

KLASIFIKASI PENYAKIT DAUN PADA TANAMAN CABAI MENGUNAKAN MODEL HYBRID CNN-VISION TRANSFORMER

Irvan Fernando, Tri Aristi Saputri

Teknik Informatika, Universitas Dharma Wacana

Jl. Kenanga No.3, Mulyojati, Kec. Metro Barat, Kota Metro, Lampung

Irvanfernanda433@gmail.com

ABSTRAK

Penyakit tanaman cabai merupakan ancaman serius yang dapat menurunkan hasil panen hingga 40%. Permasalahan utama dalam budidaya cabai adalah rendahnya akurasi identifikasi manual akibat kemiripan gejala visual antar penyakit. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi penyakit daun cabai melalui penerapan model hybrid CNN–Vision Transformer (ViT). Metode yang digunakan adalah model hybrid CNN–Vision Transformer yang menggabungkan kemampuan CNN dalam mengekstraksi fitur lokal dengan mekanisme self-attention pada ViT untuk menangkap konteks global. Eksperimen dilakukan menggunakan dataset sebanyak 1.840 citra yang mencakup kategori sehat, kuning, keriting, dan bercak melalui tahap praproses dan augmentasi data. Hasil penelitian menunjukkan bahwa model Hybrid CNN–ViT mencapai akurasi sebesar 93%, lebih tinggi dibandingkan model CNN murni yang hanya mencapai 86%. Analisis Grad-CAM menunjukkan bahwa model hybrid mampu melakukan lokalisasi lesi secara lebih akurat. Dengan demikian, pendekatan hybrid terbukti efektif dalam meningkatkan keandalan diagnosis otomatis penyakit tanaman cabai.

Kata Kunci: Daun Cabai, Self-Attention, CNN, Vision Transformer, Hybrid Model, Grad-CAM.

1. PENDAHULUAN

Sektor pertanian merupakan pilar utama ketahanan pangan dan stabilitas ekonomi global, namun tantangan besar muncul dari serangan penyakit tanaman yang mengancam produktivitas secara signifikan. Penyakit tanaman bertanggung jawab atas kehilangan hasil panen tahunan hingga mencapai 40% secara global [1]. Di Indonesia, tanaman cabai (*Capsicum annuum* L.) merupakan komoditas strategis dengan nilai ekonomi tinggi, tetapi sangat rentan terhadap serangan patogen [2]. Identifikasi gejala penyakit pada permukaan daun memerlukan ketelitian tinggi karena pola serangan antar penyakit sering kali terlihat sangat mirip jika dilihat secara kasatmata, seperti pola klorosis pada infeksi virus yang menyerupai defisiensi hara [2]. Permasalahan utama dalam budidaya ini adalah rendahnya akurasi identifikasi manual oleh petani yang bersifat subjektif dan rentan terhadap kesalahan diagnosis. Meskipun teknik *Deep Learning* melalui *Convolutional Neural Networks* (CNN) telah banyak diimplementasikan dan mampu mencapai akurasi tinggi [3], arsitektur ini memiliki kelemahan fundamental dalam menangkap hubungan spasial global karena hanya berfokus pada area piksel yang berdekatan (*local receptive field*) [4]. Keterbatasan ini menjadi kendala serius ketika model dihadapkan pada variasi citra lapangan yang kompleks di mana pola infeksi sering kali tersebar tidak merata [5].

Penelitian ini bertujuan untuk menganalisis dan membandingkan hasil akurasi model *hybrid* dengan model CNN konvensional guna mengetahui efektivitas pendekatan gabungan dalam mengklasifikasikan empat kategori kondisi daun cabai: sehat, kuning, keriting, dan bercak. Melalui perbandingan ini, penelitian diharapkan dapat membuktikan bahwa

integrasi mekanisme *self-attention* mampu menutupi kelemahan CNN dalam mengenali pola penyakit yang kompleks. Rencana penyelesaian masalah dalam penelitian ini dilakukan dengan mengembangkan arsitektur *Hybrid CNN–Vision Transformer* (ViT) yang menyinergikan kemampuan CNN dalam mengekstraksi fitur tekstur lokal dengan kemampuan ViT dalam memahami konteks spasial global [6]. Proses dimulai dengan pengumpulan 1.840 citra digital yang kemudian diproses melalui tahap augmentasi data untuk meningkatkan variasi fitur visual [7]. Model tersebut kemudian dievaluasi menggunakan metrik kuantitatif dan visualisasi *Gradient-weighted Class Activation Mapping* (Grad-CAM) untuk memastikan presisi lokalisasi lesi pada permukaan daun [6]. Dengan pendekatan ini, diharapkan tercipta sistem diagnosis otomatis yang lebih andal dan presisi untuk membantu meningkatkan produktivitas petani cabai di Indonesia.

2. TINJAUAN PUSTAKA

Tanaman cabai (*Capsicum annuum* L.) merupakan komoditas strategis di Indonesia yang memiliki nilai ekonomi tinggi namun sangat rentan terhadap serangan berbagai patogen. Penyakit tanaman secara global menjadi ancaman serius bagi ketahanan pangan karena dapat menyebabkan kerugian hasil panen tahunan hingga mencapai angka 40% [1]. Di wilayah tropis, infeksi virus jenis *Begomovirus* menjadi tantangan besar bagi petani karena sering mengakibatkan gejala klorosis dan malformasi pada bagian daun [2]. Kondisi tersebut secara signifikan menurunkan produktivitas tanaman karena proses fotosintesis terganggu akibat rusaknya klorofil secara meluas sehingga kualitas buah yang dihasilkan menjadi menurun drastis [3]. Identifikasi

gejala penyakit pada permukaan daun memerlukan ketelitian tinggi karena pola serangan antar penyakit terlihat sangat mirip jika dilihat secara kasatmata. Penyakit bercak daun yang disebabkan oleh jamur biasanya memunculkan bintik-bintik kecil berwarna coklat yang menyebar di seluruh permukaan daun cabai. Di sisi lain, serangan virus sering menyebabkan daun menjadi keriting dan berwarna kuning pekat yang sangat berbeda dengan kondisi daun sehat pada umumnya [8].

2.1. Computer Vision dan Pengolahan Citra Digital

Pengolahan citra digital merupakan sebuah disiplin ilmu yang memanipulasi citra dua dimensi menggunakan perangkat komputer untuk meningkatkan kualitas visual maupun mengekstraksi informasi tertentu secara otomatis. Proses ini dimulai dengan tahap akuisisi data di mana gambar daun cabai ditangkap melalui sensor kamera dalam format susunan piksel yang teratur untuk diproses lebih lanjut. Citra yang telah didapatkan kemudian melalui serangkaian transformasi nilai warna agar distribusi datanya menjadi lebih seragam bagi model pembelajaran mendalam [3]. Langkah-langkah dalam pengolahan citra meliputi prapemrosesan, segmentasi, hingga ekstraksi fitur yang bertujuan untuk menonjolkan karakteristik unik dari setiap jenis penyakit daun. Teknologi *computer vision* memungkinkan sistem otomatis ini untuk melakukan deteksi serta klasifikasi jenis penyakit melalui analisis pola citra secara mendalam. Pendekatan ini memberikan solusi yang jauh lebih konsisten jika dibandingkan dengan identifikasi manual yang sering kali dipengaruhi oleh faktor kelelahan maupun subjektivitas manusia di lapangan [9].

2.2. Vision Transformer (ViT) dan Klasifikasi

Citra *Vision Transformer* (ViT) merupakan sebuah arsitektur inovatif yang mengadaptasi mekanisme *self-attention* dari bidang pemrosesan bahasa alami untuk digunakan dalam tugas visi komputer. Model ini bekerja dengan cara membagi citra masukan menjadi beberapa urutan *patch* linear yang kemudian diproses sebagai urutan token seperti layaknya kata dalam sebuah kalimat [10]. Melalui pendekatan tersebut, ViT mampu menangkap dependensi spasial global yang sering kali terlewatkan oleh model konvensional yang hanya berfokus pada area lokal saja. Landasan utama dari arsitektur ini adalah penggunaan mekanisme *multi-head self-attention* yang memungkinkan model memberikan bobot perhatian lebih besar pada area lesi [11]. Klasifikasi citra sendiri bertujuan untuk menetapkan label kategori tertentu pada sebuah gambar masukan berdasarkan fitur visual yang terkandung di dalamnya. Keberhasilan klasifikasi sangat bergantung pada kemampuan model dalam mengekstraksi representasi fitur yang unik dari setiap objek yang diamati [12].

2.3. Arsitektur Hybrid CNN-Vision Transformer

Arsitektur *hybrid* dikembangkan secara khusus untuk mengatasi dilema antara ekstraksi fitur lokal dan pemahaman konteks global pada pengolahan citra digital. Integrasi jaringan saraf konvensional sebagai penyedia fitur tekstur tingkat rendah dan *transformer* sebagai pengolah hubungan spasial memberikan hasil yang lebih presisi pada citra yang kompleks [13]. Sinergi ini secara konsisten terbukti mampu mengungguli performa model tunggal dalam menangkap berbagai gejala penyakit daun yang memiliki variasi visual sangat tinggi [14]. Di dalam struktur *hybrid*, lapisan konvensional bertugas untuk mereduksi dimensi gambar sekaligus menjaga detail informasi spasial yang halus sebelum diproses oleh lapisan *transformer*. Pendekatan ini memungkinkan model untuk mempertahankan sensitivitas terhadap pola lesi kecil sambil tetap memahami struktur morfologi daun secara menyeluruh [1]. Penggunaan model gabungan ini menjadi solusi logis untuk menutupi kelemahan intrinsik dari masing-masing arsitektur konvensional [6].

2.4. Praproses, Augmentasi, dan Evaluasi Model

Praproses data merupakan langkah krusial dalam mempersiapkan citra daun cabai agar memiliki format yang seragam sebelum dimasukkan ke dalam jaringan. Proses ini dilakukan dengan mengubah ukuran seluruh citra menjadi 224×224 piksel untuk menjaga keseimbangan antara detail fitur dan efisiensi komputasi [15]. Teknik augmentasi data diterapkan secara *on-the-fly* menggunakan parameter seperti *rotation*, *zoom*, serta *horizontal flip* untuk memperkaya variasi data pelatihan tanpa memerlukan ruang penyimpanan tambahan [7]. Evaluasi kinerja model dilakukan menggunakan berbagai metrik kuantitatif seperti *accuracy*, *precision*, *recall*, dan *F1-score* [16]. Selain itu, teknik *Gradient-weighted Class Activation Mapping* (Grad-CAM) digunakan untuk memberikan interpretabilitas pada model dengan memvisualisasikan area fokus keputusan [6]. Penggunaan standar evaluasi yang komprehensif memastikan validitas dari hasil penelitian ini dapat dipertanggungjawabkan secara ilmiah [17].

2.5. Penelitian Terdahulu yang Terkait

Penelitian terdahulu mengenai klasifikasi penyakit tanaman menunjukkan bahwa penggunaan arsitektur berbasis *deep learning* telah menjadi standar utama dalam otomatisasi diagnosis di bidang pertanian. Penggunaan model CNN skala besar telah dikaji secara mendalam untuk membuktikan potensi ekstraksi fitur otomatis, meskipun masih ditemukan penurunan performa pada citra dengan latar belakang alami [3]. Studi komparatif antara berbagai arsitektur jaringan saraf juga memperlihatkan bahwa model dengan efisiensi parameter tinggi seperti DenseNet-121 tetap memiliki keterbatasan dalam memodelkan hubungan spasial jarak jauh antar piksel [5]. Selain itu, pemahaman mengenai karakteristik biologis penyakit

spesifik seperti Begomovirus di Indonesia menegaskan pentingnya detail tekstur dan bentuk daun dalam proses identifikasi secara teliti [2]. Rangkaian penelitian awal ini memberikan dasar yang kuat mengenai kebutuhan akan model yang lebih adaptif terhadap variasi lingkungan lapangan.

Transformasi metodologi penelitian mulai terlihat dengan diperkenalkannya mekanisme *transformer* yang mampu mengintegrasikan fitur lokal dan konteks global secara simultan. Penerapan arsitektur *hybrid* pada citra medis telah membuktikan bahwa sinergi antara CNN dan *transformer* menghasilkan tingkat presisi yang jauh lebih tinggi dibandingkan model tunggal [13]. Di sisi lain, penggunaan *Vision Transformer* murni pada tanaman cabai menunjukkan keunggulan dalam memahami konteks global, namun masih kurang sensitif terhadap detail tekstur mikroskopis pada gejala awal penyakit [8]. Hasil evaluasi pada komoditas lain seperti kentang memperkuat temuan bahwa model gabungan secara konsisten mengungguli arsitektur konvensional saat menghadapi variasi visual yang tinggi di lahan pertanian [14]. Hal ini menunjukkan adanya tren positif dalam pengembangan model cerdas yang lebih tangguh terhadap tantangan visual di lapangan.

Perkembangan terkini dalam deteksi penyakit daun semakin difokuskan pada peningkatan ketahanan model melalui teknik *feature fusion* untuk meminimalisir gangguan latar belakang yang rumit. Implementasi kerangka kerja *hybrid* terbaru terbukti mampu meningkatkan stabilitas deteksi pada berbagai jenis tanaman dengan kondisi pencahayaan yang tidak menentu [16]. Selain itu, aspek interpretabilitas model melalui penggunaan Grad-CAM menjadi poin krusial untuk memastikan bahwa mesin benar-benar mengenali lesi penyakit secara tepat dan akurat [6]. Integrasi berbagai teknologi ini memberikan jaminan keamanan bagi para petani dalam melakukan pemantauan kesehatan tanaman secara mandiri di lokasi perkebunan mereka. Secara keseluruhan, literatur yang ada menegaskan bahwa pendekatan *hybrid* CNN-ViT merupakan solusi paling logis untuk mencapai akurasi tinggi sekaligus transparansi keputusan model dalam klasifikasi penyakit daun cabai.

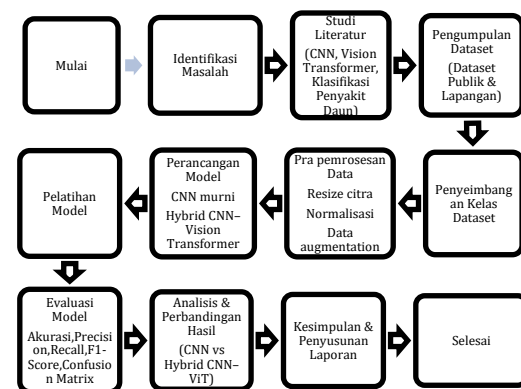
3. METODE PENELITIAN

Penelitian ini merupakan penelitian eksperimen pada bidang *image processing* dan klasifikasi citra penyakit tanaman. Penelitian dilakukan untuk menguji serta membandingkan kemampuan model *Convolutional Neural Network* (CNN) murni dengan model *Hybrid* CNN-ViT dalam mengklasifikasikan empat kondisi kesehatan pada daun cabai. Pendekatan eksperimen digunakan karena penelitian ini melibatkan perancangan alur pengolahan data, pelatihan arsitektur *deep learning*, serta evaluasi performa model berdasarkan metrik klasifikasi secara objektif [1]. Melalui metode ini, kontribusi mekanisme *self-attention* dalam menangkap hubungan spasial

global pada citra daun dapat diidentifikasi secara sistematis dibandingkan dengan penggunaan operasi konvolusi murni [6]. Pendekatan kuantitatif diterapkan dalam penelitian ini karena seluruh evaluasi kinerja model didasarkan pada metrik numerik yang terukur secara statistik. Metrik utama yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur tingkat ketepatan model dalam mengenali pola visual yang kompleks pada setiap kelas penyakit (Ramadhana dkk., 2025). Fokus utama penelitian ditekankan pada analisis efektivitas penggabungan fitur lokal dan konteks global guna meningkatkan akurasi deteksi di lingkungan lapangan yang dinamis [16].

3.1. Alur Penelitian

Tahapan penelitian ini disusun secara sistematis agar dapat dilaksanakan secara terukur dalam kurun waktu satu semester. Proses diawali dengan identifikasi permasalahan terkait kesulitan deteksi penyakit daun cabai serta pelaksanaan studi literatur mengenai arsitektur CNN dan *Vision Transformer*. Peneliti membangun fondasi teoritis yang kuat untuk memahami mekanisme kerja model dalam melakukan klasifikasi citra secara otomatis pada objek tanaman. Rangkaian aktivitas awal ini memastikan bahwa setiap langkah eksperimen memiliki landasan ilmiah yang jelas sebelum memasuki tahap pengolahan data. Secara ringkas, alur penelitian ini dapat dilihat pada Gambar 3.1 di bawah ini:



Gambar 1. Diagram alur penelitian

Tahap berikutnya difokuskan pada pengumpulan serta pengolahan dataset yang mencakup proses penyeimbangan kelas dan praproses citra secara intensif. Langkah ini meliputi penyesuaian dimensi gambar, normalisasi nilai piksel, serta penerapan teknik augmentasi data untuk meningkatkan variasi fitur visual. Setelah data siap, dilakukan perancangan arsitektur model CNN murni sebagai pembanding dan model *Hybrid* CNN-ViT sebagai arsitektur utama. Seluruh model kemudian dilatih dan divalidasi guna mendapatkan bobot parameter paling optimal dalam mengenali pola penyakit pada daun cabai. Fase akhir melibatkan evaluasi kinerja menggunakan data uji dan visualisasi Grad-CAM untuk memastikan fokus model

telah sesuai dengan karakteristik lesi pada permukaan daun.

3.2. Data dan Pembentukan Dataset

Data yang digunakan terdiri dari citra digital daun tanaman cabai yang mencakup kondisi sehat maupun terinfeksi patogen. Total citra yang dikumpulkan berjumlah 1.840 gambar yang dikelompokkan ke dalam empat kategori utama yaitu bercak, keriting, kuning, dan daun sehat. Setiap gambar diperiksa secara manual untuk memastikan kualitas visualnya layak dan tidak mengandung gangguan yang tidak relevan dengan objek penelitian. Pembentukan dataset dilakukan dengan menyusun seluruh citra ke dalam folder-folder terpisah secara sistematis sesuai dengan label kategorinya. Distribusi jumlah data pada setiap kelas diatur sedemikian rupa agar tetap seimbang guna menghindari masalah *bias* pada prediksi model di tahap pelatihan.

3.3. Praproses dan Augmentasi Data

Tahap praproses data dilakukan dengan menyeragamkan seluruh format input citra agar dapat diproses secara efisien oleh arsitektur *Hybrid CNN-ViT*. Peneliti mengubah ukuran dimensi setiap citra daun cabai menjadi 224 x 224 piksel guna menjaga keseimbangan antara detail tekstur lesi dan efisiensi waktu komputasi [15] (A. G. Howard dkk., 2017). Selanjutnya, nilai piksel pada setiap kanal warna dinormalisasi ke dalam rentang 0-1 untuk membantu mempercepat proses konvergensi bobot jaringan. Teknik augmentasi data diterapkan secara *on-the-fly* menggunakan *ImageDataGenerator* dengan parameter *rotation*, *width/height shift*, *flip*, dan *zoom* untuk memperkaya variasi informasi [7]. Peneliti juga menyetel *brightness_range* antara 0,8 hingga 1,2 untuk mengantisipasi adanya perbedaan intensitas cahaya matahari pada citra lapangan.

3.4. Perancangan dan Pelatihan Model

Perancangan model klasifikasi menerapkan arsitektur *hybrid* yang mengintegrasikan lapisan konvensional dengan blok *transformer* secara sistematis. Peneliti menggunakan lapisan CNN pada tahap awal untuk mengekstraksi fitur tekstur lokal yang kemudian dikonversi menjadi urutan token untuk diproses oleh mekanisme *self-attention* [6]. Pelatihan dilakukan dengan menetapkan *learning rate* sebesar 0,0001 dan ukuran *batch* sebanyak 32 selama maksimal 50 *epoch*. Fungsi aktivasi *softmax* ditempatkan pada lapisan keluaran untuk menghasilkan nilai probabilitas klasifikasi pada empat kategori kondisi kesehatan daun. Teknik *early stopping* diterapkan untuk menghentikan pelatihan secara otomatis jika tidak ditemukan lagi peningkatan performa pada data validasi guna mencegah *overfitting*.

3.5. Evaluasi dan Visualisasi Grad-CAM

Evaluasi model dilakukan menggunakan metrik kuantitatif yang mencakup *accuracy*, *precision*, *recall*, serta *F1-score* sebagai standar penilaian kinerja utama. *Confusion matrix* juga diterapkan untuk menganalisis secara mendalam mengenai distribusi data yang berhasil diklasifikasikan dengan benar maupun yang mengalami kekeliruan antar kategori kelas. Selain itu, teknik Grad-CAM digunakan untuk memberikan transparansi terhadap keputusan klasifikasi dengan menghasilkan peta panas (*heatmap*) pada permukaan daun. Visualisasi ini memastikan bahwa model benar-benar fokus pada bagian lesi penyakit dan bukan pada objek latar belakang lainnya yang tidak relevan. Analisis kesalahan klasifikasi dilakukan terhadap citra yang gagal diidentifikasi untuk memahami keterbatasan model dalam membedakan fitur visual yang tumpang tindih.

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Arsitektur Model

Pada tahap ini, dilakukan pembangunan dua arsitektur model sesuai dengan perancangan yang telah ditetapkan. Model pertama adalah CNN murni yang digunakan sebagai *baseline*, dan model kedua adalah *Hybrid CNN-Vision Transformer*. Implementasi model *Hybrid* melibatkan penggunaan lapisan konvolusi untuk ekstraksi fitur lokal yang kemudian diteruskan ke blok *Transformer* untuk menangkap hubungan spasial global melalui mekanisme *self-attention*. Seluruh model diimplementasikan menggunakan kerangka kerja TensorFlow/Keras dengan input dimensi citra 224 x 224 piksel guna menjaga konsistensi data.

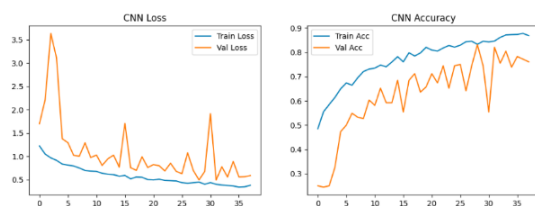
Pembangunan model CNN murni dilakukan dengan menyusun lima blok konvolusi secara sekuensial untuk mengekstraksi fitur visual pada daun cabai. Setiap blok dilengkapi dengan lapisan *Batch Normalization* untuk mempercepat proses konvergensi dan lapisan *MaxPooling* untuk mereduksi dimensi spasial citra. Pada bagian akhir jaringan, diterapkan lapisan *Global Average Pooling* (GAP) yang berfungsi untuk mengubah peta fitur menjadi vektor tunggal sebelum masuk ke lapisan klasifikasi. Model ini memiliki total parameter sebanyak 1.574.596 yang seluruhnya dioptimasi menggunakan fungsi kerugian *Categorical Crossentropy*.

Arsitektur model *Hybrid* ini menggabungkan kekuatan ekstraksi fitur lokal dari CNN dengan kemampuan pemrosesan konteks global dari *Vision Transformer*. Lapisan awal terdiri dari empat blok konvolusi yang berfungsi sebagai *backbone* untuk menghasilkan representasi fitur yang kaya sebelum dipetakan menjadi *patches*. Peta fitur tersebut kemudian diproses oleh enam lapisan *Transformer Encoder* yang dilengkapi dengan mekanisme *Multi-Head Attention* sebanyak empat kepala. Penggunaan *Label Smoothing* sebesar 0,1 diterapkan pada tahap kompilasi model untuk meminimalkan risiko

overfitting dan meningkatkan generalisasi pada data uji [6].

4.2. Hasil Pengujian Model CNN Murni

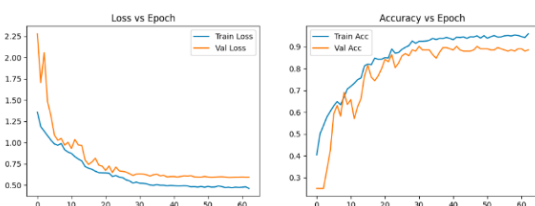
Pengujian pada model CNN murni dilakukan sebagai tolok ukur awal untuk melihat kemampuan operasi konvolusi murni dalam mengklasifikasikan penyakit daun cabai. Berdasarkan hasil pelatihan, model ini berhenti secara otomatis melalui mekanisme *early stopping* pada *epoch* ke-38 karena tidak ada penurunan nilai kerugian yang signifikan pada data validasi. Meskipun mampu mengenali gejala penyakit, model ini masih memiliki kelemahan dalam membedakan detail halus pada kelas yang memiliki kemiripan tekstur tinggi. Proses evaluasi akhir menunjukkan bahwa model *baseline* ini mencapai tingkat akurasi sebesar 86% pada data uji.



Gambar 2. Grafik accuracy model cnn murni

4.3. Hasil Pelatihan Model Hybrid CNN-Vision Transformer

Model *Hybrid* diuji menggunakan porsi 10% dari total dataset sebagai data uji yang belum pernah dilihat sebelumnya. Hasil pengujian menunjukkan bahwa penggabungan fitur lokal dari CNN dan konteks global dari *Transformer* memberikan hasil yang lebih stabil dibandingkan model tunggal. Model ini berhasil mencapai titik optimal pada *epoch* ke-63 sebelum dihentikan oleh mekanisme *early stopping* dengan akurasi akhir sebesar 93%. Penggunaan teknik *label smoothing* terbukti membantu model dalam mengurangi tingkat ambiguitas saat melakukan klasifikasi pada kelas yang memiliki kemiripan visual tinggi [16].



Gambar 3. Grafik accuracy model hybrid CNN-ViT

4.4. Analisis Kesalahan dan Confusion Matrix

Tahap analisis kesalahan dilakukan untuk memahami keterbatasan model dalam mengidentifikasi citra yang gagal diklasifikasikan secara tepat. Analisis pada model *Hybrid* menunjukkan adanya perbaikan performa yang sangat drastis jika dibandingkan dengan model *baseline* sebelumnya. Dari total 184 data uji, model *Hybrid* mampu memprediksi 171 citra dengan benar sehingga

hanya menyisakan 13 kesalahan klasifikasi saja. Hal ini membuktikan bahwa integrasi mekanisme *self-attention* sangat efektif dalam membedakan fitur visual yang kompleks pada daun cabai dibandingkan operasi konvolusi standar [8].

4.5. Perbandingan Performa dan Analisis Hasil

Berdasarkan seluruh parameter yang diamati, model *Hybrid* menunjukkan keunggulan yang konsisten baik dari sisi metrik kuantitatif maupun kualitas visualisasi fitur. Model *Hybrid* CNN-ViT berhasil mengungguli model *baseline* dengan selisih akurasi sebesar 7% yang menunjukkan peningkatan kemampuan deteksi secara nyata. Peningkatan pada nilai *F1-score* dari 0,86 menjadi 0,93 memberikan indikasi bahwa model *Hybrid* jauh lebih stabil dalam menangani seluruh kategori penyakit. Fokus yang tajam pada visualisasi Grad-CAM memastikan bahwa sistem dapat memberikan diagnosis yang lebih objektif saat diimplementasikan pada kondisi lingkungan yang bervariasi [6].

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan perbandingan yang telah dilakukan, dapat disimpulkan bahwa penerapan arsitektur Hybrid CNN-Vision Transformer terbukti lebih efektif dalam mengklasifikasikan penyakit daun cabai dibandingkan dengan model CNN konvensional. Keunggulan pendekatan gabungan ini ditunjukkan secara nyata melalui perolehan akurasi model Hybrid sebesar 93%, yang memberikan peningkatan signifikan sebesar 7% dibandingkan model *baseline* yang hanya mencapai 86%. Secara kualitatif, efektivitas arsitektur ini juga terkonfirmasi melalui visualisasi Grad-CAM yang menunjukkan kemampuan model dalam menangkap fitur lesi secara lebih presisi dan tajam daripada model konvensional yang cenderung terdistraksi oleh latar belakang citra. Dengan demikian, integrasi mekanisme *self-attention* dan penggunaan *label smoothing* terbukti mampu menangani kompleksitas visual pada data lapangan dengan stabilitas performa yang lebih baik, sehingga tujuan penelitian untuk membuktikan keunggulan model hybrid ini telah tercapai sepenuhnya.

Berdasarkan hasil penelitian, peneliti menyarankan pengembangan selanjutnya untuk fokus pada perluasan dataset dengan variasi kondisi pencahayaan ekstrem guna meningkatkan ketangguhan model di lapangan penggunaan strategi transfer learning dari model yang telah dilatih pada dataset botani yang lebih besar untuk meningkatkan kemampuan ekstraksi fitur serta implementasi model ke dalam perangkat bergerak atau aplikasi berbasis web agar dapat digunakan secara praktis oleh petani.

DAFTAR PUSTAKA

- [1] Z. Y. Younes dan E. M. T. A. Alsaadi, "A Review on Plant Leaf Disease Classification Using Deep Learning," *Iraqi J. Sci.*, hlm. 6738–

- 6752, Nov 2024, doi: 10.24996/ijis.2024.65.11.43.
- [2] M. Khalil, E. Kesumawati, dan S. Zakaria, "Begomoviral disease rates and the implications to the growth and yield of chili plants (*Capsicum annum* L.) at different elevations in Indonesia," *J. Appl. Hortic.*, vol. 22, no. 1, hlm. 71–75, Apr 2020, doi: 10.37855/jah.2020.v22i01.14.
- [3] S. P. Mohanty, D. P. Hughes, dan M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Front. Plant Sci.*, vol. 7, hlm. 1419, Sep 2016, doi: 10.3389/fpls.2016.01419.
- [4] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, dan M. Shah, "Transformers in Vision: A Survey," *ACM Comput. Surv.*, vol. 54, no. 10s, hlm. 1–41, Jan 2022, doi: 10.1145/3505244.
- [5] E. C. Too, L. Yujian, S. Njuki, dan L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Comput. Electron. Agric.*, vol. 161, hlm. 272–279, Jun 2019, doi: 10.1016/j.compag.2018.03.032.
- [6] S. Jia, G. Wang, H. Li, Y. Liu, L. Shi, dan S. Yang, "ConvTransNet-S: A CNN-Transformer Hybrid Disease Recognition Model for Complex Field Environments," *Plants*, vol. 14, no. 15, hlm. 2252, Jul 2025, doi: 10.3390/plants14152252.
- [7] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, dan A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Information*, vol. 11, no. 2, hlm. 125, Feb 2020, doi: 10.3390/info11020125.
- [8] F. Ramadhana, G. A. Mahardika, N. Day, dan E. Y. Puspaningrum, "Klasifikasi Penyakit Daun Cabai Menggunakan Vision Transformer," vol. 5, 2025.
- [9] J. Lu, L. Tan, dan H. Jiang, "Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification," *Agriculture*, vol. 11, no. 8, hlm. 707, Jul 2021, doi: 10.3390/agriculture11080707.
- [10] A. Dosovitskiy *dkk.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," 3 Juni 2021, *arXiv*: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.
- [11] A. Vaswani *dkk.*, "Attention is All you Need", doi: <https://doi.org/10.48550/arXiv.1706.03762>.
- [12] S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz, dan E. Jasińska, "Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach," *Electronics*, vol. 10, no. 12, hlm. 1388, Jun 2021, doi: 10.3390/electronics10121388.
- [13] J. Chen *dkk.*, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," 8 Februari 2021, *arXiv*: arXiv:2102.04306. doi: 10.48550/arXiv.2102.04306.
- [14] A. Saeed, "A Hybrid Deep Learning Model for Potato Leaf Disease Detection Using CNN and Vision Transformer".
- [15] A. G. Howard *dkk.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications".
- [16] K. Dhakyanabdel, "A Hybrid CNN–Vision Transformer Framework for Optimized Leaf Disease Detection Using Feature Fusion and Transfer Learning".
- [17] Y. Wang, P. Zhang, dan S. Tian, "Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion," *Front. Plant Sci.*, vol. 15, hlm. 1382802, Apr 2024, doi: 10.3389/fpls.2024.1382802.