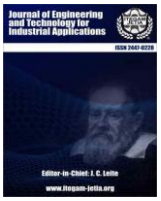
Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.58, p.732-740. March/April, 2026.
DOI: https://doi.org/10.5935/jetia.v12i58.2954



RESEARCH ARTICLE OPEN ACCESS

EVALUATING THE EFFECTIVENESS OF DEEP LEARNING MODELS FOR CHEST X-RAY IMAGE CLASSIFICATION

I Komang Somawirata*[1], Fitri Utaminingrum[2], Chikamune Wada[3], Ervin Yohannes[4]

[1]Department of Electrical Engineering, National Institute of Technology, Malang, 65145, Indonesia
[2]Faculty of Computer Science, Universitas Brawijaya, Malang, East Java, 65145, Indonesia
[3]Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Hibikino 2-4 Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan.
[4]Informatic Engineering, Faculty of Engineering, State University of Surabaya

[1]https://orcid.org/0000-0001-6625-4407, [2]https://orcid.org/0000-0002-0281-9429,
[3]https://orcid.org/0000-0002-8349-7141, [4]https://orcid.org/0000-0003-1531-9172

Email: *kmgsomawirata@lecturer.itn.ac.id

ARTICLE INFO

Article History
Received: November 15, 2025
Reviewed: December 25, 2025
Accepted: March 10, 2026
Published: April 30, 2026

Keywords:
CNN,
X-ray,
Image,
Data set,

ABSTRACT

A chest X-ray (CXR) examination is one of the radiological examinations used to help a doctor diagnose a disease in patients safely, quickly, and inexpensively. The development of Computer-Aided Diagnosis (CAD) systems has prompted numerous researchers to explore methods for detecting diseases using X-ray imaging. By implementing this research, it is hoped that it will enable medical personnel to accurately and quickly diagnose patients' diseases. This study utilizes three datasets of aortic enlargement, cardiomegaly, and COVID-19. A Convolutional Neural Network (CNN) is one method that researchers widely use to build Computer-Aided Design (CAD) systems. This study aims to compare the performance of seven CNN models with different architectures to determine which one produces the highest accuracy. The seven CNN models used include DenseNet121, DenseNet169, DenseNet201, InceptionV3, MobileNet, ResNet50, and Xception. The test results show that the DenseNet201 model, with an input size of 224 × 224 pixels, achieves the highest accuracy value for all datasets, reaching over 90% accuracy.

I. INTRODUCTION

A chest X-ray (CXR) examination can detect anatomical and physiological abnormalities of the organs and diagnose diseases in the chest. If there are abnormalities, they can be identified [1]. The results of a chest X-ray (CXR) are often used as a screening tool to detect diseases in patients quickly. CXR photos are usually chosen due to their low cost. Additionally, CXR is relatively safe for investigating these diseases in greater detail. However, reliability and accuracy in detecting abnormalities on X-ray photos often depend on medical personnel's experience and ability to read and interpret CXR results [2]. Meanwhile, the limited number of radiographic medical personnel causes delays in medical treatment.

It is one of the obstacles to the disease diagnosis process in patients, causing delays in the treatment process, which often causes delays in patient handling, which often causes the risk of the patient dying [3]. The Computer-Aided Diagnosis (CAD) system is one solution that is expected to help medical personnel diagnose various diseases through chest X-ray photos quickly and accurately [4-6]. Many researchers are still developing the CAD system. Artificial intelligence (AI) offers various methods for addressing multiple problems in the medical field. A Convolutional Neural Network (CNN) is one of the methods in Artificial Intelligence that is currently being developed in many Computer-Aided Design (CAD) systems [7].

CNN is often used in research related to medical image analysis to detect or classify diseases [8-10]. The CNN method is known to organize images with the best performance. It is often applied to medical image classification, segmentation, localization, or detection [11]. CNN models usually struggle to achieve high accuracy values when the number of datasets is limited. It resulted in a dataset with a limited amount of data, which can cause the trained model to fail to generalize properly. The resulting accuracy value is suboptimal.

Augmentation techniques are often employed to expand datasets, particularly in medical image research, as annotated medical image datasets typically have a limited number and are challenging to obtain [12-15].

The CNN method has various architectural models with layers, depths, and parameters. It dramatically affects the resulting accuracy value [16]. The CNN method offers multiple architectural models, allowing researchers to select and develop them according to the specific dataset. Hence, the accuracy results on each model can be compared to get the best accuracy value from the dataset of several diseases being tested. This research study analyzes seven CNN Architecture models, namely ResNet50, DenseNet121, DenseNet169, DenseNet201, InceptionV3, Xception, and MobileNet. These seven CNN models are recognized for their reliability in detecting and classifying images, producing high accuracy [11]. The ResNet model with residual connection has been proven to be able to detect various diseases, such as childhood glaucoma [17], COVID-19 [18], brain tumors [19], breast cancer [20], and brain disease [21].

The DenseNet model, with its densely connected structure, is also frequently used to classify various medical images, including mammogram images [20], Magnetic Resonance Imaging (MRI) scans [21], and X-ray photographs. Alzheimer's Disease [22], digital retinal images [23], thin blood smears [24], endoscopic images [25], MobileNet, and InceptionV3 models have low complexity and a light computational load, making them often used to classify various types of images [26]. Several studies have utilized these two models to analyze medical images for the detection of breast cancer [27] and skin diseases [28]. This study aims to detect six diseases based on chest X-ray (CXR) images, including those affecting the lungs, heart, and aorta. The six diseases are aortic enlargement (AE), cardiomegaly, and COVID-19. CNN research could perform well, but the computational load is also hefty. The available medical images are generally large, and one way to reduce the computational burden is to reduce the resolution before using them in the training process.

II. METHOD

Figure 1 shows a block diagram for classifying X-ray images using the Convolutional Neural Network (CNN) method to detect disease. This block diagram is divided into three steps: input, process, and output. The proposed study provides a database containing several normal chest X-ray images, as well as datasets for aortic enlargement, cardiomegaly, and COVID-19.

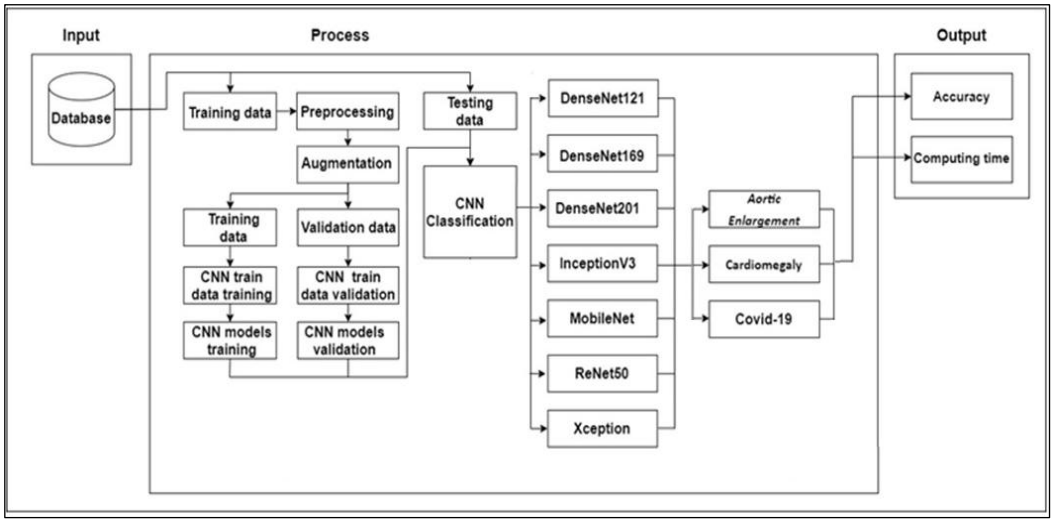


Figure 1: Block Diagram System.
Source: Authors, (2026).

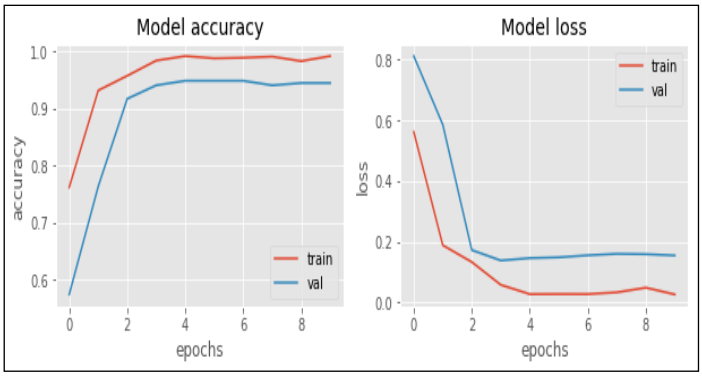


Figure 2: Example of display of loss and accuracy graph.
Source: Authors, (2026).

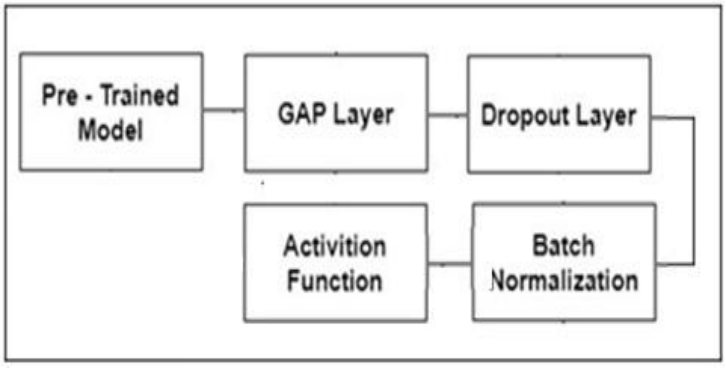


Figure 3: Convolutional Neural Network (CNN) Architecture.
Source: Authors, (2026).

The system input is taken from the X-ray image dataset. This study uses two image formats: PNG (Portable Network Graphics) and JPEG (Joint Photographic Experts Group). The PNG image format is available for the Aortic Enlargement, Cardiomegaly, and COVID-19. This research first underwent the supporting process stage before implementing the convolutional neural network (CNN) method. The available image dataset will be divided into two parts: training data and test data. The training data is used to train the seven

CNN models, while the test data is used to calculate the accuracy and computational time values for each CNN model. The ratio of training and test data from the total dataset is 3:1.

II.1 PREPROCESSING

The preprocessing step reduces the image size of the training data and test data from their original size. The image sizes used are 32x32 pixels, 75x75 pixels, 128x128 pixels, 180x180 pixels, and 224x224 pixels. Reducing the image size reduces the computational load and speeds up system performance [29].

II.2 AUGMENTATION

After preprocessing, the image results of the training data will proceed to the augmentation process. The augmentation process modifies the original images of the training data to produce new images. This study uses a geometric transformation method on the training data image, among others, by randomly rotating the image, zooming in and out, shifting the image position, and flipping the image horizontally (flip). This study aims to determine the most suitable augmentation technique for classifying each dataset with high accuracy. The augmentation stage also determined the values for each process, including the rotation process (10) and the zoom process (0.1). The flip process is done horizontally. The value of the 0.01 shift process is carried out horizontally and vertically. This study will examine the effect of augmentation techniques on three datasets, each comprising 4000 images. The three datasets include Aortic Enlargement (AE), with 1695 images, Cardiomegaly, with 2342 images, and Tuberculosis (TB), with 662 images.

II.3 CONVOLUTIONAL NEURAL NETWORK (CNN)

The results of the augmentation process from the training data will be divided into two parts, namely, training data and validation data. The distribution of validation data is based on 20% of the initial training data. To prevent overfitting, validation data will be used to determine the optimal parameter values during training, including the learning rate, batch size, and number of epochs. Figure 2 is an example of a graphical display to check for overfitting of the CNN model. A good performance model is marked by a decrease in loss value and an increase in accuracy value. However, if the opposite occurs, then overfitting happens.

The training phase will be run twice, utilizing transfer learning and fine-tuning techniques. The first training will freeze all layers of the pre-trained model, using a learning rate of 0.001 and an epoch value of 10. This stage will be connected to the more complex Convolutional Neural Network (CNN) stage of the model. The performance of the CNN model in image classification often depends on the data used during the training phase. This study will examine several layers commonly found in a CNN architecture, as illustrated in Figure 3. In the initial layer, a pre-trained model is used.

Table 1: Total training data and test data for all datasets.

Dataset	Training Data		Testing Data	
	Normal	Ab-normal	Normal	Ab-normal
Aortic Enlargement	617	651	207	218
Cardiomegaly	937	820	312	273
Covid-19	2715	2712	905	904

Source: Authors, (2026).

The pre-trained model is a CNN model that has been previously trained using the ImageNet dataset to solve the case in this study by modifying its parameters according to the new dataset. The CNN models used in the training phase are ResNet-50, DenseNet-121, DenseNet-169, DenseNet-201, MobileNet, Xception, and Inception-V3. The global average pooling (GAP) layer and the dropout layer function to minimize overfitting, where the training process is good, but the prediction results do not follow the actual class [30]. In this study, the drop layer value is 0.5 because this value is often used to maintain the output of each node in the hidden layer. Batch normalisation (BN) aims to enhance accuracy and accelerate the training process. This can happen because Batch normalisation (BN) results in fast convergence and better generalisation. The activation function can generate the probability value of an object against the available classes [31].

This study utilizes a Convolutional Neural Network (CNN) to classify chest X-ray images based on the characteristics of each dataset, as determined by the test results, including accuracy values, training time, and testing time. It will also analyze the use of an exact CNN model to build a CAD system that can later assist doctors in quickly and accurately detecting diseases from X-ray images. The Aortic Enlargement (AE) dataset used in this study is part of the VinDr-CXR or VinBigData dataset. VinBigData consists of 18,000 CXR images divided into 28 classes, namely Aortic Enlargement, Atelectasis, Cardiomegaly, Calcification, Clavicle Fracture, Consolidation, Edema, Emphysema, Enlarged PA, Interstitial Lung Disease (ILD), Infiltration, Lung Cavity, Lung Cyst, Lung Opacity, Mediastinal Shift, Nodule/Mass, Pulmonary Fibrosis, Pleural Thickening, Pleural Effusion, Rib Fracture, Lung Tumor, Chronic Obstructive Pulmonary Disease (COPD), other injuries. Meanwhile, for the image data, the name of the disease is used as a regular class. This dataset was introduced by Nguyen et al. [32] for the Detection of Chest X-ray abnormalities in VinBigData.

The Cardiomegaly dataset was derived from the ChestX-ray8 dataset, comprising 108,948 chest X-rays. The ChestX-ray8 dataset has nine classes: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, and normal. This study utilized a dataset of 2,342 Cardiomegaly images, divided into 1,249 normal and 1,093 Cardiomegaly labels [33]. The COVID-19 dataset used in this study is a subset of the COVQU dataset published by Rahman et al. and Chowdhury et al. [34]. The total COVID-19 dataset of CXR images comprises 7,236 images, divided into 3,620 regular class CXR images and 3,616 COVID-19 class CXR images. Figure 3 shows a normal chest X-ray image and a chest X-ray of a patient with COVID-19. Table 1 details the amount of training and test data on each object dataset in this study.

III. RESULT AND DISCUSSION

This test also selects the features of seven Convolutional Neural Network (CNN) models to determine the value of accuracy, training time, and test time. Seven CNN models are DenseNet121, DenseNet169, DenseNet201, InceptionV3, MobileNet, ResNet50, and Xception. Different image sizes are used, namely 32 x 32 pixels, 75 x 75 pixels, 128 x 128 pixels, 180 x 180 pixels, and 224 x 224 pixels. The exception is the InceptionV3 and Xception models, which cannot be tested using an image size of 32 x 32 pixels because they have an image size limit, as shown in Table 2. Therefore, the InceptionV3 and Xception models are resized to 75 x 75 pixels. The data sets tested were Aortic Enlargement, Cardiomegaly, and COVID-19. Figure 4 shows a sample regular (a, b) and COVID-19 (c, d) chest X-ray data set.

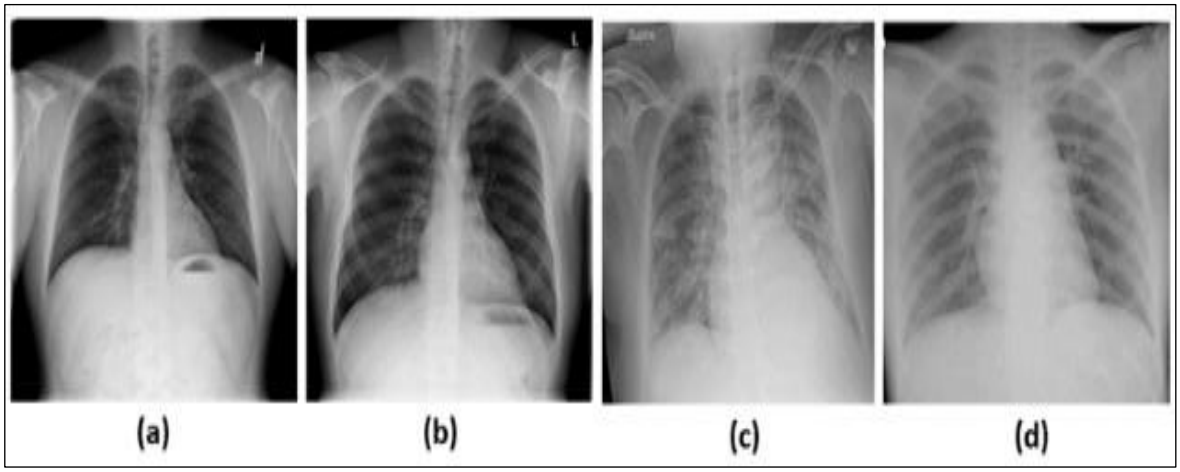


Figure 4: Normal chest x-ray (a,b) and COVID-19 chest X-ray (c, d).
Source: Authors, (2026).

Tables 2 through 6 show the accuracy of testing Aortic Enlargement, Cardiomegaly, and COVID-19 using different image sizes, 32x32, 75x75, 128x128, and 224x224, respectively. Figures 5 through 9 show the computation time for training CNN models with varying image data sizes. The larger the training data, the greater the time required for training. Tables 7 to 9 show the computation time for testing images. Table 7 presents the computation time for testing Aortic Enlargement with image sizes of 32x32, 75x75, 128x128, and 224x224. Tables 8 and 9 show the results of the computation time test for Aortic Enlargement, Cardiomegaly, and COVID-19, respectively.

The highest accuracy values for testing are obtained by the DenseNet201 model, as shown in Table 4. Using an image size of 180 x 180 results in the highest accuracy for the ResNet50 model, as shown in Table 5. Based on the Aortic Enlargement dataset test, the image size of 128 x 128 pixels using the Densenet201 model has the highest performance with an accuracy value of 0.965. The accuracy values on images measuring 224 x 224 pixels in all models are above 0.9. The InceptionV3 model has the lowest level of accuracy, as shown in Tables 2 through 6. The Cardiomegaly data has the highest accuracy value of 0.932 using the DenseNet169 model with an image size of 128 x 128 pixels. The DenseNet201 and Xception models show that the accuracy results from sizes 75 x 75, 128 x 128, and 180 x 180 have almost the same value.

DenseNet121, InceptionV3, and MobileNet models achieve the highest positioning accuracy with an image size of 180 x 180 pixels. However, the InceptionV3 model always produces the lowest accuracy compared to the other six models of the same size. The results of testing the COVID-19 dataset with the seven CNN models, using image sizes of 180 x 180 and 224 x 224, showed that each CNN model produced the highest accuracy values, with minimal differences of less than 0.005. When testing the COVID-19 dataset, it is highly suitable to use the CNN DenseNet169 model and the DenseNet201 model, as they can achieve high accuracy, especially when the desired image size exceeds 75 x 75.

Table 2: Accuracy results of CNN models using a 32x32 image size.

Model CNN	Aortic Enlargement	Cardiomegaly	Covid-19
DenseNet121	0.739	0.802	0.956
DenseNet169	0.835	0.82	0.965
DenseNet201	0.83	0.849	0.935
InceptionV3			
MobileNet	0.72	0.726	0.879
ResNet50	0.64	0.644	0.887
Xception			

Source: Authors, (2026).

Table 3: Accuracy results of CNN models using a 75 x 75 image size.

Model CNN	Aortic Enlargement	Cardiomegaly	Covid-19
DenseNet121	0.901	0.867	0.985
DenseNet169	0.929	0.891	0.982
DenseNet201	0.903	0.885	0.988
InceptionV3	0.628	0.821	0.939
MobileNet	0.821	0.829	0.957
ResNet50	0.894	0.889	0.982
Xception	0.845	0.855	0.973

Source: Authors, (2026).

Table 4: Accuracy results of CNN models using a 128 x 128 image size.

Model CNN	Aortic Enlargement	Cardiomegaly	Covid-19
DenseNet121	0.941	0.885	0.99
DenseNet169	0.936	0.932	0.99
DenseNet201	0.965	0.887	0.992
InceptionV3	0.861	0.827	0.972
MobileNet	0.932	0.882	0.988
ResNet50	0.927	0.91	0.99
Xception	0.92	0.85	0.985

Source: Authors, (2026).

Table 5: Accuracy results of CNN models using a 180 x 180 image size.

Model CNN	Aortic Enlargement	Cardiomegaly	Covid-19
DenseNet121	0.934	0.902	0.995
DenseNet169	0.948	0.899	0.996
DenseNet201	0.955	0.892	0.994
InceptionV3	0.918	0.87	0.992
MobileNet	0.934	0.9	0.989
ResNet50	0.945	0.904	0.996
Xception	0.932	0.858	0.993

Source: Authors, (2026).

Table 6: Accuracy results of **CNN** models using a 180 x 180 image size.

Model CNN	Aortic Enlargement	Cardiomegaly	Covid-19
DenseNet121	0.95	0.9	0.996
DenseNet169	0.955	0.904	0.995
DenseNet201	0.962	0.913	0.998
InceptionV3	0.924	0.779	0.989
MobileNet	0.95	0.89	0.992
ResNet50	0.932	0.91	0.996
Xception	0.936	0.884	0.993

Source: Authors, (2026).

III.1 THE CNN MODEL ON THE ACCURACY

In testing the Aortic Enlargement dataset with the ResNet50 model, the accuracy value is small compared to the MobileNet model. Although MobileNet has the fewest parameters compared to other CNN models, the accuracy value generated by depthwise separable convolution is always high when testing the entire dataset. InceptionV3 model testing performance always produces a very low accuracy value compared to other CNN models. When testing the Cardimegaly dataset, it was found that the predictions were incorrect, resulting in low accuracy rates of 0.78 and 0.79. In contrast to the Xception model, the number of parameters is slightly lower than that of InceptionV3, but the accuracy results are significantly better. The DenseNet201 model has more parameters than other DenseNets; however, its accuracy results are not always the highest compared to other DenseNets. Testing the DenseNet121 dataset with the fewest

parameters can produce the highest accuracy value compared to other DenseNet models. The three models with Dense Connection used in the test always produce higher accuracy than the other four models.

III.2 THE CNN MODEL ON TRAINING TIME

Next, the effect of differences in the CNN architecture model on training time will be determined. The test uses an X-ray image with a size of 224 x 224 pixels and does not perform augmentation techniques.

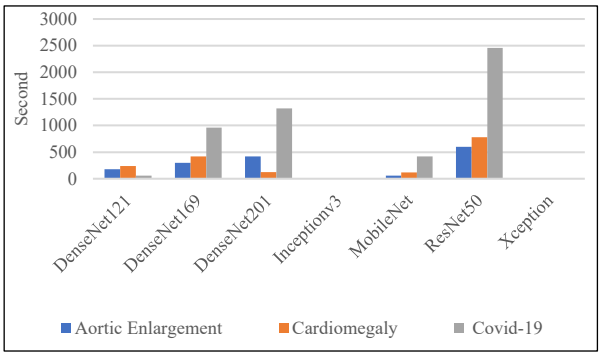


Figure 5: Computation times for training of CNN models using a 32x32 image size.
Source: Authors, (2026).

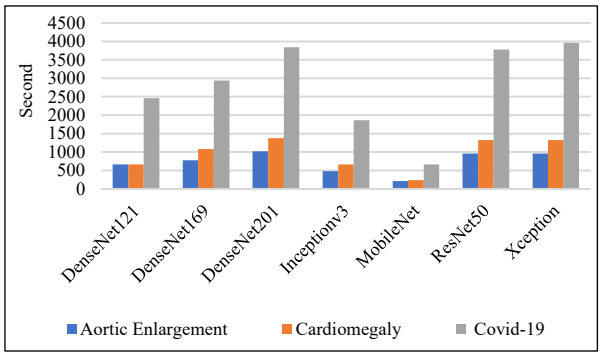


Figure 6: Computation times in seconds for training of CNN models using a 75x75.
Source: Authors, (2026).

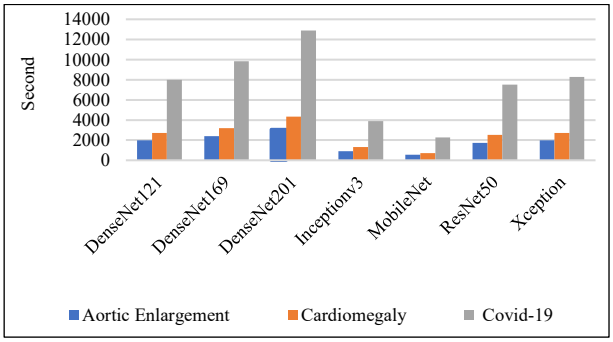


Figure 7: Computation times for training of CNN models using a 128x128 image size.
Source: Authors, (2026).

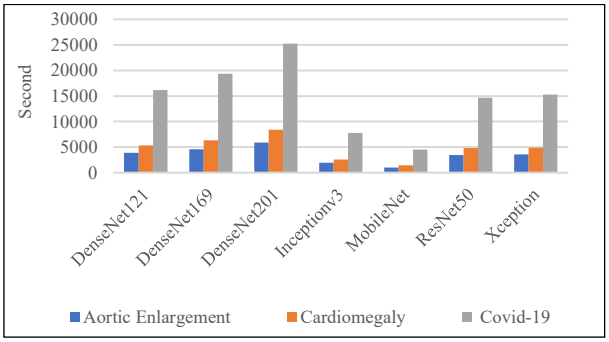


Figure 8: Computation times for training of CNN models using a 180x180 image size.
Source: Authors, (2026).

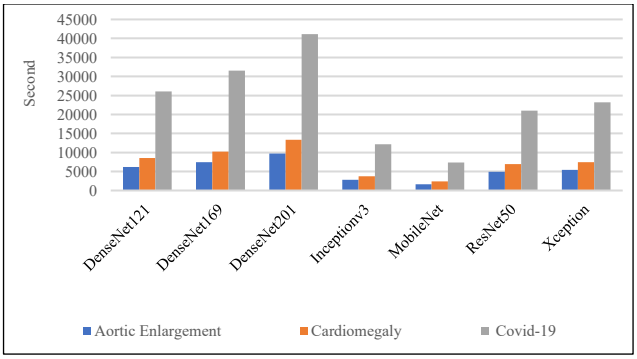


Figure 9: Computation times for training of CNN models using a 180x180 image size.
Source: Authors, (2026).

The dataset is divided into two based on the number of images: large and small. The large dataset comprises 2,061 images related to COVID-19. The small dataset consists of Aortic Enlargement (1,695 images) and Cardiomegaly (2,061 images). The two images above show the three models with dense connections, particularly the DenseNet201 model, which takes the longest to complete the training phase compared to the other models. The three Densenet models always occupy the same order according to their number of parameters, starting from DenseNet201.

III.3 DENSENET169 AND DENSENET121.

The ResNet50 and Xception models also have nearly the same training time, especially when testing datasets with a limited number of images, such as Aortic Enlargement, Cardiomegaly, and COVID-19. The two CNN models finished training with almost the same time, with a time difference of less than 9 minutes. The longest training time occurred in the COVID-19 dataset because it had the most extensive training data. The time required was 11 hours and 25 minutes using the DenseNet201 model. However, using the MobileNet model will take approximately 2 hours and 3 minutes.

III.4 THE CNN MODEL ON TEST TIME.

The influence of the CNN model on test time was analyzed using a graphical diagram with units of seconds. The test time analysis process will also be divided into two experiments: small dataset experiments and extensive dataset experiments. The test results show that the results are almost the same when calculating the training time with the DenseNet201 model, which takes the longest time to classify test data images. The MobileNet model is proven to be the fastest in completing the test step. Figures 8 and 9 show that the MobileNet and InceptionV3 models require nearly the same amount of time in the testing phase. Likewise, the Xception and ResNet50 models have almost the same testing time. After completing several tests, the research results are presented in Table 8. The model with dense connection architecture can produce high accuracy values in each dataset, especially the DenseNet201 model.

Table 7: Computation times in seconds for different sizes of the Aortic Enlargement.

Model CNN	32 x 32	75 x 75	128 x 128	180 x 180	224 x 224
DenseNet121	2	7	23	47	80
DenseNet169	2	8	28	59	94
DenseNet201	3	9	36	71	130
Inceptionv3		2	7	17	28
MobileNet	2	2	8	16	26
ResNet50	2	7	18	41	61
Xception		7	21	42	69

Source: Authors, (2026).

Table 8: Computation times in seconds for different sizes of the Cardiomegaly image.

Model CNN	32 x 32	75 x 75	128 x 128	180 x 180	224 x 224
DenseNet121	2	7	33	65	113
DenseNet169	3	11	30	76	128
DenseNet201	3	13	53	104	168
Inceptionv3		3	11	23	37
MobileNet	2	3	11	21	36
ResNet50	2	10	27	57	89
Xception		11	31	59	92

Source: Authors, (2026).

Table 9: Computation times in seconds for different sizes of COVID-19 images.

Model CNN	32 x 32	75 x 75	128 x 128	180 x 180	224 x 224
DenseNet121	5	26	62	196	324
DenseNet169	6	30	65	230	395
DenseNet201	8	39	125	302	512
Inceptionv3		10	30	68	116
MobileNet	2	9	29	65	110
ResNet50	4	26	60	170	245
Xception		30	65	185	222

Source: Authors, (2026).

The DenseNet201 model is suitable for building a CAD system that classifies X-ray photos precisely and accurately to detect six diseases in this research test. However, unlike other CNN models, the DenseNet201 model has a relatively high computational load, particularly in terms of training and testing times. The MobileNet model can be an alternative method when available resources are limited and speed in image classification is prioritized. The accuracy value is often higher than that generated by the InceptionV3, ResNet50, and Xception models. MobileNet can maximize the number of parameters while utilizing the Depthwise Separable Convolution architecture to produce high-accuracy values, all while maintaining a short computational time, including both training and testing times.

IV. CONCLUSION

Following the tests conducted in this study, the following conclusions are presented to classify X-ray images using the Convolutional Neural Network (CNN) method. The CNN model architecture used can affect the accuracy of the test results. The DenseNet201 model is most suitable for classifying normal and abnormal X-ray images, as it consistently produces the highest accuracy values in each test, specifically for Aortic Enlargement (95%), Cardiomegaly (91.8%), and COVID-19 (99.8%). The MobileNet model

requires the shortest training and testing time, especially on large datasets. MobileNet's training on the COVID-19 dataset took only 2 hours and 3 minutes, significantly faster than DenseNet201, which took 11 hours and 25 minutes.

Meanwhile, using the DenseNet201 model, the accuracy value differs by only 6%. The DenseNet201 model can be a suitable choice if available resources are limited, while maintaining high accuracy. Classification of X-ray images yields the most optimal accuracy value when using an image size of 224 x 224 pixels. However, the training will take a long time if the image size is larger.

V. AUTHOR'S CONTRIBUTION

Conceptualization: I Komang Somawirata, Fitri Utaminingrum
Methodology: Chikamune Wada, Ervin Yohannes
Investigation: Fitri Utaminingrum, Ervin Yohannes
Discussion of results: I Komang Somawirata, Fitri Utaminingrum, Chikamune Wada, Ervin Yohannes
Writing – Original Draft: I Komang Somawirata, Fitri Utaminingrum
Writing – Review and Editing: I Komang Somawirata, Fitri Utaminingrum
Resources: Fitri Utaminingrum, Ervin Yohannes
Supervision: I Komang Somawirata, Fitri Utaminingrum, Chikamune Wada
Approval of the final text: I Komang Somawirata, Fitri Utaminingrum, Chikamune Wada, Ervin Yohannes

VI. ACKNOWLEDGMENTS

The authors would like to express their gratitude to the National Institute of Technology (ITN) Malang, Brawijaya University Malang, Kyushu Institute of Technology, and State University of Surabaya for their support through research collaboration.

VII. REFERENCES

[1] Anderson PG, Tarder-Stoll H, Alpaslan M, Keathley N, Levin DL, Venkatesh S, Bartel E, Sicular S, Howell S, Lindsey RV, Jones RM. "Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays." Sci Rep. 2024 Oct 24, vol.14, no. 1, pp.25151. doi: 10.1038/s41598-024-76608-2

[2] H. H. Pham, H. Q. Nguyen, H. T. Nguyen, L. T. Le and L. Khanh, "An Accurate and Explainable Deep Learning System Improves Interobserver Agreement in the Interpretation of Chest Radiograph," in IEEE Access, vol. 10, pp. 104512-104531, 2022, doi: 10.1109/ACCESS.2022.3210468.

[3] Kulkarni, S., et al., " Artificial Intelligence, Radiology, and Tuberculosis: A Review", Academic Radiology, 2020

[4] Miyazaki A, Ikejima K, Nishio M, Yabuta M, Matsuo H, Onoue K, Matsunaga T, Nishioka E, Kono A, Yamada D, Oba K, Ishikura R, Murakami T. "Computer-aided diagnosis of chest X-ray for COVID-19 diagnosis in external validation study by radiologists with and without deep learning system". Sci Rep. 2023, vol.13, no.1,pp. 17533. doi: 10.1038/s41598-023-44818-9

[5] Kulkarni, S., et al., " Artificial Intelligence, "Radiology, and Tuberculosis: A Review. Academic Radiology", 2020

[6] Geric C, Qin ZZ, Denkinger CM, Kik SV, Marais B, Anjos A, David PM, Ahmad Khan F, Trajman A. "The rise of artificial intelligence reading of chest X-rays for enhanced TB diagnosis and elimination". Int J Tuberc Lung Dis, vol. 27, no. 5, pp.367-372, 2023, doi: 10.5588/ijtld.22.0687

[7] Marly van Assen, Mohammadreza Zandehshahvar, Hossein Maleki, Yashar Kiarashi, Timothy Arleo, Arthur E. Stillman, Peter Filev, Amir H. Davarpanah, Eugene A. Berkowitz, Stefan Tigges, Scott J. Lee, Brianna L. Vey, Ali Adibi, Carlo N. De Cecco, "COVID-19 pneumonia chest radiographic severity score: variability assessment among experienced and in-training radiologists and creation of a multireader composite score database for artificial intelligence algorithm development" , British Journal of Radiology, vol. 95, Issue 1134, 20211028, 2022, https://doi.org/10.1259/bjr.20211028

[8] Qin. Zhi Zhen, Van der Walt. Martie, Moyo. Sizulu,Ismail.Farzana, Maribe.Phaleng, Denkinger. Claudia M,Zaidi. Sarah, Barrett. Rachael, Mvusi. Lindiwe, Mkhondo. Nkateko,Zuma. Khangelani, Manda. Samuel, Koeppel. Lisa, Mthiyane. Thuli, Creswell. Jacob, "Computer-aided detection of tuberculosis from chest radiographs in a tuberculosis prevalence survey in South Africa: external validation and modelled impacts of commercially available artificial intelligence software", The Lancet Digital Health, v. 6, No. 9, pp.605-613, 2024, doi: 10.1016/S2589-7500(24)00118-3

[9] Salehi, A.W.; Khan, S.; Gupta, G.; Alabduallah, B.I.; Almjally, A.; Alsolai, H.; Siddiqui, T.; Mellit, A. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. Sustainability, vol. 15, 5930. https://doi.org/10.3390/su15075930

[10] Oloko-Oba M, Viriri S. A Systematic Review of Deep Learning Techniques for Tuberculosis Detection From Chest Radiograph. Front Med (Lausanne). 2022 doi: 10.3389/fmed.2022.830515

[11] Sarvamangala, D. R., et al., "Convolutional Neural Networks in medical image understanding: a survey", Evolutionary Intelligence, 2021

[12] Fabio Garcea, Alessio Serra, Fabrizio Lamberti, Lia Morra, "Data augmentation for medical imaging: A systematic literature review, Computers in Biology and Medicine", vol. 152, 106391, 2023,https://doi.org/10.1016/j.compbiomed.2022.106391.

[13] Goceri E. "Medical image data augmentation: techniques, comparisons and interpretations". Artif Intell Rev. pp.1-45, 2023. doi: 10.1007/s10462-023-10453-z

[14] Makhlouf, A., Maayah, M., Abughanam, N. et al. "The use of generative adversarial networks in medical image augmentation." Neural Comput & Applic 35, 24055–24068 (2023). https://doi.org/10.1007/s00521-023-09100-z

[15] Laçi, H., Sevrani, K. & Iqbal, S. "Deep learning approaches for classification tasks in medical X-ray, MRI, and ultrasound images: a scoping review" BMC Med Imaging vol. 25, pp. 156, 2025. https://doi.org/10.1186/s12880-025-01701-5

[16] Tauhidul Islam, Md. Sadman Hafiz, Jamin Rahman Jim, Md. Mohsin Kabir, M.F. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions", Healthcare Analytics, vol. 5, 2024, 100340, https://doi.org/10.1016/j.health.2024.100340.

[17] Shoukat, A.; Akbar, S.; Hassan, S.A.; Iqbal, S.; Mehmood, A.; Ilyas, Q.M. " Automatic Diagnosis of Glaucoma from Retinal Images Using Deep Learning Approach.", Diagnostics, vol. 13, 1738. https://doi.org/10.3390/diagnostics13101738

[18] Motamed S, Rogalla P, Khalvati F. "Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images", Inform Med Unlocked, 27:100779. doi: 10.1016/j.imu.2021.100779

[19] El-Dahshan EA, Bassiouni MM, Hagag A, Chakrabortty RK, Loh H, Acharya UR. "RESCOVIDTCNnet: A residual neural network-based framework for COVID-19 detection using TCN and EWT with chest X-ray images.", Expert Syst Appl, vol. 204, pp.117410, 2022 doi: 10.1016/j.eswa.2022.117410

[20] Remya, R., et al., " An Efficient Densely Connected Convolutional Networks based Detection and Classification of Breast Cancer using Mammogram Images", International Journal of Advanced Science and Technology, vol. 29, No. 4, pp. 10663–10678, 2020

[21] Talo, M., et al., " Convolutional Neural Networks for multi-class brain disease detection using MRI images, " Computerized Medical Imaging and Graphics", vol.78, pp.101673, 2019

[22] Solano-Rojas B, et al., "Low-Cost Three-Dimensional DenseNet Neural Network for Alzheimer's Disease Early Discovery", Sensors (Basel), vol.21, no.4, pp.1302, 2021

[23] Albahli, S., et al., "Recognition and detection of diabetic retinopathy using densenet-65 based faster-rcnn", Computers, Materials and Continua, vol.67, no.2, pp.1333–1351, 2021

[24] Rajaraman, S., et al., " (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ, vol.2018, no.4, pp.1–17, 2018

[25] Ghatwary, N., Ye, X., et al., " Esophageal Abnormality Detection Using DenseNet Based Faster R-CNN with Gabor Features" IEEE Access, vol.7, pp.84374–84385, 2019

[26] Bianco, S., et al., " Benchmark analysis of representative deep neural network architectures", IEEE Access, vol.6, pp.64270–64277, 2018

[27] Falconi, L. G., et al., " Transfer Learning in Breast Mammogram Abnormalities Classification with Mobilenet and Nasnet", International Conference on Systems, Signals, and Image Processing, 2019-June, pp.109–114, 2019.

[28] Velasco, J., et al., " A Smartphone-Based Skin Disease Classification Using MobileNet CNN". International Journal of Advanced Trends in Computer Science and Engineering, pp.2–8, October 2019

[29] Varshni, D., et al., " Pneumonia Detection Using CNN based Feature Extraction. Proceedings of 3rd IEEE International Conference on Electrical, Comp Velasco uter and Communication Technologies, ICECCT 2019, pp.1–7, 2019

[30] Nirthika R, Manivannan S, Ramanan A, Wang R., "Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study". Neural Comput Appl. , vol 34, no. 7, pp.5321-5347, 2022. doi: 10.1007/s00521-022-06953-8

[31] Kaifeng Lyu, Zhiyuan Li, Sanjeev Arora, "Understanding the Generalization Benefit of Normalization Layers: Sharpness Reduction", Conference on Neural Information Processing Systems (NeurIPS 2022), pp.1-20, 2022.

[32] Nguyen, H. Q., et al., "VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations, pp.1–10, 2020

[33] Wang, X., et al., " ChestX-ray: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly Supervised Classification and Localization of Common Thorax Diseases", Advances in Computer Vision and Pattern Recognition, pp.369–392, 2019

[34] Rahman, T., et al., "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-rays images. ArXiv, vol.132, pp.104319, 2020