Indonesian Applied Research Computing and Informatics
Vol. 1; No. 2, January 2026 Page 1-10;
https://jurnal.tdinus.com/index.php/iarci
e-ISSN : 3110-8806

Multi-Scale Convolutional Neural Network-Based Classification of Tuberculosis Chest X-ray Images

M. Ridwan[1*], Syafrudin[1], Sahrul Fauzan Djiaulhaq[1], Siti Mutmainah[1], Teguh Ansyor Lorosae[1]

[1] Computer Science, University of Muhammadiyah Bima, 84111, Indonesia

Email : m.ridwan180105@gmail.com[1*], syafrudinsyaffa101@gmail.com[1], fauzansahrul2@gmail.com[1], siti.mutmainah.id19@gmail.com[1], ansyorlorosae95@gmail.com[1]

Article Info
Article history:
Received:01-13-2026
Revised: 01-14 -2026
Accepted: 01-18-2026

Keywords:
Tuberculosis,
CNN,
Multi-Scale,
Thorax,
X-Ray

ABSTRACT

Tuberculosis (TB) is an infectious disease caused by the bacterium *Mycobacterium tuberculosis*, which mainly attacks the lung organs. One of the most commonly used methods of TB diagnosis is thorax X-ray imaging. The images of the examination results are visually analyzed by medical personnel to identify certain patterns or characteristics that indicate TB disease. However, the manual analysis process takes time and depends on the doctor's experience. Therefore, this study utilizes Artificial Intelligence (AI) technology as a diagnostic tool to provide alternative solutions that are faster and more efficient in determining TB status in patients. This study proposes the use of the Multi-Scale Convolutional Neural Network (CNN) method to classify tuberculosis disease based on thorax X-ray images. The data used was in the form of lung X-ray images that acted as inputs at the image processing stage. The dataset collected consisted of 790 images divided into two classes, namely normal lungs and lungs indicated by tuberculosis. The CNN architecture includes three convolutional layers with a kernel size of 3×3, three *max pooling* layers of 2×2, and one fully *connected layer with* a softmax *activation function*. Each convolutional layer uses 128 filters, and the model learning process is optimized using the Adam Optimizer algorithm. The training process was carried out for 15 epochs and resulted in an accuracy rate of 81%. Furthermore, at the model evaluation stage, an accuracy of 79% was obtained, indicating that the proposed method has sufficient performance in classifying tuberculosis disease.

*Corresponding Author:
M. RIDWAN
Computer Science, University of Muhammadiyah Bima, 84111, Indonesia
Email: m.ridwan180105@gmail.com

1. Introduction

In 2019, approximately 9.96 million people were infected with Tuberculosis (TB) worldwide. The highest number of new TB cases occurred in the World Health Organization (WHO) Southeast Asia Region,

accounting for 44% of global cases [1] . In Indonesia, TB remains a major public health concern, including in Bima City, where the prevalence of pulmonary TB shows a fluctuating yet concerning trend. Disease registry data indicate that TB case coverage increased from 0.13% in 2016 to 0.17% in 2017, followed by a decrease to 0.10% in 2018. In the working area of the Penana'e Health Center, pulmonary TB cases rose from 17.96% in 2016 to 20.57% in 2017, then slightly decreased to 18.07% (30 cases) in 2018. Furthermore, by the third quarter of 2019 (January–September), the number of cases had reached 39. These data indicate that pulmonary TB remains a significant public health problem in Bima City due to its high transmission potential and negative social impacts [2].

Tuberculosis (TB) is an infectious disease caused by acid-fast bacilli (AFB), namely *Mycobacterium tuberculosis*[3]. TB diagnosis can be performed through clinical symptom evaluation, culture examination, microscopic examination, radiological imaging, and tuberculin testing. Early detection of TB is crucial to improving treatment outcomes, reducing disease transmission, and preventing drug resistance and multidrug-resistant TB [4]. However, TB treatment requires a long duration of 6–9 months with multiple drug regimens, and many patients still underestimate prolonged coughing as a normal condition, despite it being a key symptom of TB [5].

Chest X-ray imaging is the most commonly used technique for TB diagnosis due to its wide availability and ability to visualize lung abnormalities. Traditionally, X-ray images are interpreted visually by medical experts to identify TB-specific patterns. Nevertheless, this process is time-consuming and prone to subjectivity, particularly in healthcare facilities with high patient volumes, such as those implementing government-supported free TB screening programs. These limitations highlight the need for automated and accurate computer-aided diagnostic systems.

With advances in technology, artificial intelligence (AI), particularly deep learning, has been increasingly applied in medical image analysis. Convolutional Neural Networks (CNNs) have demonstrated strong performance in extracting complex features from medical images and have been widely used in detecting pulmonary diseases such as pneumonia, lung cancer, and tuberculosis from chest X-ray images[6]. Despite their effectiveness, conventional CNN models typically employ single-scale convolutional kernels, which may limit their ability to capture TB lesions that vary significantly in size, shape, and distribution within lung regions.

Recent studies have introduced multi-scale learning approaches as a state-of-the-art solution to address this limitation. Multi-Scale Convolutional Neural Networks (MSCNN) utilize multiple convolutional kernels of different sizes within a single architecture, enabling simultaneous extraction of local and global features. This approach has been shown to improve classification performance in complex medical imaging tasks, particularly for lung disease detection, compared to single-scale CNN models [7][8].

However, most existing studies focus on general or large-scale datasets and lack validation on regional TB data, especially in areas with high local prevalence such as Bima City. This indicates a research gap in the application of MSCNN for pulmonary TB classification using localized thorax X-ray datasets, where lesion variability and epidemiological characteristics may differ from global datasets.

Therefore, this study aims to develop and evaluate a Multi-Scale Convolutional Neural Network (MSCNN) model for the classification of pulmonary Tuberculosis using chest X-ray images. The objective of this research is to improve TB detection accuracy by effectively capturing multi-scale lung lesion features and to compare the model's performance with medical expert diagnoses. The proposed approach is expected to support early TB detection, enhance diagnostic efficiency, and contribute to the development of reliable computer-aided diagnosis systems for TB, particularly in high-prevalence regions.

2. Methodology

2.1 Research Stages

The research begins by conducting a literature study then collecting the necessary datasets, after which it enters the programming process by labeling the dataset, then *splitting* or separating the dataset based on its category, then rescaling and *grayscaling* the image, then implementing the CNN model and evaluating the

results so that it can produce the expected final result [9].

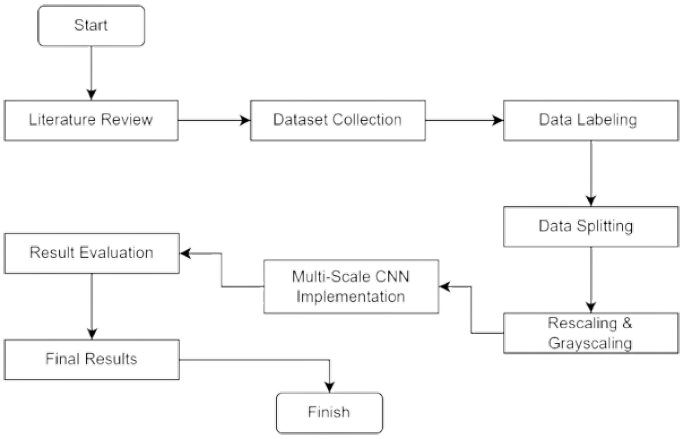


Figure 1. Stages of research

This study follows a structured methodology to develop and evaluate a Multi-Scale Convolutional Neural Network (MSCNN) for the classification of tuberculosis (TB) using chest X-ray images. The overall research workflow is illustrated in the proposed framework and consists of several sequential stages.

The research begins with the definition of the research goal, which is to design an effective deep learning model capable of accurately classifying tuberculosis from chest X-ray images. This goal guides the selection of methods, data processing techniques, and evaluation strategies throughout the study.

Next, a literature review is conducted to analyze previous studies related to tuberculosis detection, chest X-ray image analysis, and convolutional neural network architectures. This step provides theoretical foundations and identifies existing limitations, motivating the use of a multi-scale approach for feature extraction.

Following the literature review, the dataset collection stage is carried out. The dataset consists of chest X-ray images obtained from reliable and publicly available medical imaging sources. These images represent both tuberculosis-positive and normal cases, ensuring balanced and meaningful classification.

The collected images then undergo data labeling, where each X-ray image is assigned a class label based on its clinical condition. Accurate labeling is essential to ensure reliable supervised learning during model training.

After labeling, the dataset is divided during the data splitting stage into training, validation, and testing sets. This division enables effective model training, hyperparameter tuning, and unbiased performance evaluation on unseen data.

Subsequently, image preprocessing is applied, including resizing and grayscale conversion. Resizing standardizes image dimensions to match the input requirements of the MSCNN model, while grayscale conversion reduces computational complexity without removing important diagnostic features.

The core stage of the study is the implementation of the Multi-Scale CNN. The proposed architecture employs convolutional layers with different kernel sizes to extract features at multiple spatial scales. This approach allows the model to capture both fine-grained details and global patterns in chest X-ray images, which are crucial for tuberculosis detection.

Once the model is trained, performance evaluation is conducted using standard classification metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the model's effectiveness in detecting tuberculosis cases.

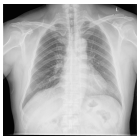
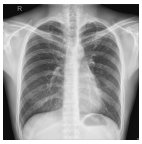
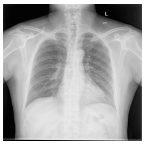
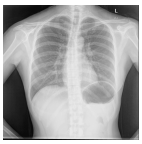
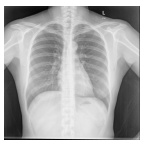
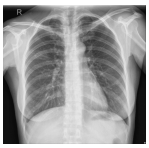
Finally, the final results are presented and analyzed. The experimental findings are discussed to

demonstrate the performance of the proposed MSCNN model and its potential advantages compared to conventional CNN approaches. This stage also highlights the applicability of the proposed method for assisting tuberculosis screening based on chest X-ray images.

2.2 Data Collection

The dataset consists of chest X-ray images representing normal lung conditions and pulmonary tuberculosis cases. The collected data serve as the primary input for training and evaluating the proposed Multi-Scale Convolutional Neural Network (MSCNN) model. Proper data collection and labeling are essential to ensure the reliability and validity of the classification results.

Table 1. Dataset Sample

X-ray Image of Normal	X-ray Image of Tuberculosis
	
	
	

The data obtained were 790 images divided into 2 categories, namely 395 images of normal lung X-rays and 395 *images of tuberculosis* lung X-rays, as seen in **Table 1.** The file is then labeled using the "d-i" format where I is the number from 0 to the total number of images available.

The chest X-ray dataset used in this study was obtained from the Kaggle platform and is publicly available at: (https://www.kaggle.com/datasets/rafadeliyati/tbc-dataset). The dataset is provided for research and educational purposes and is distributed under the license specified by the dataset owner on Kaggle.

2.3 Splitting Data

The dataset is divided into 3 parts, namely training, testing, and Val. Data sharing is done using a library of the Python programming language , namely split-folders. The dataset is divided by a ratio of 15% for data testing, 70% for data training, 15% for Val data.

```
total = len(images)
train_end = int(total * TRAIN_RATIO)
val_end = train_end + int(total * VAL_RATIO)

train_files = images[:train_end]
val_files = images[train_end:val_end]
test_files = images[val_end:]

for file in train_files:
    shutil.copy(
        os.path.join(cls_path, file),
        os.path.join(TARGET_DIR, "train", cls, file)
    )

for file in val_files:
    shutil.copy(
        os.path.join(cls_path, file),
        os.path.join(TARGET_DIR, "val", cls, file)
    )

for file in test_files:
    shutil.copy(
        os.path.join(cls_path, file),
        os.path.join(TARGET_DIR, "test", cls, file)
    )
```

Figure 2. Python Split Data Code

The **Figure 2.** shows a code snippet used to split the image dataset into training, validation, and testing subsets. The total number of images is first calculated, after which the dataset is divided according to predefined training and validation ratios. The images are then separated into three groups: training, validation, and testing. Each group of image files is copied into its corresponding directory structure (*train*, *val*, and *test*) using file operation functions. This process ensures an organized dataset distribution and supports systematic model training, validation, and performance evaluation.

2.4 Preprocessing Data

In this preprocessing stage, rescaling the image is carried out. All images are rescaled in the region of interest (ROI) with the pixel size of the image to 224x224.

```
IMG_SIZE = (224, 224)
BATCH_SIZE = 8
EPOCHS = 15
DATASET_DIR = "dataset"
MODEL_NAME = "mscnn_tb_model.h5"

```

Figure 3. Rescaling Data

The **Figure 3.** presents the main hyperparameter settings used by the authors during the training process of the Multi-Scale Convolutional Neural Network (MSCNN) model for tuberculosis classification based on chest X-ray images. The input image size was set to 224 × 224 pixels (*IMG_SIZE*) to meet CNN architecture requirements and ensure uniform data dimensions. The *BATCH_SIZE* was defined as 8, determining the number of images processed in each training iteration. The training process was conducted for 15 epochs (*EPOCHS*) to allow the model to sufficiently learn feature patterns from the dataset. The *DATASET_DIR* parameter indicates the directory where the image dataset was stored, while *MODEL_NAME* specifies the filename used to save the trained model in *.h5* format. These parameter settings were selected to achieve stable training and optimal model performance.

2.5 CNN architecture

The Multi-Scale Convolutional Neural Network (MSCNN) architecture is employed in this study for the classification of tuberculosis based on chest X-ray images. The proposed architecture is designed to extract features at multiple scales to accommodate variations in the size and patterns of tuberculosis lesions in the lungs. By integrating multiple convolutional layers and feature fusion, the MSCNN is expected to enhance classification performance compared to conventional single-scale CNN approaches.

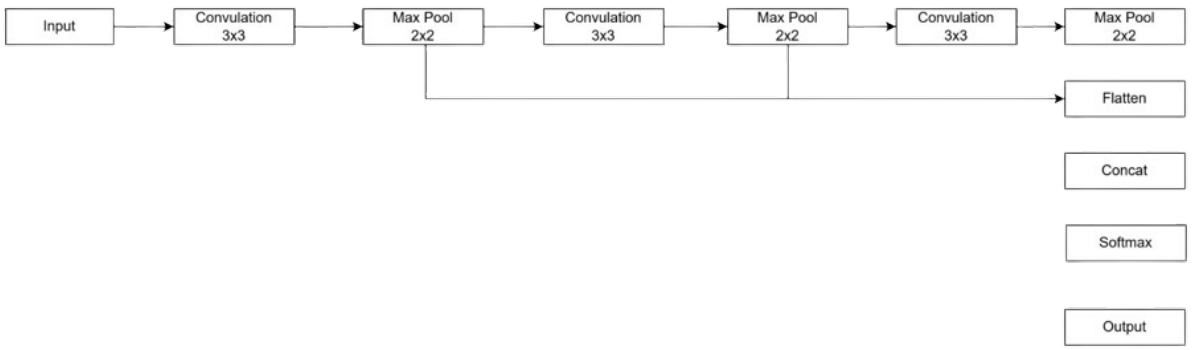


Figure 4. CNN Architecture

In 1989, Yann LeCun et al introduced a training algorithm that was used to understand the load on images and was proven to show good performance in image detection, classification, and segmentation called CNN. CNN is one of the artificial neural networks with supervised training[10]. CNNs are trained using datasets consisting of inputs and targets to which the target is affiliated with input. CNN uses targets to determine how well the training process performs and as a reference for the adaptation of the weight value to reduce the value of the loss obtained in the process[11].

CNN consists of several *layers* including *input layer*, convolutional layer, pooling layer (*maxpool*) and also *fully-connected* layer including *flatten, concat, softmax* and *output layer*. The number of filter values used in CNN is not limited, but the ones that are usually used are 8, 16, 32, 64 and so on. The filter size used is also usually 3x3, 5x5, 7x7,etc.On CNN, the initial layer is named the convolutional layer. Convolution is a mathematical form of combining two signals to form a third signal[12]. The first signal is called *the input* signal signal, while the second signal is often referred to as *a filter*. Then the third signal or the last signal is the *output signal*.

Furthermore, after the convolution process, *the output* produced will go through the next layer, namely the merging layer. The merging layer aims to achieve spatial invariability by reducing the resolution of the matrix by trimming the elements from the spatial density of money having a smaller value[13]. In this study, a merger operation was used by taking the maximum value and getting rid of the smaller value using *the maxpool*.

After passing the merging process, *the output* will be input to the next layer, which is *the fullyconnected* layer. In this layer there are several processes such as *flatten,* where the input will go through a leveling process. After going through the leveling process, a classification process will be carried out. In this process, an activation function is required, where there are various options such as *linear, step, ramp, sigmoid, hyperbolic, Gaussian, softmax,* and so on[14]. In this study, the function used in this layer is *the softmax* activation function. The *softmax* function itself will calculate the probability distribution of the existing classes[15]. After getting the probability, the error or *loss* of the calculation will be calculated based on the intended target by applying the *cross-entrophy error function*.

The training process at CNN is carried out by making adjustments and also adjustments to parameters so that the minimum loss value is achieved. Like Figure 4, this study applied 3 convolutional layers measuring 3x3. Then in the merge operation used is a merge operation that takes the maximum value and discards the smaller value (*Maxpool*)[16]. The function used to take the maximum value of spatial stability is the 2x2 window function. Then each *output* of the Maxpool merge layer will go through a *flatten* (equalization) process. After going through *the flattening* process, the resulting vectors will be combined on the *Concat* layer into a vector. After that, the vectors will be classified using *the Softmax* activation function into 2 classes[17].

3. Results and Discussion

This section will describe two subsections, the first being training and the second being testing. The programming language used is Python, utilizing the TensorFlow library. The amount of data used in this study consists of two parts with a total of 790 images. In this study, training and testing were used. The data used for the training process consisted of 552 images, comprising 276 normal images and 276 tuberculosis images, while the data used for the testing process consisted of 238 images, comprising 119 normal images and 119 tuberculosis images.

Table 2. Training Results

epoch	Accuracy(%)	Loss
1	54,50	0,7063
2	51,12	0,7062
3	57,34	0,5000
4	68,77	0,6377
5	66,99	0,6377
6	67,44	0,6253
7	71,64	0,5720
8	71,00	0,5720
9	71,23	0,5886
10	70,50	0,5857
11	73,98	0,5970
12	75,08	0,6102
13	77,53	0,5511
14	77,14	0,5787
15	81,10	0,5401

The training process using the CNN model was carried out using loss and accuracy parameters. The data training process used 552 steps with 15 epochs, and the training process took 1 hour. The CNN architecture used in this study consists of 3 convolution layers followed by a Maxpool pooling layer. The number of filters used is 128, and the filters are 3x3 in size. A 2x2 window function is used in the Maxpool layer.

Each output from the Maxpool combined layer will go through a flattening process. After going through this process, the resulting vector is then combined into a vector contained in the Concat layer. After that, the vector is classified into 2 classes using Softmax activation. It can be seen in Table 2 that the data training process in the 1st epoch has a loss value of 0.7063 and an accuracy of 54.50%. The accuracy then increases in subsequent epochs. At epoch 12, the accuracy stabilizes at 70% until the last epoch, where it achieves an accuracy of 81% and a loss of 0.5401.

3.1 Process Testing

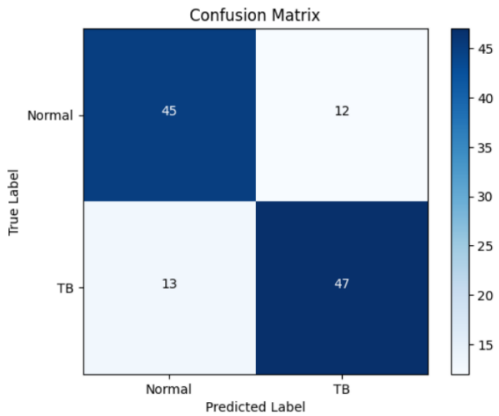


Figure 5. Convusion Matrix

The confusion matrix in the image shows the performance of the TB disease classification model with two classes, namely Normal and TB. From the tested data, the model successfully classified 45 Normal data

and 47 TB data correctly. However, there were prediction errors, namely 12 Normal data points that were incorrectly predicted as TB and 13 TB data points that were incorrectly predicted as Normal. These results show that the model is able to distinguish between the two classes quite well, although there are still classification errors that need to be minimized.

Table 3. Matrix Evaluation

Class	Precision	Recall	F1-score	Support
Normal	0.78	0.79	0.78	57
TB	0.80	0.78	0.79	60
Accuracy			0.79	117
Macro Average	0.79	0.79	0.79	117
Weighted Average	0.79	0.79	0.79	117

Based on the classification report, the TB classification model shows fairly balanced performance between the Normal and TB classes. In the Normal class, the model obtained a precision of 0.78, a recall of 0.79, and an F1-score of 0.78, while in the TB class, it obtained a precision of 0.80, a recall of 0.78, and an F1-score of 0.79. The overall accuracy value of 79% indicates that the model is capable of classifying data with a moderate level of accuracy. The macro average and weighted average values are the same (0.79), indicating that the model is not biased towards either class and its performance is relatively consistent across both categories.

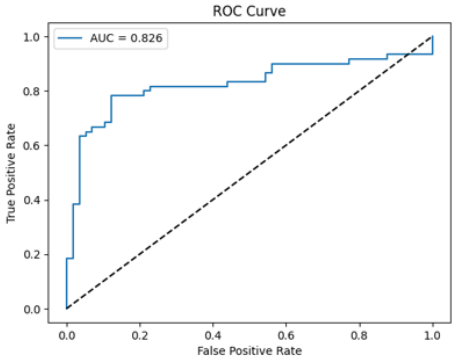


Figure 6. ROC Curve

The ROC (Receiver Operating Characteristic) curve in the figure shows the model's ability to distinguish between Normal and TB classes. An AUC value of 0.826 indicates that the model has good discriminatory ability, as the AUC value is close to 1. The further the curve is from the diagonal line (random prediction), the better the model performance. These results indicate that the model is quite effective in distinguishing TB and non-TB patients at various classification thresholds**.**

5. Conclusion

This study presented the classification of pulmonary tuberculosis using chest X-ray images based on a Multi-Scale Convolutional Neural Network (MSCNN) approach. The proposed method was designed to address the variability in size and patterns of TB lesions by extracting features at multiple scales, which is a key challenge in chest X-ray–based TB detection. The experimental results demonstrate that the MSCNN model is capable of effectively distinguishing between normal and TB-affected lungs.

The model was trained using 790 chest X-ray images and evaluated through a testing phase, achieving

stable performance across multiple epochs. The results indicate that the proposed approach achieved satisfactory classification performance, with an accuracy of 81% during training and 79% during testing, confirming the model's generalization capability. These findings suggest that multi-scale feature extraction contributes positively to improving TB classification performance compared to conventional single-scale CNN approaches.

From a practical perspective, the proposed MSCNN model has the potential to support medical personnel as a computer-aided diagnosis (CAD) tool, particularly in healthcare facilities with high patient volumes and limited radiology experts. By assisting in early TB detection, this approach may help reduce diagnostic delays and support timely treatment decisions.

However, this study has several limitations, including the size of the dataset and the use of binary classification. Future research is recommended to incorporate larger and more diverse datasets, multi-class classification, and further optimization of network architectures to enhance robustness and clinical applicability. Overall, this research contributes to the development of AI-based diagnostic systems for tuberculosis and provides a foundation for further improvements in automated TB detection.

References

[1] N. Roswita Paramata, P. Ayuningtias Mahdang, and Yusuf, "Upaya Penurunan Angka Infeksi TBC Melalui Pemberian Edukasi dan Pendampingan Kader Kesehatan dan Pengawas Menelan Obat (PMO) Efforts to reduce TB infection rates through providing education and assistance to health cadres and drug swallowing supervisors," *Jur J. Kolaboratif Sains*, vol. 7, no. 10, pp. 3659–3666, 2024, doi: 10.56338/jks.v7i10.6213.

[2] A. Shrivas and S. Singh, "Tuberculosis Diagnosis and Management: Recent Advances," *J. Glob. Infect. Dis.*, vol. 17, no. 1, pp. 3–9, 2025, doi: 10.4103/jgid.jgid_112_24.

[3] Diah Putri Kartikasari, Fiqri Dian Priyatna Sinaga, Tiara Ayu Triarta Tambak, Zahra Humaira Kudadiri, and M. Khalil Gibran, "Ekstraksi Fitur Citra Grayscale dengan Convolutional Neural Networks," *J. Tek. Inform. dan Teknol. Inf.*, vol. 5, no. 1, pp. 198–205, 2025, doi: 10.55606/jutiti.v5i1.5175.

[4] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, *A review of convolutional neural networks in computer vision*, vol. 57, no. 4. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10721-6.

[5] N. P. Sari, "Analisis Performa Algoritma CNN dalam Klasifikasi Citra Medis Berbasis Deep Learning Analysis Of CNN Algorithm In Deep Learning-Based Medical Image Classification," *J. Komput.*, vol. 2, no. 2, pp. 87–92, 2024.

[6] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," *Inf.*, vol. 16, no. 3, pp. 1–28, 2025, doi: 10.3390/info16030195.

[7] Y. G. Tejeda and H. A. Mayer, "Deep Learning with CNNs: A Compact Holistic Tutorial with Focus on Supervised Regression (Preprint)," pp. 1–42, 2024, doi: 10.3390/make6040132.

[8] H. Xiong *et al.*, "A novel method based on a multiscale convolution neural network for identifying lung nodules," *Sci. Rep.*, vol. 15, no. 1, pp. 1–15, 2025, doi: 10.1038/s41598-025-21582-6.

[9] A. R. Hakim and D. M. U. Atmaja, "Implementation of Cnn for Classifying Patchouli Leaf Images Based on Accuracy and Evaluation," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 4, pp. 728–734, 2025, doi: 10.33480/jitk.v10i4.6207.

[10] S. Oktavia, F. Hidayati, O. Lesmana, and M. B. Butar, "Hubungan Kondisi Fisik Lingkungan Rumah dengan Kejadian Tuberkulosis Paru di Indonesia (Analisis Data Survei Kesehatan Indonesia Tahun 2023)," vol. 4, no. 4, pp. 3581–3588, 2025.

[11] D. D. K. Mahardika, “No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title,” vol. 6, no. February, pp. 4–6, 2024.

[12] Rodiah, D. T. Susetianingtias, and E. Patriya, “Model Convolutional Neural Network (CNN) Custom Sequential untuk Klasifikasi Citra Ikan Hias Carassius Auratus pada Industri Akuakultur,” *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 3, pp. 813–825, 2025, [Online]. Available: https://journal.umkendari.ac.id/decode/article/view/1229

[13] M. Krichen, “Convolutional Neural Networks: A Survey,” *Computers*, vol. 12, no. 8, pp. 1–41, 2023, doi: 10.3390/computers12080151.

[14] A. Zafar *et al.*, “Convolutional Neural Networks: A Comprehensive Evaluation and Benchmarking of Pooling Layer Variants,” *Symmetry (Basel).*, vol. 16, no. 11, 2024, doi: 10.3390/sym16111516.

[15] B. S. Nusantara and M. Akbar, “Klasifikasi Penyakit Tuberculosis Berdasarkan Citra Rontgen Thorax Menggunakan Multi-Scale Convolutional Neural Network,” *sudo J. Tek. Inform.*, vol. 3, no. 1, pp. 1–8, 2024, doi: 10.56211/sudo.v3i1.483.

[16] V. M. Risma and E. Utami, “Tuberculosis Diagnosis From X-Ray Images Using Deep Learning And Contrast Enhancement Techniques,” *J. Tek. Inform.*, vol. 6, no. 2, pp. 981–994, 2025, doi: 10.52436/1.jutif.2025.6.2.4315.

[17] The Lancet Regional Health – Southeast Asia, “Eliminating tuberculosis by addressing access and equity,” *Lancet Reg. Heal. - Southeast Asia*, vol. 31, p. 100520, 2024, doi: 10.1016/j.lansea.2024.100520.