

Fine-Tuned YOLOv12 for Robust Cassava Leaf Disease Detection Under Occlusion and Complex Background Conditions

Junita Amalia*, Dolok Butar-butur, Ira Silalahi, and Sefanya Sinaga

Department of Information Systems, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Laguboti, Indonesia

*Corresponding author: junita.amalia@del.ac.id

Abstract. Cassava is a major food commodity in Indonesia, and its production is threatened by various leaf diseases that can reduce yields by up to 95%. Manual identification by farmers is often subjective and time-consuming, while conventional deep learning models frequently suffer a significant drop in accuracy under field conditions with complex backgrounds and occlusion. This study proposes a fine-tuned YOLOv12-based approach for cassava leaf disease detection under challenging field conditions. The dataset was built from two public Kaggle repositories, the main one containing 21,397 images across five classes. A whole-leaf bounding-box annotation strategy was adopted to provide stronger morphological context for detecting disease symptoms under occluded and visually complex conditions. Hyperparameters were optimized by Bayesian optimization with the tree-structured Parzen estimator (TPE) in the Optuna framework. The fine-tuned YOLOv12m model achieved a mean average precision (mAP) of 0.808 at an intersection over union (IoU) threshold of 0.5 (mAP50) and of 0.600 over IoU thresholds from 0.5 to 0.95 (mAP50-95), outperforming the YOLOv12s baseline. Although the computational load increased from 21.5 to 67.8 giga floating-point operations (GFLOPs), the model was considerably better at distinguishing disease features from field background noise. These findings indicate that the proposed approach improves detection performance and resilience for cassava leaf disease detection in complex field environments.

Keywords. YOLOv12; cassava leaf disease; occlusion; complex background; accuracy-efficiency trade-off

I. INTRODUCTION

Cassava (*Manihot esculenta*) is a staple food commodity and industrial raw material in Indonesia, and it supports progress toward the Sustainable Development Goal (SDG) of zero hunger [1]. This strategic potential, however, faces a severe threat from pathogens such as fungi, bacteria, and viruses, which attack the leaves and can reduce yields by 12% to 95% [2]. Farmers currently rely heavily on manual identification, which depends on subjective experience and on the assistance of plant pathologists. This approach is time-consuming and labor-intensive, and its subjectivity makes it prone to misdiagnosis, which often results in delayed or mistargeted disease management [3].

Although deep learning studies [4], [5] have classified plant diseases with high accuracy under controlled conditions, a significant performance gap emerges when these models are applied to field data. Accuracy can drop sharply; one study [5] reported a decline from 95% in training to 76.59% in testing, mainly because field images are far more complex than laboratory images. The main technical challenges are cluttered backgrounds, illumination variation, and occlusion, in which leaves overlap one another. These factors make it difficult for models to localize disease symptoms precisely and substantially degrade detection accuracy in real agricultural settings.

To overcome these constraints, this study applies YOLOv12, a state-of-the-art object detection architecture.

YOLOv12 introduces an attention mechanism designed to sustain real-time performance while remaining resilient to occlusion and to positional uncertainty of objects in dense images [6]. Model performance was evaluated with mean average precision (mAP) for localization accuracy and floating-point operations (FLOPs) for computational efficiency, with the aim of showing that YOLOv12 addresses the accuracy degradation caused by complex backgrounds and occlusion.

The main contributions of this study are as follows. First, it evaluates the ability of YOLOv12, particularly its area attention mechanism, to detect cassava leaf diseases under challenging field conditions involving occlusion, positional uncertainty, and cluttered backgrounds. Second, it introduces a whole-leaf bounding-box annotation strategy that uses the entire leaf, rather than only the visible symptom area, as the detection target, so that the model captures broader leaf-level context and misses fewer detections when leaves overlap. Third, it applies Bayesian optimization with the Optuna tree-structured Parzen estimator (TPE) to identify optimal hyperparameters and augmentation settings, and it analyzes the trade-off between detection accuracy and computational efficiency for potential deployment on edge devices in agriculture.

II. THEORETICAL FRAMEWORK

A. YOLOv12 Architecture

YOLOv12 is a recent development in real-time object detection that adopts an attention-centric design. Unlike earlier generations, which relied on standard convolutions, YOLOv12 uses an area attention mechanism to capture global spatial relationships efficiently [7], [8].

B. Object Detection in Field Conditions

Object detection in agricultural environments must cope with occlusion (partly hidden objects) and with complex backgrounds such as soil, weeds, or shadows. Previous research indicates that deep learning models often lose performance when visual disease features overlap with background elements [4], [9]. A model that is resilient to such visual disturbances is therefore required to minimize misdiagnosis.

C. Bayesian Optimization

Bayesian optimization is a method for finding the optimal parameters of a complex function whose internal form is unknown, often called a black-box function. It builds a probabilistic surrogate model that predicts the most promising hyperparameter combinations from previous evaluation results. This approach has proven far more efficient in computational time than conventional methods such as grid search or random search [10], [11].

D. Object Detection Evaluation Metrics

Model performance was evaluated with mAP as the primary measure of the balance between precision and recall. The mAP50 metric, computed at an intersection over union (IoU) threshold of 0.5, validates basic detection accuracy, and mAP50-95, averaged over IoU thresholds from 0.5 to 0.95, measures spatial precision more rigorously. Together, these metrics allow an objective assessment of the model's ability to distinguish types of cassava leaf disease under complex environmental conditions [12]-[14]. FLOPs were used to measure computational load. Measuring FLOPs is essential for analyzing the trade-off between higher accuracy and the greater feature representation capacity of larger architectures [15].

III. DATA COLLECTION AND ANALYSIS

A. Dataset Preparation

This study combined secondary data from two public Kaggle repositories, Cassava Leaf Disease Classification and iCassava 2019 Fine-Grained Visual Categorization Challenge. The Kaggle datasets were developed by the Makerere University AI Lab and the National Crops Resources Research Institute (NaCRRI), Uganda, from cassava leaf images captured by local farmers with mobile phones. The main dataset contains 21,397 images from five classes at a resolution of 800×600 pixels, and the iCassava 2019 dataset provides 9,436 additional images of varying resolution, which were used to address class imbalance.

B. Whole-Leaf Annotation Strategy

Bounding boxes were annotated manually on the Roboflow platform with a strategy that covers the entire leaf structure:

each bounding box encloses the complete leaf that displays disease symptoms. This strategy gives the model full morphological context and enables consistent localization under occlusion. A single image may contain several bounding boxes of different classes. After visual screening to remove blurred images and outliers, 9,511 images with a total of 18,546 bounding boxes were obtained in five classes: cassava bacterial blight (CBB), cassava brown streak disease (CBSD), cassava green mottle (CGM), cassava mosaic disease (CMD), and healthy. The dataset was split 70:20:10 into training, validation, and test sets, and the test set was kept fully isolated from training.

C. Model Architecture and Hyperparameter Optimization

Two experimental scenarios were run. In the first, a baseline YOLOv12s model was trained with the default Ultralytics hyperparameters for 100 epochs, with a batch size of 16 and an input resolution of 640×640 pixels. In the second, YOLOv12m was fine-tuned with the Adam optimizer. Hyperparameters were optimized by Bayesian optimization with the TPE algorithm in the Optuna framework over 15 trials. Each trial trained the model for 60 epochs with an early-stopping patience of 20 epochs, and the objective was to maximize mAP50-95 on the validation set.

IV. RESULTS AND DISCUSSION

Before the final performance evaluation, Bayesian optimization was used to search for the optimal hyperparameter configuration, and the resulting parameters were applied in the YOLOv12m fine-tuning scenario. The search ranges for the augmentation parameters and the training hyperparameters are listed in Tables 1 and 2.

TABLE 1. AUGMENTATION PARAMETERS AND SEARCH RANGES

Augmentation parameter	Value range
hsv_h	0.0 to 0.02
hsv_s	0.5 to 0.8
hsv_v	0.1 to 0.5
degrees	10.0 to 20.0
translate	0.05 to 0.2
scale	0.4 to 0.9
mosaic	0.5 to 1.0
fliplr	0.3 to 0.6
mixup	0.0 to 0.15
copy_paste	0.0 to 0.3

TABLE 2. TRAINING HYPERPARAMETERS AND SEARCH RANGES

Hyperparameter	Value range
Initial learning rate (lr0)	1e-4 to 5e-3
Final learning rate factor (lrf)	0.01 to 0.3
momentum	0.85 to 0.98
weight_decay	1e-5 to 1e-3
df	1.0 to 2.0
box	6.0 to 10.0
cls	0.5 to 2.0
warmup_epochs	1.0 to 5.0

The baseline model was then compared with the optimized model obtained with the TPE algorithm. The evaluation results are given in Table 3.

TABLE 3. PERFORMANCE OF THE BASELINE AND OPTIMIZED YOLOv12 MODELS

Evaluation metric	YOLOv12s baseline	YOLOv12m fine-tuned
Precision	0.765	0.792
Recall	0.721	0.748
mAP50	0.784	0.808
mAP50-95	0.582	0.600
GFLOPs	21.50	67.80

The fine-tuned YOLOv12m model improved consistently on all evaluation metrics over the YOLOv12s baseline. The mAP50 increased from 0.784 to 0.808 and the mAP50-95 from 0.582 to 0.600, a relative increase of 3.09% in mAP50-95. This gain is comparable to that reported for CFNet-YOLOv12 [16], where integrating the coordinate attention (CoordAtt) module into YOLOv12 increased mAP50-95 by 2.43% over the base model on self-collected data (and by 2.59% on open-source datasets). These results suggest that the proposed optimization strategy gives a competitive performance gain in detecting complex plant disease patterns under field conditions.

The improvement can be attributed to two main factors. First, the YOLOv12m architecture has a larger representational capacity than YOLOv12s, which allows the model to learn more complex and diverse visual features. This is particularly useful for subtle disease symptoms, such as the small necrotic spots of CBB, which may overlap with the natural texture of the leaf surface. Second, the augmentation and hyperparameter settings optimized by Bayesian optimization produced a training setup that better represents field conditions. Geometric augmentations such as scaling and translation simulate variation in camera position and distance, and mosaic augmentation simulates the density of objects within a single frame. However, the comparison set a default YOLOv12s baseline against a tuned YOLOv12m model, so the observed gains mix the effects of model size and hyperparameter tuning; no ablation study was run to separate them.

The increase in precision from 0.765 to 0.792 indicates that the fine-tuned model became more selective in confirming detections and was less likely to misclassify complex background patterns as disease features. The increase in recall from 0.721 to 0.748 shows that the model detected more of the disease objects present in the images, which matters for partial symptoms, occlusion, and overlapping leaves. The simultaneous gain in precision and recall suggests that the YOLOv12m configuration achieved a better balance between reducing false positives and minimizing missed detections.

The performance gain, however, came with a substantial increase in computational cost, from 21.5 to 67.8 GFLOPs, mainly because of the larger architectural scale of YOLOv12m. Still, the consistent improvement across the evaluation metrics indicates that the greater feature representation capacity of YOLOv12m brings practical benefits for the high visual variability of cassava leaf disease images captured in the field.

These results indicate that the fine-tuned YOLOv12m model detected disease more reliably under challenging

field conditions, particularly in images affected by occlusion, overlapping leaves, subtle disease symptoms, and complex background noise.

V. CONCLUSION

This study showed the potential of a fine-tuned YOLOv12 model, combined with a whole-leaf annotation strategy and hyperparameter tuning by Bayesian optimization, for improving cassava leaf disease detection under field conditions with occlusion and complex backgrounds. The fine-tuned YOLOv12m model achieved an mAP50 of 0.808 and an mAP50-95 of 0.600, outperforming the YOLOv12s baseline, which achieved 0.784 and 0.582, respectively. Precision and recall also improved consistently. This gain came with a computational trade-off, as the load increased from 21.5 to 67.8 GFLOPs, which must be considered for practical implementation. Future research should expand the dataset with variation in weather, cassava varieties, lighting, and field backgrounds to further improve the model's generalization to real agricultural environments.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Institute for Research and Community Service (LPPM) of Institut Teknologi Del for providing funding support and research facilities for this study.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and preparation of this work and approved the final version of the manuscript.

FUNDING

This work was supported by the Institute for Research and Community Service (LPPM) of Institut Teknologi Del, which provided funding and research facilities.

DECLARATION OF GENERATIVE AI USE

The authors declare that no generative artificial intelligence tools were used in the preparation of this manuscript. The authors take full responsibility for the content of the publication.

DATA AVAILABILITY

This study used the publicly available Cassava Leaf Disease Classification and iCassava 2019 Fine-Grained Visual Categorization Challenge datasets (Kaggle).

ETHICS STATEMENT

This study used publicly available, de-identified data and did not involve direct interaction with human participants or animals; ethical approval was therefore not required.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] A. W. Muhaimin, Hery Toiba, Dwi Retnoningsih, and L. M. Yapanto, "The impact of technology adoption on income and food security of smallholder cassava farmers: Empirical evidence from Indonesia," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 9s, pp. 699-707, 2020. [Online]. Available: <http://serc.org/journals/index.php/IJAST/issue/view/282>
- [2] Nasir Saleh, Didik Harnowo, and I Made Jana Mejaya, *Penyakit Penyakit Penting pada Ubi Kayu: Deskripsi, Bioekologi, dan Pengendaliannya*. Malang, Indonesia: Balai Penelitian Tanaman Aneka

- Kacang dan Umbi, 2016, ISBN: 978-602-95497-9-9. [Online]. Available: <https://repository.pertanian.go.id/handle/123456789/20356>
- [3] Xuwei Wang, Jun Liu, and Guoxu Liu, "Diseases detection of occlusion and overlapping tomato leaves based on deep learning," *Front. Plant Sci.*, vol. 12, Art. no. 792244, 2021, doi: 10.3389/fpls.2021.792244.
- [4] Emmanuel Moupojou, Appolinaire Tagne, Florent Retraint, Anicet Tadonkemwa, Dongmo Wilfried, Hyppolite Tapamo, and Marcellin Nkenlifack, "FieldPlant: A dataset of field plant images for plant disease detection and classification with deep learning," *IEEE Access*, vol. 11, pp. 35398-35410, 2023, doi: 10.1109/ACCESS.2023.3263042.
- [5] Sk Mahmudul Hassan and Arnab Kumar Maji, "Plant disease identification using a novel convolutional neural network," *IEEE Access*, vol. 10, pp. 5390-5401, 2022, doi: 10.1109/ACCESS.2022.3141371.
- [6] Ranjan Sapkota, Zhichao Meng, Martin Churuvija, Xiaoqiang Du, Zenghong Ma, and Manoj Karkee, "Comprehensive performance evaluation of YOLOv12, YOLO11, YOLOv10, YOLOv9 and YOLOv8 on detecting and counting fruitlet in complex orchard environments," *Agric. Commun.*, vol. 4, no. 1, Art. no. 100125, 2026, doi: 10.1016/j.agrcom.2026.100125.
- [7] Yanpeng Ji, Tianxiang Ma, Hongliang Shen, Haiyan Feng, Zizi Zhang, Dan Li, and Yuling He, "Transmission line defect detection algorithm based on improved YOLOv12," *Electronics*, vol. 14, no. 12, Art. no. 2432, 2025, doi: 10.3390/electronics14122432.
- [8] A. Chandrashekhara, B. Satyanarayana, Rajani Reddy Gorrepati, Ponduri Vasanthi, and Kothala Lakshmi Prasanna, "An efficient YOLOv12-based framework for detecting extremely small-scale objects," *Sci. Rep.*, vol. 16, Art. no. 2062, 2026, doi: 10.1038/s41598-025-31803-7.
- [9] Muhammad Shafay, Taimur Hassan, Muhammad Owais, Irfan Hussain, Sajid Gul Khawaja, Lakmal Seneviratne, and Naoufel Werghi, "Recent advances in plant disease detection: Challenges and opportunities," *Plant Methods*, vol. 21, Art. no. 140, 2025, doi: 10.1186/s13007-025-01450-0.
- [10] Ultralytics, "YOLO hyperparameter tuning," Ultralytics YOLO Docs, 2025. [Online]. Available: <https://docs.ultralytics.com/guides/hyperparameter-tuning/>
- [11] Ultralytics, "Model training with Ultralytics YOLO," Ultralytics YOLO Docs, 2025. [Online]. Available: <https://docs.ultralytics.com/modes/train/>
- [12] Qizhao Chen, "Traffic object detection using YOLOv12," *Open Access Libr. J.*, vol. 12, no. 8, pp. 1-15, 2025, doi: 10.4236/oalib.1113991.
- [13] Haitao Wu, Xiaotian Mo, Sijian Wen, Kanglei Wu, Yu Ye, Yongmei Wang, and Youhua Zhang, "DNE-YOLO: A method for apple fruit detection in diverse natural environments," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 36, no. 9, Art. no. 102220, 2024, doi: 10.1016/j.jksuci.2024.102220.
- [14] Ultralytics, "What is intersection over union (IoU)?," Ultralytics Glossary, 2025. [Online]. Available: <https://www.ultralytics.com/glossary/intersection-over-union-iou>
- [15] Jierun Chen, Shiu-hong Kao, Hao He, Weipeng Zhuo, Song Wen, Chul-Ho Lee, and S.-H. Gary Chan, "Run, don't walk: Chasing higher FLOPS for faster neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 12021-12031, doi: 10.1109/CVPR52729.2023.01157.
- [16] Laixiang Xu, Wenhao Wang, Zhaopeng Cai, Wenwen Huang, Fengyan Zhai, and Junmin Zhao, "CFNet-YOLOv12: A novel deep learning framework for apple leaf disease detection," *Hortic. Plant J.*, in press, 2026, doi: 10.1016/j.hpj.2025.09.017.