Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11711
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Classification of Tuberculosis Based on Lung X-Ray Image With Data Science Approach Using Convolutional Neural Network

Mawaddah Harahap[1)*], Alfeus P. S. Pasaribu[)], Dedy Ridoly Sinaga[3)], Romulus Sipangkar[4)], Samuel[5)]
[1,2,3,4,5)] Universitas Prima Indonesia Medan, Indonesia
[1)*]mawaddah@unprimdn.ac.id, [2)]alfeuspasaribu12345@gmail.com, [3)] dedysinaga123@gmail.com,
[4)]bangpakkar766@gmail.com, [5)]samueladamlu01@gmail.com

Submitted : Aug 21, 2022 | **Accepted** : Sep 15, 2022 | **Published** : Oct 3, 2022

Abstract: Tuberculosis (TB) is a potentially serious infectious disease in the lungs, becoming 1 of 10 causes of death. In Indonesia, the disease is ranked third after India and China with 824,000 cases and 93,000 deaths per year, equivalent to 11 deaths per hour. The increasing number of infections and deaths from TB disease is recorded as a result of its transmission, lack of early diagnosis, and inadequate professional radiologists in developing areas where TB is more common. Rapid and accurate diagnosis is essential for appropriate treatment to be initiated. Diagnosis is usually done by looking at the results of the x-ray image of the thorax and the results of the BTA test on the patient. To classify lung x-ray images detected tuberculosis or not, a study was carried out using the Convolutional Neural Network (CNN) method. The test results produce the last epochs value of 200, the accuracy obtained is 0.9892, which means the CNN accuracy is 98%, with validation the accuracy obtained is 0.9835 or 98%. So the results of the classification test using CNN are quite accurate. With the acquisition of CNN results which is quite high, it can be used as a consideration to be used in classifying TB disease.

Keywords: Tuberclosis, X-Ray Image, Lung, Data Science, CNN

INTRODUCTION

Tuberculosis (TB) is a potentially serious infectious disease in the lungs, being 1 in 10 causes of death. The bacteria that cause TB are spread from person to person through droplets released into the air through coughing and sneezing(Mitra Keluarga, 2022). If not treated, pulmonary tuberculosis can lead to complications of permanent lung damage to life-threatening(Sehat, 2021). In Indonesia, the disease ranks third after India and China with 824 thousand cases and 93 thousand deaths per year or equivalent to 11 deaths per hour(Sehat negeriku sehatlah bangasaku, 2022).

A smear positive pulmonary tuberculosis patient can infect 10-15 people around him(Tri Kristini, Rana Hamidah, 2020). The national TB control target in 2025, the morbidity rate is around 50% and the mortality rate is around 70% (Infodatin, 2018). In line with the End TB strategy, which has become a global commitment and the Indonesian government, referring to the 2020-2024 RPJMN, a National Strategy for Tuberculosis Prevention in Indonesia 2020-2024 has been prepared. A very crucial period for accelerating towards the elimination of tuberculosis by 2030. The document contains comprehensive strategies, interventions and activities as well as ambitious targets to reduce TB cases as soon possible(Germas, 2021).

The increase in the number of infections and deaths from TB disease is noted as a result of its transmission, lack of early diagnosis, and inadequate professional radiologists in developing areas where TB is more common. The computer-assisted detection model uses deep convolutional neural networks to detect TB automatically from Montgomery County (MC) Tuberculosis radiographs(Oloko-Oba, M., Viriri, S., 2020)..

The secret to managing this condition is an accurate diagnosis. A fast and accurate diagnosis is needed so that appropriate treatment can be carried out. Diagnosis is usually done by looking at the results of the chest x-ray image and the results of the smear test on the patient(Saeful Bahri, et all, 2021). With a Convolutional Neural Network (CNN) based model, lung areas are separated to solve the problem(Abdulfattah E et all, 2021). The CNN method is one of the deep learning methods that is able to carry out an independent learning process for the most significant image recognition, object extraction and classification and can be applied to high resolution images that have a nonparametric distribution model(Erlyna Nour Arrofiqoh, and Harintaka, 2018),

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11711
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(Tuti Purwaningsih et all, 2018). Data science aims to extract knowledge or information from data in the form of images or text(Glints. Data Science, 2022). Using an algorithm called machine learning, it is useful for processing images, text, audio, video and so on in order to produce an artificial intelligence system(Accurate, 2022).

In order to make it easier to help in finding the causes and solutions for early prevention of pulmonary TB disease, a method is needed through data science stages based on lung x-ray images using the CNN model. So in the research the title that will be discussed is the Classification of Tuberculosis Disease Based on X-Ray Lung Images With a Data Science Approach Using Convolutional Neural Networks.

LITERATURE REVIEW

Doodle introduction by utilizing CNN's ability use LeNet-5 architecture for introduction doodle types with 5 image objects, namely clothes, pants, chairs, butterflies and bicycles. The test results show that the first, second and fourth scenarios of bicycle objects are more recognizable with an accuracy value of 93%-98%, clothes objects are more recognizable in the third scenario with an accuracy value of 94%(Muhammad Rafly et all, 2020). Use of Convolutional Neural Network to classify the freshness of the following fruits: apples, oranges, and bananas. The test results were obtained using the Confusion Matrix method with an accuracy value of 93.3%(Femil Paraijun et all, 2022). The proposed hybrid CNN and improved Particle Swarm Optimization (CNN-ePSO) find e optimal architecture of connected layer in classification network layer especially for CXR images. The augmentation process for existing image is carried out to improve identify its features before implementing the model. The TB CXR benchmark image is used to perform binary classification, whether it is included in normal or TB(M. Yusoff, et all, 2021).

METHOD

The type of research used is experimental, where researchers analyze the level of efficiency and percentage of accuracy with the Convolutional Neural Network (CNN) model using a data science approach process in classifying normal and non-normal lung X-ray images. In this study, lung x-ray image dataset used was 3500 images, which can be found at https://drive.google.com/drive/folders/1Bcb6QSGRzVAM-UB3abkUG5qWfrUJnfLS?usp=sharing. Consisting of x-ray images of normal and TB . lungs.

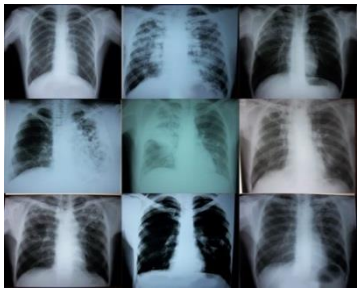

Figure 1. Overview of lung x-ray

The research was conducted to obtain information related to the processing of the lung x-ray image dataset. The following steps are carried out using a data science approach using the CNN model:

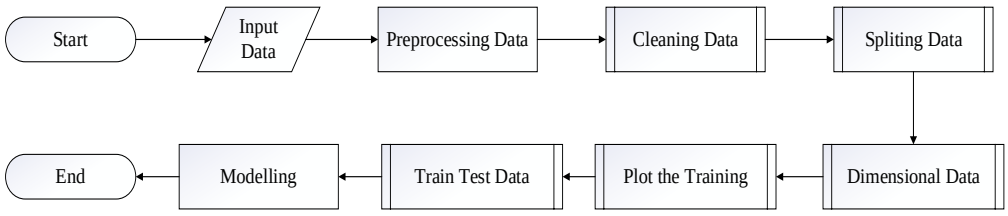

Figure 2. Work procedures

The first step, perform data input in the form of x-ray images of the lungs using Google Drive. Second step, carried out several processes, namely: cleaning dataset, splitting dataset is separation dataset of tuberculosis positive lungs and normal lungs, dimensional dataset is simplification of data from the data that has been splitting, plot training is results dataset separation and simplification, and train test dataset on results of the previous processed data. The third step, modeling uses the CNN method to detect the accuracy program.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11711
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

In this study, data collecting, pre-processing data, and data modeling were carried out by applying the Convolution Neural Network (CNN) to classify Tuberculosis (TBC) based on images of lung X-rays..

A. Data Collecting

In this study, the dataset used as the variable studied was the pulmonary TB dataset in the form of image data. There are two types of image data, namely Normal lung images and Tuberculosis lung images

B. Data Preprocessing

At this stage, the labeling process is carried out in the form of Normal and Tuberculosis labels. The value for the normal label is 0 and the value for the tuberculosis label is 1. The results of the division of these labels can be seen in Figure 4.

	image	label
	/content/drive/MyDrive/TB_Chest_Radiography_Da...	0
	/content/drive/MyDrive/TB_Chest_Radiography_Da...	0
	/content/drive/MyDrive/TB_Chest_Radiography_Da...	1
	/content/drive/MyDrive/TB_Chest_Radiography_Da...	0
	/content/drive/MyDrive/TB_Chest_Radiography_Da...	0

Figure 3. Data Labels

After labeling, then splitting the data to determine the distribution of normal and tuberculosis data. The picture of the results of the data sharing can be seen in Figure 4.

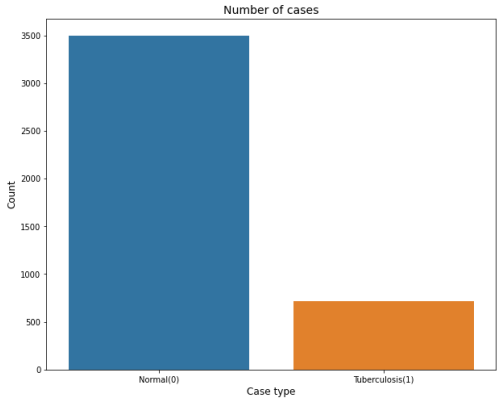

Figure 4. Data Sharing Visualization

The number of division results from the visualization in Figure 3.2 can be seen in Table 1.

Table 1. Results of Data Sharing

Label	Amount
Normal	3500
Tuberculosis	718

After the distribution of the data, the results of the image classification are displayed which can be seen in Figure 5.

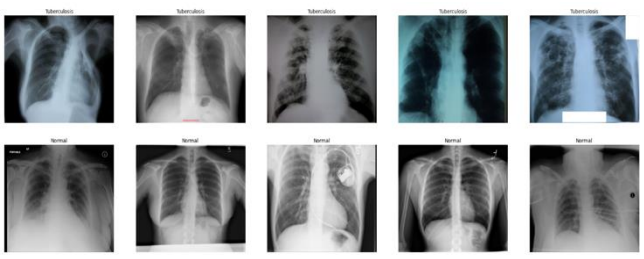

Figure 5. Classification Image

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Then Dimmension-Reduction (DR) is carried out to reduce or reduce the dimensions of the dataset used with the results obtained in the form of total validation examples and labels. The obtained results can be seen in Table 2.

Table 2. Results of Diemension-Reduction

Dimmension Reduction	**Amount**
Validation	(4218, 28, 28, 3)
Labels	(4218)

Epoch 1/200
191/191 [==============================] - 11s 49ms/step - loss: 0.3560 - accuracy: 0.8319 - val_loss: 0.1677 - val_accuracy: 0.9418
…
…
…
Epoch 195/200
191/191 [==============================] - 10s 51ms/step - loss: 0.0304 - accuracy: 0.9910 - val_loss: 0.0825 - val_accuracy: 0.9725
Epoch 196/200
191/191 [==============================] - 12s 61ms/step - loss: 0.0347 - accuracy: 0.9888 - val_loss: 0.0561 - val_accuracy: 0.9824
Epoch 197/200
191/191 [==============================] - 10s 51ms/step - loss: 0.0385 - accuracy: 0.9883 - val_loss: 0.0950 - val_accuracy: 0.9780
Epoch 198/200
191/191 [==============================] - 10s 51ms/step - loss: 0.0563 - accuracy: 0.9818 - val_loss: 0.0585 - val_accuracy: 0.9857
Epoch 199/200
191/191 [==============================] - 10s 50ms/step - loss: 0.0365 - accuracy: 0.9892 - val_loss: 0.0567 - val_accuracy: 0.9879
Epoch 200/200
191/191 [==============================] - 14s 72ms/step - loss: 0.0338 - accuracy: 0.9892 - val_loss: 0.0618 - val_accuracy: 0.9835

The results presented previously were obtained by setting the epochs limit to 200 by compiling hard to achieve the accuracy of the CNN method. Based on the last epochs, which is 200, the accuracy obtained is 0.9892, which means the CNN accuracy is 98%, with validation the accuracy obtained is 0.9835 or 98%. Thus, the results of the classification test using CNN are quite accurate. With the acquisition of CNN results which are quite high, it can be used as a consideration to be used in classifying TB disease with datasets that are normal and positive for tuberculosis.

CONCLUSION

Based on results of testing classification of the lung x-ray image dataset which is divided into 2, namely normal data 3500 images and tuberculosis data 718 images. The results of the classification of x-ray images on tuberculosis resulted in the last epoch value of 200 with an accuracy obtained of 98% and validation accuracy of 98%. The average accuracy of 200 epochs is 0.9803, with an average duration of 10.2 s.

For further research, several things are recommended, namely : this research should be carried out using other methods to improve the accuracy of the results for comparison. To achieve more optimal results, collaborative research with the medical field should be carried out, and the application of this research should be carried out in real time.

REFERENCES

Abdulfattah E. Ba Alawi, Ahmed Y. A. Saeed, Murad A. Rassam, 2021, "The Role of Pre-trained Models in Diagnosing Covid-19 Using Chest X-Ray Images", 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA), pp.1-6.

Accurate. Data Science Adalah Profesi yang Makin Dibutuhkan Perusahaan, Ini Perannya!, January 26th 2022, https://accurate.id/lifestyle/data-science adalah/#:~:text=Proses%20dalam%20melakukan%20data%20science,tujuan%20bisnis%20secara%20lebih%20lancar

Erlyna Nour Arrofiqoh dan Harintaka, Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi. 2018, Geomatika Volume 24 No 2 November: 61-68. http://dx.doi.org/10.24895/JIG.2018.24-2.810

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Femil Paraijun, Rosida Nur Aziza, Dwina Kuswardani. Implementasi Algoritma Convolutional Neural Network Dalam Mengklasifikasi Kesegaran Buah Berdasarkan Citra Buah, 2022, KILAT. Vol. 11, No. 1, April, P-ISSN 2089-1245, E-ISSN 2655-4925.

Germas, Kementrian Kesehatan Republik Indonesia. Strategis Nasional Penanggulangan Tuberkulosis di Indonesia 2020-2024. 04/06/2021. https://tbindonesia.or.id/informasi/strategi-nasional/strategis-nasional-penanggulangan-tuberkulosis-di-indonesia-2020-2024/

Glints. Data Science: Arti, Manfaat, Proses, dan Contoh Penerapannya. 12 Jan 2022. https://glints.com/id/lowongan/data-science-adalah/#.YuTiOy2l0_N

Infodatin, Pusat data dan informasi kementrian kesehatan RI. "Tuberklosis". ISSN 2442-7659. https://pusdatin.kemkes.go.id/resources/download/pusdatin/infodatin/infodatin-tuberkulosis-2018.pdf

M. Yusoff, M. S. I. Saaidi, A S. Md Afendi, A. M. Hassan, Tuberculosis X-Ray Images Classification based Dynamic Update Particle Swarm Optimization with CNN, 2021, Journal of Hunan University (Natural Sciences), Vol. 48. No. 9. September.

Mitra Keluarga. "Tuberkulosis (TBC), Kenali Gejala, Penyebab dan Cara Penularan", Rabu, 23 Maret 2022, https://www.mitrakeluarga.com/artikel/artikel-kesehatan/tuberkulosis

Muhammad Rafly Alwanda, Raden Putra Kurniawan Ramadhan, Derry Alamsyah. Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle, 2020, Jurnal Algoritme Vol. 1, No. 1, Oktober , Hal. 45–56

Oloko-Oba, M., Viriri, S., 2020. Diagnosing Tuberculosis Using Deep Convolutional Neural Network. In: El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F. (eds) Image and Signal Processing. ICISP 2020. Lecture Notes in Computer Science(), vol 12119. Springer, Cham. https://doi.org/10.1007/978-3-030-51935-3_16

Saeful Bahri, Rusda Wajhillah, Miftah Farid Adiwisastra, Diagnosa Tuberculosis Paru Berbasis Citra X-ray Menggunakan Convolutional Neural Network, 2021, IJCIT (Indonesian Journal on Computer and Information Technology) 6 (2), 181-186

Sehat negeriku sehatlah bangasaku. "Tahun ini, Kemenkes Rencanakan Skrining TBC Besar-besaran", 22 Maret 2022, https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20220322/4239560/tahun-ini-kemenkes-rencanakan-skrining-tbc-besar-besaran/

Sehat. "Kenali penyebab dan gejalaTBC paru yang menular". Kontan.co.id. Rabu, 15 Desember 2021. https://kesehatan.kontan.co.id/news/kenali-penyebab-dan-gejalatbc-paru-yang-menular.

Tri Kristini, Rana Hamidah, Potensi Penularan Tuberculosis Paru pada Anggota Keluarga Penderita. Jurnal Kesehatan Masyarakat Indonesia. Volume 15, Nomor 1, Mei 2020. https://jurnal.unimus.ac.id/index.php/jkmi, jkmi@unimus.ac.id

Tuti Purwaningsih, Imania Ayu Anjani, Pertiwi Bekti Utami. Convolutional Neural Networks Implementation for Chili Classification. 2018 International Symposium on Advanced Intelligent Informatics (SAIN). Hal 190-194. 978-1-5386-5280-0/18/$31.00 ©2018 IEEE

*name of corresponding author