

Perancangan Aplikasi Berbasis Web Menggunakan Vision Transformer (ViT) untuk Klasifikasi Penyakit Daun Tanaman

Oleh:

Zaidan Dafa Robani Hatuwe

Pembimbing: Rohman Dijaya, S.Kom., M.Kom.

Penguji 1: Sukma Aji, S.T, M.Kom

Penguji 2: Cindy Taurusta, S.ST., MT

Program Studi Teknik Informatika
Universitas Muhammadiyah Sidoarjo
2026



Pendahuluan

Urgensi Sektor Pertanian & Masalah Manual

Penyakit daun tanaman merupakan ancaman utama produktivitas pertanian global yang memicu kerugian hingga miliaran dolar setiap tahunnya.

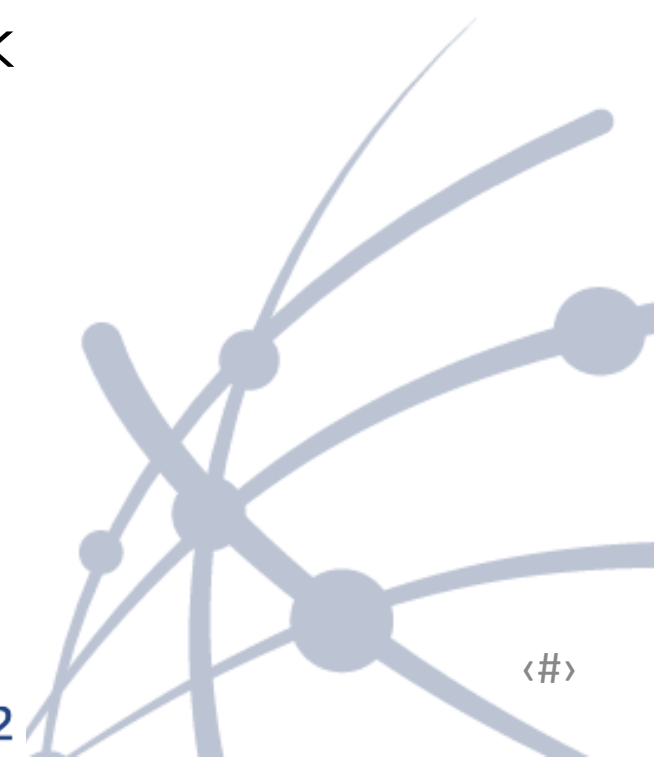
Identifikasi visual konvensional oleh petani memiliki kelemahan mendasar: bersifat subjektif, lambat, serta sangat bergantung pada ketersediaan tenaga ahli yang terbatas di daerah pedesaan.

Diagnosis yang keliru berdampak pada kekeliruan dosis pestisida, pemborosan biaya, serta kerusakan ekosistem lingkungan.

Keterbatasan CNN & Solusi ViT

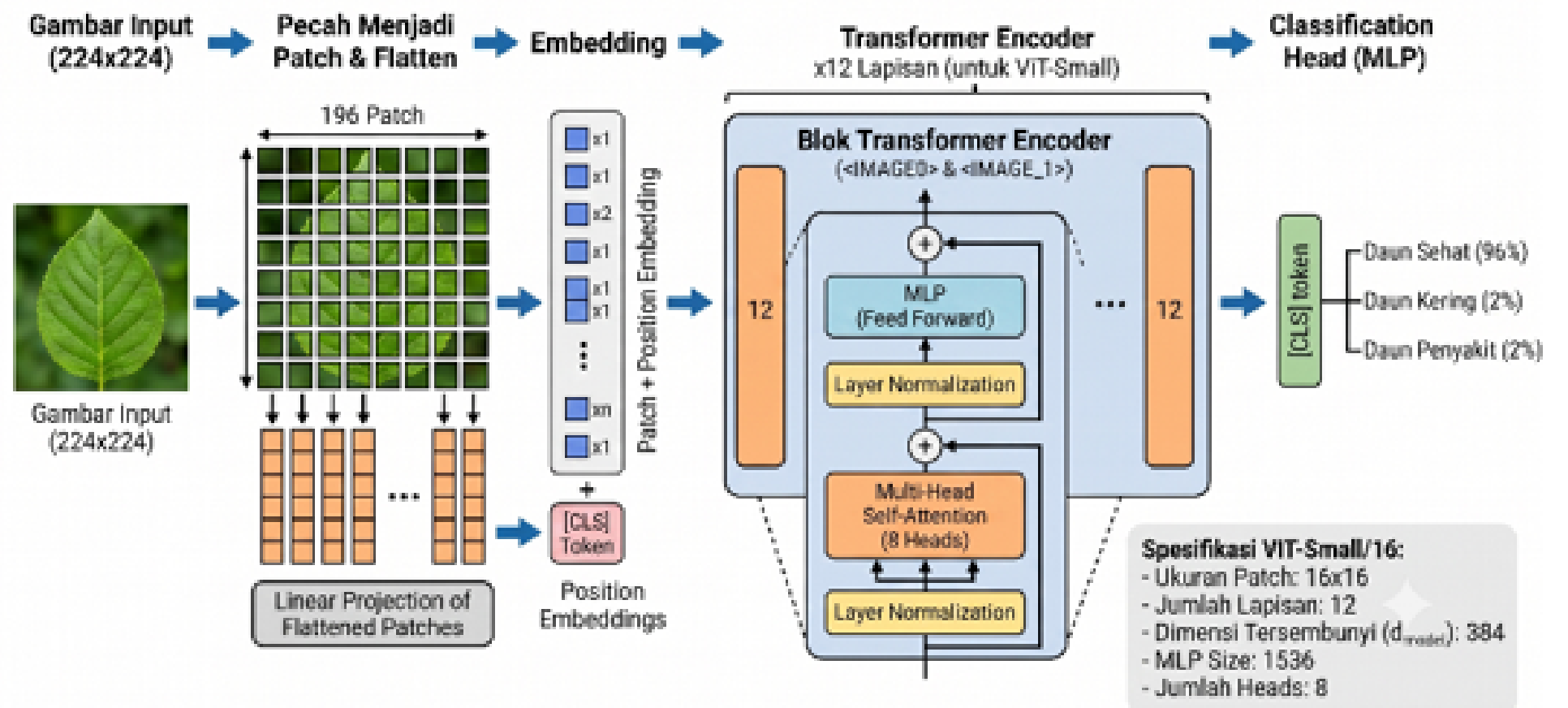
Riset terdahulu dominan menggunakan Convolutional Neural Network (CNN), namun dibatasi oleh local receptive field sehingga sulit menangkap konteks spasial global pada pola penyakit berbercak kompleks.

Solusi yang diusulkan adalah arsitektur Vision Transformer berbasis Multi-Head Self-Attention untuk memodelkan hubungan spasial jarak jauh secara simultan pada seluruh bagian citra.



Metode

arsitektur ViT-Small/16



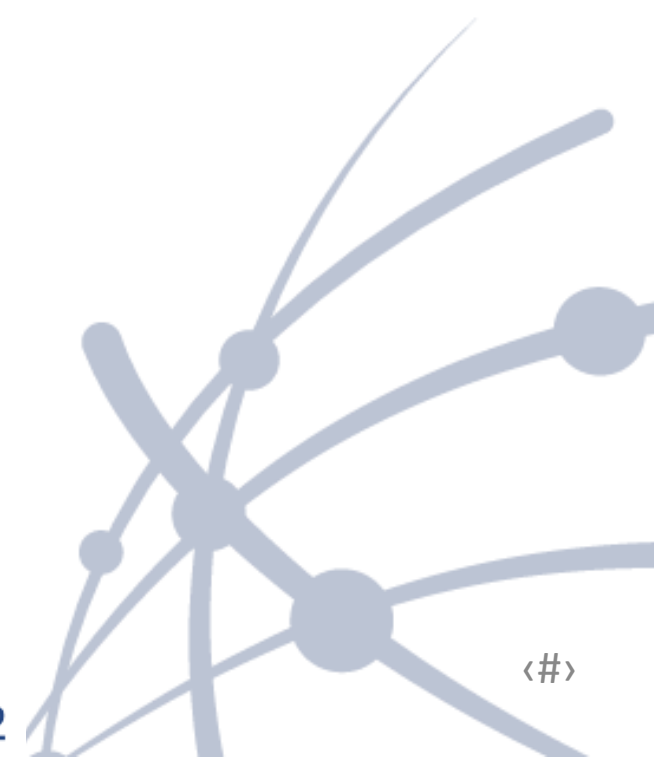
Penelitian ini menggunakan varian ViT-Small/16 (ViT-S/16) terdiri atas 22 juta parameter, 12 lapisan Transformer Encoder, 8 attention head, dan dimensi embedding sebesar 384. ViT-S/16 dipilih karena keseimbangannya antara akurasi dan efisiensi komputasi, dengan kapasitas yang memadai untuk mengklasifikasikan 57 kelas penyakit sekaligus memiliki resiko overfitting yang lebih rendah dibandingkan varian ViT yang lebih besar seperti ViT-Base/16 (86 juta parameter)

Hasil Pelatihan

Akurasi pelatihan 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 46,8% 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

tercatat Performa generalisasi terbaik dicapai pada epoch ke-17. Akurasi validasi pada titik ini berhasil menyentuh angka tertingginya yaitu 64,98%

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%



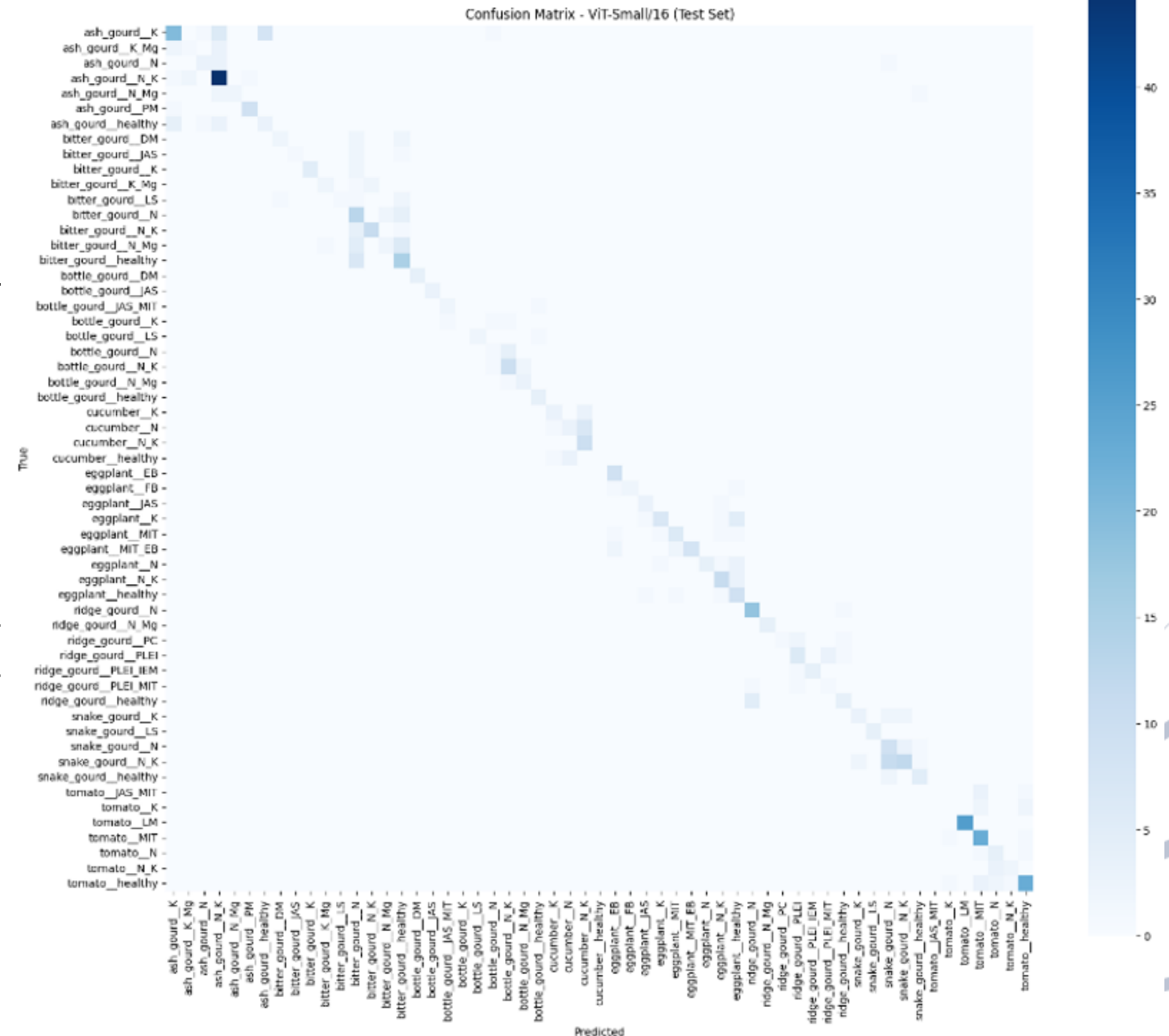
matrik evaluasi

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

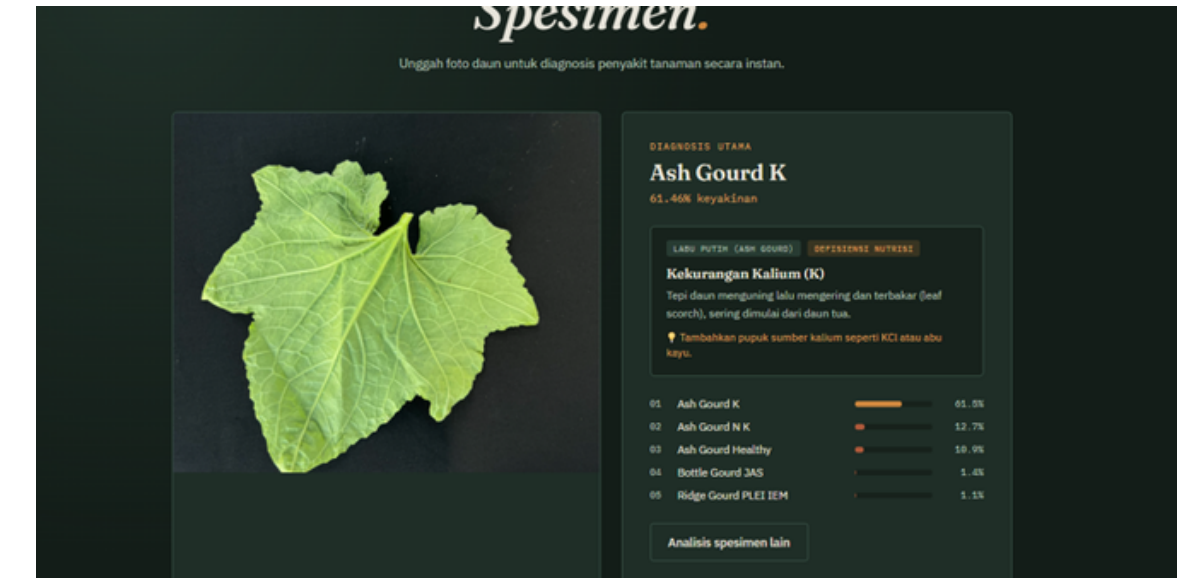
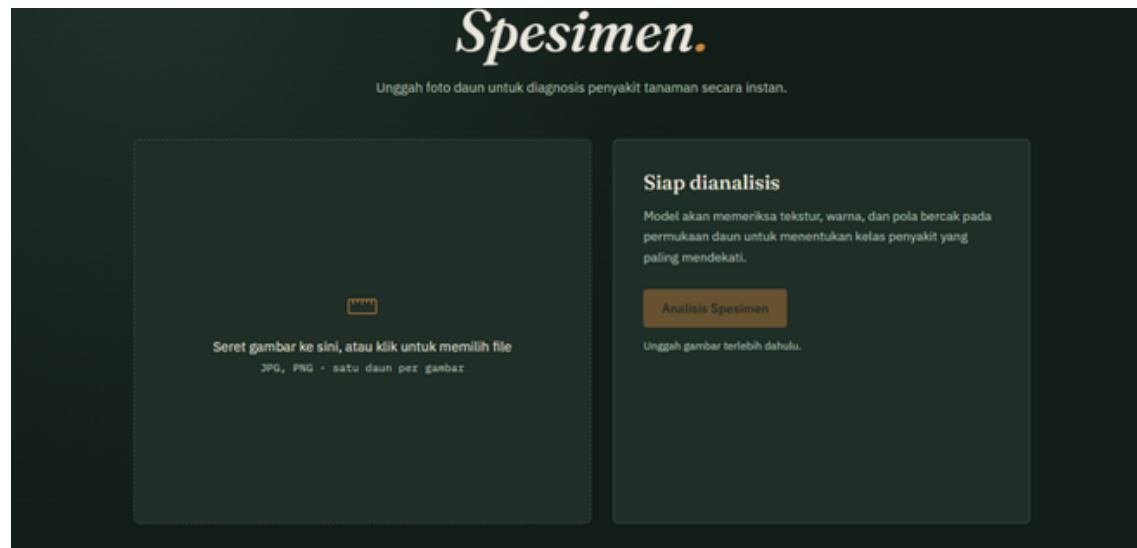
Metrik rata-rata makro secara konsisten lebih rendah dibandingkan metrik rata-rata tertimbang (Recall sebesar 57.87% dibandingkan 65.32%), yang menunjukkan bahwa model tampak berkinerja lebih buruk pada beberapa kelas minoritas, yang memiliki bobot lebih kecil pada rata-rata tertimbang tetapi dihitung 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

convusion matrix

besar terkonsentrasi pada spesies tanaman yang sama dan jarang terjadi antarspesies yang berbeda, yang menunjukkan bahwa model ini 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.



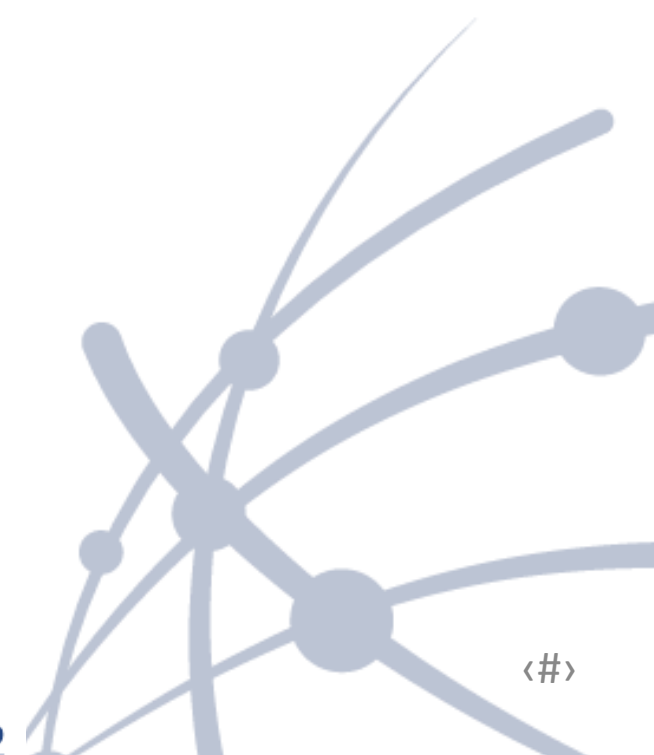
Implementasi Web



Antarmuka ini memungkinkan pengguna 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.

Pembahasan

- Akurasi 65,32% lebih rendah dari studi CNN sebelumnya, namun wajar karena OLID I mencakup 57 kelas dengan gambar 4749 yang dimana sangat kecil untuk Vision transformer
- Model dilatih from scratch tanpa pretrained weight, sehingga hasil ini merupakan baseline yang realistis untuk Vision transformer
- Kesenjangan overfitting (train 88,1% vs val ~60–65%) menunjukkan generalisasi, bukan kapasitas arsitektur, sebagai faktor pembatas utama
- Model tetap andal membedakan antar-spesies tanaman, namun masih sulit memisahkan subkelas defisiensi nutrisi yang mirip



Temuan Penting Penelitian

- Berhasil merancang dan melatih model ViT-Small/16 tanpa komponen konvolusi untuk klasifikasi 57 kelas penyakit daun
- Akurasi recall 65,32%, F1-score rata-rata makro 58.33%, F1-score rata-rata tertimbang 63.80%
- Model berhasil diintegrasikan ke aplikasi web sehingga siap digunakan
- Rekomendasi pengembangan: gunakan pretrained weight ImageNet, regularisasi lebih kuat, dan augmentasi data untuk kelas minoritas

