

Evaluating the Capability of Transfer Learning VGG16 for Detecting Tomato Diseases in Indonesia

Tegar Satriya Wiguna¹, Febri Liantoni^{2*}, Yudianto Sujana³

^{1,2,3} Information Engineering and Computer Education, Sebelas Maret University
tegarsatriya17@student.uns.ac.id¹, febri.liantoni@gmail.com², yudianto.sujana@staff.uns.ac.id³

Accepted 19 June 2026

Approved 28 June 2026

Abstract— Tomato leaf diseases pose a significant threat to agricultural productivity in Indonesia, often leading to severe yield losses. This study evaluates the effectiveness of the VGG16 convolutional neural network in detecting tomato diseases, particularly when trained on the standardized PlantVillage dataset and applied to local agricultural conditions in Sragen, Central Java. To conduct this research, the methodology included standardizing images to 256x256 pixels and applying the GrabCut method for background removal to isolate leaf regions. The model was trained via transfer learning using pre-trained ImageNet weights, and the system was deployed for practical use through a FastAPI backend and a React Native frontend. The evaluation results showed that the VGG16 model achieved an 82% accuracy, along with an 80% average precision, 81% recall, and 81% F1-score on the PlantVillage test set. However, performance exhibited a sharp decline to a 25% overall accuracy when tested on locally sourced images, failing entirely to predict samples from the Bacterial Spot, Late Blight, and Leaf Mold classes. These findings highlight the model's limited generalization capabilities and emphasize the necessity of incorporating local datasets and domain adaptation strategies to develop AI-based plant disease detection tools that are effective in real-world settings. The study underscores the importance of contextualizing AI solutions for local agricultural environments to ensure their practical applicability and reliability.

Index Terms— VGG16; Tomato; Leaf Diseases; Convolutional Neural Network.

I. INTRODUCTION

The tomato cultivation sector in Indonesia has experienced substantial growth in recent years. Data from the Indonesian Central Statistics Agency (BPS) indicated that tomato production reached approximately 1.16 million tons in 2022, reflecting a 4.88% increase compared to 1.11 million tons in the previous year. This positive trend is primarily due to comprehensive government initiatives aimed at boosting agricultural productivity and quality, alongside significant efforts from local farmers. Despite this progress, tomato farming continues to face substantial challenges, notably the prevalence of plant

diseases that severely reduce crop yield and quality [1]. Diseases affecting tomatoes include early blight (*Alternaria solani*), late blight (*Phytophthora infestans*), leaf mold (*Fulvia fulva*), and bacterial spot (*Xanthomonas campestris* pv. *vesicatoria*).

Traditional methods of disease detection rely predominantly on visual inspections by farmers [2], which are labor-intensive, subjective, and often lead to delayed identification and ineffective management responses. Consequently, these delays frequently result in substantial crop losses and increased dependency on chemical pesticides, which pose environmental and health risks.

Recent advancements in deep learning technologies have demonstrated considerable potential for automated plant disease detection through image analysis[3]. Convolutional Neural Networks (CNNs), such as VGG16, ResNet, and DenseNet, have achieved significant accuracy in classifying plant diseases in controlled research environments, primarily utilizing standardized datasets like PlantVillage. Among these architectures, VGG16 stands out due to its effective balance between architectural simplicity and robust feature extraction capabilities [4].

However, the critical question remains regarding the practical applicability and generalization of models trained exclusively on datasets such as PlantVillage, which were captured under controlled conditions [5], to local agricultural contexts with distinct environmental characteristics. Given the variability in climate, leaf morphology, and disease manifestation in Indonesian tomato plantations, it is essential to evaluate whether a VGG16 model trained on the Kaggle PlantVillage dataset can effectively detect diseases on locally collected tomato leaf images in some research it limitation cause environment [6].

This research specifically addresses this gap by evaluating the performance of a VGG16-based model trained exclusively with the Kaggle PlantVillage dataset when applied to a local dataset collected from a tomato plantation in Sragen, Central Java, Indonesia. The objective is to investigate the effectiveness and

limitations of using internationally sourced datasets for local agricultural applications, thus providing insights into necessary adaptations for enhancing the robustness and applicability of AI-driven plant disease detection systems in Indonesian agriculture.

II. METHOD

In this research, a structured methodology was employed to evaluate the performance of the VGG16 model in detecting diseases in tomato plants using local datasets, particularly focusing on its generalization capability. The research stages include dataset collection, pre-processing, model training, backend and frontend development, testing, and analysis of results.

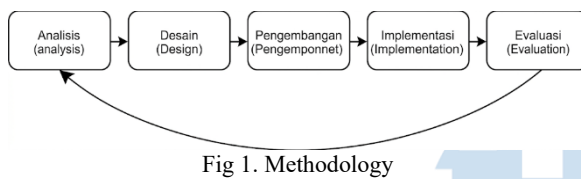


Fig 1. Methodology

The dataset used in this research was sourced from the Kaggle repository, specifically the PlantVillage dataset, which contains images categorized into bacterial spot, early blight, late blight, leaf mold, and healthy leaves. Additionally, local data was collected directly from a tomato plantation located in Sragen, Central Java, Indonesia, to test the model's ability to generalize across different environmental conditions and plant varieties.



Fig 2. Sample image from dataset kaggle

In the pre-processing stage, all collected images underwent standardization, which included resizing to a uniform size of 256×256 pixels [7] and background removal using the GrabCut method. Additionally, data augmentation techniques such as a rotation range of 20 degrees, zoom range up to 20%, and horizontal flipping were applied to enhance image variability and strengthen the model's ability to generalize across

different leaf positions and orientations. This comprehensive pre-processing ensures compatibility with the input requirements of the VGG16 architecture.

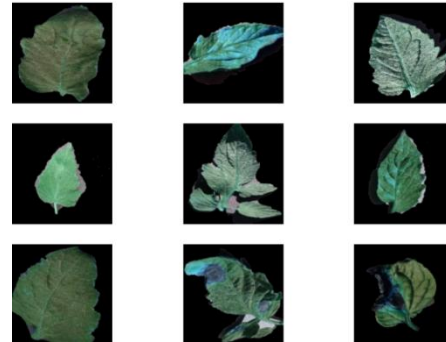


Fig 3. Sample image pre processing

The dataset was partitioned into training and testing with Kaggle dataset. The training phase involved employing the VGG16 architecture [8], a convolutional neural network known for its simplicity and strong feature extraction capabilities, using transfer learning with pre-trained weights from ImageNet.

For practical implementation, a backend system was developed using FastAPI to facilitate the integration of the trained model, allowing real-time inference[9] when processing uploaded images. A frontend interface was built using React Native, enabling users to interact easily [10] with the model through mobile devices. This user-friendly interface allows farmers to upload images and quickly receive diagnoses regarding the health status of tomato plants.

Testing involved evaluating the trained model using both the Kaggle dataset and the local dataset from Indonesia. Confusion matrices, accuracy, precision, recall, and F1-score metrics were employed to quantitatively assess model performance and highlight the model's strengths and limitations.

A. Accuracy

Measures the percentage of correct predictions out of all test data.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

B. Precision

Measures how many of the model's predictions are actually correct.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

TP is True Positive or correct prediction, and FP is False Positive or incorrect prediction.

C. Recall

Measures how many correct predictions were made by the model out of all actual positive cases.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

TP is True Positive or correct prediction, and FN is False Negative or incorrect prediction.

The results from this testing phase provided critical insights into the generalization ability of the VGG16 model when trained exclusively on standardized datasets like PlantVillage, emphasizing the importance of including locally sourced images to enhance the practical applicability of AI models in real-world agricultural scenarios.

III. RESULT AND DISCUSSIONS

This study aims to evaluate the performance of the VGG16 Convolutional Neural Network (CNN) architecture in detecting tomato leaf diseases using a model trained on the PlantVillage dataset and tested on both the same dataset and local images from a tomato plantation in Sragen, Central Java, Indonesia. Evaluation metrics include accuracy, precision, recall, and F1-score, derived from confusion matrices. In addition, local dataset testing was conducted to assess the model's generalization capabilities.

A. Model Training Result

The VGG16 model was trained using 500 images from kaggle. Before the training process began, each image underwent a series of preprocessing steps aimed at enhancing input quality and consistency. This included background removal using the GrabCut algorithm to isolate leaf regions from distracting

elements, thereby allowing the model to focus solely on disease-relevant features. Subsequently, all images were resized uniformly to a resolution of 256×256 pixels to meet the input size requirement of the VGG16 architecture. The model employed a transfer learning approach by utilizing pre-trained ImageNet weights, which were then fine-tuned specifically on the tomato leaf disease classification task by adjusting the final fully connected layers to match the five disease classes.

B. Evaluation on Planting Village Dataset

The model's initial evaluation was conducted using the PlantVillage test dataset, which served as a benchmark to measure baseline performance under controlled conditions. A confusion matrix was generated to present a detailed summary of classification outcomes across the five predefined classes: bacterial spot, early blight, late blight, leaf mold, and healthy. This matrix provides insights into the model's prediction accuracy and class-specific performance.

Table 1. Confusion matrix dataset kaggle

	Late Blight	Bacterial spot	Early blight	healthy	Leaf mold
Late blight	15	1	4	0	0
Bacterialsport	1	19	1	0	0
Early blight	3	1	11	0	4
Healthy	0	0	0	19	0
Leaf mold	1	0	1	0	18

Overall, the model achieved an accuracy of 82%, average precision of 80%, recall of 81%, and F1-score of 81%. The highest performance was observed in the Early Blight class, with all test samples correctly predicted, while Leaf Mold showed the lowest class-specific performance.

C. Evaluation on Local dataset

To assess the model's ability to generalize beyond the training domain, a set of 20 tomato leaf images collected from an actual plantation in Sragen, Central Java, was used for testing. These images represent real-world agricultural conditions, including varying lighting, backgrounds, and leaf textures, which differ significantly from the controlled settings of the PlantVillage dataset. The model's performance on this local dataset experienced a substantial decline, with an overall accuracy of just 25%. Notably, the model was only able to correctly classify 4 out of 5 images belonging to the Early Blight class. However, it failed to accurately predict any of the samples from the Bacterial Spot, Late Blight, and Leaf Mold classes, indicating a critical gap in the model's domain adaptability.

Table 2. Confusion Matrix dataset local

	Late Blight	Bacterial spot	Early blight	healthy	Leaf mold
Late Blight	0	2	0	0	1
Bacterial spot	0	0	3	0	1
Early blight	0	1	4	0	0
Healthy	0	0	1	1	1
Leaf mold	0	3	2	0	0

The high accuracy on the PlantVillage dataset demonstrates the effectiveness of the VGG16 architecture under controlled dataset conditions [11]. However, the sharp performance decline when tested on local images reveals a significant limitation [12] in generalization. This supports literature findings that AI models must incorporate diverse and localized training data to ensure robust real-world performance [13]. Future research should focus on collecting and integrating local datasets [14], applying domain adaptation techniques, and continuously fine-tuning models to match field conditions [15].



Fig 5. Prediction System

IV. CONCLUSIONS

This research demonstrated that the VGG16 convolutional neural network model, when trained exclusively on the PlantVillage dataset, is highly effective in classifying tomato leaf diseases under standardized conditions, achieving high accuracy, precision, recall, and F1-score metrics. However, when applied to a locally sourced dataset from an Indonesian tomato plantation, the model's performance declined sharply, revealing its limited ability to generalize across different domains.

The findings highlight a crucial insight: models trained on globally available datasets may not be directly applicable to real-world, localized agricultural settings without adaptation. Therefore, to develop a robust and practical disease detection system for Indonesian farmers, future implementations should incorporate localized training data, employ domain adaptation strategies, and consider continuous updates based on field data.

Ultimately, this study underscores the importance of contextualizing AI-driven agricultural tools for local environments and serves as a foundation for more inclusive, field-ready plant disease detection systems that bridge the gap between research and on-ground application.

REFERENCES

- [1] Schmey, T., Tominello-Ramirez, C. S., Brune, C., & Stam, R. (2024). Alternaria diseases on potato and tomato. *Molecular Plant Pathology*, 25(3). <https://doi.org/10.1111/mpp.13435>
- [2] Ali, M., Salma, M., Haji, M. E., & Jamal, B. (2025). Plant disease detection using vision transformers. *International Journal of Electrical and Computer Engineering (IJECE)*, 15(2), 2334. <https://doi.org/10.11591/ijece.v15i2.pp2334-2344>
- [3] Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., Su, C., & Wei, W. (2025). Advances in Deep Learning Applications for Plant Disease and Pest Detection: A Review. *Remote Sensing*, 17(4), 698. <https://doi.org/10.3390/rs17040698>
- [4] Nurafiya, N., & Chandra, A. Y. (2024). Analisis Performa Akurasi Klasifikasi Citra Jenis Sayur Salada Menggunakan Arsitektur VGG16, Xception dan NasNetMobile. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(3), 1240. <https://doi.org/10.30865/mib.v8i3.7661>
- [5] Mohanty, S.P., Hughes, D., Salathe, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. *arXiv preprint arXiv:1604.03169*.
- [6] Paymode, A. S., & Malode, V. B. (2022). Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG. *Artificial Intelligence in Agriculture*, 6, 23–33. <https://doi.org/10.1016/j.aiaa.2021.12.002>
- [7] Sinamenye, J. H., Chatterjee, A., & Shrestha, R. (2025). Potato plant disease detection: leveraging hybrid deep learning models. *BMC Plant Biology*, 25(1). <https://doi.org/10.1186/s12870-025-06679-4>
- [8] Al Sahili, Z., & Awad, M. (2022). The power of transfer learning in agricultural applications: AgriNet. *Frontiers in*

- Plant Science, 13. <https://doi.org/10.3389/fpls.2022.992700>.
- [9] Rahman, K. N., Banik, S. C., Islam, R., & Fahim, A. A. (2025). A real time monitoring system for accurate plant leaves disease detection using deep learning. *Crop Design*, 4(1), 100092. <https://doi.org/10.1016/j.crope.2024.100092>.
- [10] Govoruhina, A., & Nikiforova, A. (2022). Digital health shopping assistant with React Native: a simple technological solution to a complex health problem. In 2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA) (pp. 34–40). 2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA). IEEE. <https://doi.org/10.1109/idsta55301.2022.9923047>
- [11] Noyan, M.A. (2022). Uncovering bias in the PlantVillage dataset. arXiv preprint arXiv:2206.04374..
- [12] Fenu, G., & Mallocci, F. M. (2022). Evaluating Impacts between Laboratory and Field-Collected Datasets for Plant Disease Classification. *Agronomy*, 12(10), 2359. <https://doi.org/10.3390/agronomy12102359>
- [13] Hanna, M. G., Pantanowitz, L., Dash, R., Harrison, J. H., Deebajah, M., Pantanowitz, J., & Rashidi, H. H. (2025). Future of Artificial Intelligence—Machine Learning Trends in Pathology and Medicine. *Modern Pathology*, 38(4), 100705. <https://doi.org/10.1016/j.modpat.2025.100705>
- [14] Kouw, W.M., Loog, M. (2018). An introduction to domain adaptation and transfer learning. arXiv preprint arXiv:1812.11806.
- [15] Dong, X., Bao, J., Zhang, T., Chen, D., Gu, S., Zhang, W., Yuan, L., Chen, D., Wen, F., Yu, N. (2022). CLIP Itself is a Strong Fine-tuner: Achieving 85.7% and 88.0% Top-1 Accuracy with ViT-B and ViT-L on ImageNet. arXiv preprint arXiv:2212.06138

