a changeable architecture due to having combinations of varying numbers of hidden layers or different activation and optimizer functions. Therefore, this study uses a deep CNN architecture with a combination of Leaky ReLU activation functions and 3 different optimizers, which include Adagrad, Adadelta, and Adamax. The results showed that the combination of the Leaky ReLU activation function and the Adamax optimizer produced good and stable accuracy in the CRX and CT datasets.

Keywords. COVID-19, Image classification, Deep CNN, Leaky ReLU activation function, Optimizers

1. Introduction

According to a statement issued by the World Health Organization (WHO), the COVID-19 virus has become the largest disease-spreading phenomenon in the last 2 years and has caused a fairly high mortality rate of 3.4% [1][2][3][4]. Since COVID-19 occurrence in 2020, it has generated a lot of data in the medical world and has therefore become an interesting subject of study to support decision-making. An example of the decision-making method in image recognition is classification [5][6][7].

Furthermore, classification is an important process in image recognition learning that aims to group images. In this study, medical images are classified into certain categories and the methods used include K-Nearest Neighbor, Naïve Bayes, Decision Tree, and CNN. Among these four methods, many studies have shown that CNN provides good accuracy in the image classification process [8][9][10].

The growth of Deep Learning (DL) is very significant and has a positive impact on the medical world, mainly because of its ability to classify images. In this study, the created DL model aims to classify the CXR and CT images of COVID-19. Previous studies have classified CT and CXR datasets using different types of architectures, such as [11], SVM [12] and GoogleNet [13]. In a recent study with CT and CXR datasets

---

[1] Corresponding Author, Sugiyarto Surono, Mathematics Department, University of Ahmad Dahlan, Indonesia; Email: sugiyarto@math.uad.ac.id

using Deep CNN architecture with no feature extraction process, the analysis was performed with a combination of several activation functions and optimizers [1]. The combination of three optimizers and activation functions helps to increase the accuracy of COVID-19 detection. The performance results showed that the Depp-CNN architecture with a mini-batch size of 16, Leaky ReLU as an activation function in the hidden layer provides the best accuracy. It is important to note that one Leaky ReLU activation function and three other optimizers were proposed in this current study, which includes Adagrad, Adadelta, and Adamax. The Leaky ReLU was suggested because it does not ignore negative values, unlike the ReLU activation function that converts them to 0 [14][15][16]. Furthermore, the Adagrad, Adadelta, and Adamax optimizers are also expected to provide better accuracy in classifiying covid-19 images.

2. Proposed Approach

In this section gives several important definitions and theorems will be discussed which will be used to support further discussion which will be presented in the next section, those are facial expression datasets, Deep Convolutional Generative Adversarial Networks (DCGAN) and Convolutional Neural Networks (CNN).

2.1. Deep Convolutional Neural Network (CNN)

Recently, convolutional neural networks (CNN) are the most popular class model for image recognition and classification tasks [17][18]. This is an extension of the multilayer perceptron (MLP) for processing two-dimensional data. The CNN method consists of the image classification stage that uses feedforward and the learning stage using the backpropagation method. Furthermore, the softmax layer is used for the classification task [6][19] because it produces a well-crafted probability distribution of the output. CNN is also a very efficient method as no pre-processing or feature extraction is required before the training process [20].

The CNN architecture used in this study consists of several layers which are shown in Table 1 below.

Table 1. Deep CNN Architecture

Layer	Input Size	Filter Size	Output Size	Activation Function
Convolution	$227 \times 227 \times 3$	$3 \times 3 \times 8$	$225 \times 225 \times 8$	
Batch Norm	$225 \times 225 \times 8$		$225 \times 225 \times 8$	Leaky ReLU ReLU
Pooling	$225 \times 225 \times 8$	2×2	$112 \times 112 \times 8$	
Convolution	$112 \times 112 \times 8$	$3 \times 3 \times 16$	$110 \times 110 \times 16$	
Batch Norm	$110 \times 110 \times 16$		$110 \times 110 \times 16$	Leaky ReLU ReLU
Pooling	$110 \times 110 \times 16$	2×2	$55 \times 55 \times 16$	
Convolution	$55 \times 55 \times 16$	$3 \times 3 \times 32$	$53 \times 53 \times 32$	
Batch Norm	$53 \times 53 \times 32$		$53 \times 53 \times 32$	Leaky ReLU ReLU
Flatten	$53 \times 53 \times 32$		89888	
Dense	89888		2	SoftMax

2.2. Activation Function

In this Deep CNN architecture, the calculation of cross-entropy loss function is used.

    a.   Leaky ReLU

$$\sigma(x) = \begin{cases} \alpha x, if\ x \le 0 \\ x, if\ x > 0 \end{cases}$$

$$range = (-\infty, \infty)$$

    b.   ReLU

$$\sigma(x) = \begin{cases} 0, if\ x \le 0 \\ x, if\ x > 0 \end{cases}$$

$$range = [0, \infty)$$

    c.   Softmax

$$\sigma(\vec{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

$$range = [0,1]$$

The Softmax activation function is used only in the last layer, while the other 4 activation functions were used in the hidden layer.

2.3. Optimizer

The optimizations used include Adagrad, Adadelta, and Adamax.

    a.   Adagrad

    Adagrad is an optimization for increasing gradient descent efficiency in SGD and is applicable to large-scale artificial neural networks. [21][22]

$$w_t = w_{t-1} - \eta'_t \frac{\partial L}{\partial w_{t-1}}$$

$$\eta'_t = \frac{\eta}{\sqrt{\alpha_t + \varepsilon}}$$

$$\alpha_t = \sum_{i=1}^{t} \left(\frac{\partial L}{\partial w_{t-1}}\right)^2$$

    b.   Adadelta

    Adadelta is an extension of Adagrad optimization method, which uses Exponentially Weighted Averages. Meanwhile, this Adadelta uses the sum of all previous squared gradients [23].

$$w_t = w_{t-1} - \eta'_t \frac{\partial L}{\partial w_{t-1}}$$

$$\eta'_t = \frac{\eta}{\sqrt{S_{dw_t} + \varepsilon}}$$

$$S_{dw_t} = \beta S_{dw_{t-1}} + (1 - \beta) \left(\frac{\partial L}{\partial w_{t-1}}\right)^2$$

c.   Adamax

Adamax is a variant of Adam optimizer with infinite norm [24].

$$w_t = w_{t-1} - \left(\frac{\alpha}{(1 - \beta_1^t)}\right) \cdot \frac{m_t}{u_t}$$

$$g_t = \nabla_\theta f_t(\theta_{t-1})$$

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$$

$$u_t = max(\beta_2 \cdot u_{t-1}, |g_t|)$$

3. Material and Methods

The output of this work is a simple CNN structure design that produces good classification results in both COVID-19 and Non-COVID-19 cases. The dataset used is open source from Mendeley consisting of 2 data types, namely CXR and CT [25]. Furthermore, it has 2 folders each, which include COVID-19 and Non-COVID-19 image data. A total of 1000 CXR and CT were used respectively, and it was divided into training and validation data in the ratio of 70:30. Some samples from the dataset are shown in Figure 1 and Figure 2 below.
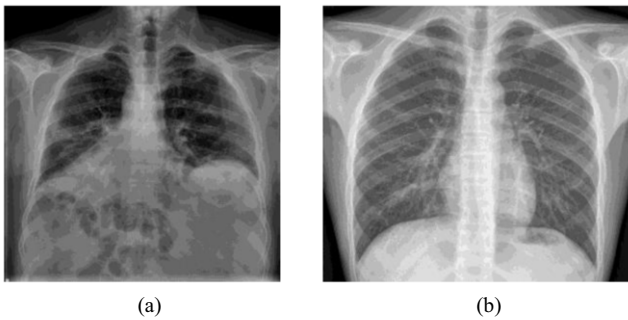

(a)   (b)

Figure 1. Sample of CXR images (a) CXR COVID-19 image (b) CXR non-COVID image

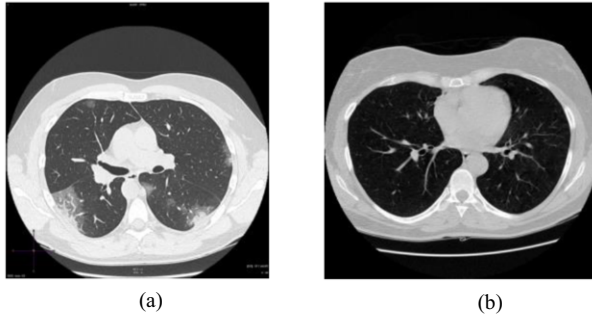| (a) | (b) |

Figure 2. Sample of CT images (a) CT COVID-19 image (b) CT non-COVID image

4. Experimental Results and Discussions

A total of 4 and 6 types of activation functions and optimizers are used respectively in the hidden layer to determine the best model in CNN. In addition, the epochs and proposed iteration numbers are 6 and 258, respectively. Tables 2 and 3 show the performance results of the above combination on these two types of datasets.
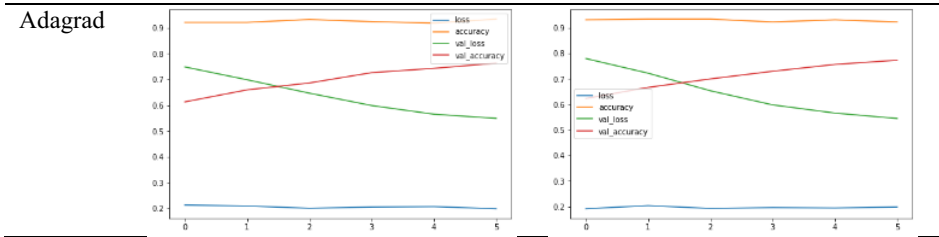
- Combination Model Performance on the CXR Dataset

Table 2. Performance results from a combination of 2 activation functions and 3 Optimizers on the CXR dataset

Activation Function	Optimizer	Accuracy	Loss
ReLU	Adamax	93.29%	0.1899
	Adadelta	92.43%	0.2052
	Adagrad	93.43%	0.1992
Leaky ReLU.	Adamax	94.43%	0.1776
	Adadelta	93.57%	0.1917
	Adagrad	92.29%	0.1987

Table 3. Plot of accuracy and loss from ReLU and Leaky ReLU

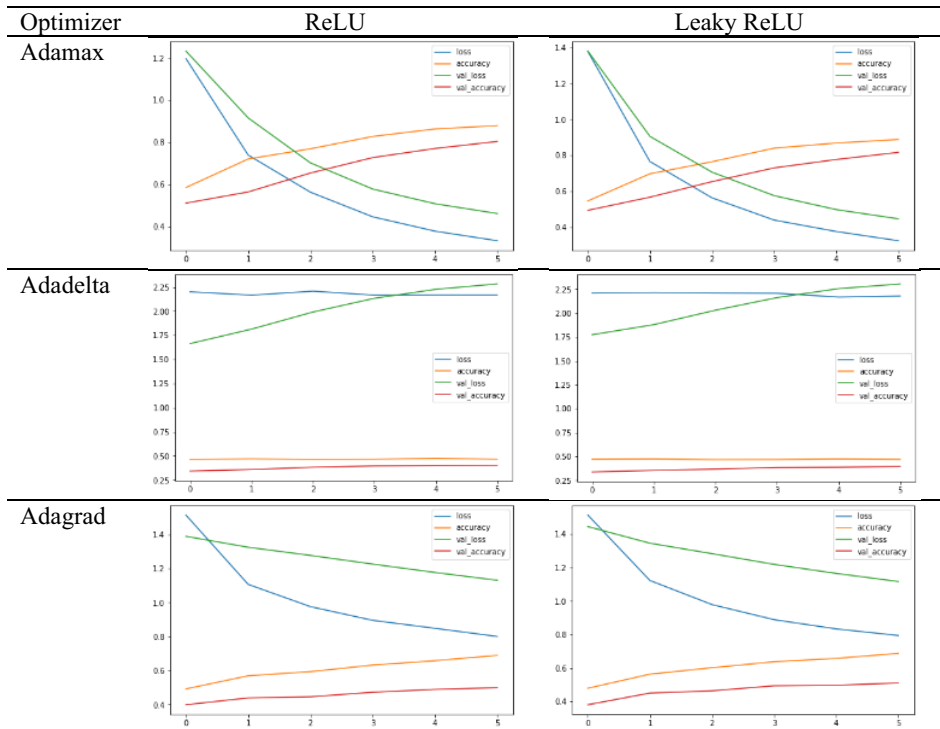| Optimizer	ReLU	Leaky ReLU
Adamax		
Adadelta		

Adagrad		
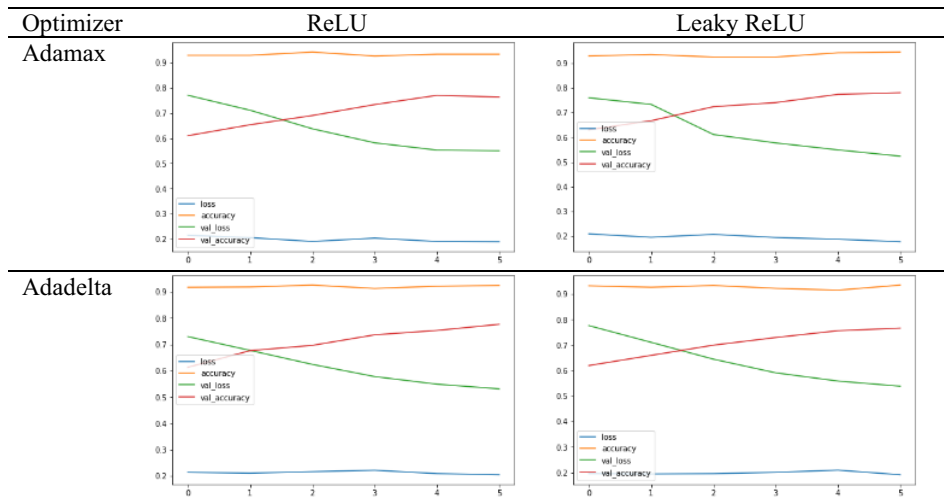
- Combination Model Performance on CT Dataset

Table 4. Performance results from a combination of 2 activation functions and 3 Optimizers on the CT dataset

Activation Function	Optimizer	Accuracy	Loss
ReLU	Adamax	87.86%	0.3312
	Adadelta	46.71%	2.1681
	Adagrad	69.00%	0.8008
Leaky ReLU.	Adamax	88.86%	0.3238
	Adadelta	46.86%	2.1796
	Adagrad	68.71%	0.7939

Table 5. Plot of accuracy and loss from ReLU and Leaky ReLU

Optimizer	ReLU	Leaky ReLU
Adamax		
Adadelta		
Adagrad		

- Optimal Combination

From the above experiment of combining several activation functions and the optimizer in the CRX dataset, the best result was shown in using Leaky ReLU and Adam Optimizer with a value of 96.29%, and the next best combination was the Leaky ReLU and Adam optimizer having a value of 94.71%. Meanwhile, in the CT dataset, the tanh and RMSprop Optimizer was the best combination with a value of 97.57% and the next best was the Leaky ReLU and RMSprop optimizer having 96.00%.

5. Conclusion and Future Work

This paper shows a comparison of combining activation functions in the hidden layer with several optimizers in two types of CRX and CT dataset in order to classify COVID-19 cases. The use of one activation function and three different optimizers from previous studies showed that the Leaky ReLU activation function provides excellent accuracy for all datasets as well as for all optimizers. From the experiment above, it shows that Leaky ReLU is an activation function that is excellent compared to ReLU in solving the problem of classifying COVID-19 CT and CRX images. It can be seen from the 2 types of existing datasets, the best activation function is shown by Leaky ReLU. This is in accordance with our initial assumption that Leaky ReLU might produce the best accuracy because Leaky ReLU does not ignore negative values in the image. For further research, it is hoped that this research can be developed with a different architecture to produce better accuracy.

References

[1]   W. El-Shafai *et al.*, "Efficient deep CNN model for COVID-19 classification," *Comput. Mater. Contin.*, vol. 70, no. 3, pp. 4373–4391, 2022, DOI: 10.32604/cmc.2022.019354.
[2]   K. Sharma, R. L. Sharma, and V. Sharma, "Coronavirus epidemiology: a review article," *Int. J. Community Med. Public Heal.*, vol. 7, no. 12, 2020, DOI: 10.18203/2394-6040.ijcmph20205207.
[3]   C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, 2020, doi: 10.1016/S0140-6736(20)30183-5.
[4]   C. C. Lai, T. P. Shih, W. C. Ko, H. J. Tang, and P. R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *International Journal of Antimicrobial Agents*, vol. 55, no. 3. 2020. DOI: 10.1016/j.ijantimicag.2020.105924.
[5]   T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian Journal of Pediatrics*, vol. 87, no. 4. 2020. doi: 10.1007/s12098-020-03263-6.
[6]   S. Khan, K. Thirunavukkarasu, R. Hammad, V. Bali, and M. R. Qader, "Convolutional neural network based SARS-CoV-2 patients detection model using CT images," *Int. J. Intell. Eng. Informatics*, vol. 9, no. 2, 2021, DOI: 10.1504/ijiei.2021.117061.
[7]   M. F. Aslan, K. Sabanci, A. Durdu, and M. F. Unlersen, "COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization," *Computers in Biology and Medicine*, vol. 142. 2022. DOI: 10.1016/j.compbiomed.2022.105244.
[8]   K. Madineni and P. Vasudevan, "Handwritten Text Recognition Based On CNN Classification," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 9s, 2020.
[9]   S. Ashraf, I. Kadery, A. A. Chowdhury, T. Z. Mahbub, and R. M. Rahman, "Fruit Image Classification Using Convolutional Neural Networks," *Int. J. Softw. Innov.*, vol. 7, no. 4, 2019, doi: 10.4018/IJSI.2019100103.
[10]   T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," 2017. DOI: 10.1109/ICBDA.2017.8078730.
[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, 2020, DOI: 10.1007/s13246-020-00865-4.

[12] M. Barstuğan, U. Özkaya, and Ş. Öztürk, "Coronavirus (Covid-19) classification using CT images by machine learning methods," in *CEUR Workshop Proceedings*, 2021, vol. 2872.

[13] N. Alsharman and I. Jawarneh, "GoogleNet CNN neural network towards chest CT-coronavirus medical image classification," *J. Comput. Sci.*, vol. 16, no. 5, 2020, DOI: 10.3844/JCSSP.2020.620.625.

[14] A. K. Dubey and V. Jain, "Comparative Study of Convolution Neural Network's Relu and Leaky-Relu Activation Functions," in *Lecture Notes in Electrical Engineering*, 2019, vol. 553. DOI: 10.1007/978-981-13-6772-4_76.

[15] J. Jebadurai, I. J. Jebadurai, G. J. L. Paulraj, and N. E. Samuel, "Super-resolution of digital images using CNN with leaky ReLU," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 11, 2019, DOI: 10.35940/ijrte.B1034.0982S1119.

[16] B. H. Nayef, S. N. H. S. Abdullah, R. Sulaiman, and Z. A. A. Alyasseri, "Optimized leaky ReLU for handwritten Arabic character recognition using convolution neural networks," *Multimed. Tools Appl.*, vol. 81, no. 2, 2022, DOI: 10.1007/s11042-021-11593-6.

[17] J. Qin, W. Pan, X. Xiang, Y. Tan, and G. Hou, "A biological image classification method based on improved CNN," *Ecol. Inform.*, vol. 58, 2020, DOI: 10.1016/j.ecoinf.2020.101093.

[18] H. Ling, J. Wu, J. Huang, J. Chen, and P. Li, "Attention-based convolutional neural network for deep face recognition," *Multimed. Tools Appl.*, vol. 79, no. 9–10, 2020, DOI: 10.1007/s11042-019-08422-2.

[19] W. S. Eka Putra, "Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101," *J. Tek. ITS*, vol. 5, no. 1, 2016, doi: 10.12962/j23373539.v5i1.15696.

[20] A. Sharma and G. Phonsa, "Image Classification Using CNN," *SSRN Electron. J.*, 2021, DOI: 10.2139/ssrn.3833453.

[21] S. Venkatesh and M. Jeyakarthic, "Adagrad Optimizer with Elephant Herding Optimization based Hyper Parameter Tuned Bidirectional LSTM for Customer Churn Prediction in IoT Enabled Cloud Environment," *Webology*, vol. 17, no. 2, 2020, DOI: 10.14704/WEB/V17I2/WEB17057.

[22] A. A. Lydia and F. S. Francis, "Adagrad-An Optimizer for Stochastic Gradient Descent," *Int. J. Inf. Comput. Sci.*, vol. 6, no. 5, 2019.

[23] F. Giannakas, C. Troussas, I. Voyiatzis, and C. Sgouropoulou, "A deep learning classification framework for early prediction of team-based academic performance," *Appl. Soft Comput.*, vol. 106, 2021, DOI: 10.1016/j.asoc.2021.107355.

[24] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 2015.

[25] W. El-shafai and F. Abd El-Samie, "Extensive COVID-19 X-Ray and CT Chest Images Dataset," *Mendeley Data*, vol. 3, 2020.