

Analisis Perbandingan CNN dan *Random Forest* dalam Klasifikasi Penyakit Tanaman Jambu Mete

Comparative Analysis of CNN and Random Forest for Cashew Plant Disease Classification

^{1,2}Mawar Ahdayani Samual*, ²Sigit Wijanarko

^{1,2}Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Jakarta, Indonesia

*e-mail: mawarsamual@gmail.com

(**received:** 2 February 2026, **revised:** 16 February 2026, **accepted:** 17 February 2026)

Abstrak

Tanaman jambu mete merupakan komoditas strategis Indonesia yang rentan terhadap penyakit, sehingga diperlukan teknik identifikasi yang cepat dan akurat untuk meminimalkan kerugian ekonomi. Studi ini mengevaluasi perbandingan kinerja CNN dan Random Forest dalam mengklasifikasikan gambar penyakit jambu mete. Dengan menggunakan data kuantitatif eksperimental, peneliti menggunakan kumpulan 6.549 gambar yang dibagi menjadi lima kelas: *anthracnose*, *gummosis*, *healthy*, *leaf miner*, dan *red rust*. Model divalidasi dengan teknik split sampling yaitu 80:10:10, dan dinilai berdasarkan F1-score dan metrik akurasi, presisi, dan recall. Sementara *Random Forest* menggunakan ekstraksi fitur manual untuk warna, tekstur (GLCM), dan bentuk (*Hu Moments*), jalur CNN menggunakan arsitektur *Sequential Custom* dengan fitur otomatis. Hasil pengujian menunjukkan bahwa CNN memiliki akurasi 85,80%, lebih tinggi 9% dibanding *Random Forest* yang akurasinya 76,95%. Kebaruan penelitian ini adalah penggabungan fitur tekstur tingkat tinggi pada *Random Forest*. Tujuan dari penggabungan ini adalah untuk menguji batas performa machine learning konvensional dibandingkan dengan fitur otomatis CNN. Penelitian ini menemukan bahwa CNN lebih baik untuk dataset ini. Namun, masih perlu mengoptimalkan latar belakang alami untuk aplikasinya.

Kata kunci: jambu mete, cnn, klasifikasi penyakit, random forest

Abstract

Cashew plants are a strategic commodity in Indonesia and are highly susceptible to various diseases, making fast and accurate identification techniques essential to minimize economic losses. This study evaluates the comparative performance of Convolutional Neural Networks (CNN) and Random Forest in classifying cashew plant disease images. Using an experimental quantitative approach, the study utilizes a dataset of 6,549 images divided into five classes: *anthracnose*, *gummosis*, *healthy*, *leaf miner*, and *red rust*. The models were validated using a split sampling technique with a ratio of 80:10:10 for training, validation, and testing, and were evaluated based on F1-score as well as accuracy, precision, and recall metrics. The Random Forest model employs manual feature extraction, including color, texture (Gray-Level Co-occurrence Matrix/GLCM), and shape (*Hu Moments*), whereas the CNN model uses a custom *Sequential* architecture with automatic feature extraction. The experimental results show that CNN achieves an accuracy of 85.80%, outperforming Random Forest by approximately 9%, which attains an accuracy of 76.95%. The novelty of this study lies in the integration of high-level texture features into the Random Forest model to evaluate the performance limits of conventional machine learning compared to CNN-based automatic feature extraction. The findings indicate that CNN performs better for this dataset. However, further optimization—particularly in handling natural background variations—is still required for practical deployment.

Keywords: cashew, cnn, disease classification, random forest

1 Pendahuluan

Tanaman jambu mete atau *Anacardium occidentale* merupakan salah satu komoditas perkebunan strategis di Indonesia. Menurut data yang dikeluarkan oleh Badan Pusat Statistik, luas lahan perkebunan jambu mete di Indonesia mencapai lebih dari 400.000 hektar pada tahun 2024. Meskipun memiliki nilai ekonomi yang tinggi, tanaman ini rentan terhadap penyakit daun seperti Antraknosa yang bisa memberikan dampak signifikan terhadap ekonomi karena mempengaruhi penampilan dan rasa dari buah. Deteksi penyakit secara manual kerap kali tidak akurat, terutama bagi petani yang kurang memiliki pengetahuan teknis [1]. Oleh karena itu, menciptakan teknik yang efisien untuk mengidentifikasi penyakit pada tanaman sangat penting dalam lingkungan pertanian saat ini. Teknologi *deep learning* telah muncul sebagai pendekatan yang sangat menjanjikan dalam bidang pertanian, terutama untuk deteksi dini penyakit tanaman melalui analisis citra digital yang memfasilitasi proses identifikasi yang cepat dan akurat [2].

Salah satu implementasi utama dari *deep learning* dalam konteks ini adalah *Convolutional Neural Network* (CNN), yang telah terbukti efektif untuk menangani kompleksitas gambar daun tanaman tropis. CNN mampu melakukan ekstraksi fitur otomatis, seperti pola lesi atau tekstur vena, melalui lapisan konvolusi yang dirancang khusus untuk mengenali variasi visual pada citra, sehingga meningkatkan akurasi klasifikasi penyakit secara keseluruhan [3]. Studi terkini menunjukkan bahwa CNN mencapai akurasi hingga 97,61% dalam deteksi penyakit tanaman mete pada dataset lokal, berkat adaptasinya terhadap variasi pencahayaan dan latar belakang alami [4]. Sebaliknya, algoritma *Random Forest* terbukti menjadi model yang paling andal dan kuat, dengan mengolah fitur-fitur yang diekstraksi secara manual seperti warna, bentuk, dan tekstur, model ini mampu mencapai akurasi sangat tinggi hingga 99,3%, menunjukkan keunggulannya yang signifikan untuk aplikasi sortasi produk pertanian [5]. Sebagai algoritma yang dikenal tahan terhadap *overfitting*, *Random Forest* sering digunakan sebagai baseline yang efisien untuk mengevaluasi performa model *deep learning* [6].

Meskipun teknologi *deep learning* seperti CNN telah menunjukkan kinerja yang mengesankan dalam penelitian sebelumnya dan *Random Forest* diakui sebagai *baseline* yang kuat dan efisien, terdapat celah penelitian di mana belum banyak studi perbandingan yang mengevaluasi efektivitas fitur otomatis CNN dibandingkan dengan fitur manual yang dirancang dengan sistematis (GLCM dan Hu Moments) pada konteks spesifik jambu mete. Berdasarkan keterbatasan penelitian ini, penelitian ini melakukan analisis yang mendalam terhadap kedua metode dengan menggunakan dataset citra jambu mete yang terdiri dari lima kelas, yaitu *anthracnose*, *gumosis*, *healthy*, *leaf miner*, dan *red rust*.

Peneliti melakukan penelitian ini dengan tujuan untuk membandingkan kinerja antara algoritma CNN dan *Random Forest* dalam menentukan metode klasifikasi yang paling efektif untuk menganalisis gambar penyakit, serta menguji seberapa jauh representasi fitur manual *Random Forest* dapat bersaing dengan fitur otomatis yang dimiliki oleh CNN dalam mengenali pola visual yang mirip. Melalui analisis ini, diharapkan hasil dari penelitian ini dapat memberikan kontribusi yang berarti bagi kemajuan sistem kecerdasan buatan di bidang pertanian, terutama dalam membantu petani untuk mendeteksi penyakit pada tanaman secara dini dengan tingkat akurasi yang tinggi.

2 Tinjauan Literatur

Evolusi metodologi identifikasi penyakit tanaman dalam satu dekade terakhir menunjukkan pergeseran dominasi dari teknik pemrosesan citra konvensional menuju pendekatan berbasis *deep learning*. Pemanfaatan CNN telah menjadi standar emas dalam literatur karena kemampuannya melakukan ekstraksi fitur hierarkis secara otomatis, yang terbukti melampaui metode tradisional dalam menangani variabilitas latar belakang alami dan kondisi pencahayaan di lapangan [7][8]. Penelitian pada komoditas jambu mete sendiri telah mengeksplorasi berbagai arsitektur, di mana model dengan kedalaman lapisan yang signifikan mampu mencapai akurasi hingga 97,61% [4]. Keunggulan ini selaras dengan tren klasifikasi citra medis yang memanfaatkan abstraksi fitur laten untuk mengenali pola-pola kompleks yang tidak tertangkap oleh mata telanjang [9][10]. Namun, efektivitas model CNN seringkali sangat bergantung pada jumlah data yang besar dan daya komputasi yang tinggi, serta memiliki risiko *overfitting* yang besar jika tidak didukung oleh teknik regulasi seperti *Batch Normalization* dan *Dropout* [7][11].

Di sisi lain, algoritma *Random Forest* muncul sebagai alternatif yang tangguh karena sifatnya yang robust terhadap data yang tidak seimbang dan kemampuannya meminimalkan varians melalui mekanisme ensemble [6][12]. Algoritma ini bekerja dengan membangun koleksi pohon keputusan yang dilatih pada subset data acak, sehingga meningkatkan kemampuan generalisasi pada data baru [6]. *Random Forest* juga terbukti imun terhadap *outliers* dan struktur error distributif yang biasanya mengganggu model parametrik [6]. Dalam berbagai aplikasi pertanian dan geologi, *Random Forest* menunjukkan performa yang stabil meskipun bekerja dengan dataset yang lebih kecil dibandingkan kebutuhan CNN [13][12]. Keunggulan teknis lainnya adalah kemampuannya menangani ribuan variabel input tanpa perlu penghapusan manual, serta mampu mengidentifikasi interaksi antar variabel terlepas dari seberapa besar pengaruh (*effect size*) variabel tersebut [6][14]. Pemilihan metode ini juga didukung oleh efisiensinya dalam menangani variabel kategorikal dibandingkan algoritma lain seperti SVM atau Neural Networks [15].

Meskipun demikian, efektivitas *Random Forest* sangat ditentukan oleh kualitas fitur yang diekstraksi secara manual. Sebagian besar penelitian yang menggunakan *Random Forest* pada citra tanaman cenderung hanya mengandalkan fitur statistik dasar atau parameter fisik sederhana [12], sehingga sering kali gagal menangkap detail tekstur lesi atau morfologi bentuk daun yang menjadi indikator utama penyakit seperti *Anthracoze* atau *Red Rust* pada tanaman mete.

Kesenjangan penelitian yang ditemukan dalam literatur saat ini adalah masih minimnya studi komparatif yang menguji apakah fitur otomatis dari lapisan konvolusi CNN benar-benar lebih superior dibandingkan fitur manual yang dirancang secara sistematis pada kasus spesifik jambu mete. Kebanyakan penelitian cenderung memisahkan kedua algoritma tersebut atau hanya berfokus pada optimasi satu metode tanpa melakukan evaluasi mendalam terhadap efisiensi fitur yang digunakan. Studi ini fokus pada bagian yang belum dikerjakan tersebut, yaitu dengan melakukan pengujian langsung antara performa CNN terhadap *Random Forest* yang didukung oleh ekstraksi fitur manual tingkat tinggi. Berbeda dengan riset sebelumnya, penelitian ini secara khusus mengintegrasikan fitur tekstur melalui *Gray Level Co-occurrence Matrix* (GLCM) dan deskriptor bentuk melalui *Hu Moments* untuk melengkapi fitur warna. Fokus utama riset ini adalah untuk membuktikan apakah representasi fitur eksplisit yang mencakup parameter kontras, homogenitas, dan invarian momen mampu memberikan akurasi yang kompetitif atau bahkan melampaui fitur laten CNN dalam mengklasifikasikan lima kelas penyakit jambu mete secara efisien.

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan membandingkan dua jalur pemrosesan citra yang berbeda, yaitu jalur *Deep Learning CNN* dengan arsitektur *Custom Sequential* dan jalur *Machine Learning* Tradisional *Random Forest*. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, ekstraksi fitur, pelatihan model, dan evaluasi.

3.1 Pengumpulan dan Persiapan Data

Data citra daun jambu mete dikumpulkan ke dalam direktori berdasarkan kelasnya. Sistem memuat citra secara otomatis dan menyimpannya dalam struktur array untuk diproses lebih lanjut.

3.2 Pra-Pemrosesan

Terdapat dua alur preprocessing yang berbeda sesuai kebutuhan algoritma:

1. Jalur CNN: Citra di *resize* menjadi 128×128 piksel dan dinormalisasi ke rentang $[0,1]$. Arsitektur dibangun menggunakan model *Custom Sequential* yang terdiri dari lapisan konvolusi untuk ekstraksi fitur otomatis, *Batch Normalization*, serta *MaxPooling*. Teknik Dropout sebesar 0,5 diterapkan pada lapisan *fully connected* untuk mencegah *overfitting*.
2. Jalur Random Forest: Citra di *resize* menjadi 128×128 piksel untuk efisiensi komputasi dalam memproses vektor fitur manual. Ekstraksi fitur dilakukan secara eksplisit meliputi fitur warna (Mean & Std Dev RGB), fitur tekstur menggunakan *Gray Level Co-occurrence Matrix* (GLCM) untuk mengambil nilai *contrast*, *homogeneity*, dan *energy*, serta fitur bentuk melalui *Hu Moments*.

Seluruh data didistribusikan menggunakan teknik split sampling dengan rasio 80% Training, 10% Validation, dan 10% Testing. Perbandingan kinerja kedua model difokuskan pada akurasi, efektivitas melalui *confusion matrix*, serta efisiensi komputasi dalam mengklasifikasikan citra penyakit tanaman jambu mete.

3.3 Ekstraksi Fitur

1. Fitur Otomatis untuk CNN: Fitur diekstraksi secara otomatis melalui lapisan *Convolutional* (Conv2D) yang dipadukan dengan *Batch Normalization* dan *MaxPooling*.
2. Fitur Manual untuk Random Forest: Dilakukan ekstraksi tiga komponen utama yaitu fitur Warna (Mean & Standar Deviasi RGB), fitur Tekstur melalui *Gray Level Co-occurrence Matrix* (GLCM) untuk mendapatkan nilai *contrast*, *homogeneity*, dan *energy*, serta fitur Bentuk menggunakan *Hu Moments*.

3.4 Arsitektur Model CNN

CNN adalah jenis arsitektur deep learning yang sangat efektif untuk tugas pengolahan data berbentuk grid, terutama citra. CNN meniru cara kerja sistem visual manusia dan mampu secara otomatis mempelajari fitur-fitur relevan dari data input secara hierarkis. Operasi dasar CNN untuk ekstraksi fitur otomatis (konvolusi) didefinisikan sebagai berikut pada rumus 1:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (1)$$

Citra di *resize* menjadi 128x128 piksel diikuti dengan normalisasi nilai piksel ke rentang [0,1]. Model menggunakan *custom top layers* yang terdiri dari:

1. Aktivasi ReLu digunakan pada lapisan konvolusi dan *dense layer* untuk memperkenalkan non-linearitas, pada rumus 2:
$$ReLU(x) = \max(0, x) \quad (2)$$
2. *Global Average Pooling & Batch Normalization* untuk menstabilkan proses pelatihan dan mengurangi dimensi.
3. *Dropout* (0.5) untuk mencegah terjadinya overfitting selama pelatihan.
4. *Output Layer (Softmax)* untuk menghasilkan probabilitas klasifikasi akhir.

Pelatihan dilakukan menggunakan optimizer Adam dan loss function Categorical Crossentropy.

3.5 Arsitektur Model Random Forest

Random Forest adalah metode *ensemble learning* yang kuat untuk klasifikasi dan regresi, bekerja dengan membangun banyak pohon keputusan (*decision trees*). Model *Random Forest* dalam penelitian ini dibangun menggunakan pendekatan fitur manual yang mengekstraksi karakteristik visual spesifik dari citra:

1. Histogram Warna: Digunakan untuk menangkap distribusi warna (RGB), yang krusial untuk membedakan objek berdasarkan intensitas warna dominan.
2. Fitur Tekstur (GLCM): Mengekstraksi informasi tekstur seperti *contrast*, *homogeneity*, dan *energy* untuk menangkap pola permukaan pada citra.
3. Fitur Bentuk (*Hu Moments*): Menangkap geometri objek yang bersifat invariant terhadap perubahan posisi dan skala.

Vektor fitur gabungan ini kemudian dilatih menggunakan algoritma *Random Forest* dengan 100 *decision trees* ($n_{estimators} = 100$). Kriteria pemisahan pada setiap *decision tree* menggunakan *Gini Impurity*, yang didefinisikan pada rumus 3:

$$Gini = 1 - \sum_{i=1}^C (P_i)^2 \quad (3)$$

Dimana P_i adalah probabilitas atau proporsi sampel kelas ke- i dalam node, dan C adalah jumlah kelas.

3.6 Evaluasi

Kinerja algoritma *Random Forest* dan CNN dievaluasi secara komparatif untuk menentukan model yang paling optimal dalam menangani klasifikasi. Evaluasi dilakukan pada dataset uji (*testing set*) guna menjamin objektivitas hasil. Parameter utama yang diukur meliputi Akurasi, Presisi, *Recall*, dan *F1-Score*. Akurasi global dihitung menggunakan rumus 4:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Dimana TP adalah *True Positive*, TN adalah *True Negative*, FP adalah *False Positive*, dan FN adalah *False Negative*.

4 Hasil dan Pembahasan

4.1 Persiapan Dataset dan Visualisasi Data

Penelitian ini mengimplementasikan dua pendekatan berbeda untuk klasifikasi citra penyakit pada tanaman jambu mete. Dataset yang digunakan terdiri dari total 6.549 citra yang terbagi ke dalam 5 kelas penyakit/kondisi: *Anthracoise*, *Gumosis*, *Healthy*, *Leaf Miner*, dan *Red Rust*.

<http://sistemasi.ftik.unisi.ac.id>

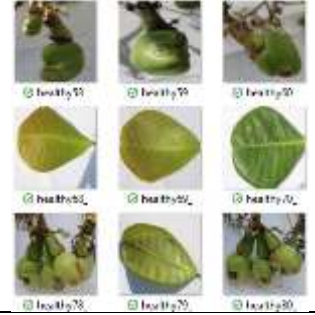
Pembagian dataset mengikuti rasio 80:10:10, di mana 80% data digunakan untuk *training*, 10% untuk validasi selama proses *training* CNN, dan 10% sisanya sebagai data *testing* untuk mengevaluasi performa model akhir. Detail pembagian jumlah data untuk setiap kelas dapat dilihat pada Tabel 1.

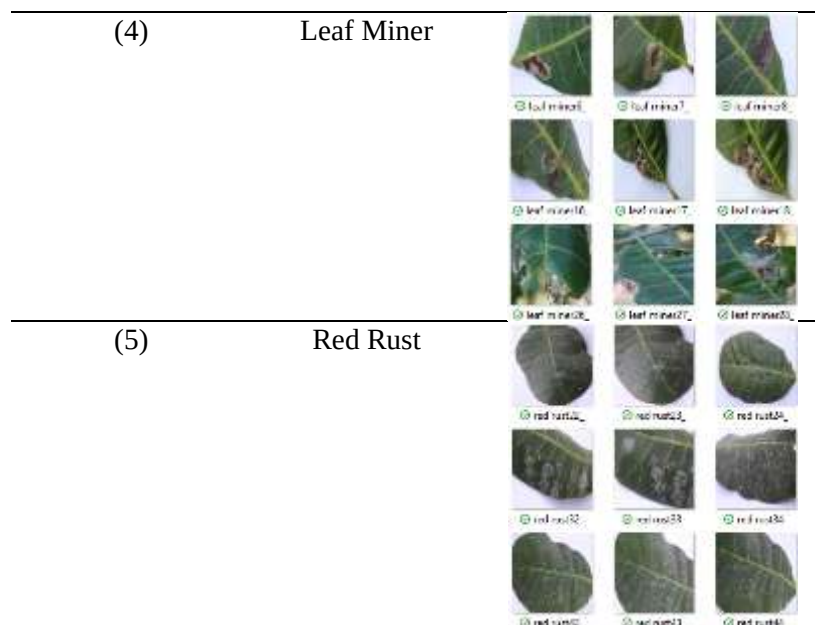
Tabel 1 Pembagian jumlah data

Nama Kelas	Total Citra	Data Training (80%)	Data Validasi (10%)	Data Testing (10%)
Anthraco nose	1.729	1.383	173	173
Gumosis	392	314	39	39
Healthy	1.368	1.094	137	137
Leaf Miner	1.378	1.102	138	138
Red Rust	1.682	1.346	168	168
Total	6.549	5.239	655	655

Visualisasi representatif dari setiap kelas disajikan pada Tabel 2 untuk memperlihatkan variasi karakteristik visual antar-kelas.

Tabel 2 Visualisasi citra

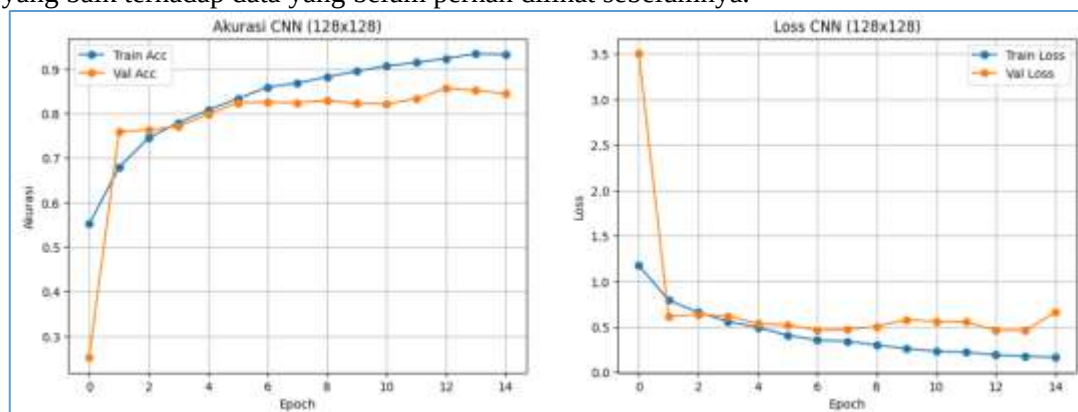
No	Nama Kelas	Visualisasi
(1)	Anthraco nose	
(2)	Gumosis	
(3)	Healthy	



4.2 Hasil Pengujian Algoritma

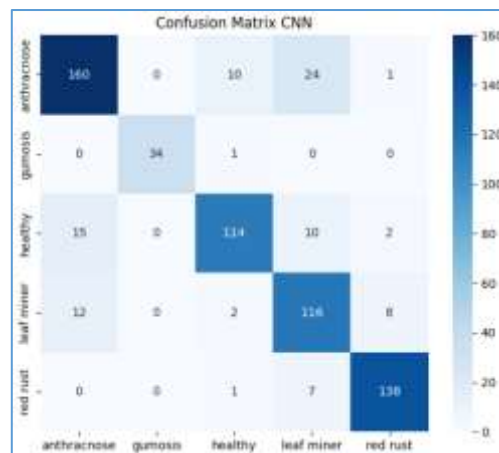
1. Kinerja *Convolutional Neural Network* (CNN)

Pengujian model CNN menunjukkan kinerja yang konsisten dengan akurasi pelatihan (*Train Acc*) yang mencapai lebih dari 90% dan akurasi validasi (*Val Acc*) berada di antara 84-86%. Tren positif tersebut dapat dilihat secara visual pada Gambar 1, di mana kurva akurasi menunjukkan peningkatan yang berkelanjutan seiring bertambahnya *epoch*. Selain itu, grafik *loss* pada Gambar 1 yang menunjukkan penurunan yang konvergen menandakan bahwa model berhasil memahami fitur penyakit jambu mete dengan baik, tanpa menunjukkan tanda-tanda *overfitting* yang signifikan. Jarak yang stabil antara kurva pelatihan dan validasi mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya.



Gambar 1 Grafik akurasi dan loss CNN

Evaluasi lebih mendalam melalui *Confusion Matrix* pada Gambar 2 menunjukkan CNN sangat akurat dalam mengidentifikasi kelas *gumosis* (34 benar) dan *red rust* (138 benar). Meskipun ada sedikit kesalahan klasifikasi antara *anthracnose* dan *leaf miner*, model ini secara keseluruhan memiliki kemampuan pemisahan kelas yang tajam melalui fitur otomatisnya.



Gambar 2 Confusion matrix CNN

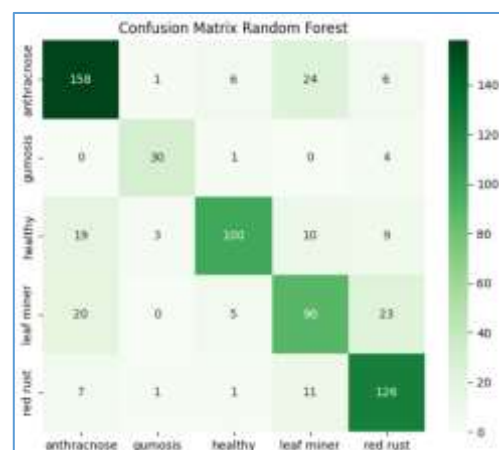
Berdasarkan laporan klasifikasi pada Gambar 3, Laporan klasifikasi mencatat akurasi total sebesar 86%. Tingginya nilai *F1-score* pada kelas *gumosis* (0.99) dan *red rust* (0.94) membuktikan bahwa arsitektur CNN sangat efektif dalam menangkap karakteristik unik dari setiap jenis penyakit.

	precision	recall	f1-score	support
anthracnose	0.86	0.82	0.84	195
gumosis	1.00	0.97	0.99	35
healthy	0.89	0.81	0.85	141
leaf miner	0.74	0.84	0.79	138
red rust	0.93	0.95	0.94	146
accuracy			0.86	655
macro avg	0.88	0.88	0.88	655
weighted avg	0.86	0.86	0.86	655

Gambar 3 Classification report CNN

2. Kinerja Random Forest

Sebaliknya, pengujian menggunakan algoritma *Random Forest* berbasis fitur manual (Warna, Tekstur GLCM, dan Hu Moments) menunjukkan performa yang lebih rendah dibandingkan model CNN. *Confusion Matrix* pada Gambar 4 memperlihatkan penyebaran prediksi yang cukup luas, terutama pada kelas *leaf miner* yang sering salah diklasifikasikan sebagai *anthracnose* (20 citra) atau *red rust* (23 citra). Hal ini mengindikasikan bahwa fitur yang dirancang secara manual (*hand-crafted features*) mengalami keterbatasan dalam menangkap kompleksitas detail tekstur yang menjadi pembeda utama pada penyakit daun jambu mete.



Gambar 4 Confusion matrix random forest

Laporan klasifikasi akhir untuk algoritma *Random Forest* pada Gambar 5 mencatatkan bahwa *Random Forest* hanya mencapai akurasi sebesar 77%. Rendahnya *F1-*

score pada kelas *leaf miner* (0.66) menegaskan bahwa algoritma *machine learning* konvensional sangat bergantung pada kualitas fitur input dan cenderung lebih lemah dibandingkan CNN dalam menangani variasi data yang rumit.

	precision	recall	f1-score	support
anthracnose	0.77	0.81	0.79	195
gumosis	0.86	0.86	0.86	35
healthy	0.88	0.71	0.79	141
leaf miner	0.67	0.65	0.66	138
red rust	0.75	0.86	0.80	146
accuracy			0.77	655
macro avg	0.79	0.78	0.78	655
weighted avg	0.77	0.77	0.77	655

Gambar 5 Classification report random forest

4.3 Komparasi dan Analisis Mendalam

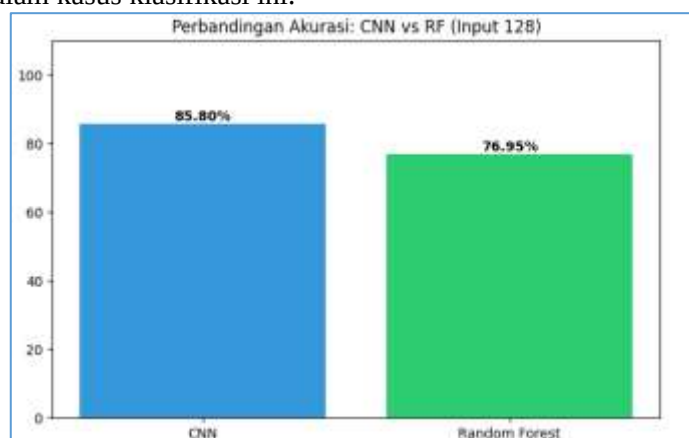
1. Perbandingan Kuantitatif

Perbandingan langsung antara kedua metode dapat disajikan pada Tabel 3, yang menunjukkan adanya selisih kinerja (*performance gap*) yang cukup jelas. Model CNN menunjukkan keunggulan dengan selisih akurasi sebesar 9% dibandingkan dengan model *Random Forest*.

Tabel 3 Perbandingan kinerja CNN dan random forest

Metrik Evaluasi	CNN (Custom Sequential)	Random Forest (Fitur Manual)
Akurasi (Accuracy)	85.80%	76.95%
Rata-rata Presisi	0.88	0.79
Rata-rata Recall	0.88	0.78
Rata-rata F1-Score	0.88	0.78

Visualisasi perbandingan pada Gambar 6 mempertegas dominasi metode *Deep Learning* CNN dalam kasus klasifikasi ini.



Gambar 6 Visualisasi perbandingan

2. Analisis Komparatif Performa Model

Dalam klasifikasi penyakit tanaman jambu mete, evaluasi kinerja antara arsitektur *deep learning* dan *machine learning* konvensional menunjukkan perbedaan yang cukup mencolok. Kinerja *Convolutional Neural Network* (CNN) dengan arsitektur *custom sequential* berhasil mencapai tingkat akurasi sebesar 85,80% pada dataset uji. Berdasarkan analisis terhadap *kurva learning rate*, terlihat penurunan nilai *loss* yang konsisten, yang menunjukkan adanya proses konvergensi yang stabil selama fase pelatihan. Keberhasilan model dalam mengurangi gejala *overfitting* yang signifikan juga didorong oleh penggunaan teknik regulasi seperti *Dropout* (0,5) dan *Batch Normalization*, sehingga generalisasi model terhadap data baru tetap terjaga.

Di sisi lain, algoritma *Random Forest* menunjukkan kinerja yang lebih rendah dengan perolehan akurasi sebesar 76,95%. Hasil *confusion matrix* mengungkapkan bahwa proses klasifikasi mengalami kendala, terutama pada kelas *leaf miner*. Sebanyak 20 gambar dalam kelas tersebut salah diprediksi sebagai *anthracnose*. Hal ini mengindikasikan bahwa penggunaan fitur manual seperti GLCM dan *Hu Moments* gagal mengekstraksi kompleksitas tekstur dan hubungan spasial yang penting untuk membedakan kedua jenis penyakit tersebut secara akurat

Rendahnya performa *Random Forest* disebabkan oleh keterbatasan representasi fitur statis yang gagal menangkap hubungan spasial kompleks antar piksel. Meskipun fitur GLCM dan *Hu Moments* telah digunakan, fitur ini tetap bersifat linear dan tidak mampu menangkap variabilitas pola penyakit yang memiliki kemiripan visual tinggi secara otomatis. Sebaliknya, keunggulan CNN terletak pada *representation learning* yang mencari fitur diskriminatif secara mandiri. Namun, perlu dicatat bahwa performa CNN sangat bergantung pada stabilitas distribusi data uji, adanya pergeseran domain pada citra lapangan nyata mungkin dapat menurunkan akurasi ini.

3. Analisis Kegagalan Fitur Manual pada Random Forest

Rendahnya performa *Random Forest* (76,95%) dibandingkan CNN (85,80%) pada klasifikasi penyakit jambu mete ini disebabkan oleh beberapa faktor teknis terkait keterbatasan fitur manual (*hand-crafted features*). Pertama, terdapat keterbatasan representasi fitur statis dimana penggunaan fitur manual seperti GLCM dan *Hu Moments* dalam penelitian ini hanya menangkap informasi tekstur dan bentuk pada tingkat dasar. Hal ini sejalan dengan teori bahwa algoritma *Random Forest* sangat bergantung pada kualitas pengumpulan data dan ekstraksi fitur awal untuk mendapatkan hasil yang akurat [6]. Dalam kasus ini, fitur warna global dan tekstur statis gagal membedakan gejala penyakit yang memiliki kemiripan visual tinggi seperti *Anthracnose* dan *Leaf Miner*.

Selain itu, kegagalan ini juga dipengaruhi oleh hilangnya informasi hierarkis. Vektor fitur yang digunakan pada *Random Forest* tidak memiliki kemampuan untuk menangkap hubungan spasial yang kompleks antar piksel sebagaimana yang dilakukan oleh operasi konvolusi pada CNN. Sebagaimana dijelaskan dalam literatur, meskipun *Random Forest* adalah metode yang andal untuk klasifikasi berdasarkan fitur fisik tertentu [12], algoritma ini cenderung kesulitan jika fitur yang diberikan tidak mampu merepresentasikan variabilitas pola yang rumit secara otomatis. Masalah generalisasi pada citra non-linear pun turut menjadi kendala, ketidakmampuan fitur manual dalam menangkap representasi data yang kompleks menyebabkan model *Random Forest* sering kali salah dalam memprediksi kelas dengan pola yang serupa. Kesalahan klasifikasi ini membuktikan bahwa tanpa rekayasa fitur yang sangat mendalam, metode *machine learning* konvensional akan tetap berada di bawah performa *deep learning* yang memiliki kemampuan *representation learning* untuk mencari fitur paling diskriminatif secara mandiri [6].

5 Kesimpulan

Penelitian ini berhasil mengevaluasi perbandingan kinerja antara algoritma *Convolutional Neural Network* (CNN) dan *Random Forest* dalam mengklasifikasikan lima kelas penyakit pada tanaman jambu mete. Hasil pengujian menunjukkan bahwa model CNN dengan arsitektur *Custom Sequential* memiliki performa yang lebih unggul dengan tingkat akurasi mencapai 85,80%, dibandingkan dengan algoritma *Random Forest* yang menggunakan fitur manual dengan akurasi sebesar 76,95%. Keunggulan CNN terletak pada kemampuannya melakukan ekstraksi fitur secara otomatis dan hierarkis yang sangat efektif dalam menangkap karakteristik unik penyakit, terutama pada kelas *gumosis* dan *red rust*. Sebaliknya, performa *Random Forest* terbatas oleh kualitas fitur statis (warna, tekstur GLCM, dan *Hu Moments*) yang kesulitan membedakan gejala dengan kemiripan visual tinggi seperti *Anthracnose* dan *Leaf Miner*. Dengan demikian, dapat disimpulkan bahwa pendekatan *deep learning* melalui CNN merupakan metode klasifikasi yang paling efektif dan akurat untuk mendukung sistem deteksi dini penyakit tanaman jambu mete. Namun, klaim keunggulan ini terbatas pada kondisi dataset eksperimental yang digunakan. Keterbatasan penelitian ini terletak pada penggunaan fitur manual standar pada *Random Forest* yang belum mencakup seluruh aspek morfologi tanaman jambu

mete. Riset mendatang disarankan untuk mengeksplorasi teknik *Hybrid* dimana CNN sebagai ekstraktor fitur untuk *Random Forest*, serta pengujian pada dataset dengan variasi latar belakang alami yang lebih kompleks untuk meningkatkan generalisasi model di lapangan.

Referensi

- [1] Syukriadi, “Sistem Pakar untuk Diagnosis Penyakit Tanaman Jagung berbasis Kecerdasan Buatan *Expert System for Diagnosing Maize Plant Diseases based on Artificial Intelligence.*”
- [2] V. Riyanto, S. Nurdianti, M. Marimin, M. Syukur, and S. N. Neyman, “*Deep Learning Approaches for Plant Disease Diagnosis Systems: A Review and Future Research Agendas.*” *Journal of Applied Agricultural Science and Technology*, Vol. 9, No. 2, pp. 185–204, May 2025, DOI: 10.55043/jaast.v9i2.308.
- [3] S. P. Mohanty, D. P. Hughes, and M. Salathé, “*Using Deep Learning for Image-based Plant Disease Detection.*” *Front. Plant SCI.*, Vol. 7, No. September, Sep. 2016, DOI: 10.3389/fpls.2016.01419.
- [4] F. Maria Bang Ritan and A. Yakobus Chandra, “Analisis Perbandingan Kinerja Model CNN Resnet-50, VGG19 dan Mobilenet dalam Klasifikasi Penyakit pada Tanaman Mete”, [Online]. Available: <https://locus.rivierapublishing.id/index.php/jl>
- [5] A. Kumari and J. Singh, “*Designing of Guava Quality Classification Model based on Anova and Machine Learning.*” *SCI. Rep.*, Vol. 15, No. 1, Dec. 2025, DOI: 10.1038/s41598-025-09684-7.
- [6] H. A. Salman, A. Kalakech, and A. Steiti, “*Random Forest Algorithm Overview.*” *Babylonian Journal of Machine Learning*, Vol. 2024, pp. 69–79, Jun. 2024, DOI: 10.58496/bjml/2024/007.
- [7] J. Lu, L. Tan, and H. Jiang, “*Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification.*” Aug. 01, 2021, MDPI AG. DOI: 10.3390/agriculture11080707.
- [8] H. Gunasekaran, K. Ramalakshmi, A. Rex Macedo Arokiaraj, S. D. Kanmani, C. Venkatesan, and C. S. G. Dhas, “*Analysis of DNA Sequence Classification using CNN and Hybrid Models.*” *Comput. Math. Methods Med.*, Vol. 2021, 2021, DOI: 10.1155/2021/1835056.
- [9] Y. AbdulAzeem, W. M. Bahgat, and M. Badawy, “*A CNN based Framework for Classification of Alzheimer’s Disease.*” *Neural Comput. Appl.*, Vol. 33, No. 16, pp. 10415–10428, Aug. 2021, DOI: 10.1007/s00521-021-05799-w.
- [10] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, “*Deep CNN for Brain Tumor Classification.*” *Neural Process. Lett.*, Vol. 53, No. 1, pp. 671–700, Feb. 2021, DOI: 10.1007/s11063-020-10398-2.
- [11] S. Kutluk, K. Kayabol, and A. Akan, “*A New CNN Training Approach with Application to Hyperspectral Image Classification.*” *Digital Signal Processing: A Review Journal*, Vol. 113, Jun. 2021, DOI: 10.1016/j.dsp.2021.103016.
- [12] A. Putra Argadianata, D. Abdul Fatah, H. Sukri, J. Raya Telang, K. Kamal, and J. Timur, “Klasifikasi Kualitas Buah Apel menggunakan Metode *Random Forest*,” 2025.
- [13] P. Behnia, J. Harris, H. Liu, T. R. C. Jørgensen, M. Naghizadeh, and E. A. Roots, “*Random Forest Classification for Volcanogenic Massive Sulfide Mineralization in the Rouyn-Noranda Area, Quebec.*” *Ore Geol. Rev.*, Vol. 161, Oct. 2023, DOI: 10.1016/j.oregeorev.2023.105612.
- [14] W. Mobley, A. Sebastian, R. Blessing, W. E. Highfield, L. Stearns, and S. D. Brody, “*Quantification of Continuous Flood Hazard using Random Forest Classification and Flood Insurance Claims at Large Spatial Scales: A Pilot Study in Southeast Texas.*” *Natural Hazards and Earth System Sciences*, Vol. 21, No. 2, pp. 807–822, Mar. 2021, DOI: 10.5194/nhess-21-807-2021.
- [15] A. J. Tesoriero, S. A. Wherry, D. I. Dupuy, and T. D. Johnson, “*Predicting Redox Conditions in Groundwater at a National Scale using Random Forest Classification.*” *Environ. SCI. Technol.*, Vol. 58, No. 11, pp. 5079–5092, Mar. 2024, DOI: 10.1021/acs.est.3c07576.