Pneumonia Detection in Chest X-ray Images Using Convolutional Neural Network

Indah Monisa Firdiantika[1)] and Yessi Jusman[1, a)]

[1]*Department of Electrical Engineering, Faculty of Engineering, Universitas Muhamadiyah Yogyakarta, Yogyakarta, Indonesia*

Author Emails
[a)] Corresponding author: yjusman@umy.ac.id

Abstract. An enormous number of children die due to pneumonia every year worldwide. Pneumonia is a large-scale cause of death amongst children, with a high prevalence rate in South Asia and Sub-Saharan Africa. Even in a developed country like the United States, pneumonia is among the top 10 causes of death. Early detection and treatment of pneumonia can significantly scale down fatality rates among children in countries with a high prevalence. Hence, this paper presents a Convolutional Neural Network model to detect pneumonia using x-ray images. A ResNet50 and VGG-16 pre-trained model was trained to classify x-ray images into two classes, viz., pneumonia and non-pneumonia, by changing various parameters, hyperparameters, and the number of convolutional layers. The VGG-16 model has better performance than ResNet50. Meanwhile, ResNet50 achieved the testing time faster than VGG-16. The CNN model had a better performance in classifying pneumonia and non-pneumonia images.

INTRODUCTION

A massive number of children die due to pneumonia every year worldwide. An estimated 1.2 million episodes of pneumonia were reported in children up to five years of age, of which 880,000 died in 2016. One of the viruses that can cause pneumonia is SARS-CoV2 (the virus that causes COVID-19). Currently, Coronavirus disease (COVID-19) has been declared as a global pandemic by WHO, almost collapsing the healthcare systems in many countries [1], [2]. The fatality rate is surging up alarmingly throughout the world, demanding an early response to recognize and inhibit the accelerated spread of this infection. Having no specific drugs and treatments, the situation has become frightening to billions of individuals [3]. Symptoms ranging from dry cough, sore throats, and fever to organ failure, septic shock, severe pneumonia, and Acute Respiratory Distress Syndrome (ARDS) are detected from COVID-19 patients [2].

Research combining the use of images and deep learning to medical images has been used as references in this study. Chest scans such as X-rays and Computerised Tomography (CT) are prescribed to all individuals with potential pneumonia symptoms for rapid diagnosis and isolation of the infected individuals. With a severe shortage of experts, while having large similarities of COVID-19 with traditional pneumonia, an artificial intelligence (AI) assisted automated detection scheme can be a significant milestone toward a drastic reduction of testing time [4].

Proposed a deep learning framework associating residual thought and dilated convolution to diagnose and detect pneumonia in children [5]. Specifically, based on an understanding of the nature of the child pneumonia image classification task, the proposed method used the residual structure to conquer the over-fitting and the degradation problems of the depth model and utilized dilated convolution to overcome the loss of feature space information caused by the increment in the depth of the model. Presented six Convolutional Neural Network models to identify pneumonia using x-ray images to classify x-ray images into two classes, viz., pneumonia and non-pneumonia, by altering diverse parameters, hyperparameters, and the number of convolutional layers [6]. The first and second models consisted of two and three convolutional layers, respectively. The other four pre-trained models: VGG16, VGG19, ResNet50, and Inception-v3. The first and second models achieved a validation accuracy of 85.26% and

2021 2nd International Conference on Information Technology, Advanced Mechanical and Electrical Engineering (ICITAMEE)
AIP Conf. Proc. 2499, 020001-1–020001-9; https://doi.org/10.1063/5.0105004
Published by AIP Publishing. 978-0-7354-4286-3/$30.00

92.31% respectively. The accuracy of VGG16, VGG19, ResNet50 and Inception-v3 were 87.28%, 88.46%, 77.56% and 70.99% respectively.

CovXNet utilizing depthwise convolution with differing dilution rates for efficiently extracting diversified features from chest X-rays was proposed. Chest X-ray images corresponding to COVID-19 causing pneumonia and other traditional pneumonia have significant similarities. Hence, at first, a large number of chest X-rays corresponding to normal and (viral/bacterial) pneumonia patients were used to train the proposed CovXNet. It provided a very satisfactory detection performance with an accuracy of 97.4% for COVID/Normal, 96.9% for COVID/Viral pneumonia, 94.7% for COVID/Bacterial pneumonia, and 90.2% for multiclass COVID/normal/Viral/Bacterial pneumonia [4].

Four different pre-trained deep Convolutional Neural Networks (CNN): AlexNet, ResNet18, DenseNet201, and SqueezeNet, were used for transfer learning. A total of 5,247 chest X-ray images consisting of bacterial, viral, and normal were pre-processed and trained for the transfer learning-based classification task. It reported three schemes of classifications: normal vs. pneumonia, bacterial vs. viral pneumonia, and normal, bacterial, and viral pneumonia. The classification accuracy of normal and pneumonia images, bacterial and viral pneumonia images, and normal, bacterial, and viral pneumonia were 98%, 95%, and 93.3%, respectively [7].

The prowess of an algorithm was applied to detect pneumonia from chest X-ray (CXR) images. Here, an entity in the CXR image could aid determine if the patient (whose CXR was used) suffered from pneumonia or not. A model of capsules (also known as Simple CapsNet) provided results comparable to the best Deep Learning models used earlier. Subsequently, a combination of convolutions and capsules was used to obtain two models that outperformed all models previously proposed. The models—integration of convolutions with capsules (ICC) and the ensemble of convolutions with capsules (ECC)—detected pneumonia with a test accuracy of 95.33% and 95.90%, respectively [8].

An approach based on a weighted classifier was introduced, combining the weighted predictions from the state-of-the-art deep learning models such as ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3 in an optimal way. Transfer learning is a supervised learning approach in which the network predicts the result based on the quality of the dataset used. It was used to fine-tune the deep learning models to obtain higher training and validation accuracy. The model achieved a test accuracy of 98.43% and an AUC score of 99.76 on the unseen data from the Guangzhou Women and Children's Medical Center pneumonia dataset [9].

Convolution neural network method was used for binary classification pneumonia-based conversion of VGG-19, Inception_V2, and decision tree model on X-ray and CT scan images dataset, which contained 360 images. Fine-tuned version VGG-19, Inception_V2, and decision tree model demonstrated a hugely satisfactory performance with a rate of increase in training and validation accuracy (91%) other than Inception_V2 (78%) and decision tree (60%) models [10].

Four different models were developed by changing the used deep learning method; two pre-trained models, ResNet152V2 and MobileNetV2, a Convolutional Neural Network (CNN), and a Long Short-Term Memory (LSTM). The proposed models were implemented and evaluated using Python and compared with recent similar research. The results demonstrated that our proposed deep learning framework improved accuracy, precision, F1-score, recall, and Area Under the Curve (AUC) by 99.22%, 99.43%, 99.44%, 99.44%, and 99.77%, respectively [11]. In 2020, we have research in classification of Chest X Ray images using NN [12].

Based on the related research, this research aims to design a deep learning-based system with chest X-ray images to ease the experts and other health professionals to quickly and accurately determine the X-ray image diagnosis.

RESEARCH METHODS

This study used Chest X-ray images (Pneumonia) from Kaggle containing 5,216 training images and 624 testing images. Program work principles are doing pre-processing image to make all the images size uniform when used for research. The image collection was divided into training and testing images, with the dataset determined based on 5-fold Cross-Validation.

This research used a Convolutional Neural Network (CNN) with a pre-trained model (transfer learning) to make the training process faster. The pre-trained model used was ResNet-50 and VGG-16. In the training process, the model was trained in 5 and 10 epochs for each dataset. In the testing process, the data tested were 50 images for each class. After training and testing, the performance of the model (Accuracy, Error Rate, Precision, Recall, Specificity, F-Measure) was calculated. The details of the methodologies are given in the below sub-sections and the stages of work can be seen in the following flowchart in Figure 1.

Dataset

The experiment conducts on a public Kaggle dataset under the name "Chest X-Ray Images (Pneumonia) ". The dataset used was chest X-ray images from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. The entire chest X-ray imaging was carried out as part of a patient's clinical care. For chest X-ray images analysis as presented in Figure 2, all chest x-ray images were filtered out initially to control image quality by eliminating low-quality and hard-to-read images. Furthermore, the diagnosis of those images was assessed by two expert doctors before processing the artificial intelligence training system. Moreover, for elaborating the error assessment, the evaluation set was examined by the third expert.

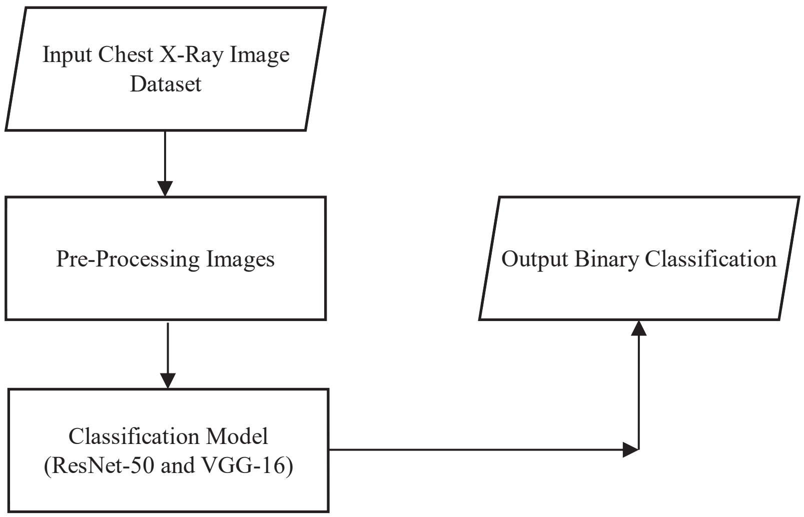

FIGURE 1. Flowchart of the proposed work.

The datasets were publicly available for other research, consisting of three folders: Train, Val, and Test. Each folder consisted of a subfolder for each category (Pneumonia and Normal) as presented in Table 1. There were 5,216 images for training data, comprising 3,875 images of the pneumonia class and 1,341 images of the normal class. In the testing data, there were 624 images, consisting of 234 images of the normal class and 390 images of the pneumonia class. Images from testing and training were combined and divided into five datasets: fold-1, fold-2, fold-3, fold-4, and fold-5, with 20% testing images and 80% training images. However, the testing data took only 50 images for each class.

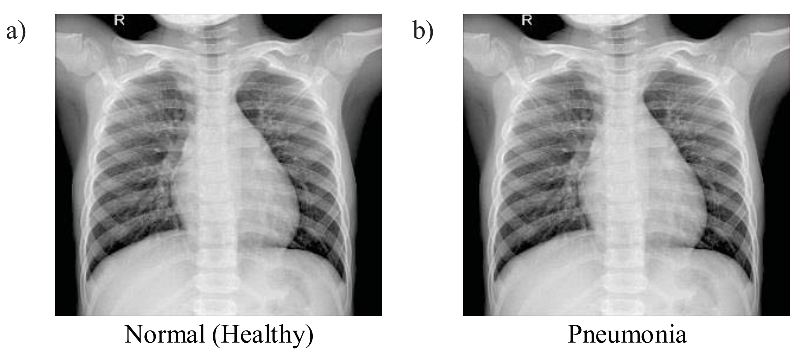

FIGURE 2. Chest X-ray of (a) is a healthy person and (b) is a person suffering from pneumonia

TABLE 1. Chest X-ray Datasets

Dataset	Number of Imagery	Information		
		Classes	Training Data	Testing Data
Dataset 1	5,840	Normal	1,260	50
		Pneumonia	3,412	50
Dataset 2	5,840	Normal	1,260	50
		Pneumonia	3,412	50
Dataset 3	5,840	Normal	1,260	50
		Pneumonia	3,412	50
Dataset 4	5,840	Normal	1,260	50
		Pneumonia	3,412	50
Dataset 5	5,840	Normal	1,260	50
		Pneumonia	3,412	50

The normal chest X-ray represents clear lungs without any areas of abnormal opacification in the image. There are mainly two types of pneumonia: bacterial and viral pneumonia. Both bacteria pneumonia and viral pneumonia were considered as one category, pneumonia infected. The dataset used in this study did not divide the case of bacterial and viral pneumonia. The result has been evaluated as pneumonia predicted and normal. The pneumonia chest X-ray depicts the area of opacity (seen as a white, gray shadow, or cloudy).

Image Labeling

For image labeling, the name of each image was based on its category: the normal image was named normal, and the pneumonia image was name pneumonia. Furthermore, those image names were then changed based on their sequences; for example, normal.1.jpg for the first normal image and pneumonia.1.jpg for the first pneumonia image, and so forth. After the labeling process, all images were divided into five datasets with 5,840 training images for each dataset, containing 1,260 normal images and 3,412 pneumonia images. The image testing used images excluded in the training image for each dataset, consisting of 315 normal images and 853 pneumonia images. For testing images, only 50 images were sampled.

Image Pre-Processing

Chest X-ray images obtained from the Guangzhou Women's and Children's Medical Center had different resolutions. Therefore, they were resized to the same resolution to accelerate and ease the computer during the learning process. The image shape input was changed to 128 x 128.

K-Fold Cross-Validation

K-Fold Cross Validation is a collection of data divided by the number of K parts/fold where each fold will be used as a test set as presented in Figure 3. For example, 80% of the dataset is used as training data, while 20% is used for testing. K-Fold Cross Validation is useful for evaluating the performance of an algorithm model by conducting experiments of K. Moreover, it is also used to improve the performance of the model and process datasets with balanced classes.

This technique is used to make model predictions and estimate how accurate a model is when implemented. In a prediction task, a model usually processes datasets known to be used in learning (training datasets) and data sets unknown or have not been studied as test data.

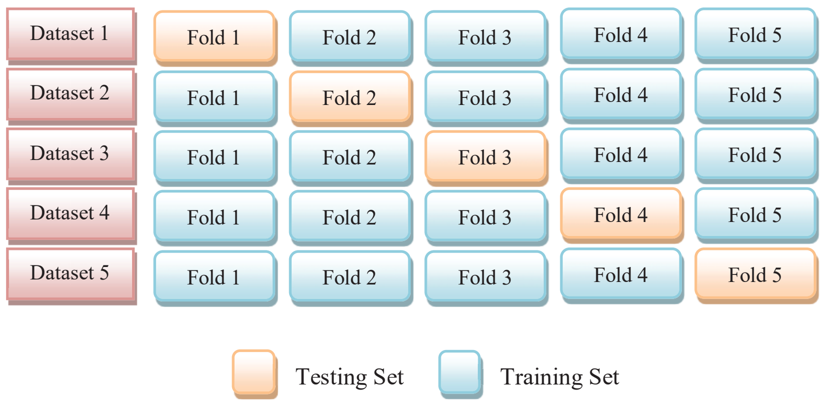

FIGURE 3. 5-Fold Cross-Validation. In the first dataset, 80% of the random datasets are used for training while the remaining 20 % will be used for testing. The experiments are repeated with a different 80% of training data and 20% of testing data.

Transfer Learning

Transfer learning is a method on the Deep Neural Network that utilizes models trained or studied before to solve other similar problems and are applied to new models with modifications to the new dataset. Transfer learning transfers the extraction features of data from pre-trained models to new models with smaller datasets as shown in Figure 4.

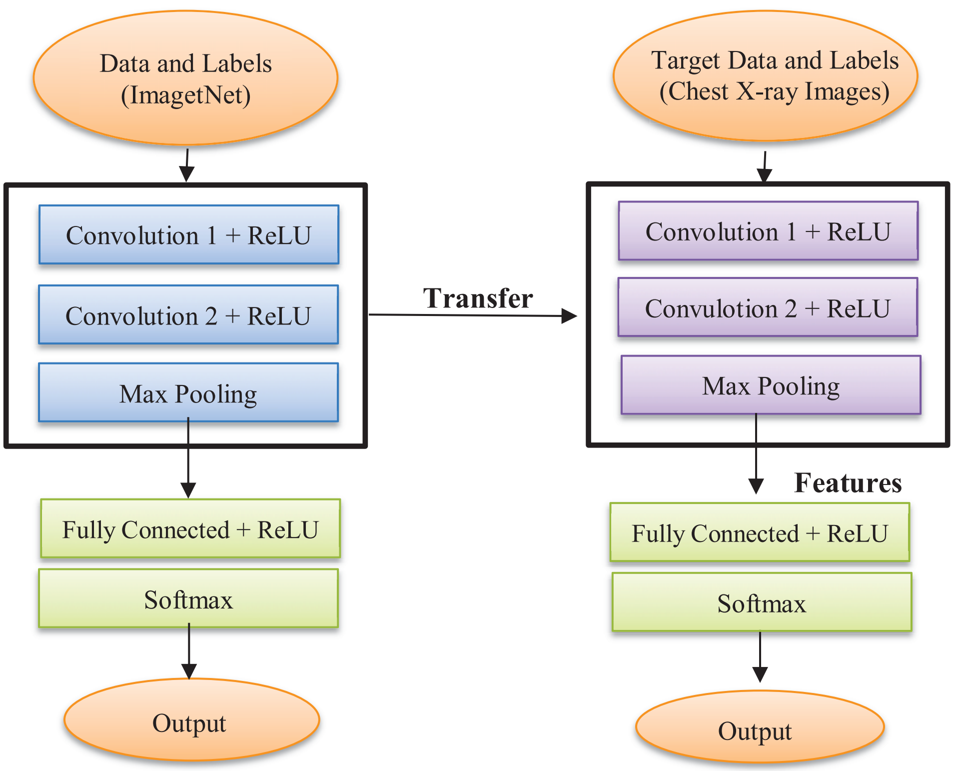

FIGURE 4. Transfer Learning Diagram. Initially, the CNN model architecture was created and trained on large datasets that produced pre-trained models. Then, at the classification stage, it was eliminated and replaced with a layer adjusted to the new task.

The architectural model used is the pre-trained ResNet50 and VGG-16 model. ResNet has many variants such as ResNet50, ResNetXt, ResNet34, etc. ResNet-50 has been trained on an ImageNet database. ResNet won the 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The ResNet pre-trained model is the Convolutional Neural Network (CNN) model integrated with image, auto-encoding, and classification. ResNet50 is a residual network consisting of 50 layers.

The VGG-16 pre-trained model is the CNN model introduced by K. Simonyan and A. Zisserman of the University of Oxford in the journal "Very Deep Convolutional Networks for Large Scale Image Recognition". The VGG-16 model has achieved 92.7% accuracy in the ImageNet dataset. The VGG-16 model is trained for a week and uses the NVIDIA Titan Black GPU. Tables 2 and 3 presents the steps in making the CNN model using transfer learning.

TABLE 2. Testing Results of ResNet-50

Iteration	Dataset	Normal		Pneumonia		Time (s)
		True (TN)	False (FP)	True (TP)	False (FN)	
Epoch 5	Dataset 1	47	3	44	6	13.2
	Dataset 2	49	1	47	3	12.6
	Dataset 3	46	4	49	1	13.3
	Dataset 4	39	11	47	3	12.3
	Dataset 5	41	9	49	1	14.1
Epoch 10	Dataset 1	46	4	48	2	14.8
	Dataset 2	47	3	47	3	14.7
	Dataset 3	50	0	48	2	12.2
	Dataset 4	43	7	45	5	13.2
	Dataset 5	44	6	43	7	13.8

TABLE 3. Testing Results of VGG-16

Iteration	Dataset	Normal		Pneumonia		Time (s)
		True (TN)	False (FP)	True (TP)	False (FN)	
Epoch 5	Dataset 1	49	1	46	4	174.03
	Dataset 2	50	0	47	3	25.94
	Dataset 3	48	2	49	1	20.89
	Dataset 4	45	5	47	3	21.14
	Dataset 5	45	5	47	3	21.14
Epoch 10	Dataset 1	50	0	40	10	161.75
	Dataset 2	49	1	48	2	79.2
	Dataset 3	48	2	48	2	23.44
	Dataset 4	46	4	43	7	27.6
	Dataset 5	50	0	42	8	19.67

RESULTS AND DISCUSSION

In the experiment, two models were trained and tested on the binary classification of Chest X-Ray Images (Pneumonia) dataset, namely ResNet-50 and VGG-16. Both models used the same technique of data pre-processing. At the training, the stage employed iterations of 5 and 10. Experimental results are shown in Table 1. The VGG-16 model achieved better performance in the normal/pneumonia binary classification task. Meanwhile, the testing time of the ResNet-50 is faster than the VGG-16 as presented in Tables 2 and 3.

After obtaining the results of the test data, then the performance evaluation value of the model was determined using the following confusion matrix as presented in Table 4. It is used to analyze the performance of the classification model after testing. Performance evaluation used to evaluate and analyzed the better performance models are Accuracy, Error Rate, Precision, Recall, Specificity, F-Measure.

TABLE 4. Chest X-ray Datasets.

	Normal Predicted	**Pneumonia Predicted**
Normal Actual	TN	FP
Pneumonia Actual	FN	TP

- TP (True Positive) = The amount of actual positive data are predicted to be positive (Pneumonia images identified as pneumonia).
- TN (True Negative) = The amount of actual negative data are predicted to be negative (Normal images identified as normal or healthy).
- FP (False Positive) = The amount of actual negative data are predicted to be positive (Normal images incorrectly identified as pneumonia).
- FN (False Negative) = The amount of actual positive data are predicted to be negative (Pneumonia images incorrectly identified as normal or healthy).

The four data types were used as a reference for evaluation metrics as follows.

1. Accuracy
   Accuracy is the ratio of correct data prediction to the overall test data. It is the most intuitive performance measure. It answers the question, "What percentage of the predicted image is correct for normal and pneumonia classes from all test data".
   $$Accuracy = (TP + TN) / (TP + FP) + (TN + FN) \tag{1}$$

2. Error Rate (ERR)
   The error rate calculates the amount of all data predicted to be incorrect of all total test images. The best error rate is 0.0, while the worst is 1.0.
   $$Error\ Rate = ((FP + FN)) / ((FP + FN + TP + TN)) \tag{2}$$

3. Precision (Positive Prediction Value)
   Precision (PREC) is a true positive prediction ratio compared to overall predicted positive results. Precision answers the question, "What percentage of the normal image of the overall predicted normal image". The best precision is 1.0, and the worst is 0.0.
   $$Precision = TP / ((TP + FP)) \tag{3}$$

4. Recall (Sensitivity)/True Positive Rate
   The recall calculates the number of positive, positive predictive images compared to the total positive and positive data. The best sensitivity value is 1.0, while the worst is 0.0.
   $$Recall = TP / ((TP + FN)) \tag{4}$$

5. Specificity
   Specificity is the truth of predicting negative compared to overall negative data. It is also often called the True Negative Rate (TNR). The best value of specificity is 1.0, while the worst is 0.0.
   $$Specificity = TN / ((TN + FP)) \tag{5}$$

6. F-Measure
   F-Measure, also known as F-Score, is a comparison of average precision and recall. F-scores are the most commonly used metric on unbalanced classification problems. If the recall and precision obtain a good value, the F-measure also obtains a good value.
   $$F - Measure = (2 \times Recall \times Precision)/(Recall + Precision) \tag{6}$$

After obtaining the results of the test data, then the performance evaluation value of the model was determined using the following confusion matrix. It is used to analyze the performance of the classification model after testing as tabulated in Tables 5 to 8. Performance evaluation used to evaluate and analyzed the better performance models are Accuracy, Error Rate, Precision, Recall, Specificity, F-Measure.

TABLE 5. Confusion Matrix Results of ResNet-50 (Epoch = 5)

Dataset	Acc	Error Rate	Precision	Recall	F-Score	Specifcity
Dataset 1	0.91	0.09	0.93	0.88	0.90	0.94
Dataset 2	**0.96**	0.04	**0.97**	0.94	**0.95**	**0.98**
Dataset 3	0.95	0.05	0.92	**0.98**	0.94	0.92
Dataset 4	0.86	0.14	0.81	0.94	0.87	0.72
Dataset 5	0.90	0.1	0.84	0.98	0.90	0.82

TABLE 6. Confusion Matrix Results of VGG-16 (Epoch = 5)

Dataset	Acc	Error Rate	Precision	Recall	F-Score	Specifcity
Dataset 1	0.95	0.05	0.97	0.92	0.94	0.92
Dataset 2	0.97	0.03	**1**	0.94	**0.97**	0.94
Dataset 3	**0.97**	0.03	0.96	**0.98**	0.96	**0.98**
Dataset 4	0.92	0.08	0.9	0.94	0.91	0.94
Dataset 5	0.92	0.08	0.9	0.94	0.91	0.94

TABLE 7. Confusion Matrix Results of ResNet-50 (Epoch = 10)

Dataset	Acc	Error Rate	Precision	Recall	F-Score	Specifcity
Dataset 1	0.94	0.06	0.92	0.96	0.93	0.92
Dataset 2	0.94	0.06	0.94	0.94	0.94	0.94
Dataset 3	**0.98**	0.02	**1**	**0.96**	**0.97**	**1**
Dataset 4	0.88	0.12	0.86	0.9	0.87	0.86
Dataset 5	0.87	0,13	0.97	0.86	0.91	0.88

TABLE 8. Confusion Matrix Results of VGG-16 (Epoch = 10)

Dataset	Acc	Error Rate	Precision	Recall	F-Score	Specifcity
Dataset 1	0.9	0.1	1	0.8	0.9	0.9
Dataset 2	**0.97**	0.03	**0.97**	**0.96**	**0.97**	**0.96**
Dataset 3	0.96	0.04	0.96	0.96	0.96	0.96
Dataset 4	0.89	0.11	0.91	0.86	0.88	0.86
Dataset 5	0.92	0.08	1	0.84	0.92	0.84

After obtaining the results of the test data, then the performance evaluation value of the model was determined using the following confusion matrix. It is used to analyze the performance of the classification model after testing. Based on Tables 5 to 8, the accuracy values are in range 0.86 to 0.98. The highest accuracy value is 0.98 in confusion matrix of ResNet-50 using epoch 10. Precision values are in range 0.81 to 1. The highest precision value is 1 in confusion matrix results of VGG-16 (Epoch = 5) and ResNet-50 (Epoch = 10). The recall values are in range 0.80 to 0.98. The highest recall value is 0.98 in confusion matrix results of ResNet-50 (Epoch = 5) and VGG-16 (Epoch = 5). Based on Tables 2 to 8, the results of Resnet in term of running time is better than VGG-16, and the results of confusion matrix of ResNet-50 (Epoch = 10) is better than VGG-16 (Epoch = 5) in term of accuracy, recall, specificity, and f-score.

CONCLUSION

This paper presented performances of classification of pneumonia and non-pneumonia images. The dataset images were used as input for Convolutional Neural Network (CNN) with the ResNet-50 and VGG-16 pre-trained model. The two different epoch nodes were used to test the CNN performance. The highest results of the ResNet-50 model were 98% of accuracy, 98% of sensitivity, and 100% of specificity, whereas the VGG-16 model is 97%, 98%, and 98% for accuracy, sensitivity, and specificity, respectively. The CNN with ResNet-50 pretrained model had a better performance in classifying pneumonia and non-pneumonia images.

ACKNOWLEDGMENT

This research is supported by Universitas Muhammadiyah Yogyakarta and a research project grant from the Ministry of Research and Technology of the Republic of Indonesia.

REFERENCES

[1] Sohrabi, C., et al., World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International journal of surgery, 2020. 76: p. 71-76.

[2] Lai, C.-C., et al., Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. International journal of antimicrobial agents, 2020. 55(3): p. 105924.

[3] Rothan, H.A. and S.N. Byrareddy, The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of autoimmunity, 2020. 109: p. 102433.

[4] Mahmud, T., M.A. Rahman, and S.A. Fattah, CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. Computers in Biology and Medicine, 2020. 122: p. 103869.

[5] Liang, G. and L. Zheng, A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Computer methods and programs in biomedicine, 2020. 187: p. 104964.

[6] Jain, R., et al., Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. Measurement, 2020. 165: p. 108046.

[7] Rahman, T., et al., Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray. Applied Sciences, 2020. 10(9): p. 3233.

[8] Mittal, A., et al., Detecting Pneumonia Using Convolutions and Dynamic Capsule Routing for Chest X-ray Images. Sensors, 2020. 20(4): p. 1068.

[9] Hashmi, M.F., et al., Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning. Diagnostics, 2020. 10(6): p. 417.

[10] Dansana, D., et al., Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm. Soft Computing, 2020.

[11] Elshennawy, N.M. and D.M. Ibrahim, Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images. Diagnostics, 2020. 10(9): p. 649.

[12] Jusman, Y., et al. (2020). "Comparison of Multi Layered Percepton and Radial Basis Function Classification Performance of Lung Cancer Data." Journal of Physics: Conference Series 1471(1): 012043.