

Multiclass Classification of Tomato Leaf Diseases Using GLCM, Color, and Shape Feature Extraction with Optimized XGBoost

Fransisko Andrade Laiskodat ^{1*}, Majid Rahardi ^{2*}

* Informatika, Universitas Amikom Yogyakarta

fransisko@students.amikom.ac.id ¹, majid@amikom.ac.id ²

Article Info

Article history:

Received 2025-09-19

Revised 2025-11-05

Accepted 2025-11-08

Keyword:

*Feature Extraction,
Gray-Level Co-occurrence
Matrix (GLCM),
Principal Component Analysis
(PCA),
Smote,
XGBoost*

ABSTRACT

Automatic classification of tomato leaf diseases is an essential component in advancing precision agriculture based on artificial intelligence. This study aims to develop a multiclass classification model for tomato leaf diseases by utilizing texture, color, and shape features, and employing an optimized XGBoost algorithm. The public PlantVillage dataset was used, with preprocessing stages including feature extraction, normalization, dimensionality reduction using PCA, and class balancing using SMOTE. The experimental results showed that the model successfully classified ten disease classes with a high accuracy of 97.63%, and both macro and weighted f1-scores of 0.98. These findings indicate that the combination of handcrafted features and XGBoost offers an effective, efficient, and applicable solution for plant disease diagnostic systems.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Penyakit daun tomat telah menjadi permasalahan signifikan dalam pertanian hortikultura global, menyebabkan penurunan hasil panen secara drastis hingga mencapai 60-80% di berbagai negara penghasil tomat utama, yang berdampak langsung pada kerugian ekonomi petani dan rantai pasok pertanian secara global [1], [2]. Deteksi penyakit secara manual oleh petani umumnya bersifat subjektif, bergantung pada pengalaman individu, serta tidak efisien ketika dilakukan pada skala lahan yang luas, sehingga sering mengakibatkan keterlambatan dalam pengendalian penyakit. Teknologi pengenalan citra berbasis kecerdasan buatan mulai banyak diadopsi untuk mendeteksi gejala penyakit tanaman lebih awal [3]. Berbagai penelitian menggunakan pendekatan deep learning seperti CNN, ResNet, hingga Vision Transformer untuk klasifikasi penyakit daun tomat [4], [5], [6]. Selain itu, metode berbasis ekstraksi fitur tekstur (GLCM), warna, dan bentuk juga terbukti efektif dalam representasi karakteristik daun [7], [8]. Kombinasi metode klasik dan pembelajaran mesin (machine learning) memberikan alternatif yang efisien dari sisi waktu pelatihan dan interpretabilitas [9].

Meskipun pendekatan deep learning menunjukkan akurasi tinggi, tantangan seperti kebutuhan komputasi besar, data imbalance, dan kurangnya interpretabilitas masih menjadi

hambatan [10]. Belum banyak penelitian yang secara komprehensif menggabungkan fitur GLCM, warna, dan bentuk bersama teknik oversampling serta reduksi dimensi [11]. Beberapa studi belum membandingkan kinerja fitur tersebut dalam skenario multikelas secara merata [12]. Selain itu, penggunaan XGBoost yang dituning secara optimal sebagai classifier utama untuk skenario multikelas penyakit daun tomat masih jarang dilakukan [13]. Sebagian besar penelitian sebelumnya berfokus pada model berbasis CNN atau deep learning, sedangkan perbandingan antara model berbasis handcrafted features dengan algoritma XGBoost yang dioptimasi secara efisien untuk lingkungan komputasi terbatas masih sangat terbatas. Oleh karena itu, terdapat celah penelitian dalam pengembangan model klasifikasi ringan dan akurat dengan strategi ekstraksi fitur terintegrasi [14].

Mengisi celah ini penting untuk menghadirkan solusi klasifikasi penyakit daun tomat yang efisien, mudah diterapkan, dan akurat tinggi [15]. Penggunaan GLCM dapat menggambarkan tekstur, warna mencerminkan pigmentasi akibat penyakit, dan bentuk menunjukkan deformasi daun, menjadikan kombinasi fitur ini sangat representatif [4]. Dengan menambahkan PCA untuk reduksi dimensi serta SMOTE untuk mengatasi data imbalance, kualitas data yang masuk ke model menjadi lebih baik [4], [5], [6]. XGBoost dipilih karena performanya yang kuat pada data tabular dan

kemampuannya dalam menangani multiklasifikasi serta overfitting [7], [8]. Optimalisasi hyperparameter XGBoost diharapkan menghasilkan model yang stabil dan akurat untuk diagnosis penyakit daun tomat [1].

Penelitian ini menggunakan dataset publik PlantVillage yang terdiri dari 10 kelas penyakit daun tomat. Fitur yang diekstrak meliputi GLCM (tekstur), histogram HSV (warna), dan fitur bentuk dasar (eccentricity, area, perimeter). Selanjutnya dilakukan reduksi dimensi menggunakan PCA dan penyeimbangan kelas dengan SMOTE. Data yang telah diproses dilatih menggunakan model XGBoost yang dioptimasi menggunakan GridSearchCV. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, F1-score, confusion matrix, dan log loss.

Model XGBoost yang telah dituning menunjukkan performa sangat tinggi dengan akurasi akhir sebesar 97.63%. Nilai precision dan recall dari seluruh kelas berkisar antara 0.95 hingga 1.00, menunjukkan kemampuan klasifikasi multikelas yang seimbang. F1-score tertinggi diraih oleh kelas Tomato Mosaic Virus dan Healthy sebesar 0.99-1.00, sedangkan kelas Late Blight dan Tomato Yellow Leaf Curl Virus menunjukkan F1-score 0.96. Secara keseluruhan, nilai macro dan weighted average untuk precision, recall, dan F1-score mencapai 0.98, membuktikan stabilitas model di seluruh kelas. Hasil ini memperlihatkan keberhasilan integrasi fitur, SMOTE, PCA, dan XGBoost dalam menghasilkan model klasifikasi yang andal.

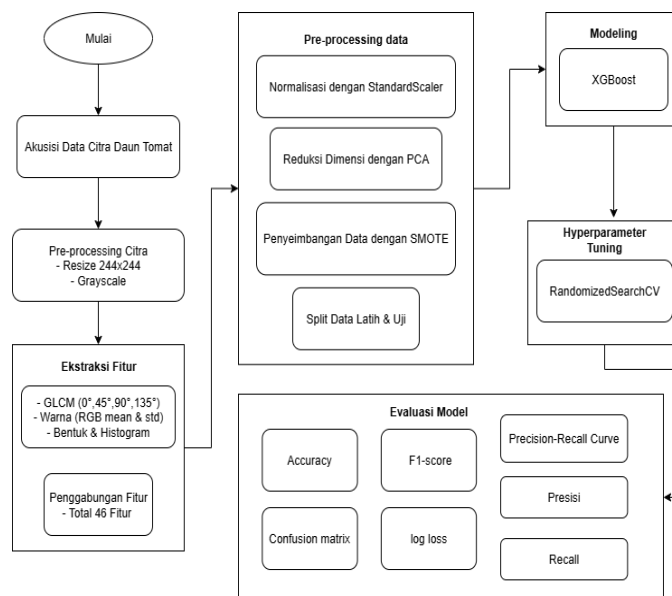
Hasil yang diperoleh menunjukkan bahwa kombinasi fitur GLCM, warna, dan bentuk mampu menangkap karakteristik penyakit secara komprehensif. Penerapan PCA membantu menurunkan kompleksitas model tanpa mengorbankan akurasi. Strategi SMOTE terbukti efektif dalam menyeimbangkan distribusi kelas dan meningkatkan generalisasi model. XGBoost yang telah dituning menunjukkan keunggulan dalam menangani klasifikasi multikelas dengan performa tinggi. Dengan demikian, penelitian ini berhasil mengisi celah yang ada dengan menyajikan pendekatan klasifikasi yang ringan namun akurat untuk penyakit daun tomat.

Keberhasilan model yang diusulkan membuka peluang penerapan sistem diagnosis otomatis dalam skala pertanian digital. Pendekatan berbasis ekstraksi fitur klasik ditambah dengan model XGBoost teroptimasi menawarkan solusi alternatif dari deep learning yang lebih ringan. Penelitian ini memberikan landasan kuat untuk pengembangan aplikasi mobile berbasis AI untuk membantu petani dalam deteksi dini penyakit tanaman. Di masa mendatang, pendekatan serupa dapat diadaptasi untuk tanaman lain dengan dataset dan fitur berbeda. Hasil ini mendukung pengembangan pertanian presisi yang murah, efisien, dan berkelanjutan.

II. METODE

Penelitian ini dilaksanakan secara daring menggunakan Google Colab untuk proses komputasi dan Google Drive sebagai media penyimpanan dataset dan hasil keluaran. Objek

penelitian adalah citra daun tomat dari dataset publik PlantVillage yang terdiri dari sepuluh kelas, termasuk satu kelas daun sehat. Dataset ini dipilih karena kelengkapan label dan kualitas citranya yang tinggi, menjadikannya representatif untuk klasifikasi penyakit tanaman. Fokus penelitian adalah pada ekstraksi fitur visual (tekstur, warna, dan bentuk) dan klasifikasi multikelas menggunakan algoritma XGBoost.



Gambar 1. Alur Penelitian

Alur penelitian yang menggambarkan keseluruhan tahapan mulai dari akuisisi data hingga evaluasi model ditunjukkan pada Gambar 1.

A. Akuisisi Data

Akuisisi data dilakukan dengan mengimpor citra yang telah terstruktur dalam folder berdasarkan label penyakit. Dataset PlantVillage terdiri dari sepuluh kelas, yaitu *Bacterial Spot*, *Early Blight*, *Late Blight*, *Leaf Mold*, *Septoria Leaf Spot*, *Target Spot*, *Tomato Yellow Leaf Curl Virus*, *Tomato Mosaic Virus*, *Two-Spotted Spider Mite*, dan *Healthy*. Pemilihan dataset ini mengacu pada penelitian terdahulu yang menggunakan sumber data serupa karena kualitas dan konsistensinya dalam pelabelan [16].

Dataset ini berisi total 14.531 citra daun tomat yang telah diklasifikasikan ke dalam sepuluh kelas penyakit. Distribusi jumlah citra pada masing-masing kelas adalah sebagai berikut: *Bacterial Spot* (1.702), *Early Blight* (800), *Late Blight* (1.528), *Leaf Mold* (762), *Septoria Leaf Spot* (1.417), *Target Spot* (1.124), *Tomato Yellow Leaf Curl Virus* (4.286), *Tomato Mosaic Virus* (299), *Two-Spotted Spider Mite* (1.341), dan *Healthy* (1.272).

B. Preprocessing Citra

Seluruh citra diseragamkan ke ukuran 224×224 piksel untuk menyatukan dimensi input sehingga proses ekstraksi fitur dan pelatihan model dapat dilakukan secara efisien dan konsisten. Citra dikonversi menjadi grayscale untuk ekstraksi fitur tekstur GLCM guna menyoroti pola spasial intensitas piksel, sedangkan informasi warna pada kanal RGB tetap dipertahankan untuk fitur warna agar seluruh karakteristik visual daun, termasuk variasi pigmentasi akibat penyakit, tetap diperhitungkan. Langkah ini mengacu pada pendekatan *preprocessing* standar dalam deteksi penyakit tanaman modern yang terbukti mampu meningkatkan kestabilan hasil ekstraksi fitur serta akurasi klasifikasi [16].

C. Ekstraksi Fitur

Ekstraksi fitur mencakup tiga aspek utama:

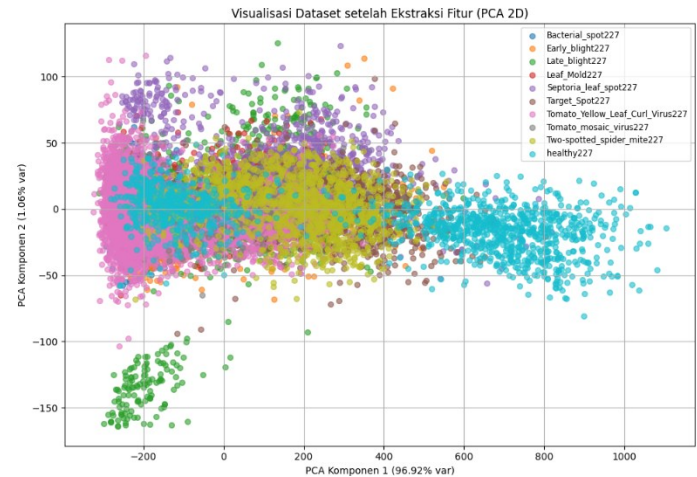
- 1) *Fitur Tekstur (GLCM)*: Menggunakan empat orientasi (0° , 45° , 90° , 135°) dan enam statistik (*contrast*, *dissimilarity*, *homogeneity*, *energy*, *correlation*, *ASM*), sehingga menghasilkan 24 fitur tekstur.
- 2) *Fitur Warna*: Menggunakan rata-rata dan deviasi standar pada tiap kanal RGB, menghasilkan 6 fitur warna.
- 3) *Fitur Bentuk*: Menggunakan histogram grayscale 16 bin untuk merepresentasikan distribusi bentuk.

Total keseluruhan fitur yang dihasilkan adalah 46 fitur numerik per citra. Pendekatan gabungan ini telah terbukti efektif dalam berbagai penelitian klasifikasi citra tanaman [17]. Kombinasi ketiga jenis fitur tersebut memberikan representasi yang lebih komprehensif terhadap karakteristik visual daun, di mana fitur tekstur mampu menangkap pola permukaan dan struktur jaringan daun, fitur warna merepresentasikan variasi pigmentasi akibat infeksi penyakit, dan fitur bentuk menggambarkan perubahan morfologi daun yang terjadi. Integrasi ketiganya memperkaya informasi citra sehingga meningkatkan kemampuan model dalam membedakan setiap kelas penyakit secara lebih akurat dibandingkan penggunaan fitur tunggal.

D. Normalisasi dan Reduksi Dimensi

Seluruh fitur dinormalisasi menggunakan StandardScaler untuk memastikan skala yang seragam. Reduksi dimensi dilakukan menggunakan Principal Component Analysis (PCA) untuk menyederhanakan komputasi, mengurangi risiko *overfitting*, dan mempercepat pelatihan model.

Hasil analisis PCA menunjukkan bahwa tiga komponen utama pertama mampu menjelaskan total 98.95% dari varians data, dengan rincian masing-masing sebesar 96.92% untuk komponen pertama (PC1), 1.06% untuk komponen kedua (PC2), dan 0.97% untuk komponen ketiga (PC3). Nilai tersebut menunjukkan bahwa sebagian besar informasi dari fitur tekstur, warna, dan bentuk terwakili oleh komponen pertama, sehingga reduksi dimensi menggunakan PCA dapat dilakukan tanpa kehilangan informasi penting secara signifikan.

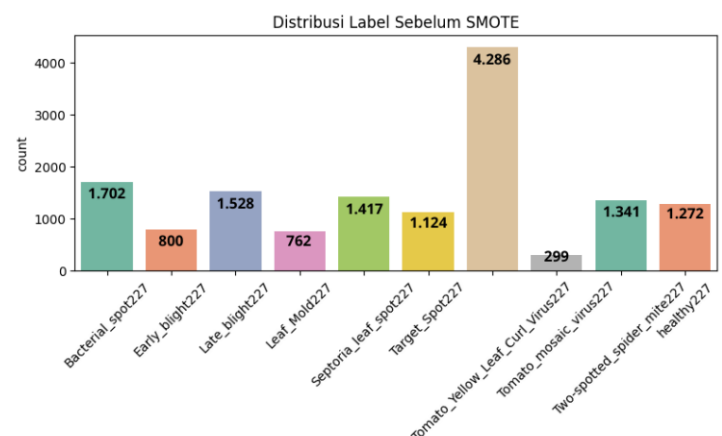


Gambar 2. Visualisasi Dataset Setelah Ekstraksi Fitur PCA 2D

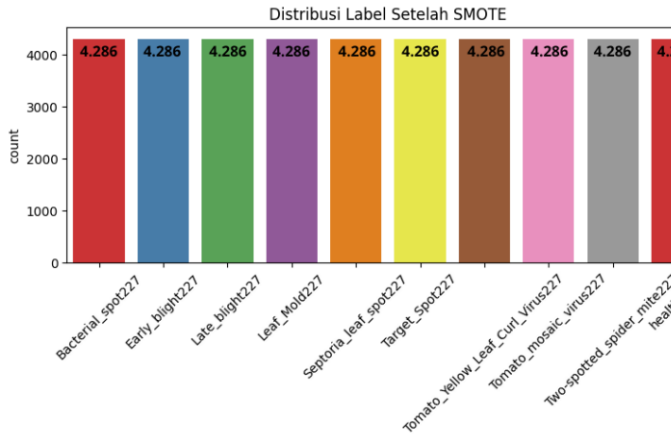
Gambar 2 memperlihatkan visualisasi data dalam ruang dua komponen utama (PCA 2D). Terlihat bahwa beberapa kelas seperti Healthy dan Tomato Mosaic Virus membentuk cluster yang lebih terpisah dibandingkan kelas lain yang memiliki gejala visual mirip seperti Early Blight dan Late Blight. Distribusi ini menunjukkan bahwa hasil ekstraksi fitur yang diikuti reduksi dimensi mampu mempertahankan perbedaan antar kelas, meskipun terdapat sedikit tumpang tindih pada kelas dengan kemiripan tekstur dan warna.

E. Penyeimbangan Data

Dataset mengalami ketidakseimbangan label sehingga diterapkan Synthetic Minority Oversampling Technique (SMOTE) untuk menyeimbangkan jumlah sampel tiap kelas. SMOTE menciptakan sampel sintesis berdasarkan tetangga terdekat dari kelas minoritas. Visualisasi distribusi data sebelum dan sesudah SMOTE ditunjukkan pada Gambar 2 dan Gambar 3. Teknik ini telah terbukti meningkatkan hasil klasifikasi pada dataset tanaman yang tidak seimbang [18].



Gambar 3. Sebelum SMOTE



Gambar 4. Sesudah SMOTE

F. Pembagian Data

Data hasil normalisasi dan SMOTE dibagi menjadi data latih (80%) dan data uji (20%) secara stratifikasi untuk mempertahankan proporsi label.

G. Pemodelan

Model klasifikasi menggunakan algoritma Extreme Gradient Boosting (XGBoost), dengan fungsi objektif sebagai berikut:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

Model dilatih terlebih dahulu dengan parameter default, kemudian dioptimasi menggunakan RandomizedSearchCV dengan validasi silang (*cross-validation*) 3-fold. Hiperparameter yang dioptimasi meliputi;

- `n_estimators`: [100, 200, 300]
- `max_depth`: [3, 5, 7, 10]
- `learning_rate`: [0.01, 0.05, 0.1, 0.2]
- `subsample`: [0.5, 0.7, 1.0]
- `colsample_bytree`: [0.5, 0.7, 1.0]

Proses optimisasi dilakukan sebanyak 25 iterasi (`n_iter` = 25) dengan seluruh inti prosesor diaktifkan (`n_jobs` = -1) untuk mempercepat pencarian kombinasi parameter yang optimal. Hasil dari *RandomizedSearchCV* menghasilkan model terbaik (`best_estimator_`) dengan kinerja tertinggi pada data uji. Model tersebut kemudian digunakan untuk melakukan prediksi terhadap data uji (`X_test`) dan menghasilkan tingkat akurasi (`accuracy_score`) sebagai metrik utama evaluasi.

H. Evaluasi Model

Evaluasi kinerja dilakukan menggunakan beberapa metrik, yaitu:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} \quad (4)$$

$$LogLoss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \cdot \log(\hat{y}_{ij}) \quad (5)$$

$$TP = \int_0^1 Precision(Recall) dRecall \quad (6)$$

Selain itu, digunakan confusion matrix yang bisa di lihat di Gambar 5. Confusion Matrix untuk melihat distribusi kesalahan prediksi pada setiap kelas. Evaluasi ini memberikan gambaran menyeluruh terhadap performa model dalam menghadapi klasifikasi multikelas dengan ketidakseimbangan data [19], [20].

III. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan model klasifikasi multikelas penyakit daun tomat menggunakan algoritma XGBoost yang telah dioptimasi. Model menunjukkan performa unggul dengan akurasi sebesar 97.63%, menunjukkan efektivitas representasi fitur gabungan berupa tekstur (GLCM), warna, dan bentuk dalam membedakan sepuluh kelas penyakit. Proses preprocessing yang mencakup normalisasi, reduksi dimensi PCA, dan penyeimbangan data dengan SMOTE terbukti meningkatkan kualitas data input serta stabilitas model. Seluruh citra daun tomat berhasil diklasifikasikan secara akurat berdasarkan fitur visual, memperlihatkan separabilitas antarkelas yang tinggi. Hasil ini konsisten dengan pendekatan hybrid yang juga digunakan dalam penelitian sebelumnya untuk klasifikasi penyakit tanaman [21].

```

=== Evaluasi Model pada Data Uji ===
Akurasi Testing : 0.9763121226196005

Classification Report:

```

	precision	recall	f1-score	support
0	0.97	0.97	0.97	861
1	0.96	0.98	0.97	861
2	0.97	0.96	0.97	861
3	0.99	0.99	0.99	862
4	0.99	0.97	0.98	861
5	0.97	0.98	0.97	861
6	0.96	0.95	0.96	862
7	0.99	1.00	0.99	861
8	0.97	0.97	0.97	861
9	0.99	1.00	1.00	861
accuracy			0.98	8612
macro avg	0.98	0.98	0.98	8612
weighted avg	0.98	0.98	0.98	8612

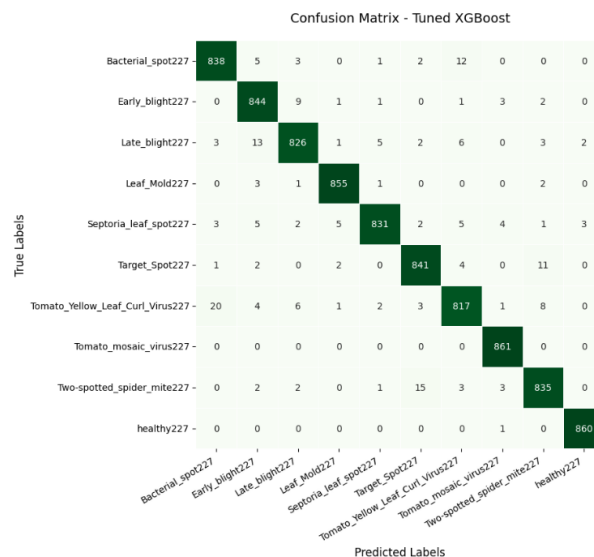
Gambar 5. Metrik Evaluasi

TABEL 1
HASIL EVALUASI MODEL XGBOOST

Class	Precision	Recall	F1-Score	Support
Bacterial Spot	0,97	0,97	0,97	861
Early Blight	0,96	0,98	0,97	861
Late Blight	0,97	0,96	0,97	861
Leaf Mold	0,99	0,99	0,99	862
Septoria Leaf Spot	0,99	0,97	0,98	861
Target Spot	0,97	0,98	0,97	861
Tomato Yellow Leaf Curl Virus	0,96	0,95	0,96	862
Tomato Mosaic Virus	0,99	1,00	0,99	861
Two-Spotted Spider Mite	0,97	0,97	0,97	861
Healthy	0,97	1,00	1,00	861
Accuracy			0,98	8612
Macro Avg	0,98	0,98	0,98	8612
Weighted Avg	0,98	0,98	0,98	8612

Evaluasi performa model dilakukan dengan menggunakan metrik akurasi, precision, recall, dan f1-score. Berdasarkan hasil evaluasi, model memiliki f1-score rata-rata sebesar 0,98, dengan skor tertinggi sebesar 1,00 pada kelas healthy dan tomato mosaic virus. Sementara itu, f1-score terendah tetap berada dalam kisaran tinggi yaitu 0,96–0,97, menunjukkan bahwa model tetap tangguh pada kelas yang secara visual mirip seperti early blight dan late blight. Performa tinggi ini menunjukkan bahwa algoritma XGBoost mampu mengatasi kompleksitas klasifikasi multikelas berbasis fitur numerik [22].

Selain itu, hasil penelitian ini membuktikan bahwa kombinasi fitur tekstur, warna, dan bentuk yang digunakan mampu menghasilkan performa klasifikasi yang kompetitif dibandingkan dengan model berbasis CNN yang dilaporkan dalam penelitian sebelumnya [5], [8], [19]. Pendekatan berbasis handcrafted features ini terbukti lebih ringan secara komputasi, tetap interpretatif, serta dapat diterapkan pada lingkungan dengan sumber daya terbatas tanpa mengorbankan akurasi. Temuan ini menunjukkan bahwa metode non-deep learning yang dioptimasi dengan baik masih memiliki relevansi tinggi untuk tugas klasifikasi citra pertanian modern.



Gambar 6. Confusion Matrix

Selanjutnya, visualisasi confusion matrix digunakan untuk menganalisis distribusi kesalahan prediksi antar kelas. Hasil confusion matrix memperlihatkan bahwa sebagian besar kelas memiliki nilai diagonal yang dominan, menunjukkan banyak prediksi yang tepat. Beberapa kesalahan minor terjadi antara kelas penyakit dengan gejala visual yang mirip, seperti antara *Bacterial Spot* dan *Target Spot*, serta antara *Early Blight* dan *Late Blight*, yang keduanya memiliki pola bercak dan perubahan warna daun yang hampir serupa. Namun, akurasi tetap tinggi di atas 95% untuk seluruh kelas, menunjukkan bahwa model mampu membedakan karakteristik visual antar penyakit dengan baik. Tidak ada kelas yang secara sistematis mengalami bias klasifikasi, menandakan keberhasilan proses penyeimbangan data menggunakan SMOTE [23].

```

=== Model Evaluation Results ===
XGBoost (tuned) Accuracy : 0.9763
Random Forest Accuracy   : 0.9662
KNN Accuracy              : 0.9646
Decision Tree Accuracy    : 0.8900

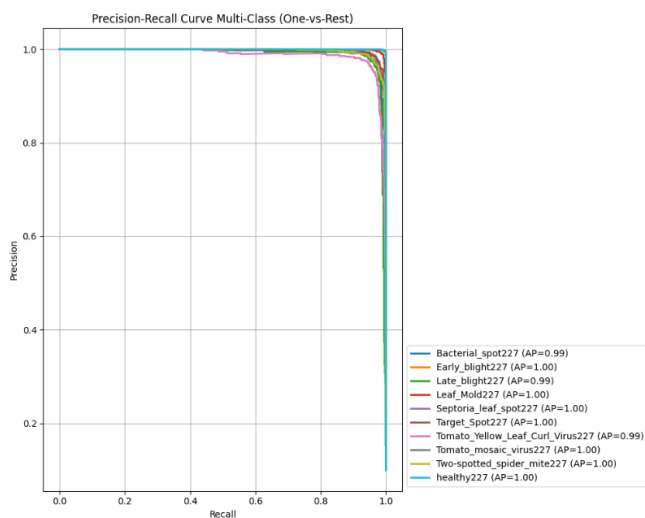
=== Perbandingan Akurasi Antar Model ===
      Model Accuracy
XGBoost (tuned) 0.976312
Random Forest   0.966210
KNN              0.964584
Decision Tree    0.890037

```

Gambar 7. Perbandingan Akurasi Model Lain

Berdasarkan hasil tersebut, model XGBoost yang telah dituning menunjukkan akurasi tertinggi sebesar 97.63%, diikuti oleh Random Forest sebesar 96.62% dan KNN sebesar 96.46%. Sementara itu, Decision Tree memiliki akurasi paling rendah yaitu 89.00%, yang menunjukkan bahwa model berbasis pohon tunggal kurang mampu menangkap kompleksitas fitur multikelas.

Performa superior XGBoost menunjukkan bahwa kombinasi fitur tekstur, warna, dan bentuk yang dioptimasi dengan teknik boosting mampu meningkatkan kemampuan diskriminatif antar kelas secara signifikan. Selain itu, XGBoost tetap lebih efisien dibandingkan pendekatan deep learning seperti CNN, karena tidak memerlukan pelatihan berbasis GPU dan waktu komputasi yang besar. Dengan demikian, model ini sangat potensial untuk diterapkan pada sistem klasifikasi citra berbasis perangkat dengan sumber daya terbatas.



Gambar 6. Kurva Precision-Recall

Untuk mengevaluasi lebih lanjut aspek presisi dan sensitivitas, digunakan Precision-Recall (PR) Curve. Seluruh kelas menghasilkan kurva PR yang tinggi, mendekati nilai sempurna dengan AUC hampir 1.0. Kurva ini memperlihatkan bahwa model dapat menjaga keseimbangan antara tingkat presisi dan recall yang baik, terutama pada kondisi data yang telah diseimbangkan sebelumnya. Ini penting karena pada klasifikasi multikelas, tidak hanya akurasi yang diperhitungkan, tetapi juga kemampuan model mendeteksi kelas minoritas secara konsisten [11].

Secara keseluruhan, pendekatan yang digunakan dalam penelitian ini terbukti efektif dan kompetitif dibandingkan studi sebelumnya. Penelitian oleh Ahmad et al. [24], menunjukkan pentingnya kombinasi fitur warna dan tekstur, sementara studi oleh Phadikar et al. [15] memperkuat peran XGBoost sebagai model yang akurat dalam klasifikasi penyakit tanaman. Dengan mengintegrasikan preprocessing yang cermat dan pemilihan fitur yang relevan, penelitian ini memberikan kontribusi nyata terhadap pengembangan sistem diagnosis penyakit tanaman yang efisien dan terjangkau.

Dari sisi efisiensi komputasi, model XGBoost menunjukkan keunggulan signifikan dibandingkan model berbasis CNN. Waktu pelatihan rata-rata XGBoost hanya membutuhkan beberapa menit pada CPU konvensional, sedangkan model CNN umumnya memerlukan waktu

pelatihan lebih lama dan sumber daya GPU untuk mencapai performa serupa [20], [15], [24]. Selain itu, waktu inferensi XGBoost relatif cepat karena model berbasis pohon keputusan mampu melakukan prediksi dengan kompleksitas komputasi yang rendah. Hal ini menjadikan XGBoost lebih efisien untuk diterapkan pada sistem klasifikasi citra di lingkungan dengan sumber daya terbatas.

Meskipun akurasi model mencapai nilai tinggi, potensi overfitting diminimalkan melalui penerapan validasi silang (3-fold cross-validation) selama proses tuning hiperparameter serta penggunaan PCA untuk reduksi dimensi dan SMOTE untuk penyeimbangan data.

IV. KESIMPULAN

Penelitian ini membuktikan bahwa kombinasi ekstraksi fitur klasik berupa tekstur (GLCM), warna, dan bentuk, yang dipadukan dengan reduksi dimensi PCA, penyeimbangan data melalui SMOTE, serta klasifikasi menggunakan XGBoost yang dioptimasi, mampu memberikan performa tinggi pada tugas klasifikasi multikelas penyakit daun tomat. Hasil eksperimen pada dataset PlantVillage yang terdiri atas 10 kelas penyakit menunjukkan capaian kinerja yang konsisten dan unggul, dengan akurasi 97.63%, presisi makro 0.98, recall makro 0.98, serta F1-score makro 0.98. Evaluasi melalui confusion matrix juga menegaskan bahwa sebagian besar kelas dapat dikenali dengan baik, dengan tingkat kesalahan klasifikasi relatif rendah meskipun terdapat kesamaan visual antarpenyakit tertentu.

Secara metodologis, penelitian ini menegaskan pentingnya integrasi fitur visual buatan dengan teknik praproses modern seperti PCA dan SMOTE untuk menghasilkan model yang tidak hanya akurat, tetapi juga efisien secara komputasi. Pendekatan ini menjadi bukti bahwa metode non-deep learning tetap memiliki relevansi tinggi ketika dirancang dengan optimalisasi yang tepat, terutama pada skenario dunia nyata yang membutuhkan sistem diagnosis cepat dengan keterbatasan sumber daya perangkat keras. Dengan demikian, penelitian ini menawarkan alternatif yang lebih praktis dan terjangkau dibandingkan pendekatan deep learning penuh, khususnya dalam konteks pertanian digital di lapangan.

Implikasi praktis dari hasil ini sangat luas, mulai dari potensi integrasi ke dalam perangkat edge seperti aplikasi mobile berbasis kamera hingga sistem IoT pertanian yang dapat melakukan diagnosis penyakit tanaman secara real-time tanpa bergantung pada infrastruktur komputasi berat. Ke depan, penelitian ini dapat dikembangkan lebih lanjut dengan menggabungkan fitur spasial lanjutan, memanfaatkan data multispektral atau hiperspektral, serta menguji generalisasi model pada berbagai jenis tanaman dan kondisi lingkungan yang berbeda. Hal ini diharapkan dapat memperkuat kontribusi penelitian ini dalam mendukung visi pertanian presisi berbasis kecerdasan buatan yang berkelanjutan, adaptif, dan inklusif.

DAFTAR PUSTAKA

- [1] P. S. Kanda, K. Xia, A. Kyslytsna, and E. O. Owoola, "Tomato Leaf Disease Recognition on Leaf Images Based on Fine-Tuned Residual Neural Networks," *Plants*, vol. 11, no. 21, Nov. 2022, doi: 10.3390/plants11212935.
- [2] S. Panno *et al.*, "A review of the most common and economically important diseases that undermine the cultivation of tomato crop in the mediterranean basin," *Agronomy*, vol. 11, no. 11, pp. 1–45, 2021, doi: 10.3390/agronomy11112188.
- [3] M. T. Rahman, S. D. Dipto, I. J. June, A. Momin, and M. R. Al Mamun, "Machine Learning-based Disease Classification in Tomato (*Solanum lycopersicum*) Plants," *J. Keteknikan Pertan. Trop. dan Biosist.*, vol. 12, no. 3, pp. 151–160, Dec. 2024, doi: 10.21776/ub.jkptb.2024.012.03.01.
- [4] C. Nyasulu *et al.*, "A comparative study of Machine Learning-based classification of Tomato fungal diseases: Application of GLCM texture features," *Heliyon*, vol. 9, no. 11, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21697.
- [5] M. Altalak, M. A. Uddin, A. Alajmi, and A. Rizg, "A Hybrid Approach for the Detection and Classification of Tomato Leaf Diseases," *Appl. Sci.*, vol. 12, no. 16, Aug. 2022, doi: 10.3390/app12168182.
- [6] A. Khan, U. Nawaz, L. Kshetrimayum, L. Seneviratne, and I. Hussain, "Early and Accurate Detection of Tomato Leaf Diseases Using TomFormer," *ArXiv*, Dec. 2023, [Online]. Available: <http://arxiv.org/abs/2312.16331>
- [7] A. Chelladurai, D. P. Manoj Kumar, S. S. Askar, and M. Abouhawwash, "Classification of tomato leaf disease using Transductive Long Short-Term Memory with an attention mechanism," *Front. Plant Sci.*, vol. 15, 2024, doi: 10.3389/fpls.2024.1467811.
- [8] O. Attallah, "Tomato Leaf Disease Classification via Compact Convolutional Neural Networks with Transfer Learning and Feature Selection," *Horticulturae*, vol. 9, no. 2, Feb. 2023, doi: 10.3390/horticulturae9020149.
- [9] Y. Borhani, J. Khoramdel, and E. Najafi, "A deep learning based approach for automated plant disease classification using vision transformer," *Sci. Rep.*, vol. 12, no. 1, pp. 1–10, 2022, doi: 10.1038/s41598-022-15163-0.
- [10] H. Ghosh *et al.*, "Advanced neural network architectures for tomato leaf disease diagnosis in precision agriculture," *Discov. Sustain.*, vol. 6, no. 1, 2025, doi: 10.1007/s43621-025-01149-1.
- [11] E. Zhang, N. Zhang, F. Li, and C. Lv, "A lightweight dual-attention network for tomato leaf disease identification," *Front. Plant Sci.*, vol. 15, no. August, pp. 1–18, 2024, doi: 10.3389/fpls.2024.1420584.
- [12] M. S. A. M. Al-gaashani, F. Shang, M. Khayyat, M. S. A. Muthanna, and A. A. Abd El-Latif, "IET Image Processing - 2022 - Al-gaashani - Tomato leaf disease classification by exploiting transfer learning and feature.pdf," 2021.
- [13] L. Tan, J. Lu, and H. Jiang, "Tomato Leaf Diseases Classification Based on Leaf Images: A Comparison between Classical Machine Learning and Deep Learning Methods," *AgriEngineering*, vol. 3, no. 3, pp. 542–558, 2021, doi: 10.3390/agriengineering3030035.
- [14] A. Tabbakh and S. S. Barpanda, "Evaluation of Machine Learning Models for Plant Disease Classification Using Modified GLCM and Wavelet Based Statistical Features," *Trait. du Signal*, vol. 39, no. 6, pp. 1893–1905, 2022, doi: 10.18280/ts.390602.
- [15] K. Kishore Kumar and E. Kannan, "An Efficient Deep Neural Network for Disease Detection in Rice Plant Using XGBOOST Ensemble Learning Framework," *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 3, pp. 116–128, 2022.
- [16] S. U. Khan, A. Alsuhaibani, A. Alabduljabbar, F. Almarshad, Y. N. Altherwy, and T. Akram, *A review on automated plant disease detection: motivation, limitations, challenges, and recent advancements for future research*, vol. 37, no. 3. Springer International Publishing, 2025. doi: 10.1007/s44443-025-00040-3.
- [17] R. Khan, N. Ud Din, A. Zaman, and B. Huang, "Automated Tomato Leaf Disease Detection Using Image Processing: An SVM-Based Approach with GLCM and SIFT Features," *J. Eng. (United Kingdom)*, vol. 2024, 2024, doi: 10.1155/2024/9918296.
- [18] C. Cuenca-Romero, O. E. Apolo-Apolo, J. N. Rodríguez Vázquez, G. Egea, and M. Pérez-Ruiz, "Tackling unbalanced datasets for yellow and brown rust detection in wheat," *Front. Plant Sci.*, vol. 15, no. May, pp. 1–11, 2024, doi: 10.3389/fpls.2024.1392409.
- [19] W. Shafik, A. Tufail, C. Liyanage De Silva, and R. A. Awg Haji Mohd Apong, "A novel hybrid inception-xception convolutional neural network for efficient plant disease classification and detection," *Sci. Rep.*, vol. 15, no. 1, p. 3936, 2025, doi: 10.1038/s41598-024-82857-y.
- [20] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Aug, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
- [21] Q. H. Cap, H. Uga, S. Kagiwada, and H. Iyatomi, "LeafGAN: An Effective Data Augmentation Method for Practical Plant Disease Diagnosis," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 2, pp. 1258–1267, 2022, doi: 10.1109/TASE.2020.3041499.
- [22] M. Bhandari, T. B. Shahi, A. Neupane, and K. B. Walsh, "BotanicX-AI: Identification of Tomato Leaf Diseases Using an Explanation-Driven Deep-Learning Model," *J. Imaging*, vol. 9, no. 2, 2023, doi: 10.3390/jimaging9020053.
- [23] M. H. Saleem, S. Khanchi, J. Potgieter, and K. M. Arif, "Image-Based Plant Disease Identification by Deep," *Plants*, vol. 9, no. 1451, pp. 1–23, 2020.
- [24] N. Ahmad, H. M. S. Asif, G. Saleem, M. U. Younus, S. Anwar, and M. R. Anjum, "Leaf Image-Based Plant Disease Identification Using Color and Texture Features," *Wirel. Pers. Commun.*, vol. 121, no. 2, pp. 1139–1168, 2021, doi: 10.1007/s11277-021-09054-2.