



Image Classification of Wounds in Diabetes Mellitus Patients Using Convolutional Neural Network (CNN) Algorithm

Sandy Andika Maulana^{1*}, Hermawan Syahputra²

^{1,2}Computer Science, Medan State University
samsandi134@gmail.com^{1*}, hermawansyahputra@unimed.ac.id²

Abstract

Diabetes Mellitus (DM) is a chronic disease that can lead to serious complications such as chronic wounds that are difficult to heal and may result in amputation. Early detection of wound types is essential to prevent worsening conditions. This study aims to develop a wound image classification system for diabetic patients using the Convolutional Neural Network (CNN) algorithm with the MobileNetV2 architecture. The dataset consists of seven classes: diabetic wounds (gangrene, blister, pressure), common wounds (abrasion, contusion, burn), and normal skin. The data were obtained from secondary sources (Roboflow) and primary sources from Rumah Sakit Umum Madani Medan. The research stages include data acquisition, augmentation, preprocessing, model training, performance evaluation, and implementation into a web-based application. Evaluation results show that the model achieved a training accuracy of 98% and a validation accuracy of 81%, along with high precision, recall, and F1-score values across most classes. The system is implemented as a web application to assist in early wound type detection and provide initial treatment recommendations, thereby supporting efforts to prevent complications in diabetic patients.

Keywords: *Diabetes Mellitus, Wound Classification, Convolutional Neural Network, MobileNetV2, Web Application.*

1. Introduction

Diabetes mellitus (DM) is classified as a chronic disease that is often referred to as a silent killer because it usually shows no symptoms in its early stages, which makes it difficult to detect and often leads to the condition going undiagnosed until more severe and life-threatening complications arise. These complications can include serious vascular problems such as angina pectoris, myocardial infarction, and stroke, which significantly increase the risk of morbidity and mortality in patients. Essentially, diabetes mellitus (DM) is characterized by hyperglycemia, a condition where blood glucose levels are elevated due to the inability of body cells to respond effectively to the hormone insulin, resulting in impaired glucose regulation and long-term damage to various organs and tissues if not properly managed [1].

According to data from the International Diabetes Federation (IDF), the number of diabetes cases worldwide continues to increase every year, reaching a total of 537 million cases in 2021. The countries with the highest number of diabetes cases are China with 140.87 million, followed by India with 74.19 million, Pakistan with 32.96 million, and the United States with 32.22 million cases, while Indonesia ranks fifth with a total of 19.47 million cases [2]. Globally, diabetes mellitus not only contributes to the rise in premature mortality rates but also stands as one of the leading causes of morbidity due to complications such as blindness, cardiovascular disease, and kidney failure.

One of the technologies that can be utilized in this case is artificial intelligence (AI). Artificial intelligence refers to systems developed to mimic or even surpass human intellectual capabilities in various fields. AI is characterized by its ability to adapt, make decisions, think cognitively, and continuously learn, making it increasingly similar to how the human brain works. This technology was initially created to support various human tasks, but over time, AI has evolved into a tool that not only assists but can also replace certain tasks that were previously performed only by humans. For example, in the medical industry, AI can be applied to classify diseases and wounds with high accuracy, thereby reducing dependence on medical personnel. Thus, the ability of AI to learn and evolve has a significant impact across multiple sectors, enhancing efficiency and precision in many aspects of human work.

Currently, machine learning methods, particularly deep learning, are increasingly being applied across various fields. The popularity of these methods is driven by the growing availability of annotated data, which allows systems to automatically extract features without the need for rules explicitly defined by humans [3]. In the past decade, with the rapid advancement of deep learning, researchers have developed various techniques to enhance object detection capabilities. These methods not only assist in object identification but also in related tasks

such as classification, localization, and segmentation by leveraging more complex and powerful deep learning models. These developments have opened up numerous opportunities for applications across various fields, including medicine and automotive industries.

2. Literature Review

In a study conducted by [4] entitled “Application of Convolutional Neural Network for Classifying the Severity of Diabetic Retinopathy in Diabetes Mellitus Patients” satisfactory results were achieved with an accuracy of 91.10%. This indicates that the Convolutional Neural Network algorithm has performed effectively in detecting diabetes based on eye images.

Diabetes Mellitus (DM) is one of the most common types of chronic diseases and has long-term impacts. This condition is characterized by an increase in blood sugar levels, known as hyperglycemia. Such a state occurs as a result of insufficient insulin supply required by the body or the body’s inability to effectively produce insulin. Consequently, the body fails to properly regulate blood sugar levels, which can lead to various serious health complications if not managed appropriately [5]. One of the most common complications experienced by diabetes mellitus patients is diabetic foot ulcers, a condition in which the health and function of the feet are impaired as a result of long-term complications of diabetes [6].

This wound condition makes patients more vulnerable to infections or ulcers that are difficult to heal and may even lead to amputation if not properly treated. Common symptoms of diabetic wounds include persistent itching, the appearance of calluses in certain areas, the development of red rashes, the presence of foul odors, and swelling in the affected area. This condition often arises as a result of diabetes complications triggered by a drastic drop in blood sugar levels within a short period of time, which affects the wound healing process and increases the risk of infection [7].

Computer Vision has become increasingly important and effective in recent years due to its wide range of applications across various fields, such as smart surveillance and monitoring, healthcare and medicine, sports and recreation, robotics, drones, and self-driving cars. Various visual recognition tasks, including image classification, object localization, and detection, form the core of this technological development. The utilization of Computer Vision has delivered optimal performance in carrying out visual recognition tasks while also supporting intelligent systems in diverse applications [8]. Image classification is one of the key tasks in the field of Computer Vision, aimed at identifying and categorizing objects or patterns in digital images into specific classes. In recent years, deep learning-based methods, particularly Convolutional Neural Networks (CNN), have become the dominant approach in addressing image classification problems due to their ability to automatically extract important features from complex visual data

Convolutional Neural Networks (CNN) are a type of artificial neural network specifically designed to process image data. CNNs work by receiving an input in the form of an image and then, through a series of convolution and pooling processes, the model learns to extract important features from the image [9]. CNN is essentially an extension of Artificial Neural Networks (ANN) specifically designed to handle spatial data such as images. This neural network-based technique is widely used in artificial intelligence approaches due to its superior ability to solve various complex problems, particularly those involving pattern recognition. The performance produced by this method has proven to be highly satisfactory, making it a primary choice in the development of machine learning and computer vision-based solutions [10]

3. Research Methods

This study utilizes secondary data sources with the aim of producing a dataset of tested quality, previously used in research, and meeting the required standards of accuracy and variation for model development. The use of validated secondary data ensures that the information collected is relevant to the needs of diabetic wound classification, thereby enhancing the effectiveness and accuracy of the developed model. With this approach, the research is expected to make a significant contribution to the development of AI-based diabetic wound classification systems that are more reliable and applicable in the medical field.

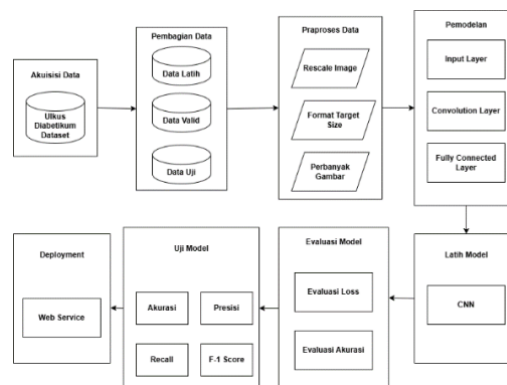



Fig. 1: Research Stages

3.1. Data Collecting

In this study, the wound image dataset was collected through secondary data obtained from the online repository Roboflow. The dataset consists of six labels, including three categories of general wounds—abrasions, bruises, and burns—and three categories of diabetic wounds—gangrene, blisters, and pressure ulcers. The secondary samples were sourced from the Roboflow website and the UCI Machine Learning Repository, both of which are open-source. The dataset used in this study comprises wound images collected through secondary data methods. It consists of six classes: three types of general wounds (abrasions, bruises, burns), three types of diabetes-related wounds (gangrene, blisters, pressure ulcers), and one additional class representing normal skin. Each class contains 200 images, all of which were obtained from secondary data sources.

Table 1: Sample Data

Type	Diabetic Wound Images	Type	Common Wound Images
Gangren Wound		Abrasions Wound	
Blister Wound		Bruises Wound	
Decubitus Wound		Burning Wound	

3.2. Split Data

Before splitting the data, a preprocessing stage was carried out on the dataset. The data collected from various sources was processed to improve quality and consistency, ensuring it was ready for use in machine learning model development. The first step in preprocessing was checking for duplicate data, which was done manually. Each image was examined one by one to identify visual similarities or repeated entries. Images detected as duplicates, whether in terms of appearance or file structure, were removed so that only one copy was retained in the dataset. This process ensured that the dataset used was clean, efficient, and free from redundancy that could affect the model's accuracy and generalization during training.

Afterward, adjustments were made to balance the number of samples across all labels. The dataset was then divided into two main parts: training and validation data. A total of 80% of the data was allocated for training, while the remaining 20% was set aside as a validation set to evaluate the final performance of the model on entirely unseen data.

3.3. Model Building

The modeling stage aims to produce a model capable of performing tasks optimally according to the required objectives. Therefore, the constructed model must meet various criteria such as functionality, behavior, and performance aligned with the system's information domain. To achieve this, an in-depth analysis of system requirements was conducted as the foundation for model design.

After the preprocessing stage was completed, the CNN architecture was designed with several interconnected layers. The initial layer applied a Convolutional Layer to extract essential features from the input images using convolution filters. This was followed by a Pooling Layer, which reduced the dimensionality of the extracted features to accelerate computation and prevent overfitting. The features obtained from pooling were then transformed into a one-dimensional vector through the Flattening Layer before being passed into the Fully Connected Layer. This layer performed the classification task based on the extracted features.

In this study, the activation function used was ReLU (Rectified Linear Unit) for the hidden layers and Softmax for the output layer, which generated probability distributions across the wound classes. Once the model architecture was finalized, the training process was conducted using the dataset divided into training, validation, and testing sets. Model training was optimized using the Adam Optimizer, combined with the Cross-Entropy Loss function to measure prediction errors. During training, parameters such as epochs, batch size, and learning rate were adjusted to achieve the best possible performance of the model.

3.4. Model Evaluation

After the model was successfully built, the next step was to evaluate its performance to ensure it could effectively perform the intended task. This evaluation was carried out by monitoring the accuracy and loss graphs at each epoch during training. The accuracy graph provided an overview of how well the model recognized patterns in the data, while the loss graph indicated the extent to which the model's errors decreased throughout the training process. By analyzing these graphs, it could be determined whether the model had learned effectively or required further adjustments, such as modifications to training parameters or changes to the model architecture.

3.5. Model Test

After the training stage was completed and the evaluation results indicated satisfactory performance, the model was saved as a file ready for testing. At this stage, the model was tested using unseen data to measure its ability to generalize patterns learned from the training data to new data. The testing process was carried out using various evaluation metrics such as accuracy, precision, recall, and F1-score. Accuracy was used to determine how well the model predicted correctly out of the total predictions made. Precision measured how accurate the model was in predicting positive categories, while recall indicated the model's ability to identify all positive categories in the data. The F1-score, as a combination of precision and recall, provided a balanced assessment between the two. Through this series of evaluations, a comprehensive understanding of the model's strengths and weaknesses was obtained, allowing a decision to be made on whether the model was ready for deployment or required further improvements.

4. Results and Discussion

4.1 Data Selection

This dataset was used as secondary data to support the training process of the wound classification model using the Convolutional Neural Network (CNN) algorithm. The data collection process was carried out with careful attention to label consistency, image quality, and relevance to the research objectives. All collected images were then processed during the preprocessing stage to ensure data consistency and quality before entering the model training phase. The results of the secondary data acquisition are presented below.

Table 2: Collecting Data

Label	Source Roboflow	Source UCI Repository	Amount Images
Abrasion Wounds	160	40	200
Bruises Wounds	135	65	200
Burning Wounds	190	10	200
Decubitus Wounds	167	33	200
Blister Wounds	173	27	200
Gangren Wounds	180	20	200
Total	1005	195	1200

4.2 Praprocessing Image

At this stage, all image data collected through crawling from various sources underwent preprocessing to ensure consistency in format and image quality before being used in the model training phase. The steps included pixel value normalization by rescaling the range from 0–255 to 0–1. This normalization was important to make the training process more stable and efficient. In addition, all images were resized to 224x224 pixels to match the input layer requirements of the pretrained MobileNetV2 architecture used in this study. A manual check for duplicate images was also carried out by comparing visuals between images and removing identical data to avoid redundancy that could affect the training results.

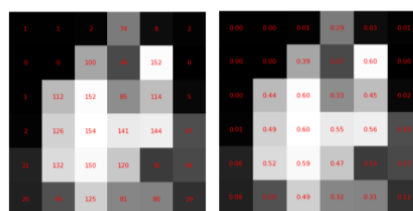


Fig. 2: Before and After Rescale

4.2.1 Image Augmentation

The original image data from each wound class was already balanced; however, to increase variability, data augmentation techniques were applied to expand and enrich the diversity of training images. The augmentation techniques used included random rotation, horizontal and vertical flipping, as well as other transformations such as zooming and translation provided by the ImageDataGenerator module. Augmentation was performed in real time during the training process, allowing the model to learn from various image variations without the need to store the augmented results as new files. The main purpose of this augmentation was to improve the model's generalization ability and reduce the risk of overfitting, particularly when working with a relatively small dataset.

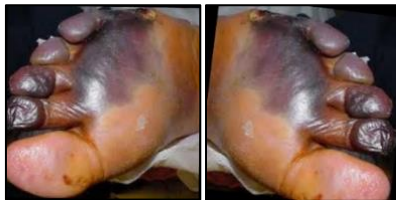


Figure 3 Horizontal Flip

With horizontal flipping, the model does not rely heavily on the position of an object on a particular side, enabling it to recognize objects with different left–right orientations. This feature is particularly important in image classification, especially when objects may appear with horizontal symmetry, such as wounds on the body, the face, or the shape of a hand.



Figure 4 Vertical Flip

When vertical flipping is applied, the top row becomes the bottom row, the second row becomes the second from the bottom, and so on. This augmentation is highly useful for increasing data variation when objects in images are not consistently positioned at the top or bottom.

4.2.2 Split Data

The collected dataset was divided into two main subsets, namely training data and validation data, with a ratio of 80:20. This ratio is considered ideal as it maintains a trade-off between bias and variance, ensuring that the model does not suffer from underfitting or overfitting. Moreover, the 80:20 split has become a widely adopted heuristic guideline in both research and industry because it is simple, effective, and well-suited for medium- to large-sized datasets. Several recent studies support this approach, noting that 20% of the data is sufficient to provide an objective evaluation of model performance without significantly reducing the amount of training data.

Table 3: Split Data Results

Wound Types	Training	Validation
Abrasion Wounds	160	40
Bruises Wounds	160	40
Burning Wounds	160	40
Decubitus Wounds	160	40
Blister Wounds	160	40
Gangren Wounds	160	40
Total	960	240

4.3 Build The Model

In this study, the MobileNetV2 model was used as the backbone architecture for the classification of diabetic wound images. MobileNetV2 is one of the Convolutional Neural Network (CNN) models developed by Google with a focus on computational efficiency and high

performance. Trained on millions of images from various object classes, the model already possesses initial capabilities in recognizing common visual features.

Layer (type)	Output Shape	Param #
mobilenetv2_1_00_224 (Functional)	(None, 7, 7, 1280)	2,257,984
dropout (Dropout)	(None, 7, 7, 1280)	0
Flatten (Flatten)	(None, 62720)	0
batch_normalization (BatchNormalization)	(None, 62720)	250,880
dense (Dense)	(None, 32)	2,007,072
batch_normalization_1 (BatchNormalization)	(None, 32)	128
activation (Activation)	(None, 32)	0
dropout_1 (Dropout)	(None, 32)	0
dense_1 (Dense)	(None, 16)	528
batch_normalization_2 (BatchNormalization)	(None, 16)	64
activation_1 (Activation)	(None, 16)	0
dropout_2 (Dropout)	(None, 16)	0
dense_2 (Dense)	(None, 6)	102

Fig. 5: Model Summary

The summary results of the constructed model provide detailed information about the network architecture used in this study. This information includes a complete list of all layers comprising the model, such as the type of layer, the output shape of each layer, and the number of parameters, both trainable and non-trainable. The model employed MobileNetV2 as a feature extractor, which had been pretrained on the ImageNet dataset, with certain parts frozen (non-trainable) to prevent their weights from being updated. Additional layers specifically designed for wound classification were retrained using the collected wound dataset.

4.4 Model Evaluation

The evaluation of the model training results in this study was carried out comprehensively using various metrics and evaluation approaches. This process not only included analyzing the accuracy and loss values on the training and validation data during the training phase but also involved testing the model's performance on unseen testing data. During training, the accuracy and loss graphs were monitored to assess the extent to which the model was able to learn progressively at each epoch. The patterns of increasing accuracy and decreasing loss served as key indicators for determining whether the model's performance was improving or facing issues such as overfitting or underfitting.

4.4.1 Accuracy

Accuracy is one of the most commonly used evaluation metrics in classification tasks to measure a model's performance in predicting the correct class. Simply put, accuracy reflects how often the model makes correct predictions compared to the total number of predictions made. In this study, accuracy indicates the percentage of wound images that were successfully classified into the correct labels by the model. The higher the accuracy value, the better the model is at recognizing the characteristics of each wound type in both the training and testing data.

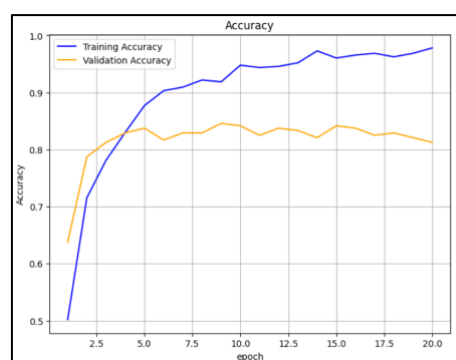


Fig 6: Accuracy Visualization

4.4.2 Loss

Loss, or loss function, is a core component in the training process of machine learning models, particularly neural networks. It measures how far the model's predictions deviate from the actual target values. In other words, loss serves as an indicator of how poor or erroneous the model's predictions are at each training iteration. Loss is not just an evaluation number but also the primary guide for optimization

algorithms (such as Adam, SGD, etc.) to update the weights and biases within the model. The smaller the loss value, the better the model’s performance in learning data patterns.

In the context of multi-class classification—such as wound classification in diabetes—the commonly used loss function is *categorical crossentropy*. This function calculates the distance between the predicted probability distribution produced by the model and the true label distribution, making it highly suitable for tasks involving more than two classes.

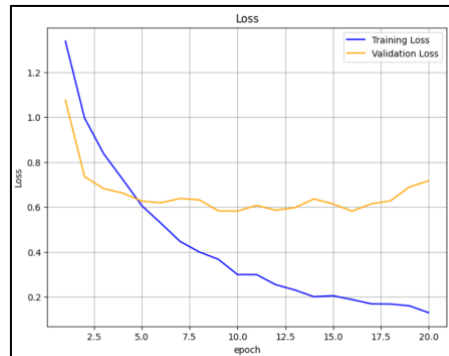


Fig. 7: Loss Visualization

4.4.3 Confusion Matrix

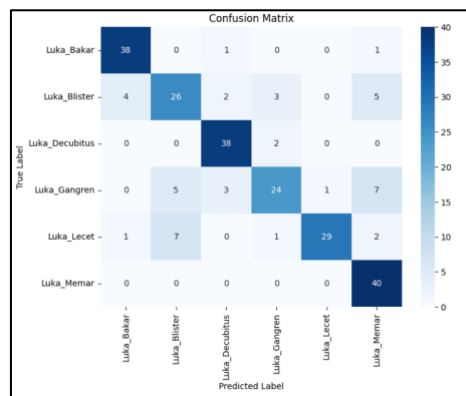


Fig. 8: Confusion Matrix

The confusion matrix above illustrates the performance of the MobileNetV2 model in classifying six types of wounds: burn wounds, blister wounds, decubitus wounds, gangrene wounds, abrasions, and bruises. In this matrix, the rows represent the actual labels (true labels), while the columns represent the labels predicted by the model (predicted labels). The main diagonal of the matrix (from top left to bottom right) shows the number of correct predictions (True Positives – TP) for each class, whereas the non-diagonal elements indicate incorrect predictions, which can be interpreted as False Positives (FP) and False Negatives (FN).

4.5 Model Test

	precision	recall	f1-score	support
Luka_Bakar	0.88	0.95	0.92	40
Luka_Bliister	0.68	0.65	0.67	40
Luka_Decubitus	0.86	0.95	0.90	40
Luka_Gangren	0.80	0.60	0.69	40
Luka_Lecet	0.97	0.72	0.83	40
Luka_Memar	0.73	1.00	0.84	40
accuracy			0.81	240
macro avg	0.82	0.81	0.81	240
weighted avg	0.82	0.81	0.81	240

Fig. 9: Classification Report

From the evaluation results of the MobileNetV2 classification model, it can be concluded that the ratio of true positive predictions to all predictions considered positive (precision) reached 0.82. This indicates that the model is fairly accurate in classifying wound images according to the predicted labels. Furthermore, the ratio of true positive predictions to all actual data belonging to the same class (recall) was 0.81, which shows that the model is sufficiently sensitive in capturing relevant cases.

5. Conclusion

This study demonstrates that the CNN model, particularly the MobileNetV2 architecture, is capable of classifying six different types of wounds—burns, blisters, decubitus, gangrene, abrasions, and bruises—with an overall accuracy of 81%, precision of 0.82, and recall of 0.81. These results indicate that the model not only achieves good accuracy but also demonstrates high sensitivity in distinguishing wound types, making it a reliable tool for automated wound classification. The high precision and recall values achieved in certain wound categories further emphasize the robustness of the model in handling complex variations in medical images.

Moreover, the use of transfer learning with MobileNetV2 allows the model to leverage pre-trained features from large-scale datasets, resulting in efficient training and effective generalization even with a relatively limited dataset. This highlights the practicality of the proposed approach for medical applications, especially in supporting healthcare professionals in the early detection and categorization of diabetic wounds.

In conclusion, the findings of this study confirm that MobileNetV2 provides an effective and efficient solution for diabetic wound classification, with promising potential for integration into real-world clinical decision support systems. Future research can focus on expanding the dataset, applying more advanced augmentation techniques, and evaluating the model in real clinical settings to further enhance its reliability and applicability in medical practice.

Acknowledgement

The author would like to express sincere gratitude to Dr. Hermawan Syahputra, S.Si., M.Si. for the guidance and support throughout this research. Special thanks are also extended to Medan State University for providing the facilities and academic environment that made this work possible.

References

- [1] Lestari, Zulkarnain, and Aisyah Sijid, "Diabetes Melitus: Review Etiologi," Nov. 2021. [Online]. Available: <http://journal.uin-alauddin.ac.id/index.php/psb>
- [2] International Diabetes Foundation, "Diabetes Facts and Figures."
- [3] D. M. Anisuzzaman, H. Barzekar, L. Tong, J. Luo, and Z. Yu, "A Deep Learning Study on Osteosarcoma Detection from Histological Images," Nov. 2020, doi: 10.1016/j.bspc.2021.102931.
- [4] F. H. Syahrul and P. S. Sasongko, "Penerapan Convolutional Neural Network Untuk Klasifikasi Tingkat Keparahan Retinopati Diabetik Pada Penderita Diabetes Melitus," 2022.
- [5] R. Nisak, "The Occurrence and Classification of Diabetic Ulcers Among Diabetes Mellitus Patients Using Wagner-Ulcer Classification Tool," 2021.
- [6] S. Widiawati and W. Kalpataria, "Implementasi Senam Kaki Diabetes Pada Penderita Diabetes Melitus di RSUD Raden Mattaher Jambi," *Jurnal Pengabdian Harapan Ibu (JPHI)*, vol. 2, no. 1, 2020, doi: 10.30644/jphi.v1i1.199.
- [7] D. Qurniawati, A. Fatikasari, J. Tafonao, E. Anggeria, and F. Keperawatan Dan Kebidanan, "Pengaruh Diabetes Self-Management Education (DSME) Terhadap Perawatan Diri Pasien Luka Diabetes Melitus The Effect of Diabetes Self-Management Education (DSME) on Self-Care among Patients with Diabetes Mellitus Wound," *Jurnal Ilmu Keperawatan*, vol. 8, p. 1, 2020.
- [8] F. Marpaung, F. Aulia, N. Suryani, and R. Cyra Nabil, *COMPUTER VISION DAN PENGOLAHAN CITRA DIGITAL*. 2022. [Online]. Available: www.pustakaaksara.co.id
- [9] R. D. Ramadhani, A. N. A. Thohari, C. Kartiko, A. Junaidi, T. G. Laksana, and N. A. S. Nugraha, "Optimasi Akurasi Metode Convolutional Neural Network untuk Identifikasi Jenis Sampah," *Jurnal RESTI*, vol. 5, no. 2, pp. 312–318, Apr. 2021, doi: 10.29207/resti.v5i2.2754.
- [10] B. P. Hartato, "Penerapan Convolutional Neural Network pada Citra Rontgen Paru-Paru untuk Deteksi SARS-CoV-2," *Jurnal RESTI*, vol. 5, no. 4, pp. 747–759, Aug. 2021, doi: 10.29207/resti.v5i4.3153.