Jurnal Penelitian Hutan Tanaman Vol. 23 No. 1, Juni 2026, 1-12
ISSN: 829-6327, E-ISSN: 2442
Terakreditasi No: 158/E/KPT/2021

Early Detection of Chili Leaf Diseases in AI-Based Agroforestry Systems Using Convolutional Neural Networks

Deteksi Dini Penyakit Daun Cabai pada Sistem Agroforestri Berbasis Kecerdasan Buatan Menggunakan Jaringan Saraf Konvolusional

Zharifah Puspita Candini[1], Dyah Aruming Tyas[2 *], and/*dan* Erianto Indra Putra[3]

[1]Undergraduate Program in Electronics and Instrumentation, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara Bulaksumur, Yogyakarta, Indonesia, 55281
[2]Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara Bulaksumur, Yogyakarta, Indonesia, 55281
[3]Department of Silviculture, Faculty of Forestry and Environment IPB University Jl. Ulin Kampus IPB Darmaga, Jawa Barat, Indonesia, 16680
* E-mail: dyah.aruming.t@ugm.ac.id

Tanggal diterima: 27 November 2025; Tanggal disetujui: 24 Desember 2025; Tanggal direvisi: 29 Januari 2026

Abstract

Early detection of diseases in chili plants is crucial to prevent yield reduction. However, its implementation in the field still faces challenges, particularly related to variations in lighting, leaf orientation, and the limitations of low-cost camera devices. This study proposes an integrated detection framework that combines image acquisition with an ESP32-CAM, transmission via MQTT, and real-time inference in Python using a fine-tuned ResNet-18 model specifically tailored to withstand noise in real-world conditions. The primary objective of this research is to evaluate how domain-aligned fine-tuning can enhance the model's generalization compared to an older, unoptimized model, especially when tested across variations in texture, pigmentation, and leaf angles. Observational evaluations over four days using two different leaf subsets demonstrate that the fine-tuned model consistently outperformed the older model in overall accuracy, inter-class stability, and resilience against leaf position changes. Real-time implementation using the Telegram Bot API successfully delivered swift classification notifications and images, demonstrating the system's viability for remote plant health monitoring. These results underscore the importance of targeted fine-tuning in bridging the gap between laboratory-based CNN models and field-ready disease detection systems. Furthermore, given that chili is widely cultivated in agroforestry systems in Indonesia, the developed early disease detection framework offers significant benefits in maintaining productivity under challenging microclimate conditions that complicate manual diagnosis.

Keywords: *Agriculture, CNN, disease detection, ESP32-CAM, ResNet-18*

Abstrak

Deteksi dini penyakit pada tanaman cabai sangat penting untuk mencegah penurunan hasil panen. Namun, implementasinya di lapangan masih terkendala oleh variasi pencahayaan, orientasi daun, dan keterbatasan perangkat kamera murah. Penelitian ini mengusulkan kerangka deteksi terpadu yang memanfaatkan ESP32-CAM untuk akuisisi citra, penggunaan MQTT untuk transmisi, dan inferensi real-time dengan Python menggunakan model ResNet-18 yang telah disesuaikan agar tahan terhadap noise dalam kondisi nyata. Tujuan utama adalah untuk mengevaluasi sejauh mana fine-tuning yang sesuai dengan domain dapat meningkatkan kemampuan generalisasi model dibandingkan model lama yang tidak dioptimasi, terutama dalam menghadapi variasi tekstur, pigmentasi, dan sudut pandang daun. Evaluasi selama empat hari dengan dua subset daun menunjukkan bahwa model fine-tuned consistently lebih akurat dan lebih stabil dalam mengklasifikasikan jenis penyakit. Implementasi *dengan Telegram Bot API menyediakan notifikasi klasifikasi dan gambar cepat, membuktikan kelayakan sistem untuk pemantauan kesehatan tanaman jarak jauh. Hasil ini menegaskan bahwa fine-tuning yang tepat sangat penting untuk menciptakan sistem deteksi penyakit yang efektif di lapangan, terutama dalam sistem agroforestri di Indonesia.*

Kata kunci: *CNN, deteksi penyakit, ESP32-CAM, pertanian, ResNet-18*

1. Introduction

Indonesia's agriculture is highly diverse and considered as an agricultural country, where agriculture has long been the backbone of its economy (Syuaib, 2016). Of the 56.5 million hectares of agricultural land, there is approximately 22.4 million hectares used for estate plantations, including chili plantations. Chili cultivation in Indonesia is also widely practiced within agroforestry systems and intercropped with multipurpose tree species such as mention some examples with supporting references. These systems enhance microclimate stability and soil fertility but introduce variability in shading and humidity that makes it difficult for manual disease detection. Successful chili-based agroforestry practices demonstrate increased income diversification but also highlight persistent vulnerability to foliar diseases such as yellowing, leaf curl, leaf spot, and whitefly infestation. Variations in the factors and symptoms causing chili plant diseases lead to false detections by farmers. To minimize the lack of distinguishability, integrating machine learning and artificial intelligence (AI) will enable early disease detection. Early detection supported by AI-based systems thus becomes essential for sustaining productivity in these heterogeneous environments

A major limitation in previous chili disease detection studies is that most proposed CNN-based models remain largely theoretical, with limited validation through hardware-based implementation in real agricultural environments. Existing research predominantly emphasizes algorithmic development while overlooking practical deployment constraints on embedded platforms, leaving many models unverified under field conditions (Mim et al., 2024; Karna et al., 2024). In addition, CNN-based classification of chili leaf diseases remains challenging due to strong visual similarities among crumpled, yellowish, and healthy leaves. Similar difficulties in distinguishing disease symptoms have long been reported in plant pathology studies, where overlapping visual characteristics and environmental influences complicate disease identification (Asmaliyah, 2015; Lelana et al., 2015). Recent computer-vision-based studies further confirm this issue, showing frequent misclassification of yellowing and leaf curl symptoms due to similarities in color and texture (Hasbollah et al., 2020; Li et al., 2024). These challenges are further exacerbated in agroforestry-based chili cultivation systems, where variable shading, diffuse illumination, and fluctuating humidity increase visual ambiguity, underscoring the need for a robust, deployable deep learning-based detection system capable of reliable operation under real-world field conditions.

1.1. Research Objectives

To support chili production within agroforestry systems, this study aims to ensure that the detection model remains reliable under complex field microclimates where illumination, leaf orientation, and background conditions vary widely. The objectives of this research are the following:

1. To evaluate the performance of the ResNet-18 model, specifically

analyzing its effectiveness in reducing misclassification between chili plant leaf diseases.
2. To develop a model on an ESP32-CAM microcontroller with Telegram Bot API for real-time monitoring.

1.2. Literature Review

Recent literature emphasizes that chili cultivation within agroforestry systems presents unique challenges and opportunities for disease dynamics. Agroforestry microclimates characterized by heterogeneous light distribution, elevated humidity under tree canopies, and multilayered vegetation structure can intensify foliar pathogen development while simultaneously reducing pest pressure through enhanced biodiversity (Hairiah et al., 2011; Roshetko et al., 2013). Studies show that Begomovirus-driven leaf curl disease and whitefly populations tend to increase in shaded, humid environments, whereas fungal leaf spots caused by *Alternaria* and *Cercospora* spp. spread more rapidly under prolonged leaf wetness. Conversely, agroforestry systems often support higher natural enemy abundance, which can suppress whitefly and other insect vectors. This dual ecological effect makes disease surveillance more complex and increases the need for automated, sensor-assisted detection frameworks. Furthermore, previous deep learning–based models for chili disease detection, while promising, often rely on homogeneous datasets lacking representation of agroforestry-related environmental variability, limiting their real-world applicability. Recent advances in CNN architectures, transfer learning, and domain-specific fine-tuning have been shown to significantly improve robustness under field conditions (Aminuddin et al., 2022; Pandey et al., 2024). Yet only a few studies have examined their performance in heterogeneous environments such as agroforestry systems, highlighting the importance of developing AI-driven detection tools capable of coping with real-world microclimatic variability.

Rahman (2025) proposed work on ResNet-18 with transfer learning for general plant disease detection across various species. This study uses the Kaggle Plant Disease Classification Merged Dataset by Aline Doborovsky, consisting of 14 sub-datasets and 88 classes with over 17,000 images. The model uses a learning rate value of 0.001, a batch size of 400 images, and several epochs. The study highlights that transfer learning from large datasets significantly reduces overfitting when training on limited datasets. The output of this study reports an accuracy of 96% with a validation loss of 4.54%.

Aminuddin et al. (2022) presented research to examine a data augmentation method known as a geometric transformation on a small chili dataset. The augmented and original images are fed into two deep learning models, Convolutional Neural Network (CNN) and Residual Network (ResNet-18). The chili leaf dataset consists of 1200 images, divided into 600 healthy chili leaves and 600 diseased chili leaves. The hyperparameters used consist of a 32 batch size, 0.0001 learning rate, 30 maximum epochs, 50 testing frequency, and Stochastic Gradient Descent with Momentum (SGDM) optimizer. The average accuracy from the original dataset is 95.8%, and the average from the original with augmented datasets using geometric transformation is 97%.

Pandey et al. (2024) proposed IoT-based crop health monitoring research. This study aims to enhance agricultural productivity and sustainability by integrating machine learning for precision agriculture techniques. They also implement a CNN for disease prediction using ESP32-CAM. The ESP32-CAM tracks the development of crops, sending all collected information wirelessly to the cloud for in-depth examination. The image is captured using the ESP32-CAM, and this information is kept in the device dataset for later analysis, which employs the CNN algorithm. The process starts by generating the data in the ESP32 microcontroller, encoding the picture data, and sending it wirelessly over a Wi-Fi network. The dataset used for the CNN is from the Kaggle dataset, consisting of five classes: healthy, leaf curl, leaf spot, whitefly, and yellowing. The model gives out the highest 63% accuracy over 100 epochs.

2. Methodology

This study adopts a systematic approach to design and evaluates a Convolutional Neural Network (CNN) model based on the ResNet-18 architecture, with four models evaluated with batch normalization and augmentation to reduce misclassification in early chili plant disease detection. This research also adopts an end-to-end machine learning and embedded system approach, combining deep learning-based image classification with real-time IoT deployment.

2.1. Research Location

This research is located in Sindangbarang, Bogor, West Java. Dataset collection was conducted for fine-tuning the ResNet-18 model at Pacet, Cianjur, West Java, and Ciherang, Bogor, West Java.

2.2. Methods

2.2.1. Research Design

This study employs a modular research design combining hardware engineering, dataset acquisition, deep learning model development, and system development. Datasets are taken from both public datasets and manually using the ESP32-CAM setup. These chili leaf images are then preprocessed through labeling, augmentation, and dataset splitting. A ResNet-18 Convolutional Neural Network (CNN) is trained using the processed dataset and optimized through regularization, validation, and fine-tuning. The model is then integrated into an ESP32-CAM Python inference pipeline and MQTT for device-to-device messaging, enabling real-time classification. Detection outputs are delivered to users via the Telegram Bot API.

2.2.2. Hardware Design and System Setup

Hardware development for the real-time disease monitoring system focuses on a controlled hardware environment to ensure consistent image capture. A black box container is assembled with dimensions of 17.5 cm x 17 cm, and the ESP32-CAM is placed inside a 3D-printed box with dimensions of 7.3 cm x 5 cm x 2.4 cm. The ESP32-CAM box is mounted right on top of the black box container, with a height of 10.5 cm above the container. The ESP32-CAM mount is positioned at 80°, with a slight 11° curvature at the tip. The ESP32-Cam is powered by a 10.0 Ah, 38.5 Wh, 3.85 V power bank.

The observation is done with a 4-day schedule to systematically compare fine-tuned and non-fine-tuned models under identical image-capture conditions. Two independent leaf batches were used: the first, consisting of five leaves per class, was used on Days 1-2, and the second, with five different leaves per class, was used on Days 3-4. Each day consisted of four 60-minute cycles, each representing a leaf position. Each cycle produced 50 captures (five classes × five leaves × two model pipelines), yielding 200 images per day.

Across all four days, the total captures will be 800 with 50 different leaves (10 per class), and 80 leaf-position combinations per batch (20 combinations x 5 classes x 5 leaves). Fine-tuned versions of Model A (augmentation) and Model B (augmentation with batch normalization) were evaluated on Days 1 and 3, while non-fine-tuned versions of the same models were observed on Days 2 and 4

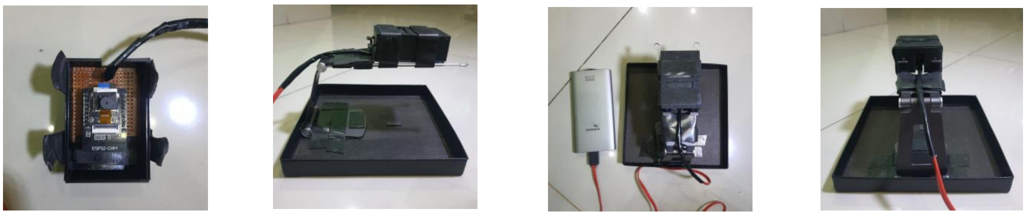


Figure (*Gambar*) 1. ESP32-CAM hardware ((a) ESP32-CAM 3D-printed box, (b) side view, © top view, and (d) back view) (*ESP32-CAM hardware ((a) Kotak 3D print ESP32-CAM, (b) tampak samping, (c) tampak atas, dan (d) tampak belakang*)

Table (*Tabel*) 1. Experimental duration, data capture summary, and leaf sampling configuration *(Durasi penelitian, ringkasan pengambilan data, dan konfigurasi sampel daun)*

Experimental duration *(Durasi penelitian)*	
Total day	4 days
Total time per day	4 hours
Total cycle per day (1 cycle = 1 leaf position)	4 cycles
Total cycle overall	16 cycles
Total time per leaf	1 minute
Total time per model	5 minutes
Total time per cycle	~ 60 minutes
Data capture summary *(Ringkasan pengambilan data)*	
Total capture taken per cycle	50 captures per cycle
Total capture taken per day	200 captures
Total capture overall	800 captures
Leaf Sampling configuration *(Konfigurasi sampel daun)*	
Total leaf per class per day	5 leaves
Total leaf per day	25 leaves/day
Total leaf per class overall	10 leaves
Total leaf overall	50 leaves
Total leaf position combination	20 leaf position combinations
Total combinations per day	100 combinations/day
Total combinations overall	800 combinations

2.2.3. Data Acquisition

The dataset used in this research was collected from several open-source platforms, primarily from Kaggle and Roboflow. One of the primary sources was the Chili Plant Disease Dataset by Dheni Dwi Prakoso on Kaggle, distributed evenly across five categories: healthy, leaf curl, leaf spot, whitefly, and yellowish. In addition, four datasets from Roboflow were used to enrich image diversity and the representation of chili plant diseases. Each dataset provides annotated images of common chili plant diseases under various conditions, lighting, and angles, offering a more general and robust dataset for training and evaluation. In total, the combined datasets yielded the following numbers of images per class: Healthy: 365; Leaf Curl: 398; Leaf Spot: 397; Whitefly: 334; and Yellowish: 338. The overall number of images collected from all sources are 1,832 images.

Chili leaf images are also collected manually using the ESP32-Cam under the configured setup, with only a single leaf placed on a black box container. The acquisition process captures leaves from multiple orientations, producing a diverse dataset. The manual collected dataset is used only for fine-tuning the ResNet-18 models. The total for the manually collected dataset is: Healthy: 629; Leaf Curl: 201; Leaf Spot: 436; Whitefly: 623; and Yellowish: 62. The overall number of manually collected images is 1,951.

2.2.4. Data Preprocessing and Analysis

The preprocessing pipeline made sure all images were standardized before model training, keeping the manually collected ESP32-CAM images consistent with the open-source dataset. Every image was resized to 224 x 224 pixels, rescaled by 1/255, and sorted by disease class. Manually collected images needed to be labeled, while the open-source ones already had labels.

Augmentation is applied during training. The augmentation configuration included: rotation (±30°), width and height shifts (0.2), shear transformations (0.2), zoom range (0.2), horizontal flipping, and nearest-pixel fill mode. These operations simulate illumination changes, leaf orientation variability, and camera misalignment, typically observed in ESP32-CAM field acquisition.

Dataset splitting followed 70% training, 15% validation, and 15% testing structure to balance learning stability and evaluation reliability. The performance analysis incorporated multiple quantitative indicators, including validation accuracy, test accuracy, class-level accuracy, and error distribution patterns as observed in the confusion matrix. Graphs of training accuracy and loss were also used to identify signals of overfitting and to evaluate the effects of augmentation strategies.

2.2.5. ResNet-18 Model Development

Eighteen versions of the ResNet-18 architecture were built to determine the optimal configuration for classifying chili diseases. All models were trained on an Acer Nitro 5 AN515-57 (Intel Core i9, NVIDIA RTX 3060) with Visual Studio Code. Different hyperparameters were tested and compared to see how they affected results. Table 2 lists the hyperparameter variations used.

Figure (*Gambar*) 2. Manual chili dataset samples ((a) healthy chili leaves, (b) leaf curl disease, (c) leaf spot disease, and (d) yellowish disease (*Sampel dataset cabai secara manual((a) daun cabai sehat, (b) penyakit keriting, (c) penyakit bintik, (d) penyakit kutu putih, dan (e) penyakit menguning*)

Table (*Tabel*) 2. Hyperparameter tuning (*Penyetelan hiperparameter)*

Parameter *(Parameter)*	Values *(Nilai)*
Batch Size	32 and 64
Initial Learning Rate	0.001; 0.01; and 0.1
Maximum Epoch	80; 40; and 10

Each hyperparameter combination was tested under four preprocessing conditions: raw, raw with batch normalization, augmented, and augmented with batch normalization. This resulted in a structured experimental grid that enabled direct comparison of augmentation effects and normalization strategies. During training, the optimizer behavior, learning curves, and early-epoch stability were monitored to identify underfitting.

All conditions will use the same ResNet18 backbone, which included a Conv-BN-ReLu stem, four residual stages with 64, 128, 256, and 512 channels (each with two basic blocks), and a two-layer classification head with global average pooling, a dense layer of 256 units, and a dense softmax layer. The training process used helper functions such as ReduceLROnPlateau scheduling, early stopping, model checkpointing, categorical cross-entropy, class weighting, L2 regularization, and optional batch normalization in the head.

The performance of the trained ResNet-18 model is evaluated using several consecutive metrics to ensure both classification effectiveness and deployment feasibility on the ESP32-CAM. The evaluation metrics include validation accuracy, test accuracy, precision, recall, F1-score, and per-class accuracy, along with confusion matrix analysis. Models were also compared using accuracy and loss curves to assess consistency between training and validation trajectories.

From the 72 trained models (generated from 18 hyperparameter configurations x 4 preprocessing variants), the two models with the highest overall test performance were selected. These best-performing models were then used as fine-tuning benchmarks for real-world deployment in manual image acquisition on the ESP32-CAM.

2.2.6. System Integration

The system uses a distributed processing setup in which the ESP32-CAM serves only as an image publisher. It sends each captured frame to an MQTT broker (HiveMQ), keeping communication light and reducing the microcontroller's workload. A Python-based subscriber then collects the images, runs TensorFlow Lite inference, and handles all further processing, such as classification, confidence scoring, and logging. This setup keeps all computation within the Python environment.

2.2.7. Real-Time Detection via Telegram Bot API

A telegram bot is created using BotFather, connected with a unique bot token and chat ID, and added to the inference pipeline to provide real-time notifications. After each prediction, the Python module sends the predicted disease class, confidence score, and timestamp to the Telegram Bot API. This automated reporting system helps monitor plant health remotely and provides users with immediate updates on chili leaf diseases.

In agroforestry systems, the fine-tuned model's robustness to variation in illumination, leaf angles, and background noise is particularly important. Tree canopy structures create heterogeneous lighting that complicates visual assessment. The model's strong performance across these conditions suggests high applicability for chili farmers practicing agroforestry. Early detection of yellowish, leaf curl, leaf spot, and whitefly symptoms supports timely management actions such as nutrient adjustment, vector control, and canopy pruning—helping preserve productivity in diversified land-use systems. (Hairiah et al., 2011; Roshetko et al., 2013; Abdoellah et al., 2006; Sanjaya et al., 2020; Rahman et al., 2016).

3. Results and Discussion

3.1. Results

Seventy-two different ResNet-18 models were created to study how preprocessing methods and hyperparameter settings affect early chili leaf disease recognition. Out of the seventy-two different models, the best model was the Augmented variant, achieving 96.74% validation accuracy and 97.83% test accuracy. The second-best model, the Augmented plus Batch Normalization reached 97.47% test accuracy. Both models showed strong per-class accuracy, addressing previously observed misclassification patterns (model performances in Table 3 through Table 6).

Table (*Tabel*) 3. Best model performance evaluation *(Evaluasi kinerja model terbaik)*

Model Scenarios *(Skenario model)*	Batch Size	LR	Max Epoch	Val Acc (%)	Test Acc (%)	Precision (%)	Recall (%)	F1-Score (%)
Augmented (best)	64	0.001	80	96.74	97.83	97.84	97.83	97.83
Augmented + BN	64	0.001	80	96.74	97.47	97.59	97.47	97.49

Table (*Tabel*) 4. Augmented model evaluation (*Evaluasi model augmented)*

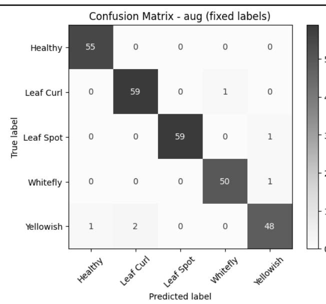
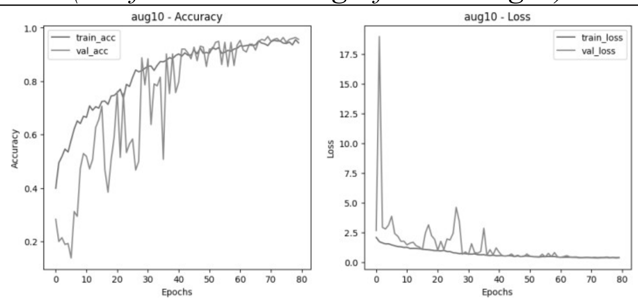
Confusion matrix	Graph accuracy and graph loss *(Grafik akurasi dan grafik kehilangan)*
	

Table (*Tabel*) 5. Augmented with batch normalization model evaluation (*Evaluasi model augmented dengan batch normalisasi)*

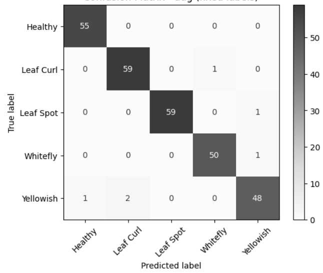
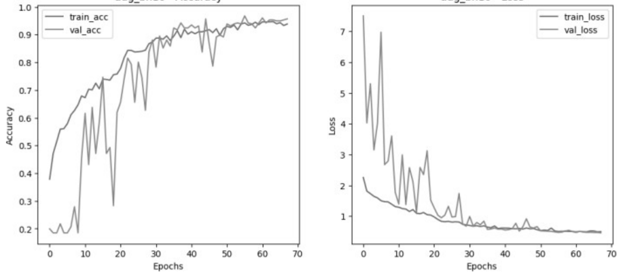
Confusion matrix	Graph accuracy and graph loss *(Grafik akurasi dan grafik kehilangan)*
	

Table (*Tabel*) 6. Per-class accuracy result (*Hasil akurasi per kelas)*

Model scenarios (*Skenario model)*	Per-class accuracy *(Akurasi per kelas)*				
	Healthy *(Sehat)*	Leaf curl *(Keriting)*	Leaf spot *(Bintik)*	Whitefly *(Kutu putih)*	Yellowish *(Kuning)*
Augmented	1.0000	0.9833	0.9833	0.9804	0.9412
Augmented with batch normalization	1.0000	0.9333	0.9833	0.9804	0.9804

When tested in real-world conditions on the ESP32-CAM, both legacy models that were trained only on public datasets made many mistakes. This difference between lab and real-world results is due to a distribution shift. To address the distribution shift, a dedicated manual ESP32-CAM dataset was collected using single-leaf images against a uniform black background. This dataset, acquired at XGA resolution (1024 x 768), aimed to minimize background interference and force the model to focus on leaf morphology rather than contextual noise. Building on this, both top models were fine-tuned using the new dataset. These results are tabulated in Table 7.

During a four-day evaluation, fine-tuned and non-fine-tuned ResNet-18 models were compared in a controlled setting to separate model improvements from sample bias. On Day 1 and 2, the same leaf samples were used. On Day 3 and 4, a different set was used. The fine-tuned models consistently outperformed their older counterparts, achieving 83.00–85.00% accuracy on the first leaf set and up to 98.00% on the second. In contrast, the non-fine-tuned models showed a significant drop in performance, with accuracy ranging from 20.00% to 30.30% and complete failures across several classes.

Results from all four days, as shown in Table 8, indicate that fine-tuned models consistently outperform older models, regardless of leaf samples, disease types, or tested angles.

Table (*Tabel)* 7. Two best fine-tuned models' evaluation (*Evaluasi dua model fine-tune terbaik)*

Model	Dataset	Accuracy	Precision	Recall	F1-Score
Augmented	Train	0.969941	0.971092	0.969941	0.970023
Augmented	Val	0.986254	0.987155	0.986254	0.986450
Augmented	Test	0.969595	0.971699	0.969595	0.970083
Augmented with BN	Train	0.989003	0.989298	0.989003	0.989055
Augmented with BN	Val	0.986254	0.986598	0.986254	0.986335
Augmented with BN	Test	0.979730	0.981193	0.979730	0.980124

Table (*Tabel)* 8. Data results on all models (*Hasil data dari semua model)*

	Model A-FT		Model B-FT		Model A-Old		Model B-Old	
	Day 1	Day 3	Day 1	Day 3	Day 2	Day 4	Day 2	Day 4
Upward 0°	84%	96%	88%	100%	33%	36%	28%	24%
Sideways 90°	96%	88%	88%	96%	24%	24%	32%	20%
Downward 180°	84%	88%	84%	96%	32%	16%	28%	16%
Sideways 270°	76%	88%	72%	100%	32%	44%	16%	20%
Healthy	85%	90%	75%	100%	0%	0%	0%	0%
Leaf Curl	100%	95%	95%	95%	0%	10%	0%	0%
Leaf Spot	45%	65%	55%	95%	65%	65%	85%	95%
Whitefly	100%	100%	90%	100%	89%	55%	45%	5%
Yellowish	100%	100%	95%	100%	0%	15%	0%	0%

3.2. Discussion

The observed performance improvements are primarily attributed to better alignment between the training process and real-world acquisition variability, including viewing angles, illumination fluctuations, and disease-feature uniformity. Models trained on augmented or batch-normalized datasets consistently outperformed their non-augmented counterparts. This confirms that data augmentation improves generalization by exposing CNNs to a variety of geometric and photometric transformations (Shorten & Khoshgoftaar, 2019). Batch normalization enhanced performance by stabilizing feature distributions and expediting convergence, thereby fortifying discriminative feature learning (Ioffe & Szegedy, 2015). However, despite high accuracy in the lab, the first time it was used in the real world, there were many misclassifications. This suggests that there was a distribution shift driven by uncontrolled factors such as uneven lighting, background complexity, and sensor noise.

Fine-tuning the models with field samples helped bridge the gap between the two domains, enabling higher-level feature representations to perform better in the target deployment environment. This process allowed the models to keep low-level visual features while adjusting class-discriminative patterns. This made them much more stable and better able to generalize when data were collected in the real world (Pan & Yang, 2010). Consequently, the proposed methodology significantly reduced misclassification between yellowing and leaf curl diseases, which have been identified as a persistent issue in previous research, thereby directly addressing the research gap related to class ambiguity and insufficient real-world validation. The results show that fine-tuning is an important step for moving CNN-based disease detection systems from controlled lab settings to reliable use in the real world.

In agroforestry-based chili cultivation systems characterized by heterogeneous microclimatic conditions, including variable shading, diffuse illumination, and fluctuating humidity, the enhanced robustness of the fine-tuned models is especially important. The reliable identification of yellowing, leaf curl, leaf spot, and whitefly indicators in these circumstances underscores the practical significance of the proposed framework for enabling prompt interventions and maintaining chili productivity across varied, environmentally dynamic agroforestry settings.

4. Conclusion and Recommendations

4.1. Conclusion

This research demonstrates that fine-tuning a ResNet-18 model for ESP32-CAM images enables reliable early detection of chili plant diseases. Over four days of real-world testing, fine-tuned models consistently outperformed non-fine-tuned models in accuracy, prediction stability, and adaptability to changes in light and viewpoint. The results underscore the need to match training data to sensor-specific noise for effective CNN-based detection systems outside the lab. The combined ESP32-CAM, MQTT, and Python framework performed well, supporting its use for continuous field monitoring.

4.2. Recommendations

Future research should test more agroforestry settings and extend real-world trials to improve reliability in changing conditions. Researchers should also use adaptive learning strategies and lighter on-device inference models to improve scalability and long-term usability.

Acknowledgments

The authors express sincere gratitude to Pak Odo, chili farmer at Pacet, Cianjur West Java and to Evi Oktaviana, nursery practitioner whose cooperation enabled the data collection. Their contributions as technical, practical, and substantive were essential to the completion of this research. This work was supported by the Department of Computer Science and Electronics, Universitas Gadjah Mada under the Publication Funding Year 2026.

References

Abdoellah, O.S., Hadikusumah, H.Y., Takeuchi, K., Okubo, S. (2006). Home gardens in West Java, Indonesia. Agroforestry Systems, 68, 1-23.

Aminuddin, N.F., Tukiran, Z., Joret, A., Tomari, R., & Morsin, M. (2022). An improved deep learning model of chili disease recognition with small dataset. In *IJACSA) International Journal of Advanced Computer Science and Applications*,13(7), 407-412. www.ijacsa.thesai.org

Asmaliyah. (2015). Penyakit-penyakit penting pada tanaman hutan rakyat dan alternatif pengendaliannya. Jurnal Penelitian Hutan Tanaman, 12(2), 141-150.

Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*.

Hairiah, K., Dewi, S., Agus, F. (2011). Agroforestry dan Perubahan Iklim. World Agroforestry Centre (ICRAF).

Islam, M.T.M., Fahim, I.A., Islam, M.N., Rahman, M., Rahman, Y., Rouf, N., Anowar, T.I., & Mondal, S. (2024). IoT-based agriculture farm monitoring system and development of plant leaf disease detection using deep neural network. In 2024 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS). IEEE.

Karna, N., Kelikualiq, R., Aditya, B., Paramartha Putra, M.A., Rahyuni, D., & Hertiana, S.N. (2024). Health monitoring system in smart greenbox for chili plant using convolutional neural network. *Proceedings of 2024 IEEE International Conference on Internet of Things and Intelligence Systems, IoTaIS 2024*, 38-43. https://doi.org/10.1109/IoTaIS64014.2024.10799433

Lelana, N.E., Anggraeni, I., & Mindawati, S. (2015). Uji antagonis *Aspergillus* sp. dan *Trichoderma* spp. terhadap *Fusarium* sp., penyebab penyakit rebah kecambah pada sengon. Jurnal Penelitian Hutan Tanaman, 12(1), 23–28.

Pan, S.J., & Yang, Q. (2010). A survey on transfer learning. In *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359). https://doi.org/10.1109/TKDE.2009.191

Pandey, V., Patil, Y., Kanagaraj, R., Dabhilkar, K., Darade, R., & Patil, H. (2024). CHMP-CNN: To implement the IoT based crops health monitoring and prediction using deep learning. *2024 International Conference on Advances in Computing Research on Science Engineering and Technology, ACROSET 2024*. https://doi.org/10.1109/ACROSET62108.2024.10743388

Rahman, R. (2025). *Using Machine Learning to Identify Crop Diseases with ResNet-18*. 311-315. https://doi.org/10.5220/0013123500003911

Rahman, S.A., Sunderland, T., Roshetko, J.M. (2016). Tree–crop interactions in agroforestry systems. Agricultural Systems, 142, 49-57.

Roshetko, J.M., Nugraha, E., Mulyoutami, E. (2013). Smallholder agroforestry systems in Indonesia. Agroforestry Systems, 87, 729-742.

Sanjaya, Y., Kholil, A., & Dewi, R. (2020). Productivity of chili in agroforestry systems in Garut, Indonesia. Jurnal Agroforestri Indonesia, 3(2), 55-63.

Shorten, C., & Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Syuaib, M.F. (2016). Sustainable agriculture in Indonesia: Facts and challenges to keep growing in harmony with environment. Agricultural Engineering International: CIGR Journal, 18(2), 170-184.