

Perancangan Aplikasi Berbasis Web Menggunakan Vision Transformer (ViT) untuk Klasifikasi Penyakit Daun Tanaman [Web-Based Application Design Using Vision Transformer (ViT) for Plant Leaf Disease Classification]

Zaidan Dafa Robani Hatuwe¹⁾, Rohman Dijaya, S.Kom. M.Kom²⁾, Sukma Aji, S.T, M.Kom³⁾, Cindy Taurusta, S.ST., MT.

¹⁾ Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: Rohman.dijaya@umsida.ac.id

Abstract. *Plant leaf disease is one of the major factors reducing agricultural productivity, yet the conventional visual-identification method used by farmers remains subjective, time-consuming, and dependent on limited expert availability. Most prior studies address this problem using Convolutional Neural Network (CNN)-based architectures, which are inherently limited by their local receptive field and therefore struggle to capture global spatial context in complex disease patterns. This study proposes a Vision Transformer (ViT), specifically the ViT-Small/16 (ViT-S/16) variant, to classify plant leaf disease using the Open Leaf Image Dataset I (OLID I), which contains 4,749 leaf images across 57 unique classes from eight major crop species. The ViT-S/16 model divides each input image into 196 non-overlapping 16×16-pixel patches and applies a Multi-Head Self-Attention mechanism to model long-range dependencies between image regions, allowing it to capture the global spatial context that CNN-based approaches typically miss. The resulting classification model is integrated into a web-based application with a user interface, allowing farmers and agricultural extension workers to obtain disease diagnoses directly in the field. On the held-out test set, the trained ViT-Small/16 model achieved an accuracy of 65.32%, with macro-averaged precision, recall, and F1-score of 65.64%, 57.87%, and 58.33% respectively (weighted averages of 67.56%, 65.32%, and 63.80%). Confusion matrix analysis shows that most misclassifications occur between closely related nutrient-deficiency sub-classes within the same crop species, while confusion across different crop species is minimal. These findings demonstrate the potential of the Vision Transformer as a context-aware alternative to CNN for large-scale, fine-grained plant disease classification, while also highlighting an overfitting gap between training and validation performance that leaves room for further optimization.*

Keywords - Vision Transformer, plant disease classification, deep learning, self-attention, image classification

Abstrak. *Penyakit daun tanaman merupakan salah satu faktor utama penurunan produktivitas pertanian, sementara metode identifikasi visual konvensional yang digunakan petani masih bersifat subjektif, memakan waktu, dan bergantung pada ketersediaan tenaga ahli yang terbatas. Sebagian besar penelitian terdahulu mengatasi masalah ini menggunakan arsitektur Convolutional Neural Network (CNN), yang secara inheren dibatasi oleh local receptive field sehingga sulit menangkap konteks spasial global pada pola penyakit yang kompleks. Penelitian ini mengusulkan Vision Transformer (ViT), khususnya varian ViT-Small/16 (ViT-S/16), untuk mengklasifikasikan penyakit daun tanaman menggunakan Open Leaf Image Dataset I (OLID I) yang terdiri atas 4.749 citra daun dari 57 kelas unik pada delapan spesies tanaman utama. Model ViT-S/16 membagi setiap citra masukan menjadi 196 patch berukuran 16×16 piksel yang tidak saling tumpang tindih, kemudian menerapkan mekanisme Multi-Head Self-Attention untuk memodelkan ketergantungan jarak jauh antarwilayah citra, sehingga mampu menangkap konteks spasial global yang umumnya terlewatkan oleh pendekatan berbasis CNN. Model klasifikasi yang dihasilkan diintegrasikan ke dalam aplikasi berbasis web yang dilengkapi antarmuka pengguna, sehingga petani dan penyuluh pertanian dapat memperoleh diagnosis penyakit secara langsung di lapangan. Data uji pada model ViT-Small/16 yang telah dilatih mencapai akurasi sebesar 65,32%, dengan precision, recall, dan F1-score rata-rata makro berturut-turut sebesar 65.64%, 57.87%, dan 58.33% (rata-rata tertimbang sebesar 67.56%, 65.32%, dan 63.80%). Analisis confusion matrix menunjukkan bahwa sebagian besar kesalahan klasifikasi terjadi antarsubkelas defisiensi nutrisi yang berkerabat dekat pada spesies tanaman yang sama, sementara kesalahan antarspesies tanaman relatif kecil. Temuan ini menunjukkan potensi Vision Transformer sebagai alternatif yang peka konteks (context-aware) terhadap CNN untuk klasifikasi penyakit daun tanaman berskala besar dan berbutir halus (fine-grained), sekaligus mengungkap adanya kesenjangan overfitting antara kinerja pelatihan dan validasi yang masih terbuka untuk dioptimalkan lebih lanjut.*

Kata Kunci - Vision Transformer, klasifikasi penyakit tanaman, deep learning, self-attention, klasifikasi citra.

I. PENDAHULUAN

Sektor pertanian menyediakan pangan, bahan baku, dan lapangan kerja yang penting bagi perekonomian suatu negara. Di Indonesia, sektor ini menghadapi berbagai tantangan signifikan, termasuk menurunnya minat generasi muda terhadap bidang pertanian dan persoalan penyakit tanaman yang terus berlangsung [1]. Penyakit daun sering menjadi indikator awal gangguan kesehatan tanaman; apabila tidak segera didiagnosis dan ditangani, penyakit tersebut dapat menyebar luas dan menimbulkan kerugian panen yang besar, mengingat berbagai kajian pertanian global menunjukkan bahwa penyakit tanaman menyebabkan kerugian ekonomi hingga miliaran dolar setiap tahun [2]. Metode konvensional yang digunakan petani untuk mengidentifikasi penyakit daun adalah pengamatan visual. Pendekatan manual ini memiliki beberapa kelemahan mendasar, yaitu subjektivitas diagnosis, waktu pemeriksaan yang lama untuk area yang luas, dan ketergantungan pada keahlian spesialis yang tidak selalu tersedia di wilayah pertanian [3]. Diagnosis yang terlambat dan tidak akurat sering kali mengakibatkan penanganan yang keliru, pemborosan penggunaan pestisida, dan kerusakan lingkungan, sehingga sektor pertanian memerlukan solusi berbasis teknologi yang mampu mendiagnosis penyakit secara otomatis, cepat, dan akurat [4].

Banyak penelitian terdahulu menerapkan arsitektur Convolutional Neural Network (CNN) untuk mendeteksi dan mengklasifikasikan penyakit daun tanaman [5]. Akan tetapi, CNN memiliki keterbatasan mendasar pada local receptive field, sehingga sering kali sulit menangkap konteks spasial global pada citra dengan pola penyakit yang kompleks. Kondisi ini mendorong perlunya arsitektur yang mampu memodelkan ketergantungan jarak jauh (long-range dependencies) pada keseluruhan citra sekaligus [6]. Solusi penelitian ini mengusulkan Vision Transformer (ViT), khususnya varian ViT-Small/16 (ViT-S/16), yang berada pada titik Pareto-optimal antara akurasi dan efisiensi komputasi: memiliki sekitar 22 juta parameter, jauh lebih ringan dibandingkan ViT-Base/16 (86 juta parameter), namun tetap memiliki kapasitas memadai untuk membedakan banyak kelas penyakit. Konvolusi standar berbeda dengan ViT-S/16 yang membagi citra menjadi patch berukuran tetap yang tidak saling tumpang tindih dan menggunakan mekanisme Self-Attention untuk mempelajari ketergantungan jarak jauh antarbagian citra [7], pendekatan yang telah terbukti mengungguli CNN konvensional pada berbagai tugas deteksi visual [8].

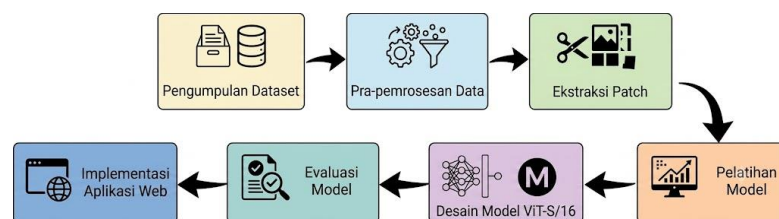
Model klasifikasi yang akurat perlu diintegrasikan ke dalam platform yang aplikatif agar bermanfaat di lapangan. Luaran penelitian ini berupa model klasifikasi berbasis Vision Transformer yang dilengkapi antarmuka pengguna untuk mendeteksi penyakit daun tanaman secara otomatis, sehingga petani dan penyuluh pertanian dapat memperoleh diagnosis yang cepat dan akurat secara langsung di lapangan [9],[10]. Latar belakang penelitian ini bertujuan untuk merancang model klasifikasi penyakit daun tanaman menggunakan Vision Transformer (ViT-Small/16) pada Open Leaf Image Dataset I (OLID I) [11], mengimplementasikan model yang dihasilkan ke dalam aplikasi berbasis web untuk mengklasifikasikan penyakit daun tanaman, dan menganalisis kinerja model berdasarkan Accuracy, Precision, Recall, dan F1-Score. Penelitian ini dibatasi pada arsitektur Vision Transformer, dataset OLID I 4.749 citra daun dari 57 kelas pada delapan tanaman yang ditanam di Bangladesh, dalam kondisi sehat, kekurangan nutrisi, dan terserang hama serta antarmuka aplikasi yang dirancang khusus untuk masukan citra dan menampilkan hasil klasifikasi model.

Penelitian mengenai klasifikasi penyakit daun tanaman hingga saat ini sebagian besar masih didominasi oleh arsitektur berbasis CNN. Penyakit daun tomat, beberapa penelitian menerapkan model ResNet-50 [12] dan CNN standar [13], serta pendekatan ensemble CNN [14]. Penyakit daun kentang telah diklasifikasikan menggunakan model gabungan CNN dan RNN [15]. Untuk penyakit daun cabai, CNN telah digunakan untuk deteksi dini [16]. Penyakit daun padi telah dikaji menggunakan KNN dengan GLCM dan Canny edge detection [17], aplikasi CNN berbasis mobile dengan integrasi Firebase secara real-time [18], serta sistem klasifikasi dan pemantauan berbasis CNN [19], [20]. Arsitektur hybrid VGG16–ResNet34 juga telah diusulkan untuk klasifikasi penyakit daun secara umum [21]. Pendekatan-pendekatan berbasis CNN tersebut melaporkan akurasi yang baik, tetapi tetap dibatasi oleh local receptive field: perluasan jangkauan konteksnya memerlukan penumpukan banyak lapisan konvolusi, yang meningkatkan biaya komputasi dan risiko overfitting, sebagaimana teramati pada beberapa penelitian tersebut. Perbandingan angka akurasi secara langsung antarpelelitian tersebut kurang bermakna, karena masing-masing menggunakan dataset yang berbeda dari segi jenis tanaman, jumlah kelas, dan kondisi citra. Injauan literatur, belum ada penelitian sebelumnya yang mengevaluasi Vision Transformer murni, tanpa komponen konvolusi sama sekali, pada tugas klasifikasi penyakit tanaman dengan jumlah kelas sebanyak dataset OLID I yang terdiri atas 57 kelas. Oleh karena itu, penelitian ini bertujuan mengisi kesenjangan tersebut dengan mengevaluasi secara langsung kemampuan ViT-Small/16 dalam memodelkan konteks citra global melalui mekanisme Self-Attention pada tugas klasifikasi multi-tanaman yang berbutir halus (fine-grained) ini, serta melaporkan temuannya sebagai referensi bagi penelitian klasifikasi penyakit tanaman berbasis deep learning selanjutnya.

II. METODE

2.1 Alur Penelitian

Penelitian ini dilaksanakan melalui beberapa tahap yang disusun secara sistematis, mulai dari pengumpulan dataset hingga evaluasi model dan implementasinya ke dalam aplikasi berbasis web, sebagaimana diilustrasikan pada Gambar 1.



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan alur penelitian secara keseluruhan. Penelitian ini diawali dengan pengumpulan Open Leaf Image Dataset I (OLID I), dilanjutkan dengan praproses citra dan ekstraksi patch untuk mempersiapkan data sesuai format masukan Vision Transformer. Model ViT-Small/16 kemudian dirancang, dilatih, dan dievaluasi menggunakan metrik klasifikasi standar, sebelum model yang dihasilkan akhirnya diintegrasikan ke dalam aplikasi berbasis web untuk penggunaan praktis.

2.2 Dataset dan Praproses Citra

Dataset yang digunakan dalam penelitian ini adalah Open Leaf Image Dataset I (OLID I) yang dikembangkan oleh Orka dan kawan kawan (2023) dan tersedia secara publik melalui Kaggle dan Zenodo [11]. OLID I terdiri atas 4.749 citra daun yang diambil pada kondisi lapangan alami, terbagi ke dalam 57 kelas unik termasuk 16 kelas multi-label yang merepresentasikan lebih dari satu kondisi stres tanaman dalam satu citra dari delapan tanaman utama yang ditanam di Bangladesh, yaitu tomat, terong, mentimun, pare, oyong, gambas, labu putih, dan labu botol. Seluruh kelas dikelompokkan ke dalam tiga kategori utama, yaitu sehat, kekurangan nutrisi, dan terserang hama.

Seluruh pelatihan citra di *preprocessing* agar data masukan Vision Transformer konsisten. Setiap citra yang semula berukuran 3024×3024 piksel diubah ukurannya menjadi 224×224 piksel agar sesuai dengan kebutuhan masukan ViT, dan nilai piksel dinormalisasi dari rentang 0–255 menjadi 0–1 dengan membagi setiap nilai piksel dengan 255. Dataset kemudian dibagi menjadi subset pelatihan, validasi, dan pengujian dengan rasio 75%:12,5%:12,5% (3.561/594/594 citra) karena ViT memiliki *inductive bias* yang lebih rendah dibandingkan CNN konvensional, sehingga model ini bersifat sangat bergantung pada kuantitas data (*data-hungry*) untuk mengoptimalkan mekanisme *self-attention* [22], dengan ukuran batch sebesar 32. Augmentasi data flip horizontal, rotasi, zoom acak, dan shear acak diterapkan pada set pelatihan untuk mengatasi ketidakseimbangan kelas dan meningkatkan generalisasi model.

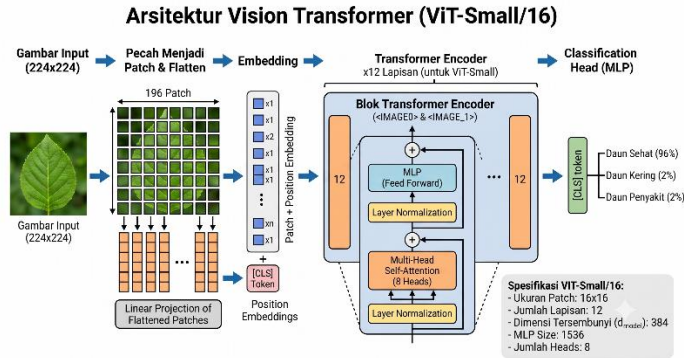
2.3 Ekstraksi Patch

Ekstraksi Patch merupakan inti dari praproses khas ViT dan menjadi pembeda mendasar dari arsitektur CNN. Setiap citra berukuran 224×224 piksel dibagi menjadi N patch berukuran tetap $P \times P$ yang tidak saling tumpang tindih, dihitung dengan $N = (H \times W) / P^2$, dengan H adalah tinggi citra, W adalah lebar citra, dan P adalah panjang sisi patch dengan $H = 224$, $W = 224$, dan $P = 16$, diperoleh $N = (224 \times 224) / 16^2 = 196$ patch. Patch berukuran $16 \times 16 \times 3$ (tinggi $\times$ lebar $\times$ kanal RGB) kemudian diratakan (flatten) menjadi vektor satu dimensi sepanjang $P^2 \cdot C = 768$ elemen, sehingga menghasilkan urutan 196 vektor patch yang selanjutnya diproyeksikan ke ruang laten model. ViT memperlakukan citra sebagai urutan 196 token visual bukan sebagai grid piksel dua dimensi dengan cara ini, sehingga mekanisme Self-Attention dapat memodelkan hubungan kontekstual pada keseluruhan citra secara simultan, tanpa dibatasi oleh local receptive field yang menjadi ciri khas konvolusi.

2.4 Arsitektur Model: Vision Transformer (ViT-Small/16)

Penelitian ini menggunakan varian ViT-Small/16 (ViT-S/16) dari Vision Transformer, yang terdiri atas sekitar 22 juta parameter, 12 lapisan Transformer Encoder, 8 attention head, dan dimensi embedding sebesar 384. ViT-S/16 dipilih karena keseimbangannya antara akurasi dan efisiensi komputasi, dengan kapasitas yang memadai

untuk mengklasifikasikan 57 kelas penyakit sekaligus memiliki resiko overfitting yang lebih rendah dibandingkan varian ViT yang lebih besar seperti ViT-Base/16 (86 juta parameter). Arsitektur yang diusulkan, sebagaimana dilustrasikan pada Gambar 2.



Gambar 2. Arsitektur Vision Transformer

Gambar 2 menjelaskan alur arsitektur Vision Transformer (ViT-Small/16) untuk klasifikasi penyakit daun tanaman, diawali dengan menerima citra masukan berukuran 224×224 piksel yang dipecah menjadi 196 *patch* berukuran 16×16 piksel dan diratakan menjadi vektor 1D. Vektor *patch* tersebut kemudian diproyeksikan melalui *Linear Projection* menjadi *embedding* berdimensi D ($d_{\text{model}} = 384$), disisipkan token [CLS] di awal urutan sebagai pembawa representasi global, serta ditambahkan *Positional Embedding* pada seluruh 197 token untuk menyertakan informasi spasial. Selanjutnya, urutan token diproses oleh *Transformer Encoder* bertumpuk sebanyak 12 lapisan masing-masing menerapkan *Multi-Head Self-Attention* (8 *heads*) dan *Multi-Layer Perceptron* (MLP size 1536) yang dilengkapi *Layer Normalization* serta *residual connection* sebelum akhirnya representasi fitur keluaran dari token [CLS] diteruskan ke *Classification Head* (MLP) dengan fungsi aktivasi *Softmax* untuk memetakan vektor menjadi distribusi probabilitas atas 57 kelas penyakit.

2.5 Pelatihan Model

Model ViT-Small/15 dilatih menggunakan *Cross-Entropy Loss* (dengan parameter *label smoothing* 0,1) sebagai fungsi kerugian (*loss function*). Nilai oterasi kerugian model dihitung untuk mengevaluasi seberapa jauh probabilitas prediksi model menyimpang dari label sebenarnya. Perhitungan *loss* baik pada fase pelatihan (*Train Loss*), validasi (*Validation Loss*) dan metrik akurasi (*Train Accuracy* dan *Validation Accuracy*) yang menghitung persentase prediksi yang tepat dari total keseluruhan data yang diuji pada fase tersebut dirumuskan sebagai berikut:

$$Loss = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (1)$$

N adalah jumlah data yang dievaluasi, C adalah total kelas penyakit, $y_{i,c}$ merupakan label target sebenarnya (bernilai 1 jika benar dan 0 jika salah), serta $p_{i,c}$ adalah probabilitas keluaran model untuk kelas c .

$$Accuracy = \frac{\sum_{i=1}^N \mathbb{I}(y_i = \hat{y}_i)}{N} \times 100\% \quad (2)$$

y_i adalah label asli, $\hat{y}_i$ adalah kelas yang diprediksi oleh model, dan $\mathbb{I}$ merupakan fungsi indikator yang bernilai 1 jika tebakan model sesuai dengan aslinya.

2.6 Metrik Evaluasi

Kinerja model diukur pada data uji menggunakan Confusion Matrix, yang darinya diturunkan empat metrik evaluasi, yaitu Accuracy, Precision, Recall, dan F1-Score, masing-masing dihitung per kelas serta dalam bentuk rata-rata makro dan rata-rata tertimbang. Perhitungan keseluruhan metrik didasarkan pada empat komponen kondisi *Confusion Matrix* [23], yaitu:

1. **True Positive (TP):** Prediksi di mana model secara tepat mengenali kelas positif (misalnya, klasifikasi penyakit daun tertentu dengan benar).
2. **True Negative (TN):** Prediksi di mana model secara tepat mengenali kelas negatif.

3. **False Positive (FP):** Prediksi keliru di mana model mengklasifikasikan data ke dalam kelas positif, padahal kondisi aslinya adalah negatif.
4. **False Negative (FN):** Prediksi keliru di mana model gagal mendeteksi kelas positif dan malah memprediksinya sebagai kelas negatif.

Keempat komponen tersebut, persamaan untuk masing-masing metrik evaluasi adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Accuracy mencerminkan proporsi keseluruhan prediksi yang benar dari total pengujian yang dihasilkan oleh model.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Precision mengukur seberapa tepat prediksi positif yang dihasilkan model; menunjukkan tingkat akurasi saat model memprediksi kelas tertentu.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Precision mengukur seberapa tepat prediksi positif yang dihasilkan model; menunjukkan tingkat akurasi saat model memprediksi kelas tertentu.

$$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (6)$$

F1-Score merupakan rata-rata harmonik dari *Precision* dan *Recall*. Metrik ini memberikan ukuran kinerja yang lebih seimbang dan informatif, yang mana hal ini sangat esensial ketika diterapkan pada dataset multi-kelas dengan distribusi yang tidak seimbang seperti OLID I.

III. HASIL DAN PEMBAHASAN

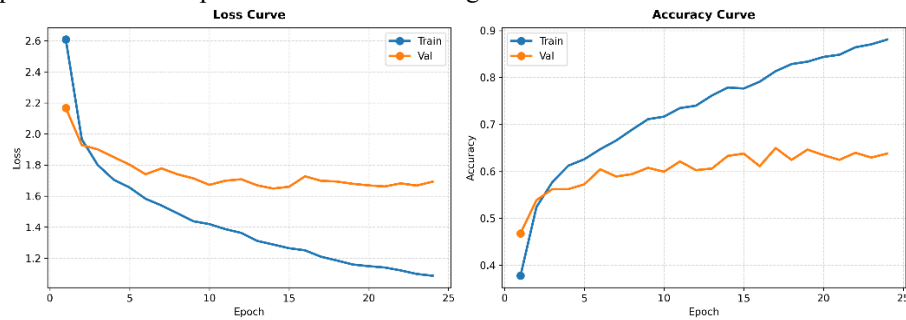
3.1 Hasil Pelatihan

Model ViT-Small/16 dilatih hingga maksimal 30 epoch menggunakan konfigurasi yang dijelaskan pada Sub bab 2.5. Mekanisme Early Stopping menghentikan pelatihan pada epoch ke-24 setelah akurasi validasi tidak lagi mengalami peningkatan, sehingga mengembalikan checkpoint dengan kinerja terbaik yang mencapai akurasi validasi sebesar 64,98%. Tabel 1 merangkum perkembangan pelatihan dan validasi pada epoch-epoch terpilih.

Tabel 1. Perkembangan Pelatihan dan Validasi pada Epoch

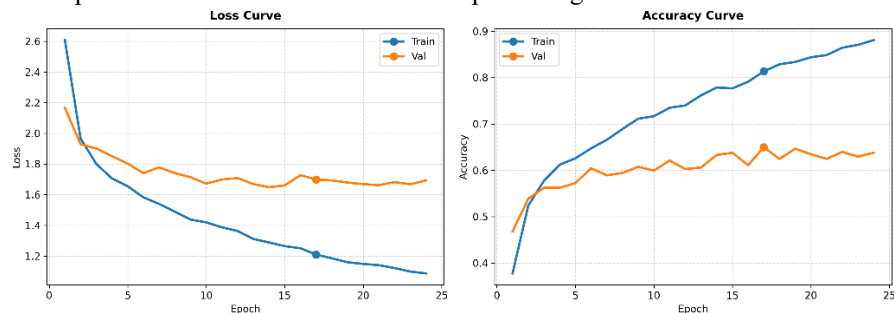
Epoch	Train Loss	Train Accuracy	Val Loss	Val Accuracy
1	2.6076	37.77%	2.1665	46.8%
6	1.5828	64.73%	1.7406	60.44%
12	1.3632	74%	1.7084	60.27%
13	1.3117	76.16%	1.6695	60.61%
14	1.2885	77.87%	1.6487	63.3%
17(terbaik)	1.2885	81.35%	1.6987	64.98%
18	1.1854	82.87%	1.6937	62.46%
19	1.1597	83.38%	1.6791	64.65%
20	1.1489	84.39%	1.6690	63.47%
24(early stop)	1.0873	88.09%	1.6929	63.8%

Akurasi pelatihan pada Tabel 1 meningkat secara konsisten dari 37,77% pada epoch pertama menjadi 88,09% pada epoch ke-24. Akurasi validasi menunjukkan bahwa melandai jauh lebih awal; meningkat dari 46,80% menjadi berkisar di angka 60–65% setelah epoch ke-6 dan terus berfluktuasi pada rentang tersebut. Pola serupa juga terlihat pada pergerakan nilai *loss* pelatihan menurun secara konsisten sepanjang iterasi, sedangkan *loss* validasi mencapai titik terendahnya di sekitar epoch ke-14 (1,6487) dan berhenti membaik secara berarti setelahnya. Kesenjangan yang semakin melebar antara metrik pelatihan dan validasi ini mengindikasikan bahwa model mulai mengalami *overfitting* terhadap data pelatihan pada paruh kedua iterasi. Detail yang lebih mendalam mengenai dinamika tersebut, perilaku model dapat dievaluasi melalui tiga fase utama berikut:



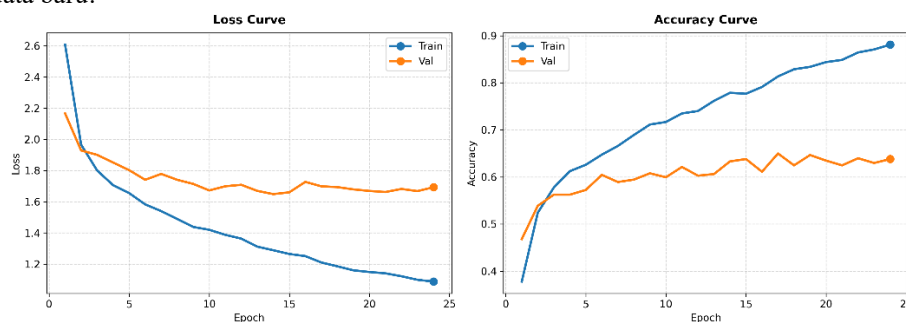
Gambar 3. Plot kurva Loss dan Akurasi pada Epoch 1

Gambar 3 mencatatkan Train Loss yang sangat tinggi sebesar 2,6076 dengan akurasi 37,77%. Kinerja pada data validasi juga masih berada di angka 46,80%. Kondisi ini merupakan hal yang wajar karena bobot model ViT-Small/15 masih dalam tahap inisialisasi awal dan belum beradaptasi dengan baik.



Gambar 4. Plot kurva Loss dan Akurasi pada Epoch 17

Gambar 4 mencatat Performa generalisasi terbaik dicapai pada epoch ke-17. Akurasi validasi pada titik ini berhasil menyentuh angka tertingginya yaitu 64,98%. Walaupun performa pada data latih masih terus berkembang (mencapai akurasi 81,35%), bobot model pada epoch inilah yang dinilai paling optimal dalam menggeneralisasi data baru.



Gambar 5. Plot kurva Loss dan Akurasi pada Epoch 24

Gambar 5 mencatat performa validasi mengalami fluktuasi dan tidak lagi mampu melampaui angka 64,98%, meskipun akurasi pelatihan terus meroket hingga 88,09% dengan *loss* 1,0873. Kondisi ini memicu mekanisme *Early Stopping* untuk menghentikan proses pelatihan secara otomatis pada epoch ke-24. Penghentian ini membuktikan bahwa mekanisme tersebut berhasil mencegah model dari kondisi *overfitting* yang semakin parah akibat terlalu menghafal data latih.

3.2 Evaluasi pada Data Uji

Data uji pada model akhir mencapai akurasi keseluruhan sebesar 65,32%, konsisten dengan akurasi validasi terbaik yang diperoleh selama pelatihan. Tabel 2 menyajikan kumpulan lengkap metrik evaluasi.

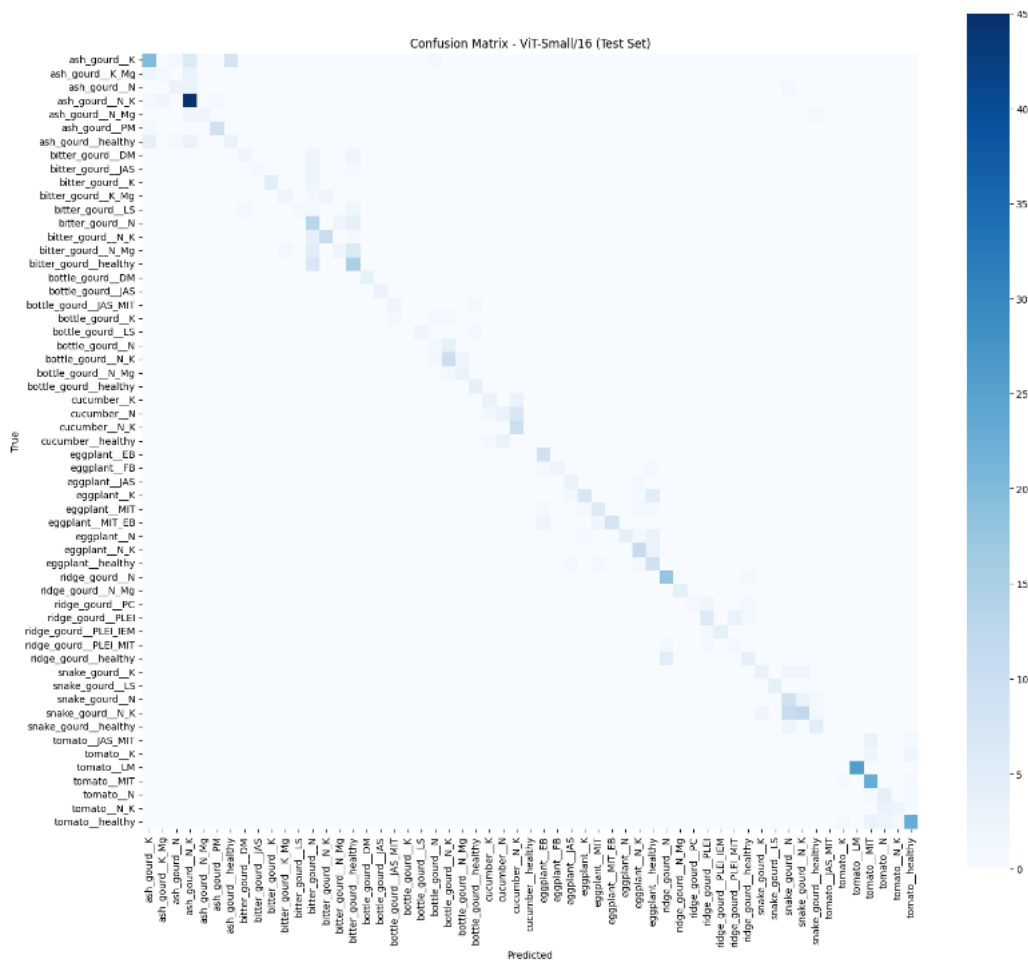
Tabel 2. Matrik Evaluasi pada Data Uji

Metrik	Rata-rata Makro	Rata-rata Tertimbang
Precision	65.64%	67.56%
Recall	57.87%	65.32%
F1-Score	58.33%	63.8%

Metrik rata-rata makro secara konsisten lebih rendah dibandingkan metrik rata-rata tertimbang (Recall sebesar 57.87% dibandingkan 65.32%), yang menunjukkan bahwa model tampak berkinerja lebih buruk pada beberapa kelas minoritas, yang memiliki bobot lebih kecil pada rata-rata tertimbang tetapi dihitung setara pada rata-rata makro. Temuan ini merupakan hal umum pada dataset dengan 57 kelas seperti OLID I, karena beberapa kelas hanya direpresentasikan oleh jumlah citra pelatihan yang relatif sedikit. Kondisi ini disebabkan karena arsitektur Vision Transformer (ViT) memiliki kebutuhan data pelatihan yang sangat besar dan cenderung kurang optimal jika diterapkan pada dataset yang terbatas [24]. Akibatnya, pada kelas minoritas dengan jumlah citra yang sedikit, ViT kesulitan mengenali pola fitur visual secara presisi.

3.3 Analisis Confusion Matrix

Analisis confusion Matrix memberikan evaluasi kinerja klasifikasi yang lebih rinci, confusion matrix model ViT-S/16 pada data uji dianalisis, sebagaimana diilustrasikan pada Gambar 6.

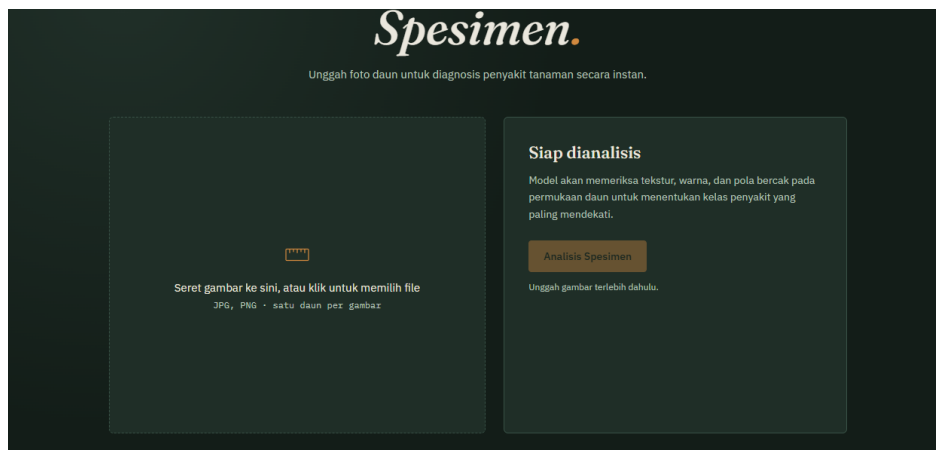


Gambar 6. Confusion Matrix

Klasifikasi pada Gambar 6 sebagian besar terkonsentrasi pada spesies tanaman yang sama dan jarang terjadi antarspesies yang berbeda, yang menunjukkan bahwa model ViT-S/16 mampu membedakan fitur visual pada level spesies tanaman secara andal. Besar kekeliruan justru terjadi antarsubkelas defisiensi nutrisi yang berkerabat dekat pada tanaman yang sama misalnya antara kelas defisiensi nitrogen (N), nitrogen-kalium (N_K), dan nitrogen-magnesium (N_Mg) pada tanaman pare dan labu putih, serta antara kelas N dan N_K pada oyong. Tumpang tindih dalam skala lebih kecil juga teramati antara kelas N_K dan kelas sehat pada terong. Pola ini mengindikasikan bahwa gejala visual tahap awal hingga menengah dari berbagai kombinasi defisiensi nutrisi pada tanaman yang sama dapat sangat mirip, sehingga menjadi subkelas yang paling sulit dipisahkan oleh model, sementara daun yang terserang hama maupun daun sehat pada umumnya diklasifikasikan dengan benar.

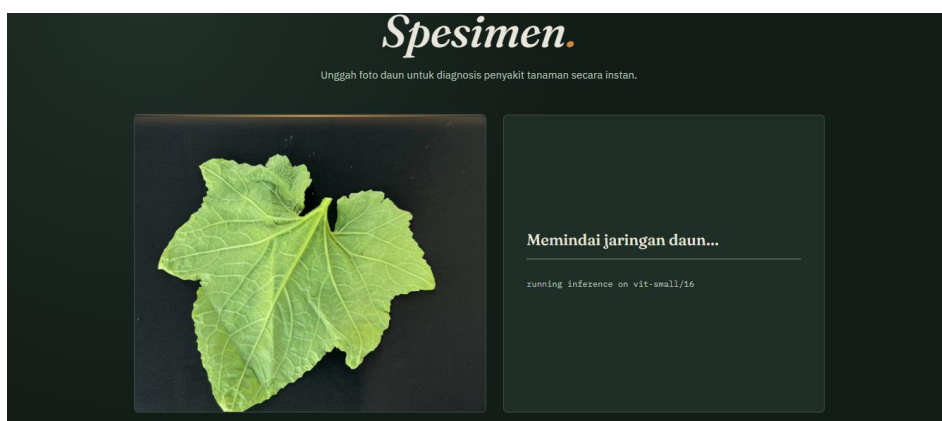
3.4 Implementasi Aplikasi Web

Model ViT-Small/16 yang dihasilkan diintegrasikan ke dalam aplikasi berbasis web yang memungkinkan pengguna mengunggah citra daun dan menerima hasil prediksi kelas penyakit sebagai keluaran, sebagaimana ditunjukkan pada Gambar 7.



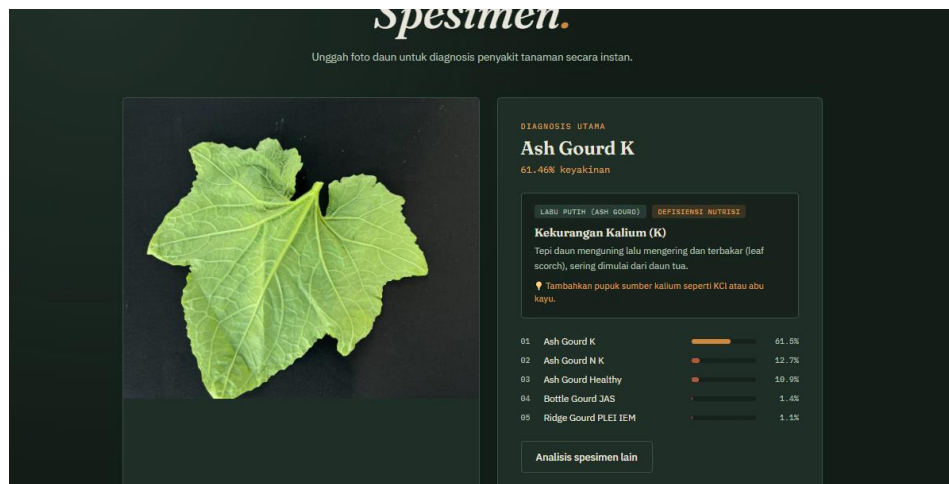
Gambar 7. Antarmuka Aplikasi Web

Antarmuka ini memungkinkan pengguna khususnya petani dan penyuluh pertanian dapat mengunggah foto daun dan memperoleh klasifikasi penyakit secara otomatis tanpa memerlukan keahlian khusus, sehingga secara langsung menjawab apa yang telah diuraikan pada bagian Pendahuluan.



Gambar 8. Pemrosesan Klasifikasi oleh Sistem

Gambar 8 adalah antarmuka yang menampilkan indikator pemrosesan (*loading screen*). Aplikasi akan mengirimkan data citra ke bagian belakang (*backend*). Citra tersebut nantinya diubah ukurannya sesuai dengan masukan model, lalu diekstraksi fitur-fiturnya oleh model ViT-Small/15 untuk menghitung probabilitas setiap kelas penyakit.



Gambar 9. Halaman Keluaran Hasil Deteksi Penyakit

Gambar 9 adalah halaman yang menampilkan hasil prediksi penyakit beserta tingkat kepercayaannya (*confidence score*). Aplikasi ini juga menampilkan informasi jenis penyakit daun dan penanganan awalnya agar dapat mudah menuntun fokus pengguna langsung pada hasil yang diminta.

3.5 Pembahasan

Penelitian Berbasis ViT(Small/16) dibandingkan dengan penelitian berbasis CNN yang dibahas pada bagian Pendahuluan, akurasi sebesar 65,32% yang dicapai dalam penelitian ini lebih rendah daripada angka yang dilaporkan oleh beberapa penelitian sebelumnya; namun, perbandingan tersebut perlu ditafsirkan secara hati-hati, Penelitian-penelitian tersebut umumnya hanya mengklasifikasikan 2–10 kelas pada satu jenis tanaman, jauh lebih sedikit dibandingkan 57 kelas berbutir halus (*fine-grained*) yang dicakup OLID I dalam penelitian ini. Kompleksitas tugas yang jauh lebih tinggi tersebut, hasil yang diperoleh mencerminkan kinerja dasar (*baseline*) yang wajar bagi Vision Transformer tanpa bobot pralatih (*pretrained*), sekaligus mengungkap adanya kesenjangan *overfitting* yang jelas akurasi pelatihan mencapai 88,1% pada saat *Early Stopping*, sementara akurasi validasi hanya melandai pada kisaran 60–65% yang menunjukkan bahwa kemampuan generalisasi model, bukan kapasitas arsitekturnya, menjadi faktor pembatas utama saat ini. Temuan ini konsisten dengan analisis *confusion matrix* pada Sub bab 3.3, yang menunjukkan bahwa model mampu memisahkan spesies tanaman yang berbeda secara andal, tetapi masih kesulitan membedakan subkelas defisiensi nutrisi yang secara visual mirip pada tanaman yang sama, yang beberapa di antaranya direpresentasikan oleh jumlah citra pelatihan yang relatif sedikit

VII. SIMPULAN

Penelitian ini telah berhasil merancang dan melatih model Vision Transformer (ViT-Small/16) untuk klasifikasi penyakit daun tanaman pada Open Leaf Image Dataset I (OLID I) yang terdiri atas 57 kelas, serta mengintegrasikan model yang dihasilkan ke dalam aplikasi berbasis web, sehingga memenuhi ketiga tujuan yang dinyatakan pada bagian Pendahuluan. Model yang diuji telah mencapai akurasi sebesar 65,32%, dengan F1-score rata-rata makro sebesar 0,5833 dan F1-score rata-rata tertimbang sebesar 0,6380, yang menegaskan bahwa arsitektur berbasis *Self-Attention* murni, tanpa komponen konvolusi sama sekali, mampu mempelajari fitur diskriminatif untuk tugas klasifikasi penyakit multi-tanaman yang berbutir halus (*fine-grained*) dengan kompleksitas setingkat ini. Kurva pelatihan dan validasi menunjukkan adanya kesenjangan *overfitting* yang jelas akurasi pelatihan mencapai 88,1% pada titik *Early Stopping* (epoch ke-24), sementara akurasi validasi hanya melandai pada kisaran 60–65% yang mengindikasikan bahwa kemampuan generalisasi model, bukan kapasitas arsitekturnya, menjadi faktor pembatas utama saat ini. *Confusion matrix* lebih lanjut menunjukkan bahwa keterbatasan ini terkonsentrasi pada subkelas tertentu: model mampu memisahkan spesies tanaman yang berbeda secara andal, tetapi masih kesulitan membedakan kombinasi defisiensi nutrisi yang berkerabat dekat pada tanaman yang sama, kemungkinan besar karena subkelas-subkelas tersebut memiliki gejala visual tahap awal yang sangat mirip dan, pada beberapa kasus, hanya direpresentasikan oleh jumlah citra pelatihan yang relatif sedikit. Kesenjangan *overfitting* yang ditemukan dalam penelitian ini dapat diatasi dengan menginisialisasi model menggunakan bobot ViT pralatih (*pretrained*) dari ImageNet alih-alih melatih model dari awal (*from scratch*), menerapkan regularisasi yang lebih kuat (misalnya *dropout* yang lebih tinggi atau *stochastic depth*), serta memperluas data pelatihan untuk subkelas yang kurang terwakili melalui augmentasi tambahan atau pengumpulan data baru. Eksperimen dengan ukuran

patch yang berbeda, varian ViT yang lebih besar, atau arsitektur hybrid CNN Transformer juga berpotensi meningkatkan kemampuan model dalam memisahkan kelas-kelas defisiensi nutrisi yang secara visual mirip.

UCAPAN TERIMA KASIH

Rasa syukur dan apresiasi mendalam penulis sampaikan kepada semua pihak yang senantiasa mendampingi dan memberikan dukungan penuh selama riset ini berlangsung. Bantuan yang diterima, mulai dari arahan akademis, evaluasi konstruktif, hingga akses ke fasilitas pendukung, telah menjadi pilar utama dalam perampungan studi ini. Motivasi tanpa henti dari berbagai pihak turut memastikan penelitian ini berjalan lancar dengan kualitas yang terjaga. Penulis berharap publikasi ini mampu memberikan dampak positif dan menjadi rujukan bagi kajian sejenis di masa depan

REFERENSI

- [1] D. A. Pangestu, O. Q. Aziz, and C. Crysdiyan, "Klasifikasi Penyakit pada Tanaman Berdasarkan Citra Daun Menggunakan Metode Convolutional Neural Network," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 10, no. 2, pp. 235–248, 2025, doi: 10.14421/jiska.2025.10.2.235-248.
- [2] Y. Gai and H. Wang, "Plant Disease: A Growing Threat to Global Food Security," Aug. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/agronomy14081615.
- [3] H. N. Ngugi, A. A. Akinyelu, and A. E. Ezugwu, "Machine Learning and Deep Learning for Crop Disease Diagnosis: Performance Analysis and Review," Dec. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/agronomy14123001.
- [4] T. Febriyanto and S. Syofian, "Implementasi Deep Learning Menggunakan Vision Transformer Untuk Klasifikasi Penyakit Daun Padi," *J. TIFDA (Technology Inf. Data Anal.)*, vol. 1, no. 2, pp. 34–39, Dec. 2024, doi: 10.70491/tifda.v1i2.47.
- [5] A. H. Foysal, F. Ahmed, and Z. Haque, "Multi-Class Plant Leaf Disease Detection: A CNN-based Approach with Mobile App Integration."
- [6] Q. Meng, J. Guo, H. Zhang, Y. Zhou, and X. Zhang, "A dual-branch model combining convolution and vision transformer for crop disease classification," *PLoS One*, vol. 20, no. 4 April, Apr. 2025, doi: 10.1371/journal.pone.0321753.
- [7] S. A. Alshammari, "TLDViT: A Vision Transformer Model for Tomato Leaf Disease Classification," 2024. [Online]. Available: www.ijacsa.thesai.org
- [8] U. Barman *et al.*, "ViT-SmartAgri: Vision Transformer and Smartphone-Based Plant Disease Detection for Smart Agriculture," *Agronomy*, vol. 14, no. 2, 2024, doi: 10.3390/agronomy14020327.
- [9] M. Niawati, F. Ardhy, T. H. Andika, P. Bintoro, U. A. Pringsewu, and J. A. Yani, "APLIKASI DIAGNOSIS PENYAKIT PADA TUMBUHAN TOMAT BERBASIS WEBSITE".
- [10] A. Fauzul, A. Kustiwa, M. Aminul Akbar, and A. Pinandito, "A Comparative Study of CNN-LSTM and LSTM Methods in Detecting Emotions in Indonesian Text Data on Twitter Social Media," vol. 1, no. 1, pp. 2548–964, 2024, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] M. S. H. Nabil Anan Orka, M. Nazim Uddin Fardeen Md. Tushique, "An Open Leaf Image Dataset of Bangladesh's Major Crops." [Online]. Available: <https://www.kaggle.com/datasets/raiaone/olid-i>
- [12] R. Raynold and Alva Hendi Muhammad, "Deep Learning Deteksi Dan Klasifikasi Penyakit Daun Tomat Menggunakan ResNet-50," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 6, no. 1, pp. 9–20, Apr. 2025, doi: 10.37859/coscitech.v6i1.8501.
- [13] R. Soekarta, N. Nurdjan, and A. Syah, "Klasifikasi Penyakit Tanaman Tomat Menggunakan Metode Convolutional Neural Network (CNN)," vol. 8, no. 2, 2023.
- [14] A. Lawi, N. S. Intizhami, R. Mukhtarom, and S. Amir, "KLASIFIKASI PENYAKIT CITRA DAUN TANAMAN TOMAT DENGAN ENSEMBLE CONVOLUTIONAL NEURAL NETWORK." [Online]. Available: <https://www.kaggle.com/datasets/emmarex/plantdisease>.
- [15] J. R. Aisyah and A. Prasetiadi, "Klasifikasi Penyakit Daun Kentang dengan Metode CNN dan RNN," *J. Tekno Insentif*, vol. 17, no. 1, pp. 1–10, Apr. 2023, doi: 10.36787/jti.v17i1.888.
- [16] A. Putra, "Klasifikasi Penyakit Daun Cabai Dengan Metode CNN Untuk Deteksi Awal.," *J. Profesi Ins. Univ. Lampung*, vol. 6, no. 1, pp. 2–7, 2025, doi: 10.23960/jpi.v6n1.147.
- [17] M. Sholihah and R. Dijaya, "Identification of Potato Plant Leaf Diseases Using a Digital Image Approach Using Algorithms K-Nearest Neighbor (KNN)," pp. 1–11, 2024, [Online]. Available: <http://dx.doi.org/10.21070/ups.5248>
- [18] A. Azis, A. Fadlil, and T. Sutikno, "Real-Time Rice Leaf Disease Diagnosis: A Mobile CNN Application with Firebase Integration," *J. Tek. Inform.*, vol. 6, no. 3, pp. 1469–1484, Jun. 2025, doi:

- 10.52436/1.jutif.2025.6.3.4452.
- [19] G. Y. Christiawan, R. A. Putra, A. Sulaiman, E. Poerbaningtyas, S. Widyayuningtias, and P. Listio, "Penerapan Metode Convolutional Neural Network (CNN) Dalam Mengklasifikasikan Penyakit Daun Tanaman Padi Informasi Artikel Abstrak," vol. 65146.
 - [20] A. W. Muzaki, B. A. Sudamto, and E. Daniati, "Sistem Identifikasi Corak Penyakit pada daun Padi dan monitoring Deteksi Dini untuk Lahan Sehat Menggunakan Metode (CNN)," 2025.
 - [21] F. B. Laksono, "Deteksi penyakit tanaman dengan convolution neural network: Kombinasi arsitektur VGG16 dan ResNet34 untuk klasifikasi daun," *J. Komput. DAN Teknol. Inf.*, vol. 2, no. 2, Jul. 2024, doi: 10.26714/jkti.v2i2.13932.
 - [22] D. W. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov *et al.*, "An Image Is Worth 16 X 16 Words :," *Int. Conf. Learn. Represent.*, 2021, [Online]. Available: <https://openreview.net/pdf?id=YicbFdNTTy>
 - [23] A. Fadlil, I. Riadi, and S. Aji, "DDoS Attacks Classification using Numeric Attribute-based Gaussian Naive Bayes," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 8, pp. 42–50, 2017, doi: 10.14569/ijacsa.2017.080806.
 - [24] H. Rumapea, "Evaluasi Kinerja CNN dan Vision Transformer pada Klasifikasi Citra Resolusi Tinggi Berbasis Deep Learning," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 9, no. 2, pp. 372–379, 2025, doi: 10.46880/jmika.vol9no2.pp372-379.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.