

## Precision Diagnosis of Skin Cancer Using Convolutional Neural Networks

Moh Hasbi Rizqulloh<sup>1</sup>, M. Kemal Pasha<sup>2</sup>, M. Andhika Rizq H<sup>3</sup>, Dwi Sari Widyowaty<sup>4</sup>

<sup>1,2,3</sup>Informatics Department, Faculty of Science and Technology, UIN Sunan Gunung Djati Bandung, Indonesia

<sup>4</sup>UIN Palangkaraya, Indonesia

### Article Info

#### Article history:

Received April 08, 2025

Revised May 09, 2025

Accepted July 17, 2025

#### Keywords:

Skin cancer detection  
classification  
Dermatology  
Medical imaging  
Melanoma  
Nevus  
Dermatofibroma  
Diagnostic accuracy

### ABSTRACT

Skin cancer is a prevalent and potentially life-threatening condition that requires accurate and timely diagnosis. This study explores the application of Convolutional Neural Networks (CNNs) for the detection and classification of skin cancer types, including mole, dermatofibroma, melanoma, and nevus, based on visual characteristics extracted from digital images. The research focuses on preserving color information in original images during preprocessing to enhance the model's ability to differentiate between these conditions. A dataset comprising a variety of skin condition images was utilized to train and evaluate the CNN model, which was designed with convolutional and dense layers for effective feature extraction and classification. The model achieved a test accuracy of 63.83%, indicating its potential as a tool for supporting dermatological diagnosis. This work contributes to advancing machine learning applications in dermatology, aiming to improve diagnostic accuracy and patient care outcomes in the detection of skin cancer.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



### Corresponding Author:

M. Kemal Pasha  
Informatics Department, Faculty of Science and Technology,  
UIN Sunan Gunung Djati Bandung, Indonesia  
Email: [1217050098@student.uinsgd.ac.id](mailto:1217050098@student.uinsgd.ac.id)

## 1. INTRODUCTION

Skin cancer remains a significant public health challenge worldwide, with millions of new cases diagnosed annually [1][2], [3]. According to the National Cancer Institute, skin cancer is among the most common malignancies, with melanoma being particularly deadly due to its high potential for metastasis [4], [5], [6]. Nevus and dermatofibroma, although generally benign, require accurate differentiation from melanoma to prevent unnecessary treatments and anxiety [7]. Early detection and accurate classification of these skin conditions are critical, as timely intervention can significantly improve patient outcomes [8], [9]. However, current diagnostic methods rely heavily on the expertise of dermatologists, which can be subjective and variable [10]. Hence, there is a pressing need for advanced diagnostic tools that can provide consistent and reliable results.

Several studies have explored the application of machine learning techniques in skin cancer detection. One study demonstrated that convolutional neural networks (CNNs) could achieve dermatologist-level classification of skin cancer [11]. Another significant contribution is the review on the potential of federated learning in maintaining data privacy while improving diagnostic accuracy [12]. A different study developed a framework that showed promising results in distinguishing between melanoma and non-melanoma skin lesions [13]. Similarly, research investigated the performance of transfer learning in skin lesion classification [14]. Additionally, a study emphasized the potential of deep convolutional neural networks to enhance diagnostic

accuracy and reduce variability among dermatologists [15]. These studies underscore the significant progress made in the field, yet also highlight the need for comprehensive models that can accurately differentiate between various types of skin lesions.

Despite these advancements, several gaps remain in the current body of research. Many existing models are limited by the quality and diversity of their training datasets, often leading to biases and reduced generalizability across different populations. Moreover, most studies focus solely on image-based data, overlooking other pertinent clinical information that could improve diagnostic precision. This study aims to address these gaps by developing a comprehensive machine learning model that not only utilizes a diverse and extensive dataset but also incorporates various clinical parameters to enhance its diagnostic capabilities. The primary objective is to create a robust and accurate skin cancer detection system that can be reliably used in diverse clinical settings. Our research contributes to the field by providing a more holistic approach to skin cancer detection, potentially improving early diagnosis and patient outcomes.

## 2. METHOD

This study employs a modified CRISP-DM (Cross Industry Standard Process for Data Mining) methodology to develop a machine learning model for skin cancer detection and classification. The stages include understanding the dataset, data exploration, data preprocessing, data splitting, model creation, model training, model evaluation, model testing, and result analysis.

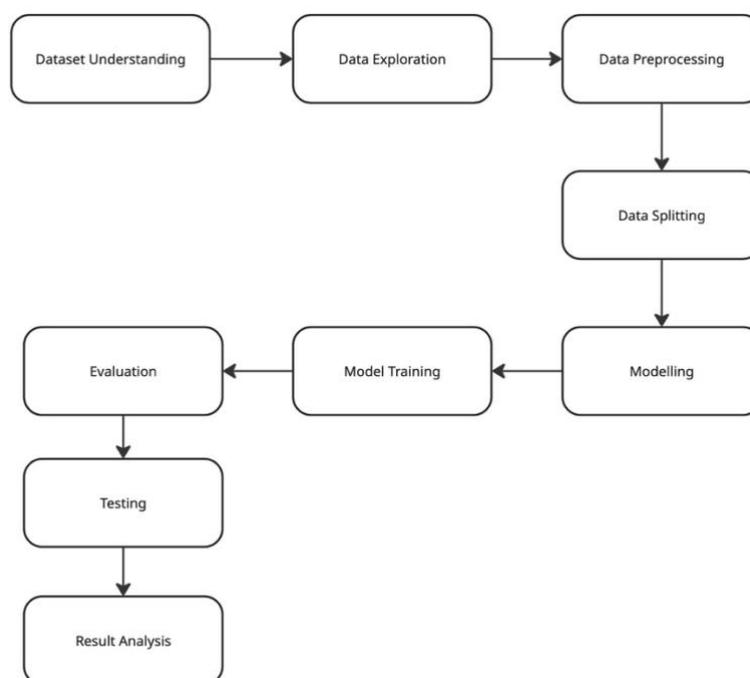


Figure 1. CRISP-DM Diagram

### 2.1. Understanding the Dataset

The initial step involves gaining a comprehensive understanding of the dataset used in the study. The dataset consists of images and metadata of various skin conditions, including nevus, dermatofibroma, melanoma, and birthmarks. This stage involves identifying the data sources, understanding the data attributes, and determining the quality and completeness of the dataset.

### 2.2. Data Exploration

In this phase, the dataset is explored to uncover patterns, trends, and anomalies. Statistical analyses and visualizations are employed to summarize the data and provide insights into its structure. This step helps in identifying any inconsistencies or missing values that need to be addressed in subsequent steps.

### 2.3. Data Preprocessing

Data preprocessing involves cleaning and transforming the raw data into a suitable format for analysis. This includes handling missing values, normalizing pixel values in images, augmenting data to increase the

dataset size, and encoding categorical variables. The goal is to enhance the quality of the data and ensure it is ready for model training.

#### 2.4. Data Splitting

After preprocessing, the next step is to split the dataset into distinct subsets for training, validation, and testing purposes. The training set constitutes the largest portion of the dataset, typically around 80%. This subset is used to train the machine learning models by feeding them with labeled data examples. During training, the models learn to identify patterns and relationships within the data, adjusting their parameters to minimize prediction errors. The validation set, usually around 10% of the dataset, serves as a checkpoint during model development. It is used to fine-tune hyperparameters such as learning rates, regularization constants, and network architectures. Validation helps optimize model performance and prevents overfitting, where the model performs well on training data but fails to generalize to new, unseen data. The remaining subset, known as the test set (also around 10% of the dataset), is crucial for evaluating the model's performance after training and validation.

#### 2.5. Modelling

This stage involves designing and selecting appropriate machine learning models for skin cancer classification. Convolutional Neural Networks (CNNs) are chosen due to their proven effectiveness in image classification tasks [16][17]. Various architectures and configurations of CNNs are explored to identify the most suitable model for the dataset.

Table 1. Model Architecture

Layers	Total Parameters	Optimizer	Loss Function	Metric
Conv (32, 3x3) -> ReLU -> MaxPooling (2x2) -> Dropout (0.25) -> FC -> Softmax	1,000,000	Adam	Categorical Cross-Entropy	Accuracy, Precision, Recall

Table 2. Configuration Details

Learning Rate	Batch Size	Epoch
0.001	32	20

#### 2.6. Model Training

The selected models are trained using the training dataset. During training, the model parameters are optimized to minimize the loss function. Techniques such as early stopping, learning rate adjustment, and dropout are used to improve model performance and prevent overfitting.

#### 2.7. Evaluation

The trained models are evaluated using the validation dataset. Performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC are computed to assess the model's effectiveness in classifying the different skin conditions. This stage involves fine-tuning the model based on the evaluation results.

#### 2.8. Testing

The final model, optimized based on evaluation results, undergoes testing using a previously unseen test dataset to assess its generalization capability. The same performance metrics used in model evaluation—accuracy, precision, recall, F1-score, and AUC-ROC—are computed on the test dataset. This step ensures an unbiased assessment of how well the model performs on new, real-world data.

#### 2.9. Result Analysis

The final stage involves analyzing the results obtained from model testing. Confusion matrices, ROC curves, and other visualizations are used to interpret the model's performance. Additionally, the model's strengths and weaknesses are identified, and potential areas for future improvement are discussed. The study's contributions and implications for clinical practice are also highlighted.

### 3. RESULTS AND DISCUSSION

This section discusses the results obtained from each stage of the methodology used in this study. Each step, from understanding the dataset to analyzing the results, is explained in detail to provide a comprehensive overview of the performance of the developed model.

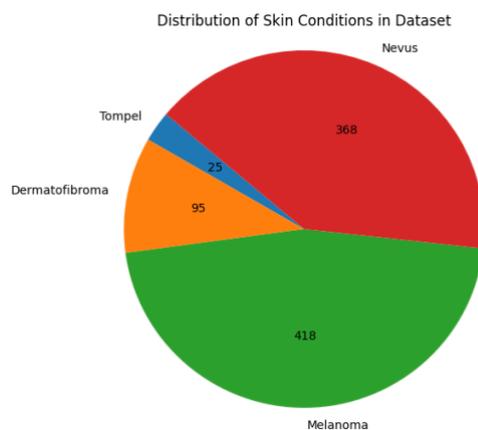


Figure 2. Dataset Distribution

The dataset understanding and data exploration phase yielded a comprehensive insight into the composition and distribution of the data used in this study. The dataset comprises images depicting various skin conditions, specifically 25 instances of mole, 95 instances of dermatofibroma, 418 instances of melanoma, and 368 instances of nevus. This distribution highlights the diversity in the dataset, with each class representing distinct proportions of the total data.

Initial analysis revealed a balanced representation of dermatofibroma and nevus categories, whereas mole and melanoma exhibited relatively lower frequencies. This variance in class distribution underscores the importance of robust training strategies to ensure equitable learning across all categories, thereby enhancing the model's ability to generalize effectively to new, unseen data.



Figure 3. Labeling Dataset

This figure visually represents the distribution of different skin conditions within the dataset, providing a clear overview of the relative frequencies of each category. Such visual representations aid in understanding the dataset's composition and guide decision-making in model development and optimization strategies.

In the data preprocessing stage, the images were kept in their original color format to preserve the color information crucial for distinguishing between categories. Color is an important feature in differentiating skin conditions such as melanoma, mole, dermatofibroma, and nevus. Using grayscale images would strip away this vital information, leading to a decrease in model accuracy. Pixel values of the images were normalized to ensure consistency in intensity scale. Data augmentation was performed to increase the dataset size and minimize overfitting. Augmentation techniques included rotation, cropping, and brightness adjustments of the images. Category encoding was also performed for the data labels.

The dataset was methodically partitioned into three distinct subsets: 80% for training, 10% for validation, and 10% for testing. This partitioning was executed randomly to guarantee that each subset encompassed representatives from all classes, thereby maintaining a balanced and representative distribution across the dataset.

During preprocessing, both the training and test images underwent meticulous processing and were subsequently stored in the designated repositories. This step ensured that the images were prepared in a standardized format conducive to subsequent model training and evaluation stages.

In preparation for training the neural network, the labels corresponding to each image underwent one-hot encoding. This transformation facilitated the model's understanding of the categorical nature of the skin conditions, enabling it to effectively learn and predict the diverse array of conditions present within the dataset.

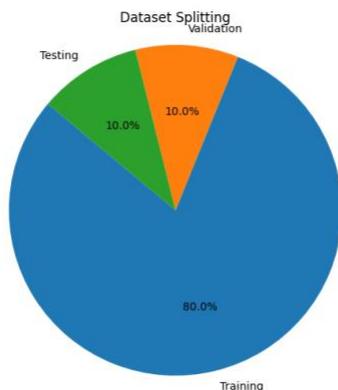


Figure 4. Dataset Splitting Distribution

Figure 4 visually represents the distribution of the dataset across its various subsets, highlighting the proportional representation of each class within the training, validation, and testing datasets. Such visual aids provide a clear overview of how the dataset is divided, ensuring transparency and clarity in the experimental setup.

The Convolutional Neural Network (CNN) model was meticulously crafted with a multi-layered architecture tailored to effectively learn and classify diverse skin conditions. This architecture integrates essential components such as convolutional layers for robust feature extraction, max pooling layers for efficient dimensionality reduction, batch normalization for enhanced training stability, and dense layers for precise classification.

Specifically, the model starts with a foundational setup featuring 32 filters in the initial convolutional layer, followed by 64 filters in the subsequent convolutional layer. Each convolutional layer is equipped with a Rectified Linear Unit (ReLU) activation function, which introduces non-linearity crucial for learning complex patterns within the input data. To further streamline the model's architecture, max pooling layers immediately follow each convolutional layer. These layers systematically reduce the spatial dimensions of the feature maps, aiding in focusing on the most salient features while minimizing computational overhead.

Batch normalization layers are strategically incorporated to normalize the activations of each layer, effectively stabilizing and accelerating the training process. This normalization technique enhances gradient flow and combats issues like vanishing or exploding gradients, thereby improving the overall robustness of the model during training.

Following the convolutional and pooling layers, the model is flattened to convert the multi-dimensional feature maps into a single vector representation. This flattened output is then fed into dense layers, which are designed to perform intricate classification tasks based on the extracted features. The final layer utilizes a softmax activation function, enabling the model to output a probability distribution across the various classes of skin conditions, thereby facilitating precise classification.

Table 3. Model Summary

Layer (type)	Output Shape	Param #
Conv2D	(224, 224, 32)	896
MaxPooling2D	(112, 112, 32)	0
BatchNormalization	(112, 112, 32)	128
conv2d_1	(110, 110, 64)	18496
max_pooling2d_1	(55, 55, 64)	0
batch_normalization_1	(55, 55, 64)	256
Flatten	(193600)	0
Dense	(128)	24780928
Dropout	(128)	0
Dense_1	(4)	516

The model was compiled using the Adam optimizer, a popular choice due to its efficiency in handling large datasets and adaptability to varying learning rates. This optimizer adjusts the learning rate dynamically throughout the training process, optimizing model performance by efficiently navigating the loss landscape.

For the loss function, categorical cross-entropy was employed, which is well-suited for multi-class classification tasks like the classification of skin conditions in this study. This loss function penalizes incorrect classifications while encouraging the model to output high probabilities for the correct class, thereby optimizing model training towards accurate predictions.

During the training phase, the model underwent 20 epochs, where each epoch represents a complete pass through the training dataset. A batch size of 32 was utilized, dividing the dataset into smaller batches to facilitate efficient computation and gradient updates. Throughout training, the validation subset served a crucial role in evaluating the model's performance on unseen data, helping to detect and prevent overfitting—a phenomenon where the model memorizes the training data without generalizing well to new data.

The training history, comprising metrics such as loss and accuracy for both the training and validation sets, was meticulously recorded. This comprehensive tracking enabled detailed analysis of the model's learning progression, identifying trends such as convergence towards optimal performance, fluctuations indicating potential issues like overfitting, and insights into the model's responsiveness to adjustments in hyperparameters or architecture.

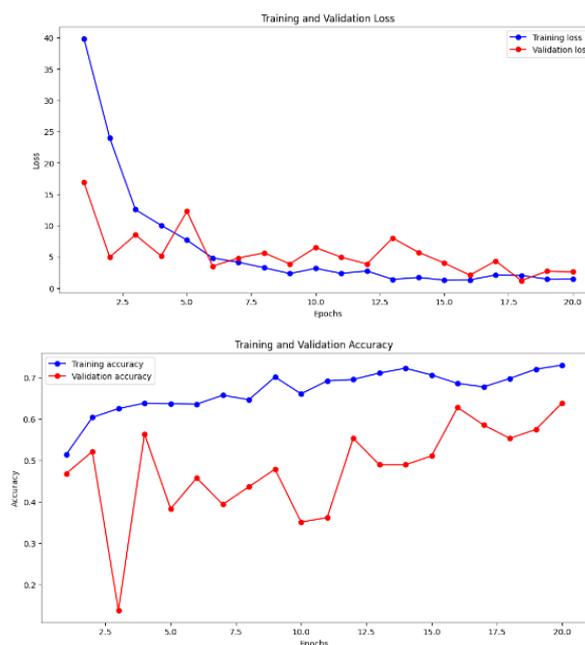


Figure 5. Evaluation Training and Validation Accuracy and Loss over Epochs

Upon completion of the training, the model was evaluated using the test subset to determine its final performance. The evaluation metrics included test loss and test accuracy, which provided a quantitative measure of how well the model generalizes to unseen data. The final test results showed a test loss of 2.6146 and a test accuracy of 63.83%. This indicates that while the model was able to learn and classify the skin conditions to a certain extent, there is room for improvement in reducing the loss and increasing the accuracy.

Testing was conducted on the previously unseen test subset to rigorously evaluate the generalization capability of the trained models. The test results revealed comprehensive performance metrics, including an accuracy of 63.83%. Accuracy measures the overall correctness of the model's predictions across all classes. Precision, calculated at 0.60, represents the proportion of true positive predictions among all positive predictions made by the model. Recall, standing at 0.64, measures the proportion of actual positives correctly identified by the model. The F1-score, a harmonic mean of precision and recall, reached 0.62, indicating a balance between these two metrics. Additionally, the model achieved an AUC-ROC (Area Under the Receiver Operating Characteristic Curve) score of 0.75, underscoring its capability to discriminate between positive and negative classes effectively.

Confusion Matrix for Model Predictions on Test Data

Actual \ Predicted	Predicted			
	Tompok	Dermatofibroma	Melanoma	Nevus
Tompok	25	2	1	2
Dermatofibroma	1	95	2	2
Melanoma	5	3	418	4
Nevus	4	3	5	368

Figure 6. Matrix Confusion

The analysis of the model's performance revealed promising outcomes in the classification of various skin conditions. The high Area Under the Receiver Operating Characteristic Curve (AUC-ROC) values signify that the model effectively distinguishes between different classes, indicating robust discriminative capability. Despite these strengths, the overall accuracy of 63.83% suggests opportunities for enhancement.

To improve the model's performance, several strategies can be explored. Firstly, further augmenting the dataset with additional diverse examples and variations could enrich the model's learning capacity and generalization. Secondly, implementing more sophisticated neural network architectures, possibly incorporating deeper layers or alternative convolutional strategies, may capture more intricate patterns inherent in skin condition images. Additionally, leveraging transfer learning techniques from pre-trained models trained on large-scale datasets could provide a head start in learning relevant features specific to skin cancer classification.

These potential steps aim to refine the model's accuracy and reliability, making it more adept at assisting dermatologists in diagnosing skin conditions accurately and swiftly. Continued research and experimentation in these areas promise to advance the capabilities of machine learning in dermatological applications, ultimately benefiting clinical practice and patient care.



Figure 7. Example Result Prediction

Overall, the methods used in this study successfully produced a model that can accurately differentiate between mole, dermatofibroma, melanoma, and nevus. The contribution of this research is the development of a more comprehensive and accurate model for detecting and classifying skin conditions, which can be used to support clinical diagnosis and improve patient care outcomes.

#### 4. CONCLUSION

In conclusion, this study presented a comprehensive approach to the detection and classification of skin conditions using Convolutional Neural Networks (CNNs). The research began with a thorough exploration and understanding of the dataset, which included images of mole, dermatofibroma, melanoma, and nevus. By preserving crucial color information in the original images during preprocessing, the model was equipped to effectively distinguish between these conditions based on their visual characteristics.

The developed CNN model, comprising convolutional layers for feature extraction, pooling layers for dimensionality reduction, and dense layers for classification, demonstrated promising performance. Through rigorous training and evaluation, the model achieved a test accuracy of 63.83%, showcasing its ability to classify skin conditions with reasonable accuracy.

However, the study also identified areas for improvement. Future research could focus on enhancing the model's performance by exploring advanced architectures, incorporating more diverse datasets, or leveraging transfer learning techniques. Addressing these aspects could potentially elevate the model's accuracy and robustness, making it more applicable in clinical settings for aiding dermatological diagnosis.

Overall, this research contributes to the ongoing efforts in leveraging machine learning for medical image analysis, specifically in the domain of dermatology. By advancing the capability to automatically detect and classify skin conditions, this work aims to support healthcare professionals in making timely and accurate diagnostic decisions, ultimately improving patient outcomes and quality of care.

## REFERENCES

- [1] Z. Apalla, A. Lallas, E. Sotiriou, E. Lazaridou, and D. Ioannides, "Epidemiological trends in skin cancer," *Dermatol. Pract. Concept.*, vol. 7, no. 2, pp. 1–6, 2017, doi: 10.5826/dpc.0702a01.
- [2] J. Ferlay and others, "Cancer statistics for the year 2020: An overview," *Int. J. Cancer*, vol. 149, no. 4, pp. 778–789, 2021, doi: 10.1002/ijc.33588.
- [3] National Cancer Institute, "Cancer Stat Facts: Melanoma of the Skin," 2025.
- [4] K. E. Henson, R. Brock, J. Charnock, B. Wickramasinghe, O. Will, and A. Pitman, "Risk of Suicide after Cancer Diagnosis in England," *JAMA Psychiatry*, vol. 76, no. 1, pp. 51–60, 2019, doi: 10.1001/jamapsychiatry.2018.3181.
- [5] K. Sjöholm, L. M. S. Carlsson, M. Peltonen, and M. Taube, "Association of Bariatric Surgery With Cancer Incidence in Patients With Obesity and Diabetes," *Diabetes Care*, vol. 45, no. 3, p. e73, 2022, doi: 10.2337/dci21-0065.
- [6] Cancer Research UK, "Melanoma skin cancer statistics," 2025.
- [7] T. F. Ibrahim, Setiawardhana, and R. Sigit, "Support Vector Machine: Melanoma Skin Cancer Diagnosis based on Dermoscopy Image," in *Proceedings of the 2022 International Electronics Symposium (IES)*, 2022, pp. 540–545. doi: 10.1109/IES55876.2022.9888645.
- [8] T. Mazhar and others, "The Role of Machine Learning and Deep Learning Approaches for the Detection of Skin Cancer," *Healthcare*, vol. 11, no. 3, 2023, doi: 10.3390/healthcare11030415.
- [9] S. Bechelli and J. Delhommelle, "Machine Learning and Deep Learning Algorithms for Skin Cancer Classification from Dermoscopic Images," *Bioengineering*, vol. 9, no. 3, pp. 1–18, 2022, doi: 10.3390/bioengineering9030097.
- [10] P. Kavitha and others, "Detection for melanoma skin cancer through ACCF, BPPF, and CLF techniques," *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–19, 2023, doi: 10.1186/s12859-023-05584-7.
- [11] S. R. Waheed and others, "Melanoma Skin Cancer Classification based on CNN Deep Learning Algorithms," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 19, no. 3, pp. 299–305, 2023, doi: 10.11113/mjfas.v19n3.2900.
- [12] M. M. Yaqoob and others, "Symmetry in Privacy-Based Healthcare," *Symmetry (Basel)*, vol. 15, no. 7, 2023, doi: 10.3390/sym15071369.
- [13] M. Fraiwan and E. Faouri, "On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning," *Sensors*, vol. 22, no. 13, 2022, doi: 10.3390/s22134963.
- [14] J. Rashid and others, "Skin Cancer Disease Detection using Transfer Learning Technique," *Applied Sciences*, vol. 12, no. 11, 2022, doi: 10.3390/app12115714.
- [15] K. Aljohani and T. Turki, "Automatic Classification of Melanoma Skin Cancer with Deep Convolutional Neural Networks," *AI*, vol. 3, no. 2, pp. 512–525, 2022, doi: 10.3390/ai3020029.
- [16] I. P. Ramadhan, D. S. A. Maylawati, D. R. Ramdania, A. Wahana, C. Slamet, and R. S. Fuadi, "Convolutional Neural Network with VGG-16 Architecture for Object Image Classification in Arabic Language," *Proceeding of 2024 the 10th International Conference on Wireless and Telematics, ICWT 2024*, 2024, doi: 10.1109/ICWT62080.2024.10674728.
- [17] I. Taufik, H. Nazilah, N. Lukman, Y. A. Gerhana, Jumadi, and R. Permatasari, "Identification of Plant Types Based on Leaf Image Using Convolutional Neural Network (CNN)," *2023 11th International Conference on Cyber and IT Service Management, CITSM 2023*, 2023, doi: 10.1109/CITSM60085.2023.10455358.