



A Good Evaluation Based on Confusion Matrix for Lung Diseases Classification using Convolutional Neural Networks

Izza Putri Kamila*, Christy Atika Sari², Eko Hari Rachmawanto³, Nur Ryan Dwi Cahyo⁴

Faculty of Computer Science, University of Dian Nuswantoro, Imam Bonjol No. 207 Semarang, Central Java, 50131, Indonesia

*111202214398@mhs.dinus.ac.id

Abstract. CNN has been widely used to detect a pattern with image classification. This study used CNN to perform a classification analysis of lung abnormality detection on chest X-ray images. The dataset consists of 5,732 2D images with dimensions of 200 x 200 x 1 divided into training data (85%) and testing data (15%). The preprocessing process includes image resizing, enhancement to increase contrast and reduce image complexity, and filtering to improve visibility and reduce noise. CNN is used to classify imagery into three categories, Normal (no abnormalities), Pneumonia, and Tuberculosis. The results showed a good level of accuracy, with an accuracy value of 97.79% and a continuously high accuracy value of 100% in 6 trials. This research provides insights into the detection of lung disorders and encourages further exploration in medical diagnosis.

Keywords: Pneumonia, Tuberculosis, TBC, CNN, Image Classification

(Received 2023-11-03, Accepted 2023-11-14, Available Online by 2023-11-20)

1. Introduction

Lungs are one of the important organs of the human respiratory system. The lungs can experience abnormalities, such as pneumonia, and tuberculosis when exposed to pathogens in the form of bacteria, viruses, fungi, or parasites [1]. A bacterial illness called pneumonia results in swelling and inflammation of the lungs' alveoli, or air sacs [2]. In Indonesia, pneumonia is one of the leading causes of mortality among children under the age of five. Pneumonia will affect up to 278,261 children under the age of five in 2021. Meanwhile, Tuberculosis (TB) is a contagious infection caused by Mycobacterium tuberculosis [3]. According to the World Health Organization (WHO), there will be 2.97 million tuberculosis (TB) patients in Southeast Asia by 2021, with Indonesia ranking second in the area. Meanwhile, the Ministry of Health of the Republic of Indonesia says that 969 thousand individuals would be infected with tuberculosis by 2023 [4][5].

If this lung condition is not detected in time, it might lead to death. Detection can be used to determine whether or not someone has a certain disease. Detecting lung abnormalities is difficult in the medical world due to the complexity of the anatomical structure of the lungs, the diversity of illnesses that can be diagnosed, and the necessity for high precision. Based on this topic, researchers suggest study using artificial intelligence, specifically convolutional neural networks, to diagnose and categorize discovered

lung illnesses such as pneumonia and TB. From these problems, researchers proposed a study using Convolutional Neural Network (CNN) to recognize, and classify lung disorders detected in the form of Pneumonia and Tuberculosis.

CNN method for predicting abnormalities in the lungs. CNN has advantages, such as being efficient in processing images, having a high level of accuracy, resistant to noise, and having automatic feature extraction [6]. However, to get a high accuracy value, it is necessary to do augmentation if the data to be processed is relatively small. In addition, CNNs are computationally intensive to train and run models, are difficult to interpret certain models, are not always suitable for processing or processing 2D and 3D images, and CNNs are classified as vulnerable to changes in images, such as light, rotation, and/or added noise [7].

In a previous study conducted by Harshvardhan GM and colleagues in 2021 on Detecting Pneumonia Using CNN and a Chest X-ray. In his study, the authors used CNN to classify binaries with chest X-ray data describing cases of impactful and non-impacted pneumonia. The accuracy resulting from the study was 93%. Meanwhile, a study conducted by Abdulfattah Alawi and colleagues in 2021 explored the Convolutional Neural Network Model for Tuberculosis Disease Screening. According to the research, the study used a CNN-based model to separate areas of the lungs in classifying chest X-ray images whether infected with tuberculosis or not. In research conducted by Abdulfatah et al, accuracy results of 98.71% were obtained [8][9].

To recognize, detect and classify the abnormalities of this disease, you can use X-rays, magnetic resonance imaging (MRI) and computed tomography (CT). X-rays are frequently used to examine a patient's chest cavity by looking for characteristics of any bone density. These images can be used to determine whether there are abnormalities in the patient's lungs or not. The approach used in this study tries to distinguish, diagnose, and categorize anomalies in patients' lungs. As a result, it is critical to recognize and detect the existence or absence of the patient's lungs as soon as possible. The CNN approach was utilized in this work to identify these lung illnesses, particularly pneumonia and TB. Thus, artificial intelligence can be used to aid in the early detection and treatment of lung illnesses, particularly in places with high incidence rates, such as Indonesia [10][11].

2. Methods

In this study, to classify lung disorders based on X-Ray images, Convolutional Neural Network was used. The presented flow diagram illustrates the steps in the proposed method.

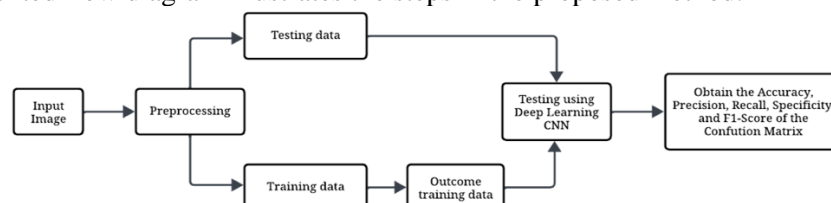


Figure 1. Research Methodology

2.1. Dataset

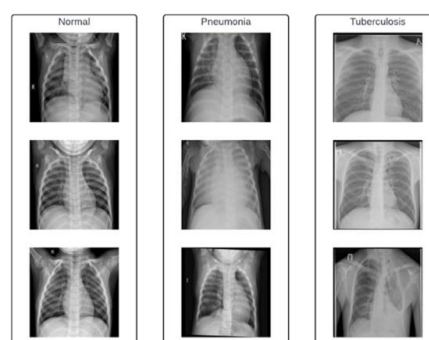


Figure 2. Normal Image Sample, Pneumonia Image, and Tuberculosis Image.

The dataset used in this study was a 2D X-ray radiography image of the lungs totaling 5,732 samples with pixel dimensions of 200 x 200 x 1. Based on the 5,732 images, categorized into two different parts, namely training data with each section including 245 images of tuberculosis, 3,054 images of pneumonia, and 1,573 images of normal lungs, and testing data totaling 860 samples. Training data is used to study and recognize the pattern of each type of lung disorder to be categorized. Data from testing is utilized to evaluate the efficacy of proposed procedures in correctly identifying the types of lung illnesses in question. This study's dataset was obtained via the Kaggle platform [12].

2.2. Preprocessing

In this step, the dataset is divided into 15% testing data and 85% training data. The first step is to resize the previous image such that all processed photos are the same size and dimensions. The image then undergoes an image enhancement process, focusing on the area that will be recognized by the next process by increasing the contrast value in the image to be processed. Then, there is a process of reducing color information's complexity and simplifying. This conversion procedure allows you to focus on pixel intensity values for simpler analysis, as well as extract characteristics from photos for additional processing or analysis. In addition, this modification is intended to boost computing efficiency and simplify some image processing processes [13][14].

2.3. Convolutional Neural Network (CNN)

Convolutional neural network is a form of artificial neural architecture that was created primarily for image processing and visual pattern detection. CNN architecture learns the internal structure of features and generalizes those features to image analysis problems, such as object recognition and computer vision. CNN is made up of three major layers: a convolutional layer, a pooling layer, and a fully connected layer [15]. Convolutional layers have filters and image maps as a list of coordinates relative to a specific image and aim to extract features from the input image by maintaining spatial relationships between pixels. The pooling layer has the effect of reducing the spatial dimension of the image features obtained from the previous process and helps reduce the number of parameters and calculations required, as well as preventing overfitting.

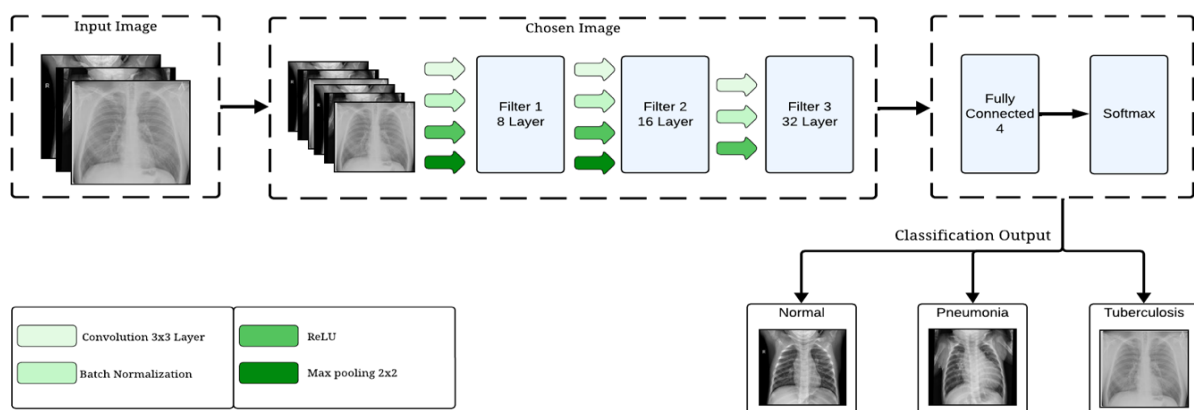


Figure 3. Chest X-Ray Classification Process with Convolutional Neural Network Method.

There are two types of pooling layers, the max pooling layer keeps the maximum value in each zone and reduces the image size and average pooling layer that maintains the average information in an image region. The fully connected layer functions as a link between the previous layers of the network and the subsequent layers. The classification process will be performed by CNN starting by processing the selected image, then applying image adjustments to provide adjustments to the image without change the intensity, in the form of increasing or decreasing contrast, adjusting brightness and color. Then, after going through the image adjustment process, the image will continue to be processed at the three major layers. The images will then be classified based on the main goal of the classification process [16][17][18][19][20][21].

2.4. Confusion Matrix

An important evaluation tool in the measurement of CNN classification models is used to understand the extent to which classification models are successful in correctly classifying data. The confusion matrix consists of four main components, True Positive (TP) denotes the value of correctly identified positive cases, True Negative (TN) denotes the value of correctly categorized negative situations, the number of negative case values classified as positive by the model is indicated by False Positive (FP), and False Negative (FN) refers to the number of cases that are truly positive but are labeled as negative by the model. To calculate the confusion matrix, you may utilize an Accuracy, Precision, Recall, and F1-Score formula [22][23].

$$Akurasi = \frac{TN + TP}{TN + TP + FN + FP} \times 100\%$$

$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$F1 - Score = 2 \times \frac{Recall * Presisi}{Recall + Presisi} \times 100\%$$

3. Results and Discussion

This study used a dataset of 5,732 images with dimensions of 200 x 200 x 1 from three different classes. The dataset is divided into 15% testing data and 85% training data. Each class in the training data amounted to 245 tuberculosis images, 3,054 pneumonia images, and 1,573 normal lung images, and testing data totaling 860 samples. Then, preprocessing is applied to each image for filtering using image adjustments, such as increasing the contrast. Furthermore, the classification uses CNN to classify lung disorders, namely Normal or not experiencing pneumonia or tuberculosis, Pneumonia, and Tuberculosis. The values of the confusion matrix's four main components are as follows.

Actual \ Predicted	Normal	Pneumonia	Tuberculosis
Normal	275	2	0
Pneumonia	20	519	0
Tuberculosis	0	0	43

False Negative (FN)

True Positive (TP)

False Positive (FP)

Figure 4. Confusion Matrix Result

To obtain the confusion matrix values, training data must be generated to provide the component values contained in the confusion matrix, namely True Positive, True Negative, False Positive, and False Negative. As indicated in Table 1, the performance of the classification model utilized may be determined or estimated using Accuracy, Recall (Sensitivity), Precision, Specificity, and F1-Score.

Table 1. Confusion Matrix.

No.	Partition (Testing-Training)	Training	Confusion Matrix				
			Accuracy	Recall	Precision	Specificity	F1-Score
1.	15% - 85%	1st Training	97,79%	96,73%	97,79%	98,89%	97,28%
2.	20% - 80%	2nd Training	97,03%	97,84%	93,54%	96,52%	95,64%
3.	25% - 75%	3rd Training	97,49%	95,45%	97,14%	98,55%	96,29%
4.	30% - 70%	4th Training	97,79%	96,58%	97,10%	95,51%	96,84%

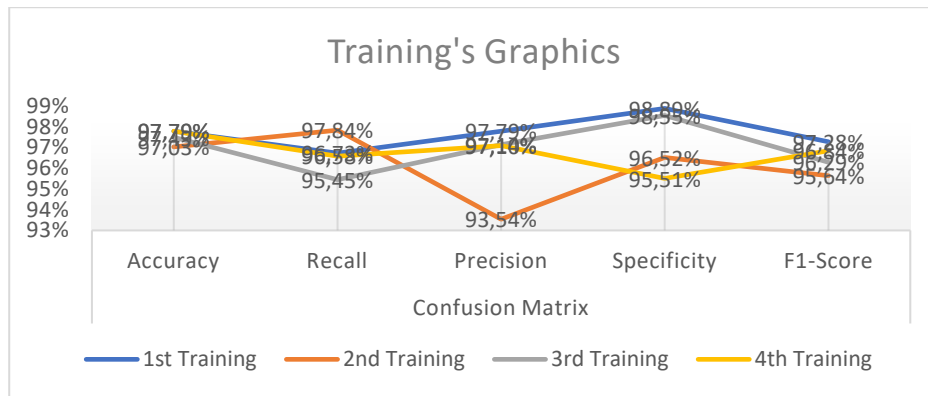


Figure 5. Training's Graphics

In Table 1 above, the author tried four divisions of testing and training data, namely the first training 15%-85%, the first training 20%-80%, the third training 25%-75%, and 30%-70%. It can be seen in Figure 5, the highest accuracy is 97.79% which is the result of 1st training and 4th training. However, for the highest Recall, Precision, Specificity, and F1 values are generated at 1st training. Thus, in this study the dataset used to detect as much as 15% of testing data and 85% of training data.

After acquiring the Confusion Matrix value, more testing is performed to evaluate the model's performance. The Confusion Matrix provides a clear picture of the extent to which the model is able to correctly classify data, comprising data on four main component of confusion matrix. With this information, we can make more informed decisions about how well the image processing model works and whether improvements or adjustments need to be made as tested in the following table.

Table 2. Dataset Testing

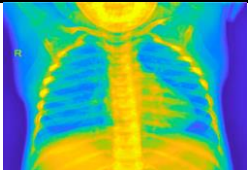
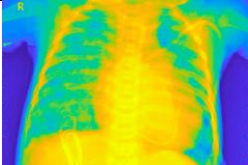
No.	Input Image	Actual Class	Classification Result	Result
1.		Normal	Normal	True
2.		Pneumonia	Pneumonia	True

Table 2. Dataset Testing Continue

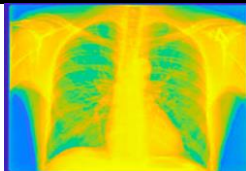
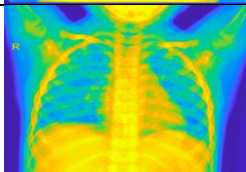
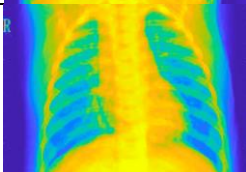
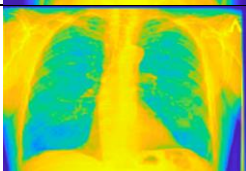
No.	Input Image	Actual Class	Classification Result	Result
3.		Tuberculosis	Tuberculosis	True
4.		Normal	Normal	True
5.		Pneumonia	Pneumonia	True
6.		Tuberculosis	Tuberculosis	True

Table 3. Comparison Results

Research	Method	Total Dataset	Precision	Accuracy	Specificity	Sensitivity (Recall)	F1 Score
T. Rahman <i>et al.</i> [13]	CNN with ensemble architecture	18.479	94.51%	95.11%	95.59%	94.56%	94.53%
H. Gm, M. Kumar Gourisaria, <i>et al.</i> [8]	CNN + ANN	5.847	-	-	92.16%	90.07%	77%
Rahman T., <i>et al.</i> [24]	CNN Chexnet	7.000	96.62%	96.47%	96.51%	96.47%	96.47%
Proposed Method	CNN	5.732	97.79%	97.79%	98.89%	96.73%	97.28%

Based on Table 3, lung anomalies examined by three research using different approaches, the writer recommends utilizing the CNN method in the next investigation. The CNN approach makes it feasible to operate more efficiently and accurately.

4. Conclusion

In the study above using a dataset of 5,732 samples with dimensions of 200 x 200 x 1 which were divided into 85% training data and 15% testing data. These samples include samples of normal lungs, lungs with pneumonia, and lungs with tuberculosis. Each sample class amounted to 1,573 normal lung samples, 3,054 lung samples with pneumonia, and 254 lung samples with tuberculosis. Samples categorized using the CNN approach had an accuracy value of 97.79% and a continuously high accuracy value of 100% in 6 trials. Overall, the CNN method can work well in the classification of lung abnormality detection.

In future study, the architectural evolution of CNN may be used to attain the same accuracy outcomes when deciding the amount of testing data and training data. And it may combine or compromise with

other designs to improve accuracy and efficacy in detecting lung problems, such as CNN ChexNet, Recurrent Neural Network, or Long-Short Term Memory (LSTM). Combining CNN with ChexNet, RNN, or LSTM can result in a more comprehensive method to medical image processing, focusing on the spatial correlations between pixels in CNN pictures and comprehending the temporal sequence in sequential images like CT-Scans and MRIs. This can assist the medical community in dealing with issues linked to changes in medical pictures during the illness detection process, such as rotation and the addition of noise.

References

- [1] H. H. Hussain, N. T. Ibraheem, N. K. F. Al-Rubaey, M. M. Radhi, N. K. K. Hindi, and R. H. K. AL-Jubori, "A review of airborne contaminated microorganisms associated with human diseases," *Medical Journal of Babylon*, vol. 19, no. 2. Wolters Kluwer Medknow Publications, pp. 115–122, Apr. 01, 2022. doi: 10.4103/MJBL.MJBL_20_22.
- [2] P. J. Shermila, A. D. Anusha, Akila. M, and Abirami. S, "Pneumonia Detection from X- Ray Images using Convolutional Neural Networks," in *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, IEEE, Apr. 2023, pp. 1–7. doi: 10.1109/ICONSTEM56934.2023.10142295.
- [3] R. Gopalaswamy, S. Shanmugam, R. Mondal, and S. Subbian, "Of tuberculosis and non-tuberculous mycobacterial infections - A comparative analysis of epidemiology, diagnosis and treatment," *Journal of Biomedical Science*, vol. 27, no. 1. BioMed Central, Jun. 17, 2020. doi: 10.1186/s12929-020-00667-6.
- [4] "World Health Organization TB Reports." Accessed: Nov. 10, 2023. [Online]. Available: <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022/covid-19-and-tb>
- [5] "TB by the Republic of Indonesia's Ministry of Health." Accessed: Nov. 10, 2023. [Online]. Available: <https://tbindonesia.or.id>
- [6] A. R. Hawas, H. A. El-Khobby, M. Abd-Elnaby, and F. E. Abd El-Samie, "Gait identification by convolutional neural networks and optical flow," *Multimed Tools Appl*, vol. 78, no. 18, pp. 25873–25888, Sep. 2019, doi: 10.1007/s11042-019-7638-9.
- [7] K. Alomar, H. I. Aysel, and X. Cai, "Data Augmentation in Classification and Segmentation: A Survey and New Strategies," *J Imaging*, vol. 9, no. 2, Feb. 2023, doi: 10.3390/jimaging9020046.
- [8] H. Gm, M. Kumar Gourisaria, S. S. Rautaray, and M. Pandey, "PNEUMONIA DETECTION USING CNN THROUGH CHEST X-RAY," 2021.
- [9] Abdulfattah E. Ba Alawi, Amer Al-basser, Amer Sallam, Amr Al-sabaeei, and Hesham Al-khateeb, "Convolutional Neural Networks Model for Screening Tuberculosis Disease," *IEEE*, pp. 1–5, Apr. 2021, doi: 10.1109/ICTSA52017.2021.9406520.
- [10] M. Siddiq, "ML-BASED MEDICAL IMAGE ANALYSIS FOR ANOMALY DETECTION IN CT SCANS, X-RAYS, AND MRIS." [Online]. Available: <http://devotion.greenvest.co.id>
- [11] R. Jain *et al.*, "Chest X-Ray as the First Pointer in Various Skeletal Dysplasia and Related Disorders," *Indographics*, vol. 01, no. 02, pp. 196–207, Dec. 2022, doi: 10.1055/s-0042-1759845.
- [12] "Chest X-Ray ," Kaggle. Accessed: Oct. 07, 2023. [Online]. Available: <https://www.kaggle.com/datasets/jtipjtj/chest-xray-pneumoniacovid19tuberculosis/data>
- [13] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput Biol Med*, vol. 132, May 2021, doi: 10.1016/j.compbiomed.2021.104319.
- [14] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltip.2022.04.020.

- [15] N. R. D. Cahyo, C. A. Sari, E. H. Rachmawanto, C. Jatmoko, R. R. A. Al-Jawry, and M. A. Alkhafaji, "A Comparison of Multi Class Support Vector Machine vs Deep Convolutional Neural Network for Brain Tumor Classification," in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep. 2023, pp. 358–363. doi: 10.1109/iSemantic59612.2023.10295336.
- [16] D. Bhatt *et al.*, "Cnn variants for computer vision: History, architecture, application, challenges and future scope," *Electronics (Switzerland)*, vol. 10, no. 20. MDPI, Oct. 01, 2021. doi: 10.3390/electronics10202470.
- [17] E. Saralioglu and O. Gungor, "Semantic segmentation of land cover from high resolution multispectral satellite images by spectral-spatial convolutional neural network," *Geocarto Int*, vol. 37, no. 2, pp. 657–677, Jan. 2022, doi: 10.1080/10106049.2020.1734871.
- [18] H. Gholamalinezhad and H. Khosravi, "Pooling Methods in Deep Neural Networks, a Review," Sep. 2020.
- [19] N. Akhtar and U. Ragavendran, "Interpretation of intelligence in CNN-pooling processes: a methodological survey," *Neural Comput Appl*, vol. 32, no. 3, pp. 879–898, Feb. 2020, doi: 10.1007/s00521-019-04296-5.
- [20] S. R. D. V. P. S. M. S.H. Shabbeer Basha, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *IEEE*, vol. 378, pp. 112–119, 2020, Accessed: Oct. 22, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231219313803>
- [21] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, "A Novel Fast Single Image Dehazing Algorithm Based on Artificial Multiexposure Image Fusion," *IEEE Trans Instrum Meas*, vol. 70, pp. 1–23, 2021, doi: 10.1109/TIM.2020.3024335.
- [22] D. S. V. J. G. Chicco, "The Benefits of the Matthews Correlation Coefficient (MCC) Over the Diagnostic Odds Ratio (DOR) in Binary Classification Assessment," *IEEE Access*, vol. 9, pp. 47112–47124, 2021, Accessed: Oct. 17, 2023. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9385097&isnumber=9312710>
- [23] C. Cao, D. Chicco, and M. M. Hoffman, "The MCC-F1 curve: a performance evaluation technique for binary classification," Jun. 2020.
- [24] T. Rahman *et al.*, "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020, doi: 10.1109/ACCESS.2020.3031384.