

PENINGKATAN ROBUSTNESS MODEL KLASIFIKASI PENYAKIT DAUN SEMANGKA MENGGUNAKAN FINE-TUNING VGG16 DAN AUGMENTASI DATA SINTETIS BERBASIS GENERATIVE ADVERSARIAL NETWORK

Suhesi¹; Dian Ade Kurnia²; Yudhistira Arie Wijaya³; Saeful Anwar⁴; Umi Hayati⁵;

Program Studi Teknik Informatika¹⁴⁵
Program Studi Manajemen Informatika²
Program Studi Sistem Informasi³

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
noorazizhesi@gmail.com

(*) Corresponding Author : noorazizhesi@gmail.com
Published : 31 Desember 2025

Abstract—This study aims to develop a more robust watermelon leaf disease classification model by integrating VGG16 fine-tuning, GAN-based synthetic augmentation, and Stratified K-Fold Cross-Validation. The dataset comprises four classes with significant imbalance, requiring strategies to enhance minority class representation. The research stages include image preprocessing, basic augmentation, targeted GAN augmentation, and model training using transfer learning combined with customized classification layers. The model's performance is evaluated using advanced metrics, including accuracy, precision, recall, F1-Score, AUC-ROC, and AUC-PR. Experimental results indicate that targeted GAN augmentation improves model stability and reduces bias toward majority classes. Training with Stratified K-Fold also provides more consistent performance compared to a single train-test split. On the unseen test set, the model demonstrates notable improvements in minority-class F1-Score and increased AUC-PR, suggesting enhanced sensitivity to imbalanced data. Grad-CAM visualizations reveal that the model focuses on disease-relevant lesion regions, confirming improved generalization after applying targeted augmentation. Overall, the integration of VGG16 fine-tuning, GAN-based synthetic augmentation, and Stratified Cross-Validation effectively enhances the robustness and performance of plant disease classification models. This approach offers substantial potential for application in computer vision-based crop disease diagnosis systems within digital agriculture.

Keywords : Deep Learning; VGG16; GAN Augmentation; Stratified K-Fold; Plant Disease Classification

Abstrak— Penelitian ini bertujuan mengembangkan model klasifikasi penyakit daun semangka yang lebih tangguh melalui integrasi *fine-tuning* VGG16, augmentasi sintetis berbasis GAN, dan *Stratified K-Fold Cross-Validation*. Dataset terdiri atas empat kelas dengan distribusi yang tidak seimbang, sehingga memerlukan strategi untuk meningkatkan representasi kelas minoritas. Tahapan penelitian meliputi pra-pemrosesan citra, augmentasi dasar, augmentasi terarah berbasis GAN, serta pelatihan model menggunakan *transfer learning* dan penyesuaian lapisan akhir. Evaluasi performa dilakukan menggunakan metrik lanjutan seperti akurasi, presisi, recall, F1-Score, AUC-ROC, dan AUC-PR. Hasil penelitian menunjukkan bahwa augmentasi sintetis GAN mampu meningkatkan stabilitas model dan mengurangi bias terhadap kelas mayoritas. Pelatihan menggunakan *Stratified K-Fold* juga memberikan performa yang lebih konsisten dibandingkan pendekatan *single split*. Pada *unseen test set*, terjadi peningkatan yang signifikan pada F1-Score kelas minoritas dan kenaikan AUC-PR, yang menandakan sensitivitas lebih baik pada distribusi data tidak seimbang. Visualisasi Grad-CAM memperlihatkan bahwa model mampu memfokuskan atensi pada pola lesi yang relevan, sehingga mengonfirmasi peningkatan kemampuan generalisasi setelah augmentasi terarah. Secara keseluruhan, perpaduan *fine-tuning* VGG16, augmentasi sintetis GAN, dan *Stratified Cross-Validation* terbukti meningkatkan ketahanan dan performa model klasifikasi penyakit daun tanaman. Pendekatan ini berpotensi dikembangkan lebih lanjut dalam sistem diagnosis tanaman berbasis visi komputer untuk mendukung pertanian digital.

Kata Kunci : *Deep Learning*; VGG16; Augmentasi GAN; Stratified K-Fold; Klasifikasi Penyakit Tanaman

INTRODUCTION

Kemajuan teknologi kecerdasan buatan (Artificial Intelligence/AI) dan visi komputer (computer vision) telah membawa transformasi signifikan dalam bidang pertanian digital, khususnya dalam deteksi dan klasifikasi penyakit tanaman berbasis citra. Klasifikasi penyakit tanaman berbasis citra memiliki peran krusial dalam pertanian digital karena memungkinkan deteksi penyakit tanaman secara dini, akurat, dan berskala besar dengan memanfaatkan data visual serta teknik pembelajaran mesin (machine learning) dan pembelajaran mendalam (deep learning) (Demilie, 2024). Penerapan metode berbasis citra memungkinkan pengurangan beban tenaga kerja, mempercepat proses diagnosis di area pertanian yang luas, serta menekan potensi kerugian hasil panen akibat keterlambatan identifikasi penyakit. Hal ini sejalan dengan pandangan (Bagga & Goyal, 2024) yang menegaskan bahwa deteksi berbasis citra pada gejala daun merupakan aspek fundamental dalam mekanisme manajemen penyakit tanaman di era kebutuhan pangan global yang terus meningkat.

Selain itu, sistem klasifikasi berbasis citra juga menjadi fondasi penting untuk membangun model yang tangguh dan dapat digeneralisasi terhadap berbagai jenis tanaman dan kondisi lingkungan yang berbeda (Sajitha, 2024). (R. Upadhyay & Mehta, 2025) menambahkan bahwa penerapan teknologi visi komputer modern pada data citra memungkinkan diagnosis penyakit secara non-destruktif, presisi, dan waktu nyata di lapangan. Oleh karena itu, pendekatan berbasis citra dianggap mendasar dalam mendukung pertanian presisi (*precision agriculture*), terutama ketika digabungkan dengan teknik peningkatan performa model seperti transfer learning dan augmentasi data berbasis Generative Adversarial Network (GAN).

Meskipun demikian, klasifikasi penyakit tanaman berbasis deep learning masih menghadapi sejumlah tantangan yang signifikan. Pertama, keterbatasan dataset publik yang berkualitas tinggi dengan catatan yang lengkap membatasi kemampuan generalisasi model dan meningkatkan risiko *overfitting* (Tugrul et al., 2022). Kedua, kompleksitas latar belakang citra, variasi pencahayaan, serta tumpang tindih gejala penyakit dapat menurunkan akurasi model ketika

diterapkan pada kondisi lapangan yang sesungguhnya (Mustofa et al., 2023; Tugrul et al., 2022). Ketiga, ketidakseimbangan kelas akibat dominasi kelas mayoritas seperti daun sehat dibandingkan kelas penyakit langka menyebabkan bias pada model kecerdasan buatan (Acar, 2024). Keempat, keterbatasan sumber daya komputasi dan kebutuhan interpretabilitas model menjadi hambatan dalam penerapan nyata di lapangan (Acar, 2024; Tugrul et al., 2022). Selain itu, masalah *robustness* lintas domain atau transferabilitas model dari satu kondisi citra ke kondisi lainnya masih menjadi tantangan utama (Mustofa et al., 2023).

Untuk mengatasi tantangan tersebut, dua pendekatan utama yang efektif adalah *fine-tuning* model convolutional neural network (CNN) pra-latih dan augmentasi data sintesis menggunakan *Generative Adversarial Network* (GAN). Teknik *fine-tuning* pada jaringan pra-latih seperti VGG16 memungkinkan model memanfaatkan representasi fitur yang kaya dari dataset berskala besar dan menyesuaikannya dengan domain target, bahkan ketika jumlah data minoritas terbatas (ValizadehAslani et al., 2024; Zhai, 2022). Di sisi lain, augmentasi berbasis GAN mampu menghasilkan citra sintesis berkualitas tinggi yang mewakili kelas minoritas, memperluas ruang fitur (*feature manifold*), dan mengurangi bias kelas (Guo et al., 2024; Khan et al., 2022). Kombinasi kedua strategi ini dapat meningkatkan keseimbangan data pada level dataset sekaligus memperkuat kemampuan generalisasi model pada level pembelajaran, sehingga menghasilkan sistem klasifikasi penyakit daun semangka yang lebih robust dan adaptif terhadap variasi kondisi lapangan.

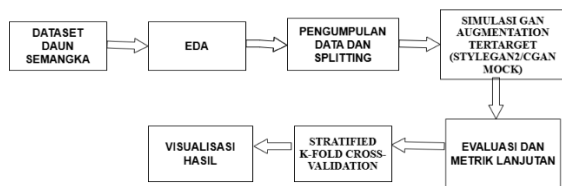
Dengan demikian, penelitian ini berfokus pada penerapan *Generative Adversarial Network* (GAN) untuk augmentasi data dalam klasifikasi citra penyakit daun semangka menggunakan model *Deep Learning* berbasis VGG16 yang di-*fine-tune*. Pendekatan ini diharapkan dapat memberikan kontribusi signifikan terhadap peningkatan ketahanan model terhadap ketidakseimbangan kelas, variasi citra alami, dan generalisasi lintas domain dalam konteks pertanian cerdas yang berkelanjutan.

MATERIALS AND METHODS

Tahapan Penelitian

Tahapan penelitian ini disusun untuk memastikan proses pengembangan model klasifikasi penyakit daun semangka berjalan terarah dan sesuai prinsip metodologis. Penelitian dimulai dari penyusunan Dataset dan analisis awal melalui EDA untuk memahami karakteristik citra. Berikutnya, dirancang Arsitektur Model berbasis *transfer learning* dan dilakukan Pra-Pemrosesan guna meningkatkan kualitas data.

Penelitian kemudian menerapkan Simulasi GAN Augmentation Terarget untuk memperkuat representasi kelas minoritas. Kinerja model dievaluasi melalui Evaluasi dan Metrik Lanjutan, serta diuji konsistensinya menggunakan Stratified K-Fold Cross-Validation. Tahap akhir berupa Visualisasi Hasil, termasuk Grad-CAM, digunakan untuk menilai fokus model terhadap fitur penyakit. Tahapan-tahapan tersebut tertuang pada Gambar 1 membentuk kerangka penelitian yang ringkas namun komprehensif untuk menghasilkan model klasifikasi yang akurat dan robust.



Gambar 1 Tahapan Penelitian

Dataset

Dataset yang digunakan dalam penelitian ini adalah **Watermelon Leaf Disease Image Dataset** yang diperoleh dari platform Kaggle dapat diakses melalui tautan berikut : <https://www.kaggle.com/datasets/nirmalsankalan/a/watermelon-disease-dataset>, sebuah repositori data publik yang banyak dimanfaatkan dalam riset berbasis kecerdasan buatan. Dataset ini terdiri atas **1.155 citra daun semangka** sebelum augmentasi sintetis, yang diklasifikasikan ke dalam empat kategori: *anthracnose*, *downy mildew*, *mosaic virus*, dan *healthy*. Seluruh citra telah melalui proses pemilahan internal untuk memastikan kesesuaian label dan kualitas visual sehingga layak digunakan dalam pipeline

deep learning. pemilihan *Watermelon Leaf Disease Image Dataset* dari Kaggle didasarkan pada pertimbangan ilmiah yang menempatkannya sebagai sumber data paling relevan dan representatif untuk penelitian klasifikasi penyakit daun semangka berbasis *deep learning*. Dataset ini memuat empat kategori penyakit yang umum dijumpai pada tanaman semangka, disertai keragaman visual yang mencakup variasi kondisi pencahayaan, tekstur daun, dan bentuk lesi. Keragaman tersebut memberikan konteks realistis yang diperlukan untuk mengevaluasi kemampuan

model dalam melakukan generalisasi pada kondisi lapangan yang tidak terstandarisasi. EDA

Tahap Exploratory Data Analysis (EDA) digunakan untuk memperoleh gambaran menyeluruh mengenai karakter dan kualitas dataset citra penyakit daun semangka sebelum proses pelatihan model dilakukan. Analisis ini membantu mengungkap pola distribusi kelas, variasi kondisi visual, serta potensi bias yang dapat memengaruhi rancangan model maupun kemampuan generalisasinya. Sebagaimana dijelaskan oleh (Xu et al., 2024), EDA berperan penting dalam menilai kesesuaian dataset dengan lingkungan nyata, termasuk mendeteksi ketidakseimbangan kelas, latar belakang citra yang tidak konsisten, dan perbedaan setelan pengambilan gambar. Selain itu, penggunaan histogram, analisis warna, dan distribusi intensitas piksel sebagaimana direkomendasikan oleh (Shoaib et al., 2023) memberikan landasan dalam menentukan kebutuhan preprocessing seperti normalisasi dan strategi augmentasi tambahan.

Di sisi lain, EDA juga membantu menentukan arah pemilihan arsitektur serta metode pelatihan. (A. Upadhyay, 2025) menunjukkan bahwa kompleksitas dataset dapat memengaruhi keputusan penelitian untuk menggunakan transfer learning atau model CNN yang lebih ringan. Temuan EDA lainnya, seperti statistik per kelas (resolusi, ukuran, dan variasi warna), juga dapat mengungkap bias yang berpotensi memicu overfitting, sebagaimana dicatat (Zhao & Sun, 2025). Lebih lanjut, tinjauan (Nyawose, 2025) menegaskan bahwa pemeriksaan kualitas data misalnya tingkat noise, konsistensi resolusi, dan tingkat representasi tiap kelas menjadi faktor penting sebelum membangun pipeline pelatihan. Dengan demikian, EDA pada penelitian ini menjadi langkah fundamental untuk merumuskan strategi pemrosesan, augmentasi, serta desain model yang lebih tangguh dan adaptif.

Pengumpulan Data dan Splitting
Pengumpulan data dalam penelitian ini dilakukan dengan memanfaatkan *Watermelon Leaf Disease Image Dataset* dari Kaggle, yang memuat empat kategori citra daun semangka berpenyakit. Prosedur kurasi mencakup penilaian kualitas visual, pengecekan ulang akurasi label, serta normalisasi resolusi citra ke format 224×224 piksel. Langkah ini mengikuti prinsip kurasi data yang direkomendasikan dalam literatur, termasuk standarisasi prosedur pencitraan, anotasi berbasis pakar, pemantauan kualitas, serta pelacakan asal-usul data untuk menjamin integritas dataset (Moupojou et al., 2023; Xu, 2024). Tingkat ketimpangan distribusi kelas diperoleh melalui perhitungan rasio:

yang menunjukkan nilai 2{,}68, mengonfirmasi bahwa dataset perlu diseimbangkan sebelum memasuki tahap pelatihan. Dengan demikian, kurasi dataset menghasilkan basis data yang valid secara metodologis untuk proses *fine-tuning* pada arsitektur VGG16.

Untuk mengurangi ketimpangan kelas tersebut, penelitian ini menerapkan augmentasi sintetis menggunakan pendekatan simulatif berbasis prinsip *conditional generation* dan *style modulation* pada GAN. Formulasi cGAN digunakan untuk memastikan citra sintetis sesuai dengan distribusi kelas target, yang diwujudkan agar distribusi kelas menjadi lebih merata. Pendekatan ini selaras dengan rekomendasi studi mutakhir yang menekankan pentingnya pemeriksaan realisme citra sintetis dan evaluasi distribusi untuk mencegah kesalahan anotasi dan *label leakage* (Ahmad et al., 2023).

Pembagian data dilakukan menggunakan *Stratified K-Fold Cross-Validation* untuk menjamin konsistensi distribusi kelas pada setiap fold. Secara matematis, stratifikasi memastikan bahwa atau dengan kata lain, proporsi setiap kelas y_c pada fold F_k setara dengan proporsi di seluruh dataset. Nilai evaluasi keseluruhan diperoleh dengan menghitung rata-rata performa antar-

Literatur mengenai pembelajaran pada data tidak seimbang menunjukkan bahwa stratifikasi memberikan estimasi performa yang lebih stabil, mengurangi bias evaluasi, serta meningkatkan kemampuan generalisasi model (Mahesh, 2023; Prusty & al., 2022; Szeghalmy & Fazekas, 2023). Melalui penerapan metode ini, penelitian menghasilkan proses evaluasi yang lebih dapat dipercaya dan representatif untuk seluruh kelas dalam klasifikasi penyakit daun semangka. Simulasi GAN Augmentation Tertarget (StyleGAN2/CGAN Mock)

Untuk mengatasi ketidakseimbangan kelas yang teridentifikasi pada dataset daun semangka dengan rasio ketimpangan mencapai **2,68:1** penelitian ini menerapkan pendekatan *simulasi augmentasi* berbasis prinsip *Conditional GAN* (cGAN) dan modulasi gaya pada StyleGAN2. Meskipun model GAN tidak dilatih secara penuh, formulasi teoretis cGAN digunakan untuk memandu proses generatif dalam menghasilkan citra sintetis kelas minoritas. Secara matematis, cGAN memodelkan hubungan antara noise z dan label kelas y melalui fungsi objektif berikut, yang memastikan generator menghasilkan citra sesuai distribusi kelas target (Wang & Cao, 2023; Bi & Hu, 2020):

Model simulatif ini membantu mensintesis variasi tekstural yang relevan pada daun semangka berpenyakit, seperti pola nekrosis, mosaik, atau

perubahan warna yang lazim terjadi pada kelas *Downy Mildew*, *Mosaic Virus*, dan *Anthraco*.

Dalam augmentasi tertarget, penelitian ini mengintegrasikan mekanisme *style modulation* dari StyleGAN2 melalui **Adaptive Instance Normalization (AdaIN)** untuk menyesuaikan distribusi fitur dan tekstur sintetis agar menyerupai ciri visual kelas minoritas. Mekanisme ini dituliskan sebagai:

di mana f merepresentasikan fitur citra dan parameter gaya μ, σ mengendalikan intensitas serta pola warna. Pendekatan ini telah terbukti efektif dalam mempertahankan struktur semantik citra sambil memperkaya keragaman tekstur (Xu et al., 2022; Y. Zhang & al., 2023), sehingga sangat relevan untuk memperluas representasi penyakit daun semangka yang minoritas.

Untuk menentukan jumlah citra sintetis yang diperlukan demi mencapai distribusi yang seimbang, penelitian ini menerapkan formula perhitungan penyeimbangan berbasis selisih populasi antar kelas, yang secara umum digunakan dalam GAN-based resampling (Lee et al., 2022; Miftahushudur et al., 2025):

Formula ini memastikan bahwa setiap kelas minoritas seperti *Anthraco* (155 sampel) dan *Healthy* (205 sampel) diperkuat hingga mendekati jumlah kelas mayoritas (415 sampel), sehingga distribusi menjadi lebih merata. Pendekatan ini mendukung stabilitas pelatihan, meningkatkan kemampuan generalisasi model, dan mengurangi bias terhadap kelas mayoritas sebagaimana direkomendasikan dalam literatur GAN augmentasi untuk penyakit tanaman (Bi & Hu, 2020; Kang & al., 2025; Wang & Cao, 2023).

Evaluasi dan Metrik Lanjutan

Evaluasi performa model dilakukan menggunakan *unseen test set* serta Stratified 5-Fold Cross-Validation untuk memastikan stabilitas hasil setelah proses *fine-tuning* VGG16, *class weighting*, dan augmentasi GAN. Pada setiap lipatan validasi, nilai metrik dihitung dan dirata-ratakan menggunakan $\bar{M} = \frac{1}{5} \sum_i^5 M_i$, sedangkan tingkat variabilitasnya diestimasi melalui $\sigma_M = \sqrt{\frac{1}{5} \sum_{i=1}^5 (M_i - \bar{M})^2}$. Karena dataset memiliki ketidakseimbangan kelas, penelitian ini menggunakan metrik yang lebih representatif dibandingkan akurasi, yaitu AUC-ROC, AUC-PR, Macro-F1, dan Weighted-F1.

AUC-ROC memberikan evaluasi pemisahan kelas yang robust (Rainio et al., 2024) dan dihitung melalui integral kurva hubungan TPR dan FPR:

Untuk menilai sensitivitas terhadap kelas minoritas, AUC-PR digunakan karena integral hubungan

precision-recall lebih mencerminkan performa pada data jarang (Journal of Big Data, 2023): Pada konteks multi-kelas, Macro-F1 menilai semua kelas secara setara melalui rata-rata F1-score: sementara Weighted-F1 memasukkan bobot proporsional kelas sehingga lebih stabil terhadap distribusi tidak seimbang (Harbecke & al., 2022; Hinojosa Lee et al., 2024):

Analisis *confusion matrix* juga dilakukan untuk mengidentifikasi pola kesalahan prediksi, terutama pada kelas dengan gejala mirip dan kelas minoritas. Hubungan yang telah lazim digunakan dalam penelitian klasifikasi multi-kelas pada penyakit tanaman (Aboelenin et al., 2025; Albattah et al., 2022; Ashurov et al., 2024; Sundhar et al., 2025; Ullah et al., 2023). Evaluasi matriks sebelum dan sesudah augmentasi GAN juga membantu menilai pengaruh citra sintesis terhadap pola prediksi dan generalisasi model.

Pendekatan ini menghasilkan evaluasi menyeluruh, memastikan bahwa peningkatan performa tidak bias terhadap kelas mayoritas dan bahwa model mampu melakukan generalisasi secara konsisten di seluruh kelas penyakit daun semangka (Adegbenjo & al., 2024).

RESULTS AND DISCUSSION

Simulasi GAN Augmentation Tertarget (Stylegan2/CGAN Mock)

Sebagai bagian dari upaya memperbaiki ketidakseimbangan data, penelitian ini memilih tiga kelas dengan jumlah sampel rendah *Anthraco*se, *Healthy*, dan *Downy Mildew* untuk diperkaya menggunakan teknik GAN. Melalui proses augmentasi tersebut dihasilkan 1.200 citra sintesis, yang kemudian digabungkan dengan data asli sehingga total sampel yang digunakan pada tahap *K-Fold Cross-Validation* mencapai 2.124 citra. Penambahan data buatan ini dirancang untuk menambah keragaman visual pada kelas minoritas sehingga model memperoleh representasi yang lebih komprehensif selama proses pelatihan.

Tabel 1 Distribusi Kelas pada Data Gabungan

Label	Jumlah
<i>Mosaic Virus</i>	332
<i>Downy Mildew</i>	704
<i>Healthy</i>	564
<i>Anthraco</i> se	524

Berdasarkan Tabel 1, setelah integrasi data asli dan hasil augmentasi dilakukan, distribusi kelas menunjukkan pola yang lebih seimbang. Kelas *Downy Mildew* kini berjumlah 704 citra, diikuti oleh *Healthy* sebanyak 564 citra dan *Anthraco*se sebanyak 524 citra, sementara *Mosaic Virus* tetap berjumlah 332 citra karena tidak termasuk dalam kelas yang diaugmentasi. Komposisi data yang lebih proporsional ini memungkinkan model mempelajari karakteristik tiap

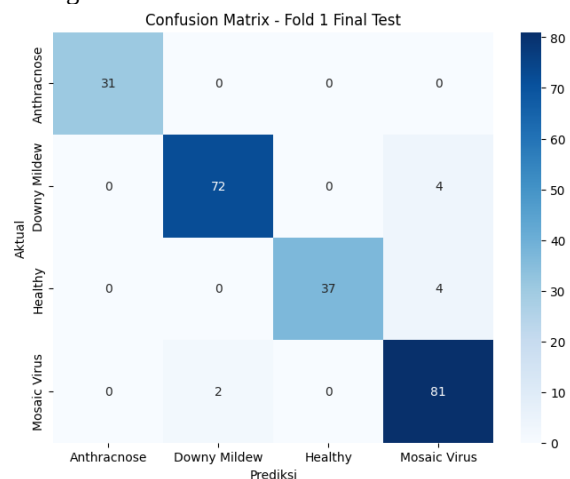
kelas secara lebih merata. Dengan demikian, penerapan augmentasi berbasis GAN memberikan kontribusi penting dalam meningkatkan kualitas dataset sekaligus memperkuat kemampuan generalisasi model klasifikasi.

Evaluasi Dan Metrik Lanjutan

Untuk memperoleh pemahaman yang lebih menyeluruh mengenai performa model serta karakteristik visual dari dataset yang digunakan, analisis dilakukan melalui dua sumber informasi utama, yaitu *confusion matrix* hasil pengujian Fold 1 sampai Fold 5 dan tabel statistik intensitas warna RGB. Kedua representasi ini tidak hanya menampilkan kemampuan model dalam mengenali pola penyakit, tetapi juga memberikan gambaran mengenai konsistensi warna citra yang mendasari proses klasifikasi

Evaluasi Fold 1

Pada evaluasi fold 1, model menghasilkan Confusion Matrix berikut yang menggambarkan sebaran prediksi terhadap setiap kelas penyakit daun semangka.



Gambar 2 Confusion Matrix - Fold 1

Berdasarkan Gambar 2 memperlihatkan bahwa model mampu mengidentifikasi sebagian besar sampel dari masing-masing kelas dengan tepat, yang terlihat dari dominasi nilai pada diagonal utama. Meski demikian, masih terdapat sejumlah kesalahan prediksi, khususnya antar kelas *Downy Mildew*, *Healthy*, dan *Mosaic Virus*, yang mengisyaratkan adanya kesamaan ciri visual pada daun-daun tersebut. Kondisi ini menunjukkan tantangan bagi model dalam membedakan fitur warna dan tekstur yang mirip antar kategori.

Tabel 2 Metrik Area Under Curve (AUC) per Kelas Fold 1

Kelas	AUC-ROC	AUC-PR
<i>Anthraco</i> se	1	1
<i>Downy Mildew</i>	0.997	0.9947
<i>Healthy</i>	0.9997	0.9988
<i>Mosaic Virus</i>	0.9947	0.9903

Berdasarkan Tabel 2 analisis terhadap statistik intensitas warna RGB memberikan pemahaman awal

mengenai karakteristik visual yang membentuk dataset citra daun semangka. Dominasi kanal hijau disusul kanal merah, serta variasi yang lebih tinggi pada kanal biru mencerminkan sifat alami daun yang kaya pigmen klorofil namun tetap dipengaruhi kondisi pencahayaan, bayangan, dan tekstur permukaan. Keceragaman nilai median dan rata-rata pada ketiga kanal menunjukkan stabilitas distribusi warna antar citra, sementara penyebaran yang lebih luas pada kanal biru mengindikasikan adanya variabilitas visual yang dapat berperan dalam memicu kemiripan antarkelas tertentu. Kondisi ini memberikan konteks awal mengapa beberapa kelas penyakit, terutama yang menampilkan pola warna serupa, dapat lebih sulit dipisahkan oleh model.

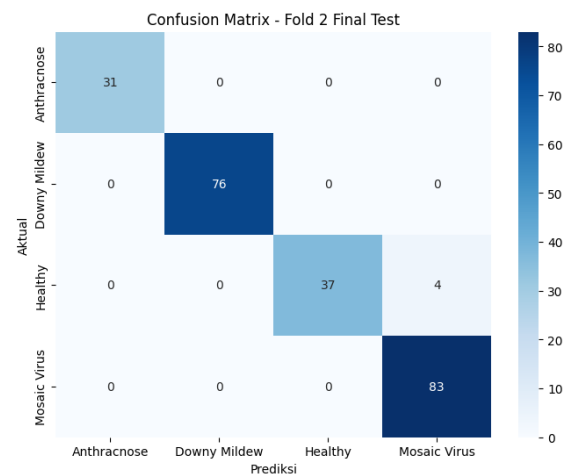
Tabel 3 Ringkasan Metrik Klasifikasi Fold 1

Kelas	precision	recall	F1-score	support
Anthraco se	1	1	1	31
Downy Mildew	0.973	0.947 4	0.96	76
Healthy	1	0.902 4	0.948 7	41
Mosaic Virus	0.9101	0.975 9	0.941 9	83
accuracy	0.9567	0.956 7	0.956 7	0.9567
macro avg	0.9708	0.956 4	0.962 6	231
weighted avg	0.9588	0.956 7	0.956 8	231

Berdasarkan Tabel 3 selaras dengan karakteristik warna tersebut, *confusion matrix* dan tabel metrik klasifikasi memperlihatkan bagaimana model merespons variasi visual pada data uji. Model menunjukkan performa yang kuat pada Fold 1, dengan kelas Anthracnose berhasil diklasifikasikan sempurna dan kelas lainnya Downy Mildew, Healthy, dan Mosaic Virus memperlihatkan tingkat ketepatan prediksi yang tinggi meskipun masih terdapat sejumlah kecil kesalahan pada kelas dengan kemiripan fitur visual. Hal ini tercermin dalam nilai *precision*, *recall*, dan *f1-score* yang sebagian besar berada di atas 0.94, dengan *f1-score* tertinggi sebesar **1.00** pada kelas Anthracnose dan nilai terendah tetap berada pada kisaran yang kuat sekitar **0.94** pada kelas lainnya. Secara keseluruhan, akurasi model mencapai **95.67%**, menunjukkan kemampuan generalisasi yang baik dan konsistensi performa terhadap variasi visual dalam dataset.

Evaluasi Fold 2

Pada evaluasi fold 2, model menghasilkan Confusion Matrix berikut yang menggambarkan sebaran prediksi terhadap setiap kelas penyakit daun semangka.



Gambar 3 Confusion Matrix - Fold 2

Berdasarkan Gambar 3 memperlihatkan bahwa model mampu mengklasifikasikan empat kelas penyakit daun semangka *Anthraco*, *Downy Mildew*, *Healthy*, dan *Mosaic Virus* dengan tingkat ketepatan yang sangat tinggi. Seluruh sampel Anthracnose (31 citra) berhasil dikenali tanpa kesalahan, begitu pula kelas Downy Mildew dengan 76 sampel yang terklasifikasi sempurna.

Untuk kelas *Healthy*, model mampu mengidentifikasi 37 sampel dengan benar, tetapi masih terdapat empat sampel yang salah diarahkan ke kategori *Mosaic Virus*. Hal ini menunjukkan adanya kedekatan kemiripan visual antara kedua kelas tersebut. Sementara itu, kelas Mosaic Virus menunjukkan akurasi penuh dengan seluruh 83 sampel diprediksi tepat tanpa ada kesalahan klasifikasi. Dari pola tersebut terlihat bahwa sebagian besar kesalahan model pada Fold 2 terkonsentrasi pada kelas *Healthy*, sedangkan tiga kelas lainnya dapat dipisahkan dengan sangat baik.

Tabel 4 Metrik Area Under Curve (AUC) Per Kelas Fold 2

Kelas	AUC-ROC	AUC-PR
Anthraco	1	1
Downy Mildew	1	1
Healthy	1	1
Mosaic Virus	0.9998	0.9997

Berdasarkan Tabel 4 metrik AUC-ROC dan AUC-PR untuk seluruh kelas berada pada rentang hampir sempurna. *Anthraco*, *Downy Mildew*, dan *Healthy* masing-masing memperoleh nilai 1.00 pada kedua jenis AUC, yang menunjukkan kemampuan model membedakan kelas positif dan negatif secara ideal.

Kelas *Mosaic Virus* juga menunjukkan performa yang hampir sempurna dengan nilai AUC-ROC sebesar 0.9998 dan AUC-PR sebesar 0.9997. Nilai yang sangat tinggi ini mencerminkan sensitivitas dan spesifisitas model yang luar biasa, sekaligus menunjukkan bahwa model tetap stabil pada berbagai ambang prediksi. Secara keseluruhan, rentang nilai AUC ini menegaskan bahwa pemisahan antar kelas

dapat dilakukan model dengan tingkat kejelasan yang nyaris sempurna.

Tabel 5 Ringkasan Metrik Klasifikasi Fold 2

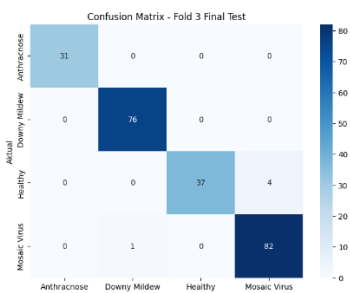
Kelas	precision	recall	F1-score	support
Anthraco se	1	1	1	31
Downy Mildew	1	1	1	76
Healthy	1	0.902 4	0.948 7	41
Mosaic Virus	0.954	1	0.976 3	83
accuracy	0.9827	0.982 7	0. 9827	0.9827
macro avg	0.9885	0.975 6	0.981 3	231
weighted avg	0.9835	0.982 7	0.982 4	231

Berdasarkan Tabel 5 hasil metrik klasifikasi memperlihatkan bahwa model menunjukkan performa unggul pada hampir seluruh kelas. *Anthraco* dan *Downy Mildew* memperoleh *precision*, *recall*, dan *f1-score* sempurna (1.00), sesuai dengan performa yang tampak pada *confusion matrix*. Untuk kelas *Healthy*, *precision* berada pada angka 1.00, tetapi *recall* menurun menjadi 0.9024 akibat empat sampel yang salah teridentifikasi sebagai *Mosaic Virus*; kondisi ini menyebabkan *f1-score* berada pada nilai 0.9487, yang tetap tergolong sangat baik.

Kelas *Mosaic Virus* memberikan hasil yang konsisten kuat dengan *precision* sebesar 0.954, *recall* sempurna 1.00, serta *f1-score* 0.9765. Temuan ini menunjukkan kemampuan model mengenali seluruh citra *Mosaic Virus* dengan benar sekaligus menjaga tingkat ketepatan prediksi yang tinggi. Secara keseluruhan, akurasi Fold 2 mencapai 98.27%, didukung oleh nilai *macro average* dan *weighted average* yang mendekati sempurna, menandakan stabilitas kinerja model di seluruh kelas meskipun terdapat perbedaan jumlah sampel antar kategori.

Evaluasi Fold 3

Pada evaluasi fold 3, model menghasilkan Confusion Matrix berikut yang menggambarkan sebaran prediksi terhadap setiap kelas penyakit daun semangka.



Gambar 4 Confusion Matrix - Fold 3

Berdasarkan Gambar 4 menunjukkan bahwa model mampu mengklasifikasikan sebagian besar sampel dengan sangat akurat. Seluruh sampel *Anthraco* (31 citra) diidentifikasi dengan benar. Kelas *Downy Mildew* juga hampir sempurna, dengan 76 sampel sesuai prediksi dan hanya satu kasus salah klasifikasi ke *Mosaic Virus*. Pada kelas *Healthy*, sebanyak 37 sampel terdeteksi benar, sementara 4 sampel kembali keliru diprediksi sebagai *Mosaic Virus*. Untuk kelas *Mosaic Virus*, model mengklasifikasikan 82 dari 83 sampel secara tepat, hanya satu yang salah diprediksi menjadi *Downy Mildew*. Pola kesalahan tetap konsisten: sebagian besar kekeliruan terjadi antara *Healthy* dan *Mosaic Virus*, sedangkan kelas lain tetap terpisah dengan jelas.

Tabel 6 Metrik Area Under Curve (AUC) Per Kelas Fold 3

Kelas	AUC-ROC	AUC-PR
Anthraco nose	1	1
Downy Mildew	0.9999	0.9998
Healthy	0.9997	0.9988
Mosaic Virus	0.9987	0.9978

Berdasarkan pada Tabel 6 nilai AUC-ROC dan AUC-PR pada Fold 3 menunjukkan performa pemisahan kelas yang sangat kuat. Kelas *Anthraco* mencapai nilai sempurna pada kedua metrik, menandakan tidak adanya tumpang tindih karakteristik antar kelas. Kelas *Downy Mildew* serta *Healthy* juga memperoleh nilai nyaris sempurna, menunjukkan kemampuan model yang sangat baik dalam mendeteksi sampel positif di berbagai ambang keputusan. Meskipun kelas *Mosaic Virus* memiliki nilai AUC sedikit lebih rendah dibanding kelas lainnya, performanya tetap berada pada level hampir ideal, mengindikasikan bahwa model mampu mempertahankan sensitivitas dan presisi yang tinggi untuk seluruh kelas.

CONCLUSION

Dari seluruh rangkaian pengujian yang dilakukan, penelitian ini menyimpulkan bahwa kombinasi metode fine-tuning VGG16 dengan augmentasi data sintetis berbasis GAN mampu meningkatkan akurasi dan ketahanan model dalam mengklasifikasikan penyakit daun semangka. Proses fine-tuning pada VGG16 memungkinkan model mempelajari karakter visual penyakit secara lebih mendalam, sedangkan augmentasi berbasis GAN terbukti efektif mengatasi ketidakseimbangan distribusi kelas dengan memperkaya representasi sampel pada kelas minoritas. Evaluasi menggunakan Stratified K-Fold Cross-Validation menunjukkan performa model yang stabil pada semua kategori penyakit. Selain itu, analisis Grad-CAM memastikan bahwa model menitikberatkan perhatian pada area lesi yang benar, sehingga

prediksinya sejalan dengan kondisi visual pada citra asli. Secara keseluruhan, pendekatan ini berhasil meningkatkan kualitas klasifikasi dan dapat dipertimbangkan sebagai solusi diagnosis berbasis citra yang handal dalam konteks pertanian.

REFERENCE

- Acosta-Prado, J. C., Rojas Rincón, J. S., Mejía Martínez, A. M., & Riveros Tarazona, A. R. (2024). Trends in the Literature About the Adoption of Digital Banking in Emerging Economies: A Bibliometric Analysis. *Journal of Risk and Financial Management*, 17(12). <https://doi.org/10.3390/jrfm17120545>
- Apau, R., Titis, E., & Lallie, H. S. (2025). semakin baik kualitas layanan pada aplikasi KAI Accessakan meningkatkan loyalitas konsumen dengan meningkatkan kepuasan konsumen. *Computers*, 14(4).
- Aung, S. T., Funabiki, N., Aung, L. H., Kinari, S. A., Mentari, M., & Wai, K. H. (2024). A Study of Learning Environment for Initiating Flutter App Development Using Docker. *Information (Switzerland)*, 15(4). <https://doi.org/10.3390/info15040191>
- Bangun, N., Intarti, K., Karo, S. B., Dewiningsih, S., & Tahar, S. (2023). *System quality , information quality , system design quality website PT KCI berpengaruh terhadap user satisfaction*. 9(2), 944–958.
- Berihun, N. G., Dongmo, C., & Van der Poll, J. A. (2023). The Applicability of Automated Testing Frameworks for Mobile Application Testing: A Systematic Literature Review. *Computers*, 12(5). <https://doi.org/10.3390/computers12050097>
- Chen, H., Chu, Y. C., & Lai, F. (2023). Mobile time banking on blockchain system development for community elderly care. *Journal of Ambient Intelligence and Humanized Computing*, 14(10), 13223–13235. <https://doi.org/10.1007/s12652-022-03780-6>
- Ilham Tri Maulana. (2022). Penerapan Metode Sdlc (System Development Life Cycle) Waterfall Pada E-Commerce Smartphone. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 2(2), 1–6. <https://doi.org/10.55606/juisik.v2i2.162>
- Jafri, J. A., Mohd Amin, S. I., Abdul Rahman, A., & Mohd Nor, S. (2024). A systematic literature review of the role of trust and security on Fintech adoption in banking. *Heliyon*, 10(1), e22980. <https://doi.org/10.1016/j.heliyon.2023.e22980>
- Marpid, N. N., Kurniawan, Y. I., & Rahayu, S. P. (2025). Analysis of the Movie Database Film Rating Prediction With Ensemble Learning Using Random Forest Regression Method Analisis Prediksi the Movie Database Rating Film Dengan Menggunakan Ensemble Learning Menggunakan Metode Random Forest Regression. *Jurnal Teknik Informatika (JUTIF)*, 6(1), 1–10. <https://doi.org/10.52436/1.jutif.2025.6.1.1563>
- Meneses, B., & Varajão, J. (2022). A Framework of Information Systems Development Concepts. *Business Systems Research*, 13(1), 84–103. <https://doi.org/10.2478/bsrj-2022-0006>
- Ramayanti, R., Rachmawati, N. A., Azhar, Z., & Nik Azman, N. H. (2024). Exploring intention and actual use in digital payments: A systematic review and roadmap for future research. *Computers in Human Behavior Reports*, 13(October 2023), 100348. <https://doi.org/10.1016/j.chbr.2023.100348>
- Saifudin, A., Saputra, A., Saputra, B., Subhan, F., Maulana, F., & Kusyadi, I. (2022). Pengembangan Aplikasi Sistem Informasi Persediaan Barang Menggunakan Model Waterfall. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 5(4), 247–254. <https://doi.org/10.32493/jtsi.v5i4.21197>
- Saravanos, A., & Curinga, M. X. (2023). Simulating the Software Development Lifecycle: The Waterfall Model. *Applied System Innovation*, 6(6). <https://doi.org/10.3390/asi6060108>
- Sasmoko, Indrianti, Y., Manalu, S. R., & Danaristo, J. (2024). Analyzing Database Optimization Strategies in Laravel for an Enhanced Learning Management. *Procedia Computer Science*, 245(C), 799–804. <https://doi.org/10.1016/j.procs.2024.10.306>
- Souha, A., Benaddi, L., Ouaddi, C., & Jakimi, A. (2024). Comparative analysis of mobile application Frameworks: A developer's guide for choosing the right tool. *Procedia Computer Science*, 236, 597–604. <https://doi.org/10.1016/j.procs.2024.05.071>