Classification SARS-CoV-2 Disease based on CT-Scan Image Using Convolutional Neural Network

Kelvin Leonardi Kohsasih[1*], B. Herawan Hayadi[2]

[1,2]Department of Computer Science, Faculty of Engineering and Computer Science,
Universitas Potensi Utama, Indonesia

Abstract.

Purpose: Convolutional Neural Network (CNN) is one of the most popular and widely used deep learning algorithms. These algorithms are commonly used in various applications, including image processing in medical and digital forensics, speech recognition, and other academic disciplines. SARS-CoV-2 (COVID-19) is a disease that first appeared in Wuhan, China, and has symptoms similar to pneumonia. This study aims to classify the covid-19 virus by proposing a deep learning model to prevent infection rates.

Methods: The dataset used in this study is a public dataset originating from a hospital in Sao Paulo, Brazil. The data images consisted of 1252 infected with covid and 1230 data classified as non-covid but have other lung diseases. The classification method proposed in this research is a CNN model based on Resnet 50.

Result: The experimental results show that the proposed Resnet 50-based convolutional neural network model works well in classifying SARS-CoV-2 disease using CT-Scan images. Our proposed model obtains 95% accuracy, precision, recall, and f1 values on the Epoch 500.

Novelty: In this experiment, we utilized the Resnet50-based CNN model to classify the SARS-CoV-2 (COVID-19) disease using CT-Scan images and got good performance.

Keywords: Convolutional Neural Network, Resnet50, Classification, COVID-19, SARS-CoV-2

INTRODUCTION

Deep learning is an evolution of machine learning, and its adaptability to computer vision has made great strides in problem solving [1], [2]. Deep learning, in general, has proven to be an advanced technology widely used in a wide range of applications, including image processing in medical and digital forensics [3], speech recognition [4], [5] and other academic disciplines [6]. There are many studies conducted to classify or predict different illnesses using deep learning techniques, such as the prediction of cardiovascular disease [7], lung disease [8], [9], brain tumor disease [10], detection of leukemia cancer [11], and SARS-CoV-2 disease [12], [13].

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a virus that was detected in late 2019 at a traditional market in Wuhan, Hubei Province, China, when many people were experiencing pneumonia-like symptoms [14], [15]. After clinical examination, virus infection has confirmed by analyzing the patient's sputum using the polymerase chain reaction (PCR) method [14]. The virus was then given the name "COVID-19" on the World Health Organization (WHO) recommendation and was officially declared a pandemic in March 2020 [16]. In addition to using the RT-PCR test, there are several ways to diagnose SARS-CoV-2, using radiographic imaging such as x-rays or computer tomography (CT) scans of the lungs [17]. The COVID-19 pandemic has considerably impacted the global economy, from large industries such as airlines to small businesses [18], [19].

A convolutional neural network (CNN) is a deep learning model inspired by human neural networks and is the most widely used method in classification cases. Previous research by Chaniago M et al. A study was performed to detect SARS-CoV-2 on lung x-rays using a convolutional neural network (CNN) method. The built model with CNN achieved an accuracy of 97.87% [20]. According to [21], their study used x-ray images to classify SARS-CoV-2 disease. The classification process was performed using the VGG16

---

model. The results show that the model built with VGG16 achieved an accuracy of 79.58. In the following research [22], detected SARS-CoV-2 using CT images of the lungs. The detection process in this study uses the AlexNet network model to apply a convolutional neural network (CNN). The model training process was performed three times using epochs 30, 70, and 100. As a result, the model built on the AlexNet network achieved an accuracy of 90.90%.

Based on several previous studies on the prediction and classification of the SARS-CoV-2 disease, the dataset used is mostly an X-ray image of the lungs. This study aims to create a classification model using a convolutional neural network by applying the Resnet50 network model. The model built will be trained using Lung CT-Scan images from infected SARS-CoV-2 and non-Infected SARS-CoV-2 patients. The results of this research activity are expected to produce models that can predict SARS-CoV-2 infection and improve performance, such as accuracy, precision, and recall.

METHODS
The research method begins with data collection, data normalization, CNN model building, model training, model testing, and evaluating the performance of the built model. Figure 1 shows the procedure performed in this study.
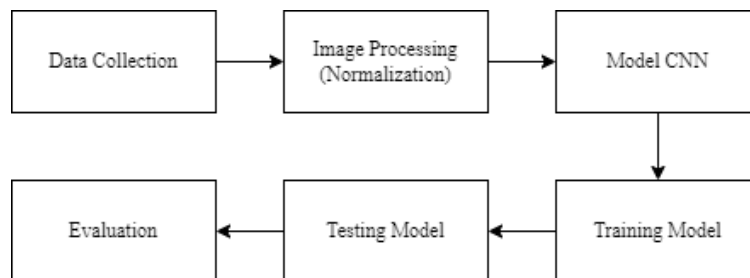

Figure 1. Flowchart research methodology

Data Collection
The dataset used in this study is a public dataset, which has been collected from actual patients in a hospital in Sao Paulo, Brazil. This dataset contains 2482 CT Scan images, consisting of 1252 infected with covid and 1230 data classified as non-covid but who presented other pneumonia diseases [23]. Figure 2 shows the data from datasets of each class used in the study.
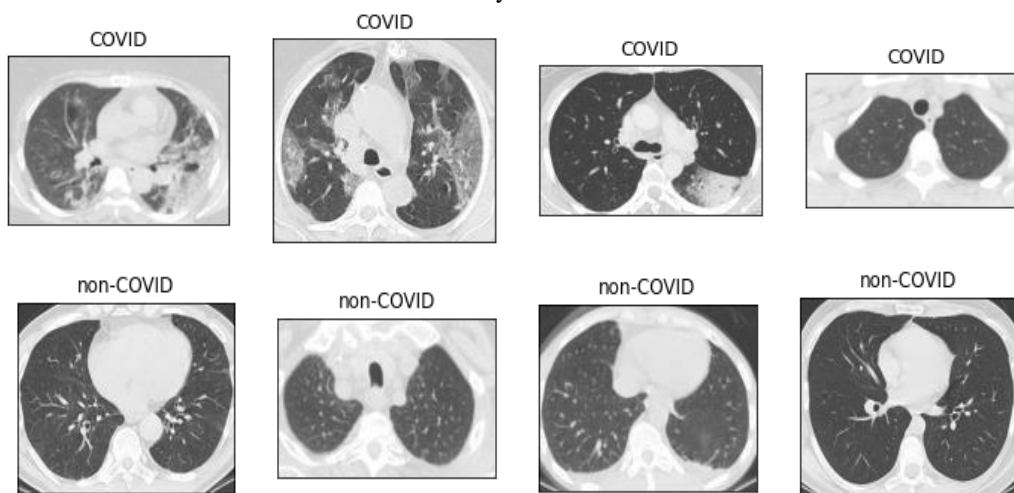

Figure 2. CT scan images of a patient infected with COVID-19 and a non-infected patient.

Image Processing (Normalization)
Normalization is typically used in the preprocessing process. In image processing, normalization changes the pixel intensity values range [24]. Normalization changes the dimension of a gray image $I: \{ \mathbb{X} \subseteq \mathbb{R}^n \} \rightarrow \{Min, .., Max\}$ with intensity values in the oven $\{Min, Max\}$ to a new idea $I_n: \{ \mathbb{X} \subseteq \mathbb{R}^n \} \rightarrow$

$\{newMin,..,newMax\}$ with intensity value in the range (new Min, New Max) [25], [26]. Linear normalization of grayscale digital images is performed by Equation (1).

$$I_n = (I - Min)\frac{newMax - newMin}{Max - Min} = newMin \tag{1}$$

For example, if the image intensity range is 60 to 190 and the desired coverage is 0 to 255, the process must subtract 40 from each pixel intensity, resulting in a range of 0 to 140. Then multiply each pixel intensity by 255/140. create a range of 0 to 255. In this study, we perform image preprocessing (normalization) by changing the range of pixel intensity values from the original size of 348x256 pixels to 64x64 pixels. Figure 3 shows an example of the data preprocessing on the CT scan image.
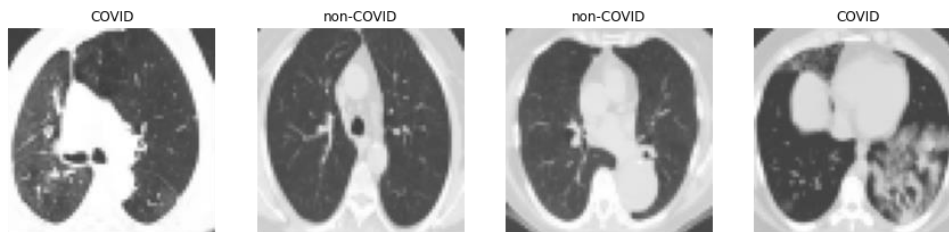

Figure 3. Example CT scan after image preprocessing (normalization)

Convolutional Neural Network

Convolutional Neural Network (CNN) is part of a deep learning algorithm developed with Multilayer Perceptron (MLP) designed to process data in two-dimensional forms [27], such as sound and images [28]. This network belongs to a deep neural network widely used for image analysis and recognition [29]. This network structure consists of several convolution layers, a pooling layer including max pooling or mean pooling and a fully connected layer [30]–[32]. Several medical applications that apply the CNN network model include brain tumor detection [33], breast cancer classification [34], skin diseases [35], and cardiovascular disease [36].

This study proposes Covid-19 disease classification to CNN using the Resnet50 architecture. Layers, output shapes, and a number of parameters in the proposed architecture are shown in Table 1.

Table 1. Layers of proposed architecture

Type	Layer	Parameter	Output Shape
Input layer	input_2	0	[(None, 64, 64, 3)]
Conv2Dx	conv2d	84	(None, 64, 64, 3)
Functional	Resnet50	23587712	(None, None, None, 2048)
GlobalAveragePooling2D	global_average_pooling2d	0	(None, 2048)
Batch Normalization	batch_normalization	8192	(None, 2048)
Dropout	dropout	0	(None, 2048)
Dense	dense	524544	(None, 256)
Batch Normalization	batch_normalization_1	1024	(None, 256)
Dropout	dropout_1	0	(None, 256)
Dense	root	514	(None, 2)

The proposed layer has a total parameter of 24,122,070, consisting of 24,064,342 trainable parameters and 57,728 untrainable parameters.

Evaluation Measures

To determine the model's performance, we evaluate the already built model. The classification model will be evaluated using evaluation criteria such as accuracy, precision, recall, and f1 score. Confusion Matrix is used to represent the performance of the classification model [37], [38]. Table 2 shows the Confusion Matrix that visualizes the model's performance [39].

Table 2. Confusion matrix

	Covid	Non-Covid
Covid	TP	FP
Non-Covid	FN	TN

Accuracy is the correct prediction ratio to the number of aggregate estimates. The formula for calculating the accuracy is shown in Equation (2).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \qquad (2)$$

Precision is the ratio of the correct positive prediction to the overall result of positive predictions. Precision is calculated using Equation (3).

$$Precision = \frac{TP}{TP+FP} \times 100 \qquad (3)$$

Recall It's the ratio between a positive prediction and positive global data. Recall is calculated using Equation (4).

$$Recall = \frac{TP}{TP+FN} \times 100 \qquad (4)$$

F1-score is a kind of balance between accuracy and recall in the system. It is the harmonized average of the precision and recovery values. F1-score is calculated using Equation (5).

$$F1 - Score = \frac{2*Precision*recall}{Precision+recall} \times 100 \qquad (5)$$

RESULT AND DISCUSSION

This section describes some of the results obtained from the experimental results. This study classifies CT scan images to help diagnose covid infection. Based on 2482 CT scan images consisting of 1252 infected with covid and 1230 data classified as non-covid, this dataset will split into 80% used as training data and 20% used as test data. The results of the study are described as follows.

In this research, we conducted several experiments at different times. In the first experiment, we run the training data for an epoch with a value of 50 and slow down the learning speed when the metric growth stops. In the first experiment, the result was an accuracy rate of 90%. in the following test, we increased the number of epochs to increase the accuracy of the model. on the 100-epoch model, we achieved 94% accuracy. When the epoch was increased again to 200 and 500, our model saw a 95% improvement in accuracy to 94%. Based on several experiments that we have done, the proposed model obtains the best accuracy on epoch 500 with 95% accuracy. Figure 4 shows a graph of the results of our proposed model training.
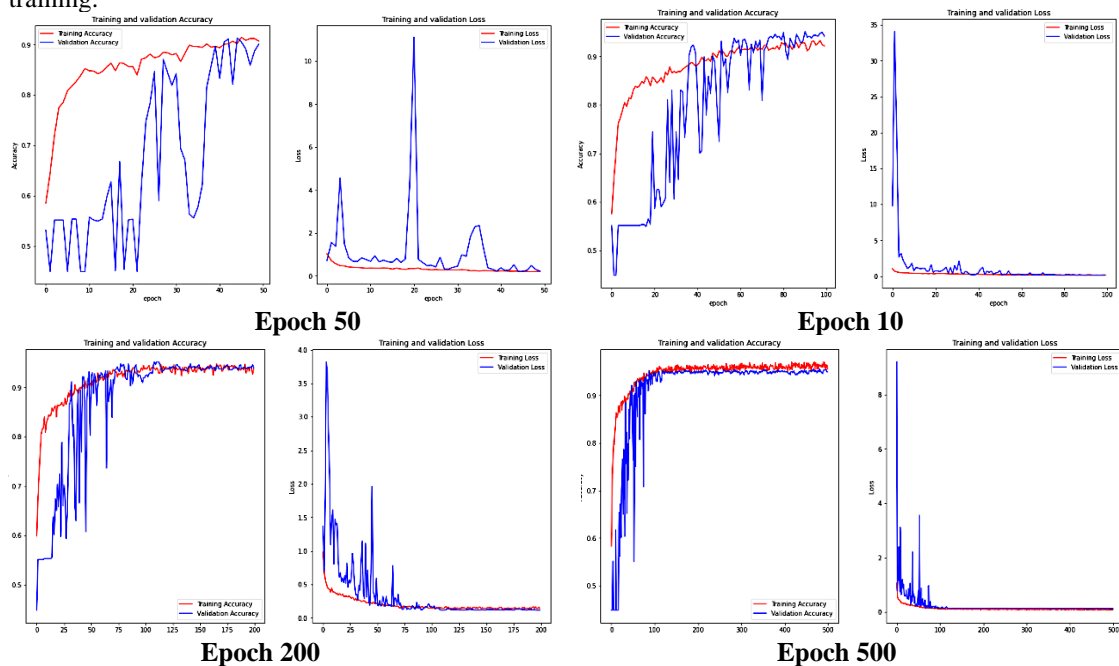

Figure 4. Model training results

In addition to assessing the model's accuracy, we also consider the confusion matrix of the trained model [40]. Figure 5 shows the performance evaluation of the Confusion matrix.

Epoch 50

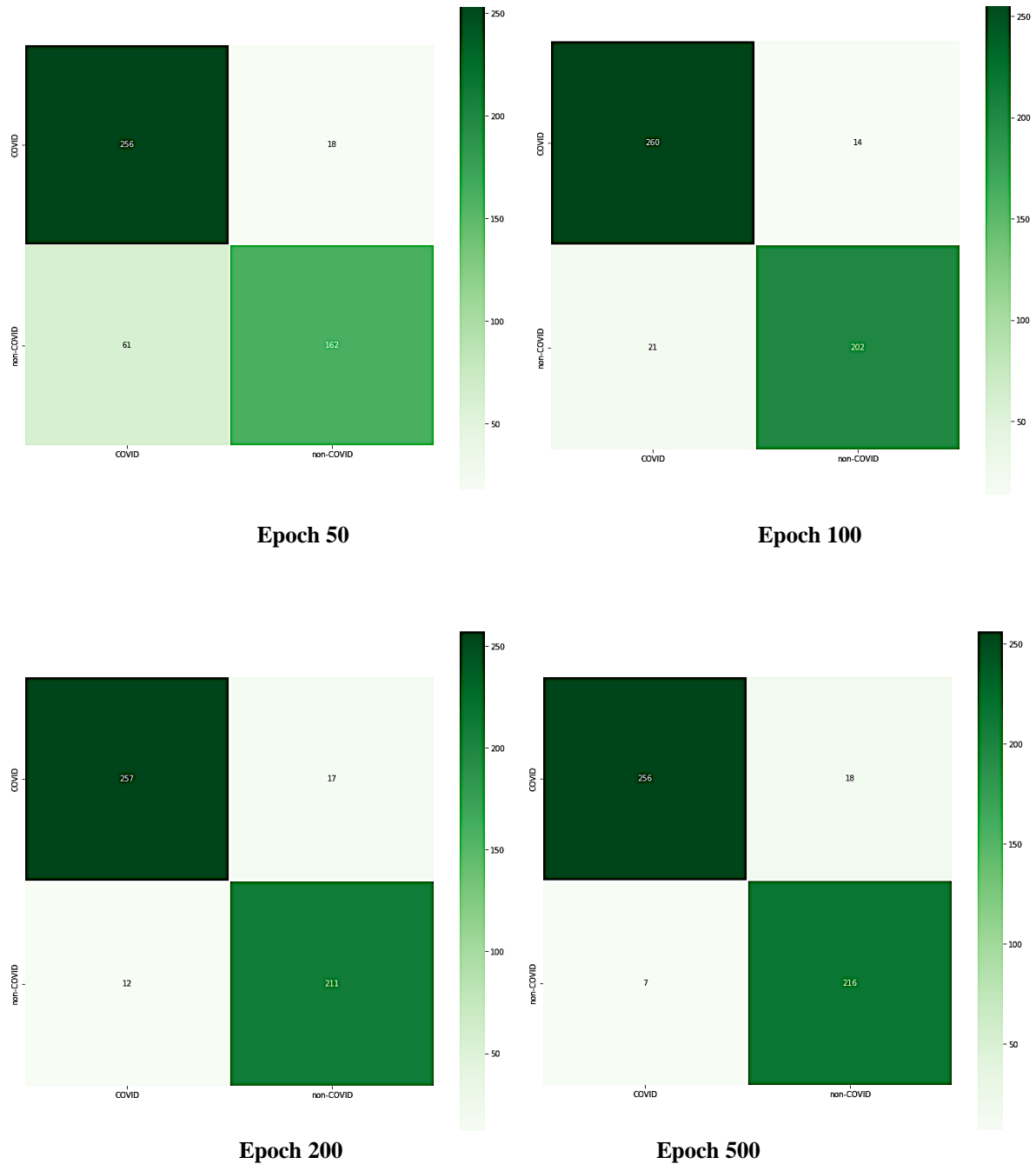**Epoch 100**

Epoch 200

Epoch 500

Figure 5. Confusion matrix evaluation

As shown in Figure 5, Epoch 500 classifies 472 correct data and 25 incorrect data when evaluating the confusion matrix. Epoch 200 has categorized 468 accurate data and 29 inaccurate dates. Its result shows the improved model performance from Epoch 100, correctly classified data 462, and classified data 418 from Epoch 50. Table 3. shows the results of the evaluation of the model's performance at various epochs.

Table 3. Results performance evaluation

Epoch	Accuracy	Precision	Recall	F1-Score
50	90%	90%	91%	90%
100	9 4%	94%	94%	94%
200	95%	95%	94%	95%
500	95%	95%	95%	95%

The findings of this study indicate that our proposed model can outperform several previous studies in terms of performance evaluation, with 95% precision, accuracy, memory, and f1-score values. Compared with research [22], using the Alexnet model to classify SARS-CoV-2 only achieved an accuracy of 90.90%, and the study [21] used the VGG16 model and achieved an accuracy of 79.58%.

Based on the confusion matrix information, it shows that the performance of the proposed model is improved and can classify SARS-CoV-2 CT scan images correctly. However, this study still has limitations where the model is only trained using images. Therefore, future research can make predictions using other data and improve performance with other models.

CONCLUSION

This paper proposes a method for classifying SARS-CoV-2 (COVID-19) disease on lung CT images using a convolutional neural network model based on Resnet50. This study consists of data collection, image processing (normalization), network structure development, and model performance testing. Based on the results of our experiment, we found that the accuracy, f1-score, precision and recall of the SARS-CoV-2 classification can reach 95%. Its shows that the Resnet50-based CNN model can be used to classify SARS-CoV-2 diseases and work well.

REFERENCES

[1] V. Maeda-Gutiérrez *et al.*, "Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases," *Appl. Sci.*, vol. 10, no. 4, pp. 1–15, 2020.

[2] H. Alaskar and T. Saba, "Machine Learning and Deep Learning: A Comparative Review," in *Proc. Integr. Intell. Enable Netw. Comput.*, 2021, pp. 143–150.

[3] B. Diallo, T. Urruty, P. Bourdon, and C. Fernandez-Maloigne, "Robust Forgery Detection for Compressed Images Using CNN Supervision," *Forensic Sci. Int. Reports*, vol. 2, p. 100112, 2020.

[4] A. Alsobhani, H. M. A. ALabboodi, and H. Mahdi, "Speech Recognition Using Convolution Deep Neural Networks," *J. Phys. Conf. Ser.*, vol. 1973, no. 1, p. 12166, 2021.

[5] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019.

[6] A. A. Reshi *et al.*, "An Efficient CNN Model for COVID-19 Disease Detection Based on X-Ray Image Classification," *Complexity*, vol. 2021, p. 6621607, 2021.

[7] S. N. Pasha, D. Ramesh, S. Mohmmad, A. Harshavardhan, and Shabana, "Cardiovascular Disease Prediction Using Deep Learning Techniques," in *IOP Conf. Ser.: Mater. Sci. Eng.*, 2020, vol. 981, no. 2, p. 22006.

[8] N. Zakaria, F. Mohamed, R. Abdelghani, and K. Sundaraj, "VGG16, ResNet-50, and GoogLeNet Deep Learning Architecture for Breathing Sound Classification: A Comparative Study," in *2021 Int. Conf. Artif. Intell. Cyber Secur. Syst. Priv. (AI-CSP)*, 2021, pp. 1–6.

[9] D. M. Midyanti, S. Bahri, and R. Hidayati, "Diagnosis of Lung Disease Using Learning Vector Quantization 3 (LVQ3)," *Sci. J. Informatics*, vol. 7, no. 2, pp. 173–179, 2020.

[10] D. C. Febrianto, I. Soesanti, and H. A. Nugroho, "Convolutional Neural Network for Brain Tumor Detection," in *IOP Conf. Ser.: Mater. Sci. Eng.*, 2020, vol. 771, no. 1, p. 12031.

[11] M. Bukhari, S. Yasmin, S. Sammad, and A. A. Abd El-Latif, "A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning," *Math. Probl. Eng.*, vol. 2022, p. 2801227, 2022.

[12] M. Ghaderzadeh and F. Asadi, "Deep Learning in the Detection and Diagnosis of COVID-19 Using Radiology Modalities: A Systematic Review," *J. Healthc. Eng.*, vol. 2021, p. 6677314, 2021.

[13] T. B. Alakus and I. Turkoglu, "Comparison of Deep Learning Approaches to Predict COVID-19 Infection," *Chaos, Solitons & Fractals*, vol. 140, p. 110120, 2020.

[14] N. Chen *et al.*, "Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: a Descriptive Study," *Lancet*, vol. 395, no. 10223, pp. 507–513, 2020.

[15]   F. Rustam *et al.*, "COVID-19 Future Forecasting Using Supervised Machine Learning Models," *IEEE Access*, vol. 8, pp. 101489–101499, 2020.

[16]   L. Kranthi Kumar and P. J. A. Alphonse, "COVID-19 Disease Diagnosis with Light-Weight CNN Using Modified MFCC and Enhanced GFCC from Human Respiratory Sounds," *Eur. Phys. J. Spec. Top.*, pp. 1–18, 2022.

[17]   T. Ai *et al.*, "Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases.," *Radiology*, vol. 296, no. 2, pp. E32–E40, 2020.

[18]   V. Puspita *et al.*, "Optimization of Digital Marketing Integration Through The Role of The Millenial Generation as MSMEs Effort To Survive During The Covid-19 Pandemic," in *2021 9th Int. Conf. Cyber IT Serv. Manag. (CITSM)*, 2021, pp. 1–5.

[19]   B. H. Hayadi, J. Kim, K. Hulliyah, and H. T. Sukmana, "Predicting Airline Passenger Satisfaction with Classification Algorithms," *Int. J. Inform. Inf. Syst.,* vol. 4, no. 1, pp. 82–94, 2021.

[20]   M. Chaniago, A. A. Sugiharto, Q. D. Khatulistiwa, Z. Sari, and A. Eko, "Covid-19 Detection Using Convolutional Neural Networks (CNN) Classification Algorithm," *J. Rekayasa Sist. dan Teknol. Inf.*, vol. 5, no. 158, pp. 190–197, 2022.

[21]   R. Anand, V. Sowmya, Vijaykrishnamenon, E. A. Gopalakrishnan, and K. P. Soman, "Modified Vgg Deep Learning Architecture For Covid-19 Classification Using Bio-Medical Images," in *IOP Conf. Ser.: Mater. Sci. Eng.*, 2021, vol. 1084, no. 1, p. 12001.

[22]   T. Wang, Y. Zhao, L. Zhu, G. Liu, Z. Ma, and J. Zheng, "Lung CT Image Aided Detection COVID-19 based on Alexnet Network," in *2020 5th Int. Conf. Commun. Image Signal Process. (CCISP)*, 2020, pp. 199–203.

[23]   E. Soares, P. Angelov, S. Biaso, M. H. Froes, and K. Abe, "SARS-CoV-2 CT-scan Dataset : A Large Dataset of Real Patients CT Scans for SARS-CoV-2 Identification," *medRxiv*, pp. 1–8, 2020.

[24]   M. B. Patel, S. M. Parikh, and A. R. Patel, "Global normalization for fingerprint image enhancement," in *Int. Conf. Comput. Vis. Bio Inspired Comput.g*, 2020, pp. 1059–1066.

[25]   L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization Techniques in Training DNNs : Methodology , Analysis and Application," *arXiv*, pp. 1–20, 2020.

[26]   R. C. Gonzales and R. E. Woods, *Digital Image Processing*, Third Edit. Prentice Hall, 2017.

[27]   A. K. Nugroho, I. Permadi, and M. Faturrahim, "Improvement of Image Quality Using Convolutional Neural Networks Method," *Sci. J. Informatics*, vol. 9, no. 1, pp. 95–103, 2022.

[28]   D. R. Hartawan, T. W. Purboyo, and C. Setianingsih, "Disaster Victims Detection System Using Convolutional Neural Network (CNN) Method," in *2019 IEEE Int. Conf. Ind. 4.0 Artif. Intell. Commun. Technol. (IAICT)*, 2019, pp. 105–111.

[29]   T. Purwaningsih, I. A. Anjani, and P. B. Utami, "Convolutional Neural Networks Implementation for Chili Classification," in *2018 Int. Symp. Adv. Intell. Inform. (SAIN)*, 2018, pp. 190–194.

[30]   A. Saxena, S. P. Singh, and V. H. Gaidhane, "A Deep Learning Approach for the Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks," *arXiv*, 2022.

[31]   M. Lin, Q. Chen, and S. Yan, "Network In Network," *arXiv*, 2013.

[32]   N. F. Nissa, A. Janiati, N. Cahya, and P. Astuti, "Application of Deep Learning Using Convolutional Neural Network (CNN) Method for Women ' s Skin Classification," *Sci. J. Informatics*, vol. 8, no. 1, pp. 144–153, 2021.

[33]   A. Chattopadhyay and M. Maitra, "MRI-based Brain Tumour Image Detection Using CNN Based Deep Learning Method," *Neurosci. Informatics*, vol. 2, no. 4, p. 100060, 2022.

[34]   N. Saffari *et al.*, "Fully Automated Breast Density Segmentation and Classification Using Deep Learning.," *Diagnostics (Basel, Switzerland)*, vol. 10, no. 11, Nov. 2020.

[35]   Y. N. Fu'adah, N. K. C. Pratiwi, M. A. Pramudito, and N. Ibrahim, "Convolutional Neural Network (CNN) for Automatic Skin Cancer Classification System," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 982, no. 1, 2020, p. 12005.

[36]   J. D. Dormer, M. Halicek, L. Ma, C. M. Reilly, E. Schreibmann, and B. Fei, "Convolutional Neural Networks for the Detection of Diseased Hearts Using CT Images and Left Atrium Patches.," in *Proc. SPIE--Int. Soc. Opt. Eng.*, 2018, vol. 10575.

[37]   B. Korbar *et al.*, "Deep Learning for Classification of Colorectal Polyps on Whole-slide Images," *J. Pathol. Inform.*, vol. 8, no. 1, p. 30, 2017.

[38]   T. L. Nikmah, M. Z. Ammar, Y. R. Allatif, R. M. P. Husna, P. A. Kurniasari, and A. S. Bahri, "Comparison of LSTM , SVM , and Naive Bayes for Classifying Sexual Harassment Tweets," *J. Soft Comput. Explor.*, vol. 3dend, no. 2, pp. 131–137, 2022.

[39]   A. Çı, M. Yıldırı, and Y. Eroğlu, "Traitement du Signal Classification of Pneumonia Cell Images Using Improved ResNet50 Model," *Int. Inf. Eng. Technol. Assoc. (IIETA)*, vol. 38, no. 1, pp. 165–

173, 2021.

[40]   C. A. Sari, W. Shinta, and H. Rahmalan, "A Combination of K-Means and Fuzzy C-Means for Brain Tumor Identification," *Sci. J. Informatics*, vol. 8, no. 1, pp. 76–83, 2021.