

Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
<https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index>
 Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
 Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
 e-ISSN: 2685-0893

Optimasi *MobileNetV2* untuk Klasifikasi Hama dan Simulasi Dampak Finansial

DOI: [Http://Dx.Doi.Org/10.35889/Jutisi.V15i4.3897](http://Dx.Doi.Org/10.35889/Jutisi.V15i4.3897)

Creative Commons License 4.0 (CC BY – NC)



Sapto Utomo¹, Dwi Agus Diartono²

Informatika, Universitas BPD, Semarang, Indonesia

*e-mail Corresponding Author: saptoutomo@edu.ubpd.ac.id

Abstract

Inaccurate pest identification can delay control measures and increase the risk of agricultural losses. This study evaluated MobileNetV2 for image-based pest classification and simulated the financial impact of prediction errors. The Dangerous Farm Insects Dataset, containing 1,591 images in 15 classes, was used in four scenarios: baseline, augmentation, baseline fine-tuning, and augmentation fine-tuning. The models were assessed using accuracy, precision, recall, macro F1-score, and confusion matrices. The results showed that the baseline achieved the best performance, with an accuracy of 0.750 and a macro F1-score of 0.744. Fine-tuning produced the same accuracy but a slightly lower macro F1-score, while the two augmentation scenarios achieved accuracies of only 0.528 and 0.536. Most errors involved visually similar classes. The financial simulation indicated that the baseline had the lowest relative risk, although the cost values remained illustrative parameters requiring field validation.

Keywords: *MobileNetV2; image classification; agricultural pests; data augmentation; financial impact*

Abstrak

Identifikasi hama yang tidak tepat dapat menunda pengendalian dan meningkatkan risiko kerugian pertanian. Penelitian ini mengevaluasi MobileNetV2 untuk klasifikasi citra hama serta menyimulasikan dampak finansial kesalahan prediksi. Dangerous Farm Insects Dataset yang berisi 1.591 citra dalam 15 kelas digunakan pada empat skenario: *baseline*, *augmentation*, *baseline fine-tuning*, dan *augmentation fine-tuning*. Model dinilai menggunakan *accuracy*, *precision*, *recall*, *macro F1-score*, dan *confusion matrix*. Hasil pengujian menunjukkan bahwa *baseline* memberikan performa terbaik dengan *accuracy* 0,750 dan *macro F1-score* 0,744. *Fine-tuning* menghasilkan *accuracy* yang sama, tetapi *macro F1-score* sedikit lebih rendah, sedangkan dua skenario *augmentation* hanya mencapai *accuracy* 0,528 dan 0,536. Kesalahan terbanyak terjadi pada kelas yang serupa secara visual. Simulasi finansial menunjukkan bahwa *baseline* memiliki risiko relatif terendah, meskipun nilai biaya masih berupa parameter ilustratif dan memerlukan validasi lapangan.

Kata kunci: *MobileNetV2; klasifikasi citra; hama pertanian; data augmentation; dampak finansial*

1. Pendahuluan

Sektor pertanian sangat rentan terhadap ancaman hama yang dapat merusak hasil panen dan menimbulkan kerugian ekonomi [1], [2]. Selama ini, identifikasi manual oleh petani rentan keliru akibat subjektivitas pengamat dan tingginya kemiripan morfologi visual antar spesies hama (seperti *Armyworms* dan *Fall Armyworms*) [2], [3]. Padahal, kesalahan deteksi dapat berujung pada penanganan yang terlambat serta pemborosan biaya pestisida [1], [3]. Oleh karena itu, pengembangan sistem klasifikasi hama berbasis citra yang efisien, seperti pemanfaatan Dangerous Farm Insects Dataset [4], menjadi sangat krusial guna mendukung pengambilan keputusan yang tepat pada sistem pertanian presisi.

Pendekatan *deep learning*, khususnya *Convolutional Neural Network* (CNN) dan *transfer learning*, telah terbukti efektif dalam mengenali objek pertanian secara otomatis [5],[6],[7],[8],[9]. Namun, masih terdapat celah (*gap*) besar pada penelitian-penelitian terdahulu:

evaluasi model hampir seluruhnya hanya bertumpu pada metrik teknis (seperti *accuracy* dan *F1-score*). Sebaliknya, analisis mendalam terkait pola kelas hama yang sering tertukar (*multiclass error*) beserta potensi konsekuensi finansial dari kesalahan prediksi tersebut belum dibahas secara eksplisit.

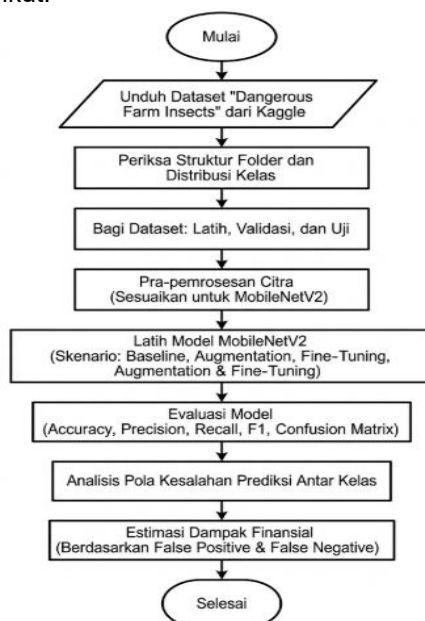
Berangkat dari *gap* tersebut, penelitian ini mengusulkan evaluasi menggunakan MobileNetV2, sebuah arsitektur CNN ringan yang sangat ideal untuk diimplementasikan pada perangkat *mobile* atau *edge computing* di lingkungan lapangan [10],[11]. Kebaruan (*novelty*) dari penelitian ini terletak pada pendekatan evaluasi hibridanya: membandingkan empat skenario optimasi (*baseline*, *augmentation*, dan *fine-tuning*), lalu menilai keberhasilannya tidak sekadar dari angka akurasi, melainkan dengan mengintegrasikan analisis pola kesalahan (*confusion matrix*) dan simulasi dampak kerugian finansial. Melalui pendekatan ini, evaluasi sistem klasifikasi hama menjadi jauh lebih kontekstual dan relevan bagi kebutuhan praktis di bidang pertanian.

Kontribusi utama penelitian ini adalah penyajian evaluasi MobileNetV2 pada *dataset* hama serangga dengan pendekatan yang tidak hanya berbasis metrik teknis, tetapi juga mempertimbangkan konsekuensi finansial dari kesalahan prediksi.

2. Metodologi

2.1 Alur Penelitian

Penelitian ini menggunakan pendekatan eksperimental kuantitatif. Alur penelitian terdiri dari beberapa tahap, yaitu pengumpulan *dataset*, pra-pemrosesan data, pembagian data, pembangunan model MobileNetV2, pelatihan model pada beberapa skenario, evaluasi performa, analisis kesalahan prediksi, dan estimasi dampak finansial. Secara umum, tahapan penelitian adalah sebagai berikut:



Gambar 1. Tahapan Penelitian

2.2 Dataset

Dataset yang digunakan adalah *Dangerous Farm Insects Dataset* yang tersedia secara publik pada Kaggle [4]. *Dataset* ini terdiri dari 1591 citra yang terbagi ke dalam 15 kelas. *Dataset* ini dipilih karena memuat berbagai jenis hama dan gangguan pertanian yang memiliki karakteristik visual beragam, termasuk beberapa kelas yang memiliki kemiripan morfologi.

Kelas pada *dataset* meliputi: *Africanized Honey Bees (Killer Bees)*, *Aphids*, *Armyworms*, *Brown Marmorated Stink Bugs*, *Cabbage Loopers*, *Citrus Canker*, *Colorado Potato Beetles*, *Corn Borers*, *Corn Earworms*, *Fall Armyworms*, *Fruit Flies*, *Spider Mites*, *Thrips*, *Tomato Hornworms*, *Western Corn Rootworms*.

Perlu dicatat bahwa kelas *Citrus Canker* dalam *dataset* merupakan penyakit tanaman, bukan serangga hama. Kelas tersebut tetap dipertahankan karena merupakan bagian dari struktur *dataset* asli, tetapi interpretasi hasil diarahkan pada klasifikasi citra gangguan pertanian secara umum. *Dataset* dibagi menjadi data latih, data validasi, dan data uji dengan rasio 70:15:15. Pembagian dilakukan agar model dapat dilatih, divalidasi, dan diuji pada data yang berbeda. Berdasarkan pembagian tersebut, data uji yang digunakan dalam evaluasi berjumlah 252 citra.

2.3 Pra-pemrosesan Data

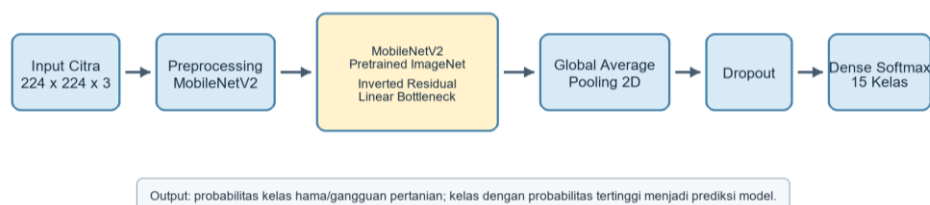
Setiap citra diubah ukurannya menjadi 224 x 224 piksel karena MobileNetV2 menerima input citra dengan ukuran tersebut [10]. Selanjutnya, nilai piksel diproses menggunakan fungsi *preprocessing* bawaan MobileNetV2. Proses ini bertujuan menyesuaikan distribusi nilai piksel dengan format yang digunakan saat MobileNetV2 dilatih pada *ImageNet*.

Data augmentation diterapkan hanya pada skenario MobileNetV2 + *Augmentation* dan MobileNetV2 + *Augmentation* + *Fine-Tuning*, serta hanya digunakan pada data latih. Data validasi dan data uji tidak diberikan *augmentation* agar evaluasi tetap objektif. *Augmentation* yang digunakan meliputi *random horizontal flip*, *random rotation* sebesar 0,04, *random zoom* sebesar 0,05, dan *random contrast* sebesar 0,05. *Data augmentation* umum digunakan untuk meningkatkan variasi data latih, tetapi pemilihan jenis transformasi perlu disesuaikan dengan karakteristik data [12]. Intensitas *augmentation* dibuat rendah karena objek serangga memiliki detail visual yang kecil dan dapat berubah makna apabila transformasi terlalu agresif.

2.4 Arsitektur MobileNetV2

Model yang digunakan adalah MobileNetV2 *pretrained ImageNet* tanpa lapisan klasifikasi bawaan [10]. Lapisan atas model diganti dengan *GlobalAveragePooling2D*, *Dropout*, dan *Dense* dengan aktivasi *softmax* sebanyak 15 neuron sesuai jumlah kelas. Penggunaan *pretrained ImageNet* bertujuan memanfaatkan kemampuan ekstraksi fitur umum yang telah dipelajari sebelumnya, seperti tepi, tekstur, pola warna, dan bentuk dasar objek. Strategi pemanfaatan model *pretrained* ini sejalan dengan konsep *transfer learning*, yaitu memindahkan pengetahuan dari domain sumber ke domain target untuk mengurangi kebutuhan data latih yang besar [13].

Diagram Arsitektur MobileNetV2 untuk Klasifikasi Hama



Gambar 2. Diagram Arsitektur MobileNetV2 untuk Klasifikasi Hama

Diagram tersebut menunjukkan alur pemrosesan citra pada model yang digunakan. Citra masukan berukuran 224 x 224 x 3 terlebih dahulu melalui *preprocessing* MobileNetV2, kemudian fitur visual diekstraksi oleh MobileNetV2 *pretrained ImageNet*. Hasil ekstraksi fitur diringkas menggunakan *GlobalAveragePooling2D*, diberi regularisasi *Dropout*, dan diklasifikasikan melalui *Dense softmax* menjadi 15 kelas hama atau gangguan pertanian.

Tabel 1. Konfigurasi Arsitektur MobileNetV2 untuk Klasifikasi Hama

Tahap	Komponen	Keterangan
1	Input	Citra berukuran 224 × 224 × 3
2	<i>Preprocessing</i>	Fungsi <i>preprocessing</i> bawaan MobileNetV2
3	<i>Feature extractor</i>	MobileNetV2 <i>pretrained ImageNet</i> tanpa <i>top classifier</i>
4	<i>Pooling</i>	<i>GlobalAveragePooling2D</i>
5	Regularisasi	<i>Dropout</i>
6	Klasifikasi	<i>Dense</i> dengan aktivasi <i>softmax</i> sebanyak 15 kelas

2.5 Skenario Eksperimen

Empat konfigurasi MobileNetV2 diuji untuk menilai pengaruh *augmentation* dan *fine-tuning* terhadap performa klasifikasi. Pada skenario *baseline*, *base model* MobileNetV2 dibekukan dan hanya *classifier head* baru yang dilatih menggunakan *optimizer* Adam, *learning rate* 1e-3, *batch size* 32, dan maksimum 25 *epoch*. Konfigurasi ini digunakan sebagai pembandingan dasar. Pada skenario *augmentation*, transformasi citra hanya diterapkan pada data latih. *Base model* tetap dibekukan dan *classifier head* dilatih menggunakan konfigurasi yang sama dengan *baseline* untuk menguji pengaruh peningkatan variasi data terhadap generalisasi model [12]. Pada skenario *baseline fine-tuning*, model *baseline* dilanjutkan dengan membuka lapisan MobileNetV2 mulai indeks ke-130. Lapisan tersebut dilatih ulang menggunakan *learning rate* 3e-6 selama maksimum 20 *epoch* untuk menyesuaikan fitur *pretrained* dengan karakteristik *dataset* target [13]. Pada skenario *augmentation fine-tuning*, model *augmentation* dilanjutkan dengan konfigurasi *fine-tuning* yang sama. Skenario ini digunakan untuk menilai apakah kombinasi *augmentation* dan penyesuaian fitur *pretrained* dapat meningkatkan performa.

Tabel 2. Parameter Pelatihan pada Setiap Skenario Eksperimen

Skenario	Base Model	Augmentati on	Layer Dilatih	Optimiz er	Learning Rate	Epoch Maks.
<i>Baseline</i>	MobileNetV2 pretrained ImageNet	Tidak	<i>Classifier head</i>	Adam	1e-3	25
<i>Augmentatio n</i>	MobileNetV2 pretrained ImageNet	Ya: <i>flip</i> , <i>rotation</i> 0,04, <i>zoom</i> 0,05, <i>contrast</i> 0,05	<i>Classifier head</i>	Adam	1e-3	25
<i>Baseline Fine-Tuning</i>	MobileNetV2 pretrained ImageNet	Tidak	<i>Layer</i> MobileNetV2 mulai <i>indeks</i> 130 + <i>classifier head</i>	Adam	3e-6	20
<i>Augmentatio n Fine-Tuning</i>	MobileNetV2 pretrained ImageNet	Ya: <i>flip</i> , <i>rotation</i> 0,04, <i>zoom</i> 0,05, <i>contrast</i> 0,05	<i>Layer</i> MobileNetV2 mulai <i>indeks</i> 130 + <i>classifier head</i>	Adam	3e-6	20

Parameter pada Tabel 2, digunakan agar setiap skenario eksperimen dapat dibandingkan secara konsisten. Pada tahap *baseline* dan *augmentation*, *feature extractor* MobileNetV2 dibekukan sehingga pelatihan hanya dilakukan pada *classifier head*. Pada tahap *fine-tuning*, sebagian *layer* MobileNetV2 mulai indeks ke-130 dibuka kembali dengan *learning rate* lebih kecil agar penyesuaian fitur tidak merusak bobot *pretrained* secara drastis.

2.6 Evaluasi Model

Evaluasi model dilakukan menggunakan data uji. Metrik evaluasi yang digunakan dihitung dari *confusion matrix*, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). TP menunjukkan data yang benar diklasifikasikan sebagai kelas target, TN menunjukkan data bukan kelas target yang benar diklasifikasikan bukan sebagai kelas target, FP menunjukkan data bukan kelas target yang keliru diprediksi sebagai kelas target, sedangkan FN menunjukkan data kelas target yang gagal dikenali oleh model.

Pada klasifikasi multikelas (15 kelas pada penelitian ini), perhitungan metrik dilakukan menggunakan pendekatan *macro average*, di mana metrik dihitung untuk setiap kelas lalu dirata-ratakan untuk memperoleh nilai evaluasi keseluruhan tanpa terlalu dipengaruhi oleh ketidakseimbangan distribusi data. *Accuracy* digunakan untuk mengukur proporsi prediksi benar terhadap seluruh data uji, *precision* digunakan untuk mengetahui ketepatan model ketika

memprediksi suatu kelas, sedangkan *recall* menunjukkan kemampuan model mengenali data aktual pada kelas tersebut. *F1-score* digunakan karena mampu menyeimbangkan *precision* dan *recall*. Rumus metrik evaluasi yang digunakan adalah sebagai berikut [14]:

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

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

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

$$F1 - Score = 2x \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Selain metrik kuantitatif, *confusion matrix* digunakan untuk melihat pola kesalahan prediksi. *Confusion matrix* membantu mengidentifikasi kelas aktual yang sering diprediksi sebagai kelas lain. Informasi ini penting karena kelas yang sering tertukar dapat menunjukkan adanya kemiripan visual atau ketidakseimbangan kualitas data.

2.7 Estimasi Dampak Finansial

Estimasi dampak finansial dilakukan dengan menghubungkan kesalahan prediksi model dengan potensi biaya yang muncul dalam pengendalian hama. Dua jenis kesalahan yang diperhatikan adalah *false positive* dan *false negative*.

False positive merupakan kondisi ketika model memprediksi adanya kelas hama tertentu secara keliru. Dalam konteks pengendalian hama, kesalahan ini dapat menyebabkan tindakan pengendalian yang tidak diperlukan, seperti penggunaan pestisida atau tenaga kerja tambahan. Penggunaan pestisida memiliki konsekuensi ekonomi dan lingkungan sehingga kesalahan prediksi perlu dipertimbangkan dalam evaluasi sistem [1], [3]. Dampak finansial *false positive* dihitung sebagai biaya pengendalian yang tidak diperlukan.

False negative merupakan kondisi ketika model gagal mengenali hama yang sebenarnya ada atau memprediksi hama sebagai kelas lain yang menyebabkan tindakan pengendalian tidak tepat. Kesalahan ini dapat menyebabkan keterlambatan penanganan dan potensi kehilangan hasil panen. Dampak finansial *false negative* dihitung sebagai potensi kehilangan hasil.

$$\text{Kerugian FP} = \text{jumlah FP} \times \text{biaya pengendalian per kasus} \quad (5)$$

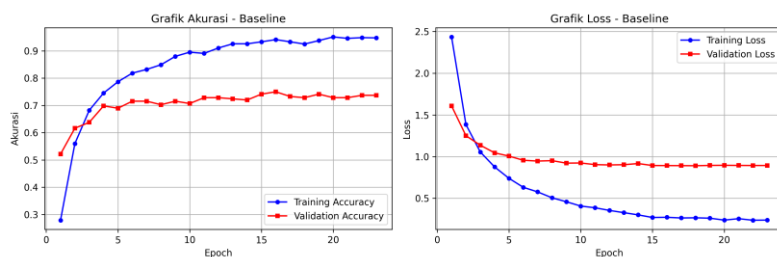
$$\text{Kerugian FN} = \text{jumlah FN} \times \text{estimasi kehilangan hasil per kasus} \quad (6)$$

$$\text{Total kerugian simulatif} = \text{Kerugian FP} + \text{Kerugian FN} \quad (7)$$

Dalam eksperimen ini, biaya FP diasumsikan sebesar Rp25.000 per kasus, biaya FN diasumsikan sebesar Rp100.000 per kasus. Nilai biaya digunakan sebagai parameter skenario yang dapat disesuaikan menurut komoditas, luas lahan, dan kondisi lokal. Data BPS menunjukkan bahwa biaya pestisida per hektar pada 2024 bervariasi signifikan antar-komoditas, yaitu sekitar Rp798 ribu untuk padi, Rp655 ribu untuk jagung, Rp15,59 juta untuk bawang merah, dan Rp12,91 juta untuk cabai merah besar [15]. Karena *dataset* penelitian memuat beragam kelas dan tidak menyediakan luas lahan maupun tingkat serangan, biaya aktual per kesalahan tidak dapat diturunkan langsung. Parameter Rp25.000 untuk *false positive* dan Rp100.000 untuk *false negative* digunakan sebagai skenario ilustratif, bukan pengukuran kerugian aktual.

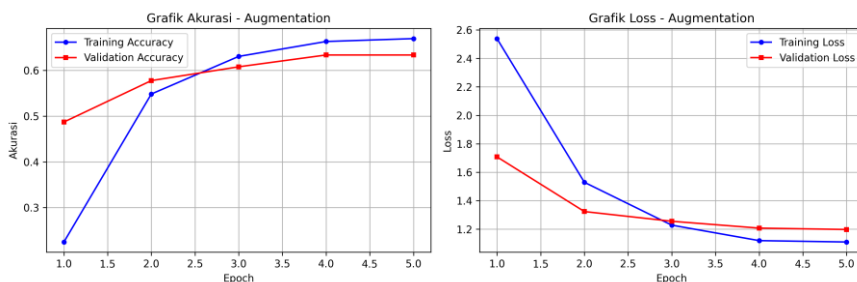
3. Hasil Dan Pembahasan

3.1 Hasil



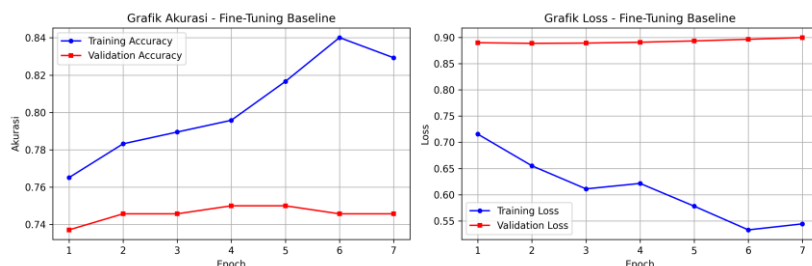
Gambar 3. Kurva pembelajaran akurasi dan *loss* pada skenario MobileNetV2 *Baseline*.

Visualisasi akurasi dan *loss* pada MobileNetV2 *Baseline* (Gambar 3) menunjukkan pola pelatihan yang sangat stabil. Walaupun rentang antara *training accuracy* (sekitar 0,95) dan *validation accuracy* (sekitar 0,74) mengindikasikan *overfitting* moderat akibat kecilnya *dataset*, model tetap mencapai titik konvergensi yang optimal. Ketiadaan lonjakan (*spiking*) pada kurva *validation loss* setelah *epoch* ke-10 mengonfirmasi bahwa model tidak mengalami degradasi generalisasi, menjadikannya skenario terbaik dibandingkan yang lain.



Gambar 4. Kurva pembelajaran akurasi dan *loss* pada skenario MobileNetV2 dengan *Augmentation*.

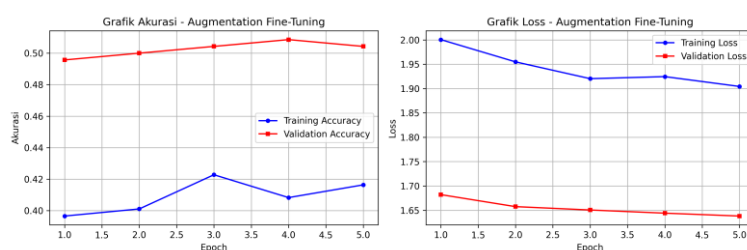
Kurva pembelajaran pada skenario *Augmentation* (Gambar 4) memperlihatkan jarak yang sangat sempit antara garis pelatihan dan validasi, menandakan keberhasilan model dalam menekan *overfitting*. Namun, akurasi maksimal yang dicapai tergolong rendah (sekitar 0,63–0,68) dan kurva dengan cepat mendatar (*underfitting*). Secara visual, grafik ini mempertegas bahwa transformasi citra justru merusak detail visual hama, membuat model kesulitan mengekstrak fitur penting untuk mencapai tingkat akurasi yang tinggi.



Gambar 5. Kurva pembelajaran akurasi dan *loss* pada skenario *Fine-Tuning Baseline*.

Visualisasi kurva pembelajaran pada skenario *Fine-Tuning Baseline* (Gambar 5) menunjukkan performa yang cukup stabil selama 7 *epoch* pelatihan. Meskipun model tidak mengalami kerusakan bobot yang parah (garis kurva tetap bergerak naik dan stabil), proses *fine-tuning* ini tertahan dan gagal mengangkat nilai *validation accuracy* untuk melampaui capaian *baseline* utama. Hal ini membuktikan bahwa membuka lapisan (*unfreezing*) pada

pretrained model tidak selalu memberikan peningkatan performa apabila ukuran *dataset* terlalu kecil untuk mengarahkan kembali bobot fitur secara optimal.



Gambar 6. Kurva pembelajaran akurasi dan *loss* pada skenario *Augmentation Fine-Tuning*.

Kurva pembelajaran pada skenario *Augmentation Fine-Tuning* (Gambar 6) menunjukkan performa terburuk, di mana akurasi langsung terperosok dan mendatar di kisaran 0,40 serta *loss* tertahan tinggi di angka 2,0. Tingkat stagnasi yang parah ini disebabkan oleh kegagalan ganda: sulitnya penangkapan fitur akibat distorsi visual dari *augmentation*, yang kemudian diperparah oleh terjadinya catastrophic forgetting saat lapisan *pretrained* dilatih ulang. Akibat distorsi gambar dan kecilnya *dataset*, model kehilangan memori fitur lamanya dan sepenuhnya gagal beradaptasi untuk membentuk bobot baru, sehingga konvergensi tidak tercapai.

3.1.1 Distribusi Dataset

Dataset yang digunakan terdiri dari 1591 citra dalam 15 kelas hama serangga. Setelah dilakukan pembagian data dengan rasio 70:15:15, jumlah data uji yang digunakan untuk evaluasi adalah 252 citra. Jumlah citra per kelas pada data uji relatif kecil, yaitu sekitar 14 sampai 19 citra per kelas. Kondisi ini perlu diperhatikan karena kesalahan pada beberapa citra dapat memengaruhi nilai metrik secara cukup signifikan. Jumlah data yang terbatas juga menjadi salah satu tantangan utama dalam penelitian ini. Pada *dataset* kecil, model dapat mengalami kesulitan dalam membedakan kelas yang memiliki karakteristik visual mirip. Selain itu, strategi *augmentation* dan *fine-tuning* perlu dilakukan secara hati-hati agar tidak menyebabkan *overfitting* atau penurunan performa.

3.1.2 Hasil Pelatihan Model

Evaluasi terhadap empat skenario pelatihan menunjukkan bahwa MobileNetV2 *baseline* (hanya melatih *classifier head*) merupakan konfigurasi paling stabil dan terbaik dengan *accuracy* 0,750 dan *macro F1-score* 0,744. Sebaliknya, penambahan *data augmentation* justru menurunkan performa (*accuracy* 0,528; *macro F1-score* 0,508) karena transformasi citra (rotasi, zoom) diduga merusak fitur morfologi halus yang penting pada identifikasi hama. Strategi *baseline fine-tuning* juga tidak memberikan peningkatan berarti (*accuracy* 0,750; *macro F1-score* 0,742), mengindikasikan bahwa pembaruan bobot *pretrained* pada *dataset* kecil rentan tidak optimal. Hal serupa terlihat pada skenario gabungan *augmentation fine-tuning* yang performanya tetap di bawah *baseline* (*accuracy* 0,536; *macro F1-score* 0,513), mempertegas bahwa kombinasi optimasi tidak otomatis berhasil jika ukuran dan variasi data kurang merepresentasikan kondisi objek.

Tabel 3. Ringkasan hasil pelatihan dan evaluasi akhir setiap skenario

Skenario	Strategi Pelatihan	Augmentation	Fine-Tuning	Accuracy Uji	Macro F1-Score
MobileNetV2 Baseline	<i>Classifier head</i> dilatih; <i>feature extractor</i> dibekukan	Tidak	Tidak	0,750	0,744
MobileNetV2 + Augmentation	<i>Classifier head</i> dilatih dengan variasi citra latih	Ya	Tidak	0,528	0,508

Skenario	Strategi Pelatihan	Augmentation	Fine-Tuning	Accuracy Uji	Macro F1-Score
MobileNetV2 + <i>Baseline Fine-Tuning</i>	<i>Classifier head</i> dan <i>layer</i> mulai <i>indeks</i> 130 dilatih ulang	Tidak	Ya	0,750	0,742
MobileNetV2 + <i>Augmentation</i> + <i>Fine-Tuning</i>	<i>Augmentation</i> digabung dengan <i>fine-tuning layer</i> akhir	Ya	Ya	0,536	0,513

Ringkasan hasil pelatihan tersebut menunjukkan bahwa konfigurasi *baseline* merupakan skenario paling stabil, sedangkan *augmentation* dan *fine-tuning* perlu disesuaikan lebih lanjut agar dapat memberikan peningkatan performa.

3.1.3 Perbandingan Performa Model

Hasil evaluasi menunjukkan bahwa MobileNetV2 *baseline* memperoleh performa terbaik dengan *accuracy* sebesar 0,750, *precision macro* sebesar 0,747, *recall macro* sebesar 0,748, dan *F1-score macro* sebesar 0,744. Skenario *baseline fine-tuning* menghasilkan *accuracy* yang sama, tetapi *F1-score* tanpa pembulatan sedikit lebih rendah, yaitu 0,742 dibandingkan 0,744 pada *baseline*. *Data augmentation* menurunkan *accuracy* menjadi 0,528, sedangkan penambahan *fine-tuning* pada skenario *augmentation* hanya meningkatkannya menjadi 0,536.

Tabel 4. Perbandingan performa model

Skenario	Accuracy	Precision Macro	Recall Macro	F1-Score Macro
MobileNetV2 <i>Baseline</i>	0,750	0,747	0,748	0,744
MobileNetV2 + <i>Augmentation</i>	0,528	0,557	0,520	0,508
MobileNetV2 + <i>Baseline Fine-Tuning</i>	0,750	0,745	0,748	0,742
MobileNetV2 + <i>Augmentation</i> + <i>Fine-Tuning</i>	0,536	0,559	0,527	0,513

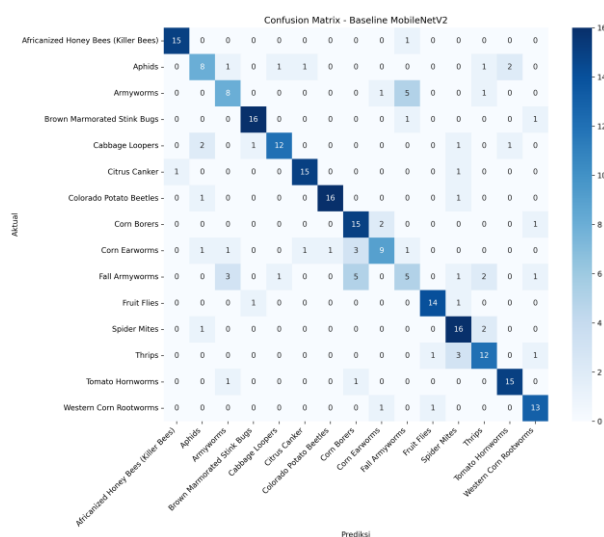
Hasil penelitian ini berbeda dari Thenmozhi dan Reddy [7], yang melaporkan bahwa *transfer learning* dan *augmentation* mendukung akurasi tinggi pada tiga *dataset* serangga yang lebih besar dan memiliki 24-40 kelas. Pada *dataset* penelitian ini yang hanya berisi 1591 citra, *augmentation* justru menurunkan *accuracy* dari 0,750 menjadi 0,528. Temuan ini memperkuat survei Shorten dan Khoshgoftaar [12] bahwa manfaat *augmentation* bergantung pada kesesuaian transformasi dengan karakteristik data. *Fine-tuning* juga tidak melampaui *baseline*, sehingga optimasi tidak dapat diasumsikan selalu meningkatkan kinerja.

3.1.4 Analisis Performa per Kelas

Pada MobileNetV2 *baseline*, model berkinerja tinggi mengenali kelas dengan ciri visual khas (*F1-score*: *Africanized Honey Bees* 0,94; *Colorado Potato Beetles* 0,91; *Brown Marmorated Stink Bugs* 0,89). Sebaliknya, model kesulitan membedakan kelas yang secara visual mirip, ditunjukkan dengan *F1-score* rendah pada *Fall Armyworms* (0,32), *Armyworms* (0,55), *Aphids* (0,59), dan *Corn Earworms* (0,60). Sementara itu, skenario augmentasi justru menurunkan performa beberapa kelas (seperti *F1-score Western Corn Rootworms* 0,12 dan *Armyworms* 0,20) karena transformasi citra diduga mengaburkan fitur visual penting pembeda objek.

3.1.5 Analisis Kesalahan Prediksi

Analisis kesalahan prediksi dilakukan dengan melihat pasangan kelas aktual dan kelas prediksi yang paling sering tertukar pada skenario terbaik, yaitu MobileNetV2 *baseline*. Berdasarkan *confusion matrix* pada Gambar 7, kesalahan terbesar terjadi pada *Armyworms* yang diprediksi sebagai *Fall Armyworms* dan *Fall Armyworms* yang diprediksi sebagai *Corn Borers*, masing-masing sebanyak 5 kasus. Kesalahan berikutnya adalah *Corn Earworms* yang diprediksi sebagai *Corn Borers*, *Thrips* sebagai *Spider Mites*, dan *Fall Armyworms* sebagai *Armyworms*, masing-masing sebanyak 3 kasus.



Gambar 7. Confusion matrix skenario MobileNetV2 *baseline*

Tabel 5. Kesalahan prediksi terbesar pada MobileNetV2 *baseline*

Kelas Aktual	Kelas Prediksi	Jumlah Kesalahan
Armyworms	Fall Armyworms	5
Fall Armyworms	Corn Borers	5
Corn Earworms	Corn Borers	3
Thrips	Spider Mites	3
Fall Armyworms	Armyworms	3
Corn Borers	Corn Earworms	2
Spider Mites	Thrips	2
Fall Armyworms	Thrips	2
Cabbage Loopers	Aphids	2
Aphids	Tomato Hornworms	2

Pola kesalahan tersebut menunjukkan bahwa model cenderung tertukar pada kelas yang memiliki kemiripan visual. *Fall Armyworms*, *Armyworms*, *Corn Borers*, dan *Corn Earworms* merupakan kelas yang secara visual dapat memiliki bentuk tubuh, warna, atau konteks citra yang serupa. Dengan demikian, kesalahan prediksi tidak terjadi secara acak, tetapi berkaitan dengan kedekatan karakteristik visual antar kelas. Pada skenario MobileNetV2 dengan *data augmentation*, kesalahan terbesar terjadi pada *Armyworms* yang diprediksi sebagai *Fall Armyworms* sebanyak 7 kasus, diikuti *Western Corn Rootworms* yang diprediksi sebagai *Thrips* sebanyak 6 kasus dan sebagai *Fruit Flies* sebanyak 4 kasus. *Aphids* juga diprediksi sebagai *Thrips* sebanyak 4 kasus. Pola ini menunjukkan bahwa *augmentation* dapat memperbesar kebingungan model apabila transformasi citra tidak sesuai dengan karakteristik objek.

3.1.6 Estimasi Dampak Finansial Kesalahan Prediksi

Kesalahan prediksi pada klasifikasi hama memiliki implikasi praktis. *False positive* dapat menyebabkan tindakan pengendalian yang tidak diperlukan, seperti penggunaan pestisida, tenaga kerja, atau biaya operasional lain [3]. Sebaliknya, *false negative* dapat menyebabkan

hama tidak terdeteksi secara tepat sehingga potensi kerusakan tanaman dan kehilangan hasil panen meningkat [1], [2].

Dalam penelitian ini, simulasi dampak finansial dihitung berdasarkan jumlah *false positive* dan *false negative* dari *confusion matrix*. Parameter Rp25.000 per *false positive* dan Rp100.000 per *false negative* digunakan sebagai skenario ilustratif, bukan biaya universal. Pendekatan skenario diperlukan karena data BPS menunjukkan variasi biaya pestisida yang lebar antar-komoditas [15], sedangkan *dataset* citra tidak memuat luas lahan, intensitas serangan, atau komoditas terdampak untuk setiap observasi. Pada klasifikasi multikelas satu-label, setiap prediksi yang salah menghasilkan satu *false negative* dan satu *false positive* sehingga memiliki biaya simulasi Rp125.000. Nilai ini membandingkan risiko relatif antarskenario dan tidak dapat diperlakukan sebagai kerugian aktual tanpa validasi lapangan.

Tabel 6. Estimasi dampak finansial kesalahan prediksi

Skenario	Total False Negative	Total False Positive	Estimasi Kerugian
MobileNetV2 <i>Baseline</i>	63	63	Rp7.875.000
MobileNetV2 + <i>Augmentation</i>	119	119	Rp14.875.000
MobileNetV2 + <i>Baseline Fine-Tuning</i>	63	63	Rp7.875.000
MobileNetV2 + <i>Augmentation + Fine-Tuning</i>	117	117	Rp14.625.000

Berdasarkan Tabel 6, MobileNetV2 *baseline* dan *baseline fine-tuning* mencatat kerugian simulatif terendah (Rp7.875.000 dari 63 kesalahan), jauh lebih baik dibandingkan skenario berbasis *augmentation* yang menyentuh angka kerugian hingga ~Rp14,8 juta. Karena bobot kerugian nyata dapat berbeda, analisis ini perlu disempurnakan ke depannya menggunakan parameter komoditas dan biaya pengendalian aktual di lapangan. Secara keseluruhan, capaian akurasi 0,750 membuktikan bahwa MobileNetV2 cukup potensial sebagai model dasar klasifikasi hama. Meskipun demikian, pola kesalahan yang ada menunjukkan perlunya evaluasi khusus pada kelas yang rentan tertukar, seperti *Fall Armyworms*, *Armyworms*, *Corn Borers*, *Corn Earworms*, *Thrips*, dan *Spider Mites*. Berdasarkan temuan riset terdahulu terkait kompleksitas pertanian [6],[16],[8],[9], perbaikan di masa depan dapat difokuskan pada pengayaan *dataset*, penerapan *augmentation* spesifik, pemanfaatan mekanisme attention, atau komparasi dengan arsitektur ringan lain (seperti MobileNetV3 dan EfficientNetB0).

3.2 Pembahasan

Penelitian ini bertujuan untuk mengevaluasi performa arsitektur ringan MobileNetV2 dalam klasifikasi citra hama serangga serta menyimulasikan konsekuensi finansial dari kesalahan prediksi. Hasil pengujian membuktikan bahwa konfigurasi MobileNetV2 *baseline* mampu memberikan performa paling optimal dan stabil dengan *accuracy* mencapai 0,750 dan *macro F1-score* 0,744. Keberhasilan ini menunjukkan bahwa *pretrained weights* dari ImageNet pada MobileNetV2 sudah memiliki ketahanan ekstraksi fitur visual umum yang memadai untuk mengenali objek hama pada dataset skala kecil [10]. Temuan ini sekaligus menjawab tujuan penelitian bahwa arsitektur ringan dapat diandalkan sebagai fondasi awal sistem deteksi di lingkungan pertanian.

Secara teoritis, rasionalisasi dari keunggulan model *baseline* dibandingkan skenario *fine-tuning* dan *data augmentation* terletak pada rasio antara ukuran dataset dan kompleksitas model. Pada dataset berukuran kecil, upaya *unfreezing* (membuka lapisan ekstraksi fitur) justru memicu *catastrophic forgetting* di mana model kehilangan bobot umumnya tanpa mampu mempelajari pola baru secara sempurna [13]. Lebih lanjut, klasifikasi hama merupakan permasalahan *Fine-Grained Visual Categorization* (FGVC), di mana perbedaan antar kelas sangat tipis dan hanya bergantung pada detail morfologi mikroskopis [17]. Oleh karena itu, transformasi citra standar seperti rotasi atau *zoom* yang diterapkan pada

skenario *augmentation* justru mengaburkan dan merusak fitur morfologi halus tersebut, yang pada akhirnya menurunkan akurasi model secara drastis [12], [17].

Hasil ini memberikan temuan yang kontras atau berbeda dibandingkan dengan penelitian-penelitian sebelumnya. Riset dari Thenmozhi dan Reddy [7] serta Zhao dk [6] melaporkan bahwa penerapan *transfer learning* lanjutan dan *data augmentation* sukses melambungkan akurasi model. Namun, perbedaan hasil ini sangat wajar karena penelitian mereka menggunakan dataset raksasa dengan puluhan ribu citra yang mampu menyerap variasi dari *augmentation*. Penelitian ini membuktikan bahwa strategi optimasi model (seperti augmentasi) tidak bersifat mutlak dan justru dapat merugikan performa (berbalik menjadi distorsi) apabila diterapkan secara naif pada dataset agrikultur yang sangat terbatas dan spesifik [18].

Dari segi kontribusi terhadap pengembangan ilmu pengetahuan, penelitian ini menawarkan kebaruan (*novelty*) dengan menjembatani evaluasi metrik teknis (komputer) dengan evaluasi kelayakan finansial (agronomi). Berdasarkan analisis *confusion matrix*, pola kesalahan tertinggi terjadi pada hama dengan kemiripan visual ekstrem seperti *Armyworms* dan *Fall Armyworms* [2]. Dalam konsep *Smart Agriculture*, *false positive* pada kelas-kelas ini akan memicu penyemprotan pestisida yang berujung pada kerugian finansial dari biaya *overhead*, sementara *false negative* akan mengancam *Economic Threshold* (ambang batas ekonomi) yang memicu gagal panen [1], [3]. Integrasi simulasi kerugian seperti yang ditawarkan dalam riset ini sejalan dengan tren terbaru dalam *Precision Agriculture*, di mana kecerdasan buatan harus dievaluasi berdasarkan risiko ekonomi dan ekologi, bukan hanya metrik *accuracy* semata [19].

Meskipun model *baseline* menunjukkan hasil yang menjanjikan, penelitian ini tidak lepas dari sejumlah keterbatasan. Pertama, penggunaan *dataset* publik masih memuat bias visual (seperti adanya kelas penyakit *Citrus Canker* di dalam *dataset* serangga) dan kurang merepresentasikan kondisi iklim atau pencahayaan pertanian lokal. Kedua, nilai Rp25.000 dan Rp100.000 untuk penghitungan biaya kesalahan masih bersifat parameter simulatif. Untuk peluang penelitian lebih lanjut, sangat disarankan untuk membangun *dataset* hama endemik lokal dengan anotasi pakar, menerapkan teknik augmentasi yang dikhususkan untuk objek *fine-grained* (seperti *Copy-Paste augmentation*), serta memvalidasi simulasi kerugian finansial menggunakan data biaya produksi lapangan yang sesungguhnya. Eksplorasi menggunakan arsitektur yang lebih mutakhir seperti MobileNetV3 atau EfficientNetB0 juga terbuka lebar untuk dibandingkan dengan hasil *baseline* ini.

4. Simpulan

Penelitian ini berhasil mengevaluasi performa MobileNetV2 dalam klasifikasi citra hama sekaligus menyimulasikan dampak finansial dari kesalahan prediksi. Dari empat skenario yang diuji pada 1.591 citra (15 kelas), konfigurasi MobileNetV2 *baseline* terbukti paling optimal (*accuracy* 0,750; *macro F1-score* 0,744), mengungguli skenario *fine-tuning* maupun *data augmentation*. Kegagalan skenario optimasi tambahan tersebut dipicu oleh sulitnya model membedakan kelas-kelas dengan kemiripan morfologi visual yang sangat tinggi (seperti *Armyworms* dan *Fall Armyworms*).

Kajian ini menegaskan bahwa model klasifikasi pertanian tidak cukup dievaluasi murni dari akurasi teknis, tetapi wajib mempertimbangkan pola kesalahan dan risiko finansial praktisnya. Prediksi hama yang keliru (*false positive*) akan memicu pemborosan biaya pestisida, sedangkan kegagalan deteksi (*false negative*) berisiko menunda penanganan dan merusak hasil panen. Pendekatan yang menggabungkan teknis dan finansial inilah yang membuat evaluasi model menjadi jauh lebih relevan untuk kondisi lapangan.

Meskipun demikian, studi ini memiliki keterbatasan berupa penggunaan *dataset* publik berukuran kecil (termasuk adanya kelas penyakit *Citrus Canker*), estimasi finansial yang masih simulatif, dan belum adanya uji lapangan. Ke depan, penelitian lanjutan sangat disarankan untuk melakukan validasi dengan citra lokal, menyempurnakan kalkulasi finansial menggunakan parameter biaya aktual, serta menguji penerapan model ringan ini langsung pada sistem identifikasi berbasis aplikasi atau perangkat *mobile* [3],[10],[11]. Pendekatan pengenalan pola cerdas berbasis CNN semacam ini diharapkan dapat terus direplikasi secara

luas pada berbagai sistem visual maupun teks lainnya, sebagaimana keberhasilan pada pengolahan deteksi kemiripan dokumen [20] serta klasifikasi kondisi infrastruktur jalan [21].

Daftar Referensi

- [1] S. Savary et al., "The global burden of pathogens and pests on major food crops," *Nat. Ecol. Evol.*, vol. 3, pp. 430-439, 2019, doi: 10.1038/s41559-018-0793-y.
- [2] D. Popescu, A. Dinca, L. Ichim, and N. Angelescu, "New trends in detection of harmful insects and pests in modern agriculture using artificial neural networks: A review," *Front. Plant Sci.*, vol. 14, p. 1268167, 2023, doi: 10.3389/fpls.2023.1268167.
- [3] S. M. Venkateswara and J. Padmanabhan, "Deep learning based agricultural pest monitoring and classification," *Sci. Rep.*, vol. 15, p. 8684, 2025, doi: 10.1038/s41598-025-92659-5.
- [4] T. Dalal, "Dangerous Farm Insects Dataset," Kaggle, 2024. [Online]. Tersedia: <https://www.kaggle.com/datasets/tarundalal/dangerous-insects-dataset>
- [5] A. Kamilaris and F. X. Prenafeta-Boldu, "Deep learning in agriculture: A survey," *Comput. Electron. Agric.*, vol. 147, pp. 70-90, 2018, doi: 10.1016/j.compag.2018.02.016.
- [6] S. Zhao, J. Liu, Z. Bai, C. Hu, and Y. Jin, "Crop pest recognition in real agricultural environment using convolutional neural networks by a parallel attention mechanism," *Front. Plant Sci.*, vol. 13, p. 839572, 2022, doi: 10.3389/fpls.2022.839572.
- [7] K. Thenmozhi and U. S. Reddy, "Crop pest classification based on deep convolutional neural network and transfer learning," *Comput. Electron. Agric.*, vol. 164, p. 104906, 2019, doi: 10.1016/j.compag.2019.104906.
- [8] D. Xia, P. Chen, B. Wang, J. Zhang, and C. Xie, "Insect detection and classification based on an improved convolutional neural network," *Sensors*, vol. 18, no. 12, p. 4169, 2018, doi: 10.3390/s18124169.
- [9] C. Li, T. Zhen, and Z. Li, "Image classification of pests with residual neural network based on transfer learning," *Appl. Sci.*, vol. 12, no. 9, p. 4356, 2022, doi: 10.3390/app12094356.
- [10] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.
- [11] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017. [Online]. Tersedia: <https://arxiv.org/abs/1704.04861>
- [12] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, 2019, doi: 10.1186/s40537-019-0197-0.
- [13] F. Zhuang et al., "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43-76, 2021, doi: 10.1109/JPROC.2020.3004555.
- [14] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: An Overview," *arXiv preprint arXiv:2008.05756*, 2020. [Online]. Tersedia: <https://arxiv.org/abs/2008.05756>
- [15] Badan Pusat Statistik, "Hasil Survei Ekonomi Pertanian (SEP) 2024: Nilai dan Biaya Produksi Beberapa Komoditas Utama pada Subsektor Tanaman Pangan, Hortikultura, dan Perkebunan," Jakarta, Publikasi No. 05100.25001, 2024. [Online]. Tersedia: <https://www.bps.go.id/id/publication/2025/01/24/8faa8436999bb9b54ba0b7a9/hasil-survei-ekonomi-pertanian-sep-2024-nilai-dan-biaya-produksi-beberapa-komoditas-utama-pada-subsektor-tanaman-pangan-hortikultura-dan-perkebunan.html>
- [16] B. Wang, "Identification of crop diseases and insect pests based on deep learning," *Sci. Program.*, vol. 2022, p. 9179998, 2022, doi: 10.1155/2022/9179998.
- [17] X. Wei, J. Wu, and H. Su, "Fine-Grained Image Analysis with Deep Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8927-8948, 2022, doi: 10.1109/TPAMI.2021.3126648.
- [18] T. Kasinathan, D. Singaraju, and S. R. Uyyala, "Insect classification and detection in field crops using modern machine learning techniques," *Inf. Process. Agric.*, vol. 8, no. 3, pp. 446-457, 2021, doi: 10.1016/j.inpa.2020.09.006.
- [19] R. Sharma, A. Kamble, and R. Raman, "Deep Learning Applications for Agriculture: A Review," *Sensors*, vol. 22, no. 18, p. 7004, 2022, doi: 10.3390/s22187004.

-
- [20] S. Utomo, I. M. I. Subroto, and A. Riansyah, "Deteksi Plagiat Tugas Akhir dengan Metode Jaccard Similarity," *Jurnal Transistor Elektro dan Informatika (TRANSISTOR EI)*, vol. 4, no. 2, pp. 132-141, 2022.
- [21] S. Utomo, M. N. W. Affandi, and G. F. Shidik, "Asphalt Road Quality Classification Based on CNN Architecture VGG16," in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, 2023.