

## Classification of Pneumonia Disease Based on Web Application Using Convolutional Neural Networks (CNNs)

Tri Suhartono<sup>1</sup>, Diah Rahayu Ningtias<sup>2</sup>

<sup>1</sup>Electromedical Engineering School Of Health Science Semarang, Semarang, Indonesia

<sup>2</sup>Electromedical Engineering School Of Health Science Semarang, Semarang, Indonesia

Corresponding Author : [diahrahayu@atemsemarang.ac.id](mailto:diahrahayu@atemsemarang.ac.id)

### Abstract

*Artificial Intelligence can be used as a medical image diagnosis model by training a deep learning Convolutional Neural Networks (CNNs) image classification model to classify X-Ray images of the lungs. In this study, X-Ray images of the lungs classified as pneumonia or normal were used by the CNNs model that had been built. In the process of diagnosis there are difficulties, one of which is the lack of predictability of a follow up examination procedure only through X-Ray image. We need a method that can help diagnose X-Ray image objectively, quantitatively, and has high accuracy. There are used the CNNs model is then applied to Streamlit web-based application systems. The aim of this research is to develop a pneumonia disease classification application to simplify the diagnosis process. In this study, 5,855 X-Ray images of the lungs were used. The model is given epoch values of 10, 15 and 20 with 326 steps per epoch and the highest accuracy value is obtained at epoch 20. The result of accuracy value is 91.504%. The accuracy value can be affected from the epoch value, but the addition of the epoch value cannot fully increase the resulting accuracy value. The CNNs program model that has been created is then deployed to a web application using Streamlit and can be used for X-Ray image classification. The results of using this web application can be used for the initial detection process of X-Ray images with a diagnosis of pneumonia.*

**Keywords:** CNN, Streamlit, Pneumonia, Image Classification.

## INTRODUCTION

Pneumonia is a lung infection that can be caused by various types of microorganisms such as viruses, bacteria, and fungi (Torres et al., 2021). One of the causes of pneumonia is the pathogen *Klebsiella Pneumoniae* (Wang et al., 2020). Pneumonia can be prevented by vaccination with pneumococcal conjugate vaccines (PCVs) (Raguindin, 2020). Pneumonia can be known using medical images. This is done by means of radiographic examination, the results of which can then be analyzed by a doctor so that a diagnosis can be obtained. However, in making a diagnosis there are several difficulties, one of which is the lack of predictability of a follow-up examination procedure only through X-Ray images (Nanammal & Venugopalakrishnan, 2022). This is because in diagnosing X-Ray images only subjectively, so the level of accuracy is still low. We need a method that can help diagnose X-Ray images objectively, quantitatively, and has high accuracy (Qi et al., 2021).

Along with the development of technology in the health sector, medical image diagnostics can be carried out by utilizing artificial intelligence technology (Ghaffar et al., 2023). By training the deep learning (DL) system, it is hoped that it can be used for the benefit of disease diagnosis (Bozorgtabar et al., 2019). One that can be used to classify images is to build an image classification model (Vermeire et al., 2022). Image classification is a deep learning computer vision method that can be used for visual recognition by reading the features in the image (Chen et al., 2021). With deep learning it is possible to use some classification models using X-Ray image (P & Annavarapu, 2021). One of the artificial intelligence models that can be used as a medical image diagnostic model is Convolutional Neural Networks (CNNs) (Ruan et al., 2023). CNNs are a type of deep learning architecture for processing big data and are increasingly being applied to address the challenges of medical imaging analysis (Litjens et al., 2017).

CNNs are better when compared to other machine learning models such as Radio Frequency (RF), Gradient Boosting Machine (GBM), Support Vector Classifier (SVC), Learning Rate (LR), and K-Nearest Neighbor (KNN) (Reshi et al., 2021). The accuracy value can be influenced by the epoch value given during training data and the number of layers (Ayumi & Nurhaida, 2021). The CNNs program model has features that are able to identify patterns in image objects, CNNs are able to overcome the problem of overfitting by using regulation and dropout tactics (Ismail et al., 2019). In addition, the augmentation method can also be used to overcome the problem of overfitting (Zhang et al., 2022). The CNNs machine learning model also allows the CNNs model to be combined with other machine learning methods to get a better machine learning model (Nair & Domnic, 2022). In the CNNs model,

there are 4 main layers, consisting of the convolutional layer, the activation layer, the pooling layer, and the fully connected layer (Hibatullah & Maliki, 2019).

## Convolutional layer

The calculation was performed on the convolutional layer is the convolution operation of the input matrix with the filter matrix. The formula for the convolution operation is as an equation (1) :

$$FM[i]_{j,k} = (\sum_m \sum_n N_{[j-m,k-n]} F_{[m,n]}) + bF \quad (1)$$

with :

**FM[i]** : the first feature map matrix

**N** : input image matrix

**F** : convolution filter matrix

**bF** : nilai bias pada filter

**j,k** : pixel position in the input image matrix

**m,n** : pixel position in the convolution filter matrix

## Activation layer

In this layer, calculation of the activation function was performed, usually using a Rectified Linear Unit (ReLU) or a sigmoid to add non linearity to the network.

## Pooling layer

In this layer, size of the feature map is reduced. The type of poll in this study is max pooling, which is taking the maximum value. An example of feature max pooling can be seen in Table 1.

**Table 1. Max Pooling**

|        |        |        |        |
|--------|--------|--------|--------|
| 1.2028 | 1.7564 | 0.7353 | 1.0968 |
| 1.0999 | 0.2335 | 2.9597 | 0.2345 |
| 1.8787 | 0.8875 | 0.1585 | 1.6467 |
| 1.2552 | 1.8785 | 2.6575 | 0.7245 |



|        |  |
|--------|--|
| 1.7564 |  |
|        |  |

## Fully connected layer

This layer consists of *input layer*, *hidden layer*, and *output layer*.

a. Input layer

It is a combination of the entire feature map from the pooling layer, then the pixels are stretched into a vector that will be used in the hidden layer.

b. Hidden layer

Calculations on this layer multiply the values in the input layer with the weights that were initialized before adding the bias value. The calculation formula is as an equation (2) :

$$Z_{in_i} = \sum_{j=1}^n X_j * V_{j,i} + V_{0,i} \quad \text{with :} \quad (2)$$

$Z_{in_i}$  : input for the i-th hidden layer nodes with the number of m nodes

$x_j$  : node X to j

$V_{j,1}$  : value of V for  $X_j$  and node  $Z_i$

$V_0$  : bias V for  $Z_{in_i}$

c. Output layer

Calculations at this layer are multiplying the calculated values in the hidden layer with the weights that have been initialized plus the bias. The calculation formula is as an equation (3) :

$$Y_{in_i} = \sum_{j=1}^m Z_j * V_{j,i} + W_{0,i} \quad (3)$$

With :

$Y_{in_i}$  : input for node *hidden layer* Z ke-i with the number node m

$x_j$  : node Z to j

$V_{j,1}$  : value W for  $Z_j$  and node  $Y_i$

$V_0$  : bias W for  $Z_{in_i}$

From the CNNs model that was created then displayed on a web application with the Streamlit framework (Qin, 2021). The aim of this research is to develop a pneumonia disease classification application to simplify the diagnosis process.

## METHOD

### Data Set

In this study there used seconder data of lung's X-Ray image a numbers 5,765 as beginning data. In this beginning data consist of lung's X-Ray image with format .jpeg, and .png. The attribute and data types are presented in Table 2.

**Table 2. Data Set**

| Number | Folder | Total Data |
|--------|--------|------------|
| 1      | Train  | 5,215      |
| 2      | Test   | 624        |
| 3      | Val    | 16         |

The data used 5,855 divided into 1,582 normally image of lungs and 4,273 pneumonia image of lung from X-Ray. There are used 5,855 data set, consisting of 5,215 training data, 624 testing data, and 16 val data. The following in an example of normal X-Ray image data set shown as in Figure 1 and an X-Ray image of pneumonia shown as in Figure 2 below.



Figure 1. Normal Image



Figure 2. Pneumonia Image

## Research Procedure

CNN is deep learning algorithm modeling was used for a pneumonia disease classification system using X-Ray images of the lungs. Training data of 5,215 is used in the early stages of calculating the CNNs algorithm for the convolution process. In this study, CNNs modeling was used with 14 layers which can be seen through the flowchart in Figure 3.

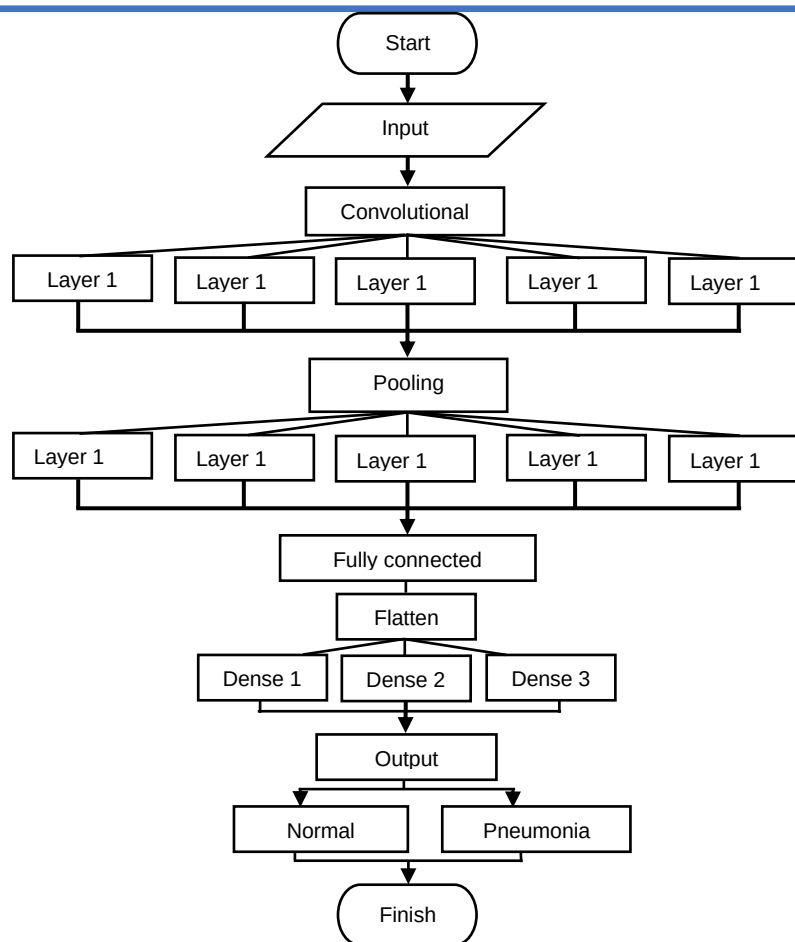


Figure 3. Flowchart Algorithm CNN

## Deploying Model

The CNNs algorithm program model has been tested and has a best accuracy value, was then deployed into a web application using Streamlit. The appearance of the classification application using Streamlit can be seen in Figure 4.



Figure 4. Streamlit web APP

In the interface display of the application, it shows brief information about pneumonia, and then click "Mulai Klasifikasi", and enter the X-Ray image you want to classify.

## RESULTS AND DISCUSSION

In this study there used 5,215 data of X-Ray image as data training in CNNs algorithm by providing epoch values and steps per epoch to produce accuracy values. The step values per 326 epochs were used with the epoch values changed to find good accuracy results. The results of the accuracy values with some epoch values can be seen in Table 3.

Table 3. Accuracy Value

| Epoch | Accuracy |
|-------|----------|
| 10    | 91.025 % |
| 15    | 90.865 % |
| 20    | 91.504 % |

According to the Table 3. It can be seen that in the CNNs research model with training data of 5.215 X-Ray image and a step per epoch value of 326, any different accuracy values were obtained seen from Table 2. The highest accuracy value is obtained when using an epoch value of 20 with an accuracy value of 91.504%. This can be interpreted that the epoch value given will affect the results of the accuracy value obtained, but adding the epoch value cannot fully increase the accuracy value. As shown in Table 2, the CNNs model with an epoch value 20 has a higher accuracy value compared to epoch value of 10 and 15, but epoch value 15 has a lower accuracy value than epoch value 10.

## Simulation

After data training using X-ray image data 5,215 with a step per epoch value of 326 and the epoch value changed as shown in Table 2, the highest accuracy results were obtained at the epoch value of 20 which resulted in an accuracy value of 91.504%. From this model, it will be deployed to become a web-based application using Streamlit to be used as a media for classification of pneumonia. The results of classification of the Streamlit application can be seen in Figure 5 and Figure 6.



Figure 5. Normal Classification

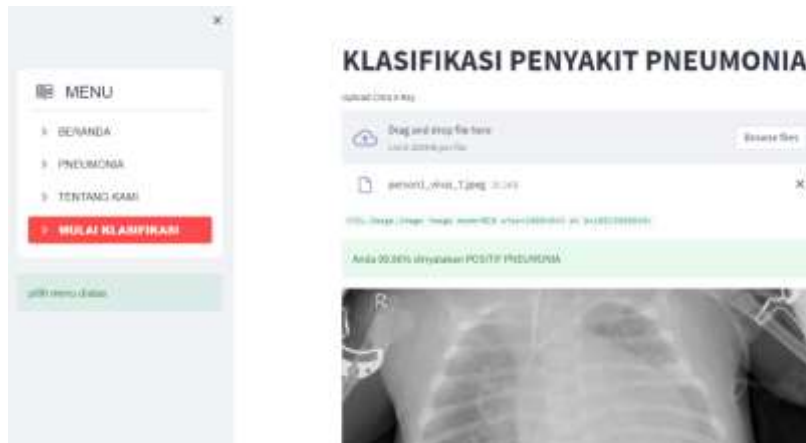


Figure 6. Pneumonia Classification

## CONCLUSION

In this study it was concluded that the epoch value can affect the resulting accuracy value, but the addition of the epoch value cannot fully increase the resulting accuracy value. The highest accuracy value is when an epoch value of 20 is used with a step per epoch value of 326 with 5,215 image data used as training data, with an accuracy value of 91.504%. This Convolutional Neural Network (CNNs) algorithm can be used for X-Ray image classification and deployed using the Streamlit web application, so that it can be used by doctors for initial classification used for the initial process of X-Ray image detection with a pneumonia diagnosis.

## REFERENCES

- Ayumi, V., & Nurhaida, I. (2021). Klasifikasi Chest X-Ray Images Berdasarkan Kriteria Gejala Covid-19 Menggunakan Convolutional Neural Network. JSAI (Journal Scientific and Applied Informatics). <https://doi.org/10.36085/jsai.v4i2.1513>
- Bozorgtabar, B., Mahapatra, D., von Tengg-Kobligk, H., Poellinger, A., Ebner, L., Thiran, J.-P., & Reyes, M. (2019). Informative sample generation using class aware generative adversarial networks for classification of chest Xrays. *Computer Vision and Image Understanding*, 184, 57–65. <https://doi.org/10.1016/j.cviu.2019.04.007>
- Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. In *Remote Sensing* (Vol. 13, Issue 22). <https://doi.org/10.3390/rs13224712>
- Ghaffar, N., Erkan, N., & Ahad, K. (2023). Evaluation of artificial intelligence techniques in

- disease diagnosis and prediction. Discover Artificial Intelligence. <https://doi.org/10.1007/s44163-023-00049-5>
- Hibatullah, A., & Maliki, I. (2019). Penerapan Metode Convolutional Neural Network Pada Pengenalan Pola Citra Sandi Rumpit. 1–8.
- Ismail, A., Rahmat, T., & Aliman, S. (2019). Chest X-Ray Image Classification Using Faster R-Cnn. Malaysian Journal of Computing. <https://doi.org/10.24191/mjoc.v4i1.6095>
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- Nair, R. S., & Domnic, S. (2022). A combination of learning and non-learning based method for enhancement, compression and reconstruction of underwater images. Aquaculture and Fisheries, 7(2). <https://doi.org/10.1016/j.aaf.2021.10.006>
- Nanammal, V., & Venugopalakrishnan, J. (2022).>NNLGBM: Medical Image Classification through Secure Collaboration in Pneumonia Detection by Blending NN and LGBM. International Journal of Intelligent Systems and Applications in Engineering, 10(3), 201–212.
- P, S. A. B., & Annavarapu, C. S. R. (2021). Deep learning-based improved snapshot ensemble technique for COVID-19 chest X-ray classification. Applied Intelligence, 51(5). <https://doi.org/10.1007/s10489-021-02199-4>
- Qi, X., Brown, L. G., Foran, D. J., Noshier, J., & Hacihaliloglu, I. (2021). Chest X-ray image phase features for improved diagnosis of COVID-19 using convolutional neural network. International Journal of Computer Assisted Radiology and Surgery, 16(2), 197–206. <https://doi.org/10.1007/s11548-020-02305-w>
- Qin, Y. (2021). A Cancer Cell Image Classification Program: Based on CNN Model. Proceedings - 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology, AINIT 2021. <https://doi.org/10.1109/AINIT54228.2021.00037>
- Raguindin, P. F. (2020). The changing epidemiology of pneumococcal diseases: new challenges after widespread routine immunization. International Journal of Public Health, 65(6), 709–710. <https://doi.org/10.1007/s00038-020-01403-5>
- Reshi, A. A., Rustam, F., Mehmood, A., Alhossan, A., Alrabiah, Z., Ahmad, A., Alsuwailem, H., & Choi, G. S. (2021). An Efficient CNN Model for COVID-19 Disease Detection Based on X-Ray Image Classification. Complexity, 2021.



<https://doi.org/10.1155/2021/6621607>

- Ruan, D., Wang, J., Yan, J., & Gühmann, C. (2023). CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis. *Advanced Engineering Informatics*, 55. <https://doi.org/10.1016/j.aei.2023.101877>
- Torres, A., Cilloniz, C., Niederman, M. S., Menéndez, R., Chalmers, J. D., Wunderink, R. G., & van der Poll, T. (2021). Pneumonia. *Nature Reviews Disease Primers*, 7(1). <https://doi.org/10.1038/s41572-021-00259-0>
- Vermeire, T., Brughmans, D., Goethals, S., de Oliveira, R. M. B., & Martens, D. (2022). Explainable image classification with evidence counterfactual. *Pattern Analysis and Applications*, 25(2). <https://doi.org/10.1007/s10044-021-01055-y>
- Wang, G., Zhao, G., Chao, X., Xie, L., & Wang, H. (2020). The characteristic of virulence, biofilm and antibiotic resistance of klebsiella pneumoniae. In *International Journal of Environmental Research and Public Health* (Vol. 17, Issue 17). <https://doi.org/10.3390/ijerph17176278>
- Zhang, Y. D., Jiang, X., & Wang, S. H. (2022). Fingerspelling Recognition by 12-Layer CNN with Stochastic Pooling. *Mobile Networks and Applications*. <https://doi.org/10.1007/s11036-021-01900-8>

