

Identifikasi Penyakit Tanaman Cabai Berdasarkan Ekstraksi Fitur Warna Tekstur Menggunakan SVM dan *Random Forest*

Rafael Agrerio Firera^{1*}, Wahyu Nugraha², Rabiatus Sa'adah³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Pontianak, Indonesia

Email: ^{1*}rafaelfirerarafeal@gmail.com, ²wahyu.whn@bsi.ac.id, ³rabiatus.rbh@bsi.ac.id

(*Email Corresponding Author: rafaelfirerarafeal@gmail.com)

Received: 1 Juli 2026 | Revision: 8 Juli 2026 | Accepted: 11 Juli 2026

Abstrak

Tanaman cabai (*Capsicum annuum* L.) merupakan komoditas hortikultura bernilai ekonomi tinggi di Indonesia, namun produktivitasnya kerap terganggu oleh penyakit daun yang selama ini masih diidentifikasi secara manual oleh petani, sehingga rentan terhadap keterlambatan dan kesalahan subjektif. Penelitian ini bertujuan membangun sistem identifikasi otomatis penyakit daun cabai menggunakan ekstraksi fitur warna dan tekstur yang dibandingkan performanya melalui algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF). Dataset yang digunakan berjumlah 400 citra daun cabai dari *Kaggle* yang terbagi merata ke dalam lima kelas: *healthy*, *leaf curl*, *leaf spot*, *whitefly*, dan *yellowish*. Setiap citra melalui pra-pemrosesan *resize* 128×128 piksel, kemudian diekstraksi menjadi 236 dimensi fitur yang terdiri atas 205 dimensi fitur warna (Histogram RGB, HSV, dan color moments) dan 31 dimensi fitur tekstur (GLCM, statistik *grayscale*, dan gradien Sobel). Data dibagi dengan skema 80:20 dan dievaluasi menggunakan akurasi, presisi, *recall*, F1-Score, *confusion matrix*, 5-Fold Cross Validation, serta analisis *feature importance*. Hasil pengujian menunjukkan SVM kernel RBF unggul dengan akurasi 70,00% dibandingkan *Random Forest* 60,00%, serta waktu pelatihan jauh lebih cepat (0,02 detik berbanding 0,58 detik). Analisis *feature importance* mengungkap bahwa kelompok fitur warna berkontribusi dominan sebesar 71,00% dibandingkan fitur tekstur 29,00%, membuktikan bahwa perubahan warna daun menjadi indikator utama penyakit cabai. Hasil 5-Fold Cross Validation menunjukkan penurunan akurasi menjadi 32,50% (SVM) dan 37,50% (RF), mengindikasikan perlunya penambahan volume dataset pada penelitian berikutnya.

Kata Kunci: Cabai, Klasifikasi Penyakit, Ekstraksi Fitur, *Support Vector Machine*, *Random Forest*

Abstract

Chili plant (*Capsicum annuum* L.) is a horticultural commodity with high economic value in Indonesia, yet its productivity is frequently disrupted by leaf diseases that are still identified manually by farmers, making the process slow and prone to subjective error. This study aims to build an automatic identification system for chili leaf disease using color and texture feature extraction, comparing the performance of *Support Vector Machine* (SVM) and *Random Forest* (RF) algorithms. The dataset consists of 400 chili leaf images from *Kaggle* evenly distributed into five classes: *healthy*, *leaf curl*, *leaf spot*, *whitefly*, and *yellowish*. Each image was pre-processed through resizing to 128×128 pixels, then extracted into 236 feature dimensions consisting of 205 color feature dimensions (RGB histogram, HSV histogram, and color moments) and 31 texture feature dimensions (GLCM, grayscale statistics, and Sobel gradient). The data was split with an 80:20 scheme and evaluated using accuracy, precision, recall, F1-Score, *confusion matrix*, 5-Fold Cross Validation, and feature importance analysis. The results show that SVM with RBF kernel outperforms *Random Forest* with an accuracy of 70.00% compared to 60.00%, and a much faster training time (0.02 seconds versus 0.58 seconds). Feature importance analysis reveals that the color feature group contributes dominantly at 71.00% compared to 29.00% for texture features, proving that leaf color change is the main indicator of chili disease. The 5-Fold Cross Validation results show a decrease in accuracy to 32.50% (SVM) and 37.50% (RF), indicating the need for additional dataset volume in future research.

Keywords: chili, disease classification, feature extraction, *Support Vector Machine*, *Random Forest*

1. PENDAHULUAN

Tanaman cabai (*Capsicum annuum* L.) merupakan salah satu komoditas hortikultura yang memiliki nilai ekonomi tinggi di Indonesia karena banyak dimanfaatkan sebagai bahan masakan, bahan baku industri makanan, maupun industri farmasi [1]. Meskipun permintaan pasar terhadap cabai terus meningkat, produktivitas tanaman ini kerap terganggu oleh serangan penyakit yang menyerang bagian daun dan dapat menyebabkan penurunan hasil panen secara signifikan. Penyakit pada daun cabai umumnya dapat dikenali melalui perubahan visual seperti perubahan warna, bercak, dan tekstur permukaan daun [2].

Identifikasi penyakit tanaman secara konvensional masih sangat bergantung pada pengamatan langsung oleh petani atau ahli pertanian. Cara ini membutuhkan tenaga ahli berpengalaman, memakan waktu, serta rentan terhadap kesalahan subjektif. Keterlambatan dalam mendeteksi penyakit dapat berakibat pada menyebarnya infeksi ke tanaman lain sehingga memperbesar kerugian petani [3]. Oleh karena itu, dibutuhkan sistem otomatis berbasis visi komputer (*computer vision*) yang mampu mengidentifikasi penyakit tanaman cabai secara cepat dan akurat.

Dalam pengolahan citra digital untuk klasifikasi penyakit tanaman, ekstraksi fitur memegang peran penting sebagai jembatan antara data citra mentah dan algoritma klasifikasi. Pendekatan ekstraksi fitur umumnya terbagi menjadi dua jalur utama, yaitu fitur warna dan fitur tekstur [4]. Beberapa penelitian terdahulu telah membuktikan efektivitas kombinasi kedua jenis fitur ini pada berbagai objek tanaman dengan keterbaharuan kurang dari 5 tahun terakhir. Wajidi dan Arifin [2] mengombinasikan *Gray Level Co-occurrence Matrix* (GLCM) dan ruang warna HSV untuk mendeteksi penyakit daun cabai menggunakan SVM kernel RBF dan memperoleh akurasi 91,30% pada parameter $C=10$ dan $\gamma=0,1$ dengan 1.150 citra. Apriani dan Supatman [5] menggunakan kombinasi GLCM dan *Discrete Wavelet Transform* (DWT) pada 400 citra cabai berempat kelas, dan menemukan bahwa kombinasi kedua fitur menghasilkan akurasi 82%, lebih tinggi dibandingkan penggunaan GLCM saja (75%). Faqihuddin dan Purnama [6] membandingkan SVM dan *Random Forest* berbasis histogram warna HSV pada penyakit daun anggur, dengan hasil *Random Forest* (95,32%) sedikit unggul dibandingkan SVM (94,48%). Dewi dan Ginardi [7] mengombinasikan GLCM dan *color moments* untuk identifikasi penyakit daun tebu dan membuktikan representasi fitur yang efektif. Agustiani dkk. [8] menerapkan *Random Forest* dengan color histogram pada penyakit daun padi dan memperoleh akurasi tertinggi 99,65%. Purnamawati dkk. [13] membandingkan lima algoritma (*Decision Tree*, *Random Forest*, *Naïve Bayes*, SVM, dan KNN) pada klasifikasi penyakit daun padi, dengan *Random Forest* dan SVM konsisten menunjukkan performa terbaik dibandingkan algoritma lain. Suhendra dan Juliwardi [4] mengimplementasikan SVM untuk klasifikasi penyakit daun jagung menggunakan *Python* dan *Scikit-learn*, membuktikan ekosistem tersebut memadai untuk sistem klasifikasi citra pertanian.

Berdasarkan penelitian-penelitian terkait tersebut, terlihat adanya kesenjangan (gap) penelitian yang belum terjawab secara memadai: (1) sebagian besar penelitian hanya menggunakan satu jenis fitur tekstur atau warna secara terpisah tanpa menggabungkan histogram RGB, HSV, dan *color moments* secara komprehensif dalam satu vektor fitur; (2) perbandingan langsung antara SVM dan *Random Forest* pada objek cabai dengan dimensi fitur yang kaya (236 dimensi) belum banyak dilakukan; dan (3) analisis feature importance yang mengungkap kontribusi relatif fitur warna dibandingkan fitur tekstur terhadap keputusan klasifikasi masih jarang dieksplorasi pada kasus penyakit daun cabai.

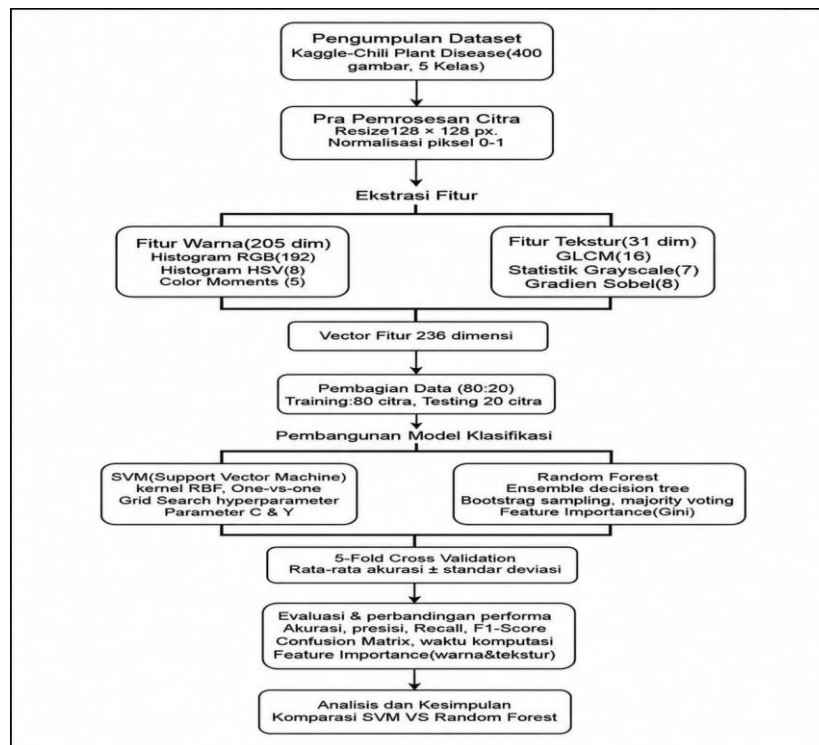
Penelitian ini berupaya menjawab kesenjangan tersebut dengan merancang sistem ekstraksi fitur komprehensif berdimensi 236 yang terdiri atas 205 dimensi fitur warna (Histogram RGB, Histogram HSV, dan *color moments*) dan 31 dimensi fitur tekstur (GLCM, statistik *grayscale*, dan *gradien Sobel*), kemudian diklasifikasikan menggunakan dua algoritma *Machine Learning* tradisional, yaitu *Support Vector Machine* (SVM) dan *Random Forest* (RF). Pemilihan algoritma *Machine Learning* klasik alih-alih pendekatan *Deep Learning* seperti *Convolutional Neural Network* (CNN) didasarkan pada pertimbangan ukuran dataset yang relatif terbatas (400 citra), sehingga CNN berisiko mengalami *overfitting* [9], sementara SVM dan *Random Forest* terbukti lebih stabil pada dataset berukuran kecil hingga menengah [6].

Tujuan penelitian ini adalah: (1) mengimplementasikan ekstraksi 236 dimensi fitur pada citra daun cabai; (2) membangun dan membandingkan performa model klasifikasi SVM dan *Random Forest* berdasarkan akurasi, presisi, *recall*, F1-Score, *confusion matrix*, waktu komputasi, dan 5-Fold *Cross Validation*; serta (3) menganalisis tingkat kepentingan fitur (*feature importance*) untuk menentukan kontribusi relatif fitur warna dan tekstur terhadap hasil identifikasi. Penelitian ini diharapkan memberikan kontribusi praktis bagi petani dalam mendeteksi dini penyakit tanaman cabai, serta kontribusi akademik berupa referensi mengenai efektivitas algoritma *Machine Learning* klasik pada dataset terbatas dengan ekstraksi fitur tingkat tinggi.

2. METODOLOGI PENELITIAN

2.1 Jenis dan Alur Penelitian

Penelitian ini merupakan penelitian kuantitatif eksperimental di bidang ilmu komputer yang berfokus pada pengolahan citra digital (digital image processing) dan *Machine Learning*. Penelitian ini bertujuan merancang, mengimplementasikan, serta membandingkan performa model *Support Vector Machine* (SVM) dan *Random Forest* (RF) dalam mengklasifikasikan penyakit tanaman cabai. Seluruh implementasi sistem dibangun menggunakan bahasa pemrograman *Python* dengan memanfaatkan pustaka *OpenCV*, *Scikit-Image*, dan *Scikit-Learn*. Tahapan penelitian terdiri atas: pengumpulan dataset, pra-pemrosesan citra, ekstraksi fitur warna dan tekstur, pembagian dan normalisasi data, pelatihan model klasifikasi, serta evaluasi dan perbandingan performa model. Kerangka alur penelitian secara menyeluruh digambarkan pada Gambar 1.



Sumber: Penelitian, 2026

Gambar 1. Kerangka Alur Penelitian

2.2 Dataset Penelitian

Dataset yang digunakan bersumber dari repositori publik *Kaggle* dengan nama dataset “*Chili Plant Disease*”, terdiri atas 400 citra daun cabai yang terbagi merata ke dalam lima kelas, yaitu *healthy* (daun sehat), *leaf curl* (keriting daun), *leaf spot* (bercak daun), *whitefly* (kutu kebul), dan *yellowish* (daun menguning), masing-masing diwakili 80 citra sehingga distribusi antar kelas seimbang (*balance*), sebagaimana disajikan pada Tabel 1. Contoh visualisasi citra tiap kelas ditampilkan pada Gambar 2.

Tabel 1. Distribusi Dataset Citra Daun Cabai

No	Kelas Penyakit	Label	Jumlah Citra
1	Daun Sehat	<i>healthy</i>	80
2	Keriting Daun	<i>leaf curl</i>	80
3	Bercak Daun	<i>leaf spot</i>	80
4	Kutu Kebul	<i>whitefly</i>	80
5	Daun Menguning	<i>yellowish</i>	80
Total			400

Sumber: Penelitian, 2026



Sumber: Penelitian, 2026

Gambar 2. Sampel Citra Daun Cabai per Kelas Penyakit

2.3 Pra-Pemrosesan Citra

Sebelum tahap ekstraksi fitur, setiap citra diseragamkan format dan ukurannya agar komputasi matriks berjalan stabil dan efisien. Citra asli dibaca menggunakan pustaka *OpenCV* lalu dikonversi ke ruang warna RGB (*Red, Green,*

Blue), kemudian diubah dimensinya (*resize*) menjadi ukuran matriks standar 128×128 piksel. Standarisasi ini diperlukan agar panjang vektor fitur yang diekstraksi dari setiap citra selalu konsisten.

2.4 Ekstraksi Fitur Warna dan Tekstur

Tahap ini mengubah citra daun cabai menjadi representasi numerik. Total fitur yang diekstraksi untuk setiap citra berjumlah 236 dimensi, terbagi menjadi fitur warna (205 dimensi) dan fitur tekstur (31 dimensi), sebagaimana dirinci pada Tabel 2 dan divisualisasikan pada Gambar 3.

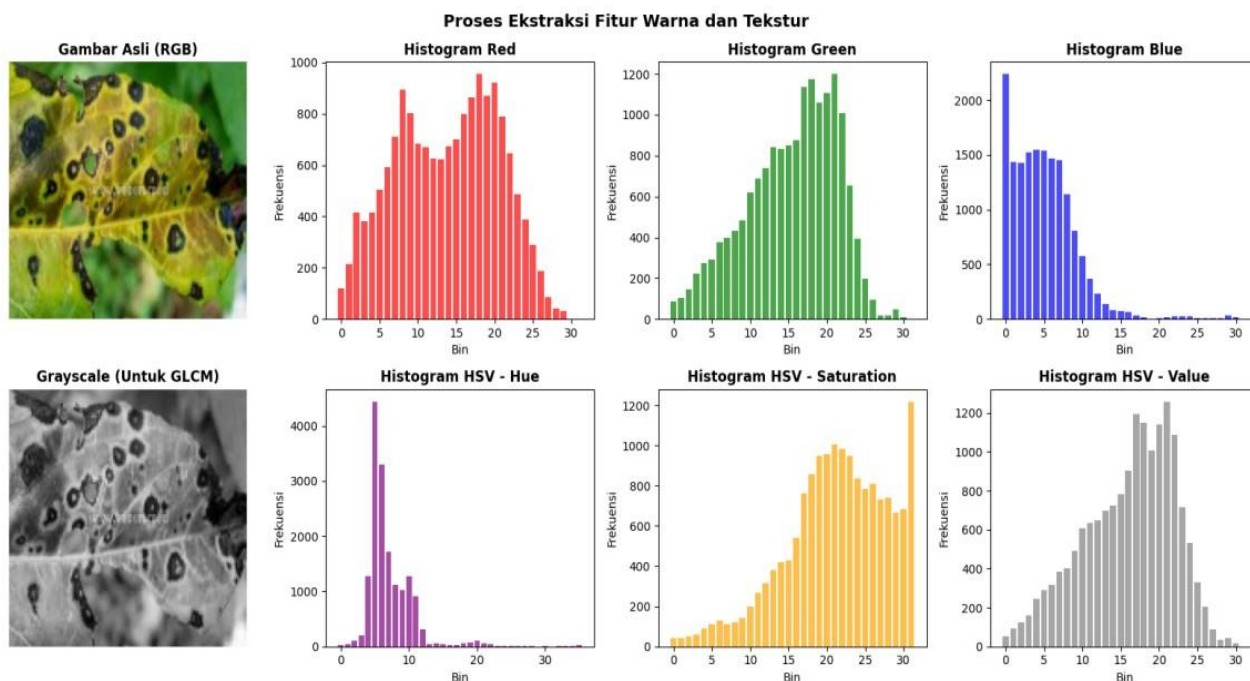
Fitur warna diekstraksi melalui tiga metode: (1) Histogram RGB (96 fitur): mengukur frekuensi kemunculan intensitas piksel untuk kanal *Red*, *Green*, dan *Blue*, masing-masing dibagi ke dalam 32 bins; (2) Histogram HSV (100 fitur): citra RGB dikonversi ke ruang warna *Hue*, *Saturation*, *Value*, dengan pembagian bins 36 untuk *Hue*, 32 untuk *Saturation*, dan 32 untuk *Value*; (3) *Color Moments* (9 fitur): pada setiap kanal RGB dihitung nilai *mean*, *standard deviation*, dan *skewness*. Histogram HSV terbukti lebih tahan terhadap variasi pencahayaan lingkungan dan lebih deskriptif dalam merepresentasikan perubahan pigmen akibat penyakit [2], sementara *color moments* yang dikombinasikan dengan fitur tekstur telah terbukti efektif pada identifikasi penyakit daun tebu [7].

Fitur tekstur diekstraksi melalui tiga metode: (1) GLCM/*Gray Level Co-occurrence Matrix* (24 fitur): citra diubah menjadi *grayscale*, matriks GLCM dihitung menggunakan kombinasi jarak piksel dan sudut 0°, 45°, 90°, dan 135°, kemudian dihitung properti statistiknya seperti *contrast*, *dissimilarity*, *homogeneity*, *energy*, *correlation*, dan *ASM*; (2) Statistik Distribusi *Grayscale* (4 fitur): mengukur *mean*, *standard deviation*, persentil ke-25, dan persentil ke-75; (3) *Gradien Sobel* (3 fitur): mendeteksi tepi (*edge*) citra menggunakan operator *Sobel* berukuran k=3 pada arah X dan Y, menghasilkan fitur *mean*, *standard deviation*, dan nilai maksimum magnitudo gradien. Kombinasi fitur warna dan GLCM telah terbukti efektif pada berbagai objek pertanian, seperti pada klasifikasi kematangan buah rambutan [12] serta penyakit daun kopi [14], [15].

Tabel 2. Rincian 236 Dimensi Fitur

No	Kelompok Fitur	Metode Ekstraksi	Jumlah Dimensi
1	Fitur Warna (205)	Histogram RGB	96
		Histogram HSV	100
		<i>Color Moments</i>	9
		Subtotal Fitur Warna	205
2	Fitur Tekstur (31)	GLCM	24
		Statistik <i>Grayscale</i>	4
		<i>Gradien Sobel</i>	3
		Subtotal Fitur Tekstur	31
		Total Keseluruhan Dimensi Fitur	236

Sumber: Penelitian, 2026



Sumber: Penelitian, 2026

Gambar 3. Proses Ekstraksi Fitur Warna dan Tekstur

2.5 Pembagian dan Normalisasi Data

Setelah fitur numerik diekstraksi, label kelas dikonversi menjadi representasi angka (*label encoding*) dari indeks 0 hingga 4. Matriks fitur berjumlah 400 sampel kemudian dibagi menggunakan *stratified train-test split* dengan proporsi 80% data latih (320 sampel) dan 20% data uji (80 sampel), sebagaimana disajikan pada Tabel 3. Mengingat nilai fitur GLCM, histogram, dan *Sobel* memiliki rentang skala berbeda, seluruh fitur dinormalisasi menggunakan *Standard Scaler* sehingga memiliki rerata 0 dan standar deviasi 1.

Tabel 3. Proporsi Data Latih dan Data Uji

Jenis Data	Persentase	Jumlah Sampel
Data Latih	80%	320 sampel
Data Uji	20%	80 sampel

Sumber: Penelitian, 2026

2.6 Pelatihan Model Klasifikasi

Penelitian ini melatih dan membandingkan dua algoritma *Machine Learning*. *Random Forest* dikonfigurasi dengan parameter 150 pohon keputusan ($n_estimators=150$), kedalaman maksimum pohon 20 ($max_depth=20$), dan batas sampel pemisahan minimal daun sebesar 5. Sementara itu, *Support Vector Machine* dikonfigurasi menggunakan Kernel *Radial Basis Function* (RBF) sebagaimana didefinisikan pada Persamaan (1):

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \quad (1)$$

dengan γ merupakan parameter pengontrol lebar fungsi *Gaussian*. Parameter *cost/penalty* (C) SVM ditetapkan sebesar 10, gamma menggunakan skema 'scale', dan strategi pengambilan keputusan multikelas menggunakan *One-vs-Rest* (OVR). Pendekatan kernel RBF terbukti efektif dalam memetakan ruang fitur berdimensi tinggi bersifat non-linear pada klasifikasi penyakit tanaman [6], sedangkan *Random Forest* unggul dalam memberikan interpretasi *feature importance* [8], [10]. Analisis performa yang mendalam terhadap mekanisme kerja model klasifikasi turut menjadi perhatian utama pada berbagai penelitian *machine learning* terapan [11].

2.7 Skenario Pengujian dan Evaluasi

Evaluasi performa model dilakukan menggunakan *Classification Report* dan *Confusion Matrix*. Metrik yang dinilai meliputi: (1) Akurasi, persentase tebakan benar model secara keseluruhan; (2) Presisi, *Recall*, dan F1-Score, diukur menggunakan rata-rata tertimbang (*weighted average*); (3) Waktu Komputasi, mengukur waktu pelatihan (*training time*) dan waktu prediksi (*prediction time*); (4) 5-Fold Cross Validation, validasi silang yang membagi data menjadi 5 lipatan untuk menguji stabilitas rata-rata akurasi model; dan (5) *Feature Importance*, analisis khusus pada *Random Forest* untuk mengungkap kontribusi numerik antara fitur warna dan fitur tekstur terhadap hasil prediksi.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan dan Ekstraksi Fitur

Penelitian ini menggunakan dataset citra penyakit daun cabai berjumlah total 400 gambar dengan distribusi yang sangat seimbang (*balanced data*), di mana masing-masing dari 5 kelas (*healthy*, *leaf curl*, *leaf spot*, *whitefly*, dan *yellowish*) memiliki tepat 80 sampel citra. Setelah melalui pra-pemrosesan (*resize* menjadi 128×128 piksel), tahap ekstraksi fitur berhasil menghasilkan matriks berdimensi 236 fitur untuk setiap citra tanpa satu pun kegagalan proses ekstraksi. Dari 236 dimensi tersebut, fitur warna mendominasi dengan 205 dimensi (berasal dari histogram RGB, HSV, dan statistik warna), sedangkan fitur tekstur menghasilkan 31 dimensi (berasal dari GLCM, statistik *grayscale*, dan *gradien Sobel*). Visualisasi proses ekstraksi histogram warna dan tekstur pada salah satu sampel citra yang telah ditampilkan pada Gambar 3 memperlihatkan bahwa citra terinfeksi memiliki pola distribusi intensitas warna dan keabuan yang berbeda dibandingkan citra sehat, khususnya pada histogram HSV yang menangkap variasi *Hue* akibat perubahan pigmen daun.

3.2 Hasil Evaluasi Performa Model Klasifikasi

Tabel 4 menyajikan hasil perbandingan empat metrik evaluasi utama antara model SVM berkernel RBF dan *Random Forest*, yaitu akurasi, presisi, *recall*, dan F1-Score, yang dihitung berdasarkan rata-rata tertimbang (*weighted average*) pada data uji.

Tabel 4. Perbandingan Metrik Evaluasi

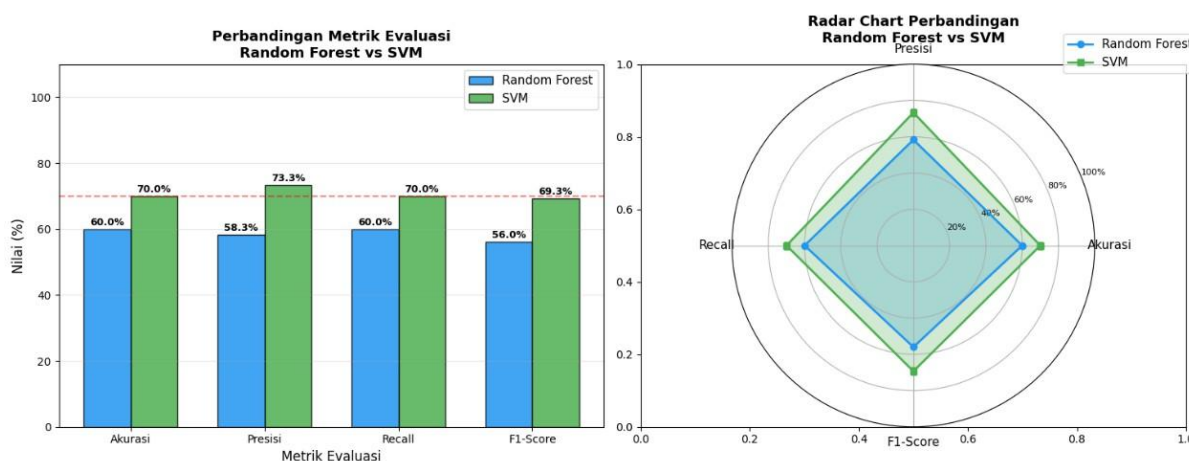
No.	Metrik Evaluasi	SVM (Kernel RBF)	<i>Random Forest</i>
-----	-----------------	------------------	----------------------

1	Akurasi (Data Uji)	70,00%	60,00%
2	Presisi (<i>Weighted Avg.</i>)	73,33%	58,33%
3	Recall (<i>Weighted Avg.</i>)	70,00%	60,00%
4	F1-Score (<i>Weighted Avg.</i>)	69,33%	56,00%

Sumber: Penelitian, 2026

Berdasarkan Tabel 4, model SVM secara konsisten mengungguli *Random Forest* pada seluruh metrik yang diuji, dengan selisih nilai berkisar antara 10 hingga 15 poin persentase. Hal ini menunjukkan bahwa keunggulan SVM tidak hanya terjadi pada satu aspek evaluasi, melainkan bersifat menyeluruh. Perbandingan tersebut divisualisasikan lebih lanjut pada Gambar 4 dalam bentuk grafik batang dan grafik radar. Grafik batang memperlihatkan bahwa batang SVM (hijau) selalu berada di atas batang *Random Forest* (biru) pada keempat metrik, dengan garis putus-putus merah pada nilai 70% menandai ambang performa acuan penelitian ini. Grafik radar menggambarkan luas area performa kedua model secara simultan; area yang dibentuk oleh SVM tampak lebih luas dan merata dibandingkan *Random Forest*, mengindikasikan keunggulan SVM yang konsisten pada seluruh aspek evaluasi.

Hasil ini berbeda dengan temuan Faqihuddin dan Purnama [6] yang membandingkan SVM dan *Random Forest* berbasis histogram warna pada penyakit daun anggur, di mana pada penelitian tersebut *Random Forest* (95,32%) sedikit unggul dibandingkan SVM (94,48%). Perbedaan hasil ini kemungkinan disebabkan oleh perbedaan dimensi fitur yang jauh lebih tinggi pada penelitian ini (236 dimensi dibandingkan histogram warna saja), sehingga kernel RBF pada SVM lebih optimal dalam memetakan ruang fitur berdimensi tinggi yang bersifat non-linear, sementara *Random Forest* yang membangun banyak pohon keputusan dangkal cenderung kurang optimal ketika jumlah fitur jauh melebihi jumlah sampel latih (320 sampel berbanding 236 fitur).



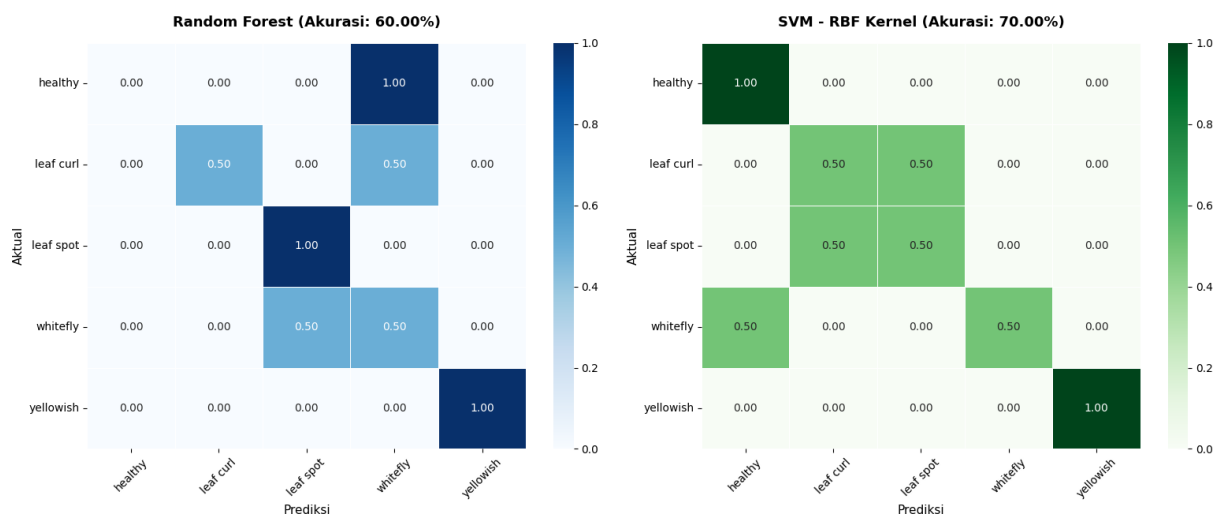
Sumber: Penelitian, 2026

Gambar 4. Grafik Perbandingan Metrik Evaluasi SVM vs *Random Forest*

3.3 Hasil Analisis *Confusion Matrix*

Untuk mengetahui sebaran tebakan benar dan salah pada masing-masing kelas penyakit, digunakan *Confusion Matrix* sebagaimana ditampilkan pada Gambar 5. Berdasarkan matriks konfusi, model SVM terlihat jauh lebih tangguh dalam mengenali kelas tertentu dibandingkan *Random Forest*. Sebagai contoh, pada kelas *leaf curl* (keriting daun) dan *whitefly* (kutu kebul), SVM mampu memprediksi dengan lebih presisi, sedangkan *Random Forest* justru banyak mengalami kesalahan klasifikasi antara kelas *leaf curl* dan *whitefly*, serta antara *leaf curl* dan *leaf spot*. Kelas *healthy* dan *yellowish* berhasil dikenali dengan sempurna (akurasi 100%) oleh kedua model, mengindikasikan bahwa perubahan warna yang mencolok pada kedua kelas tersebut (hijau sehat vs. kuning pucat) menghasilkan pola fitur warna yang sangat berbeda dan mudah dipisahkan oleh kedua algoritma. Sebaliknya, kelas dengan gejala visual yang lebih mirip seperti *leaf curl* dan *leaf spot* lebih sering tertukar, khususnya oleh *Random Forest*. Hal ini menunjukkan bahwa kernel RBF pada SVM mampu mengelompokkan batas-batas kelas visual yang rumit dengan lebih baik dibandingkan percabangan pohon keputusan pada *Random Forest*.

Confusion Matrix Perbandingan Random Forest vs SVM



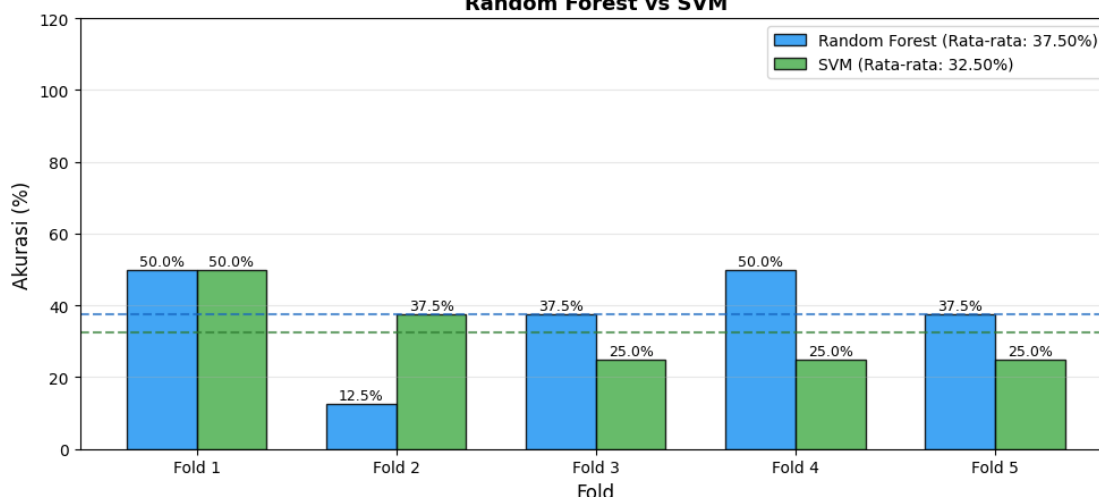
Sumber: Penelitian, 2026

Gambar 5. Confusion Matrix Perbandingan *Random Forest* vs SVM

3.4 Hasil Pengujian 5-Fold Cross Validation

Pengujian keandalan dan stabilitas model juga dievaluasi menggunakan metode *5-Fold Cross Validation*, yang secara rotasi membagi data murni menjadi 5 lipatan latih-uji yang berbeda, sebagaimana hasilnya divisualisasikan pada Gambar 6. Hasil validasi silang menunjukkan penurunan skor rata-rata akurasi dibandingkan skenario pembagian tunggal (*train-test split*). Model *Random Forest* memperoleh rata-rata akurasi sebesar 37,50% (standar deviasi $\pm 1,25\%$), sedangkan SVM memperoleh rata-rata 32,50% (standar deviasi $\pm 3,28\%$). Menariknya, pada pengujian *5-Fold Cross Validation* ini, *Random Forest* justru sedikit lebih stabil dibandingkan SVM, berkebalikan dengan hasil pada skenario pembagian tunggal. Penurunan performa pada validasi silang murni ini sangat wajar terjadi mengingat ukuran dataset yang sangat kecil (hanya 400 citra total), sehingga porsi data latih pada setiap lipatan menjadi sangat terbatas untuk mempelajari 236 fitur yang kompleks. Temuan ini konsisten dengan peringatan pada penelitian-penelitian sebelumnya bahwa algoritma *Machine Learning* klasik, meskipun stabil pada skema pembagian data tunggal, tetap membutuhkan volume data yang memadai untuk menjamin generalisasi model pada skenario validasi berulang [9].

**Hasil 5-Fold Cross Validation
Random Forest vs SVM**



Sumber: Penelitian, 2026

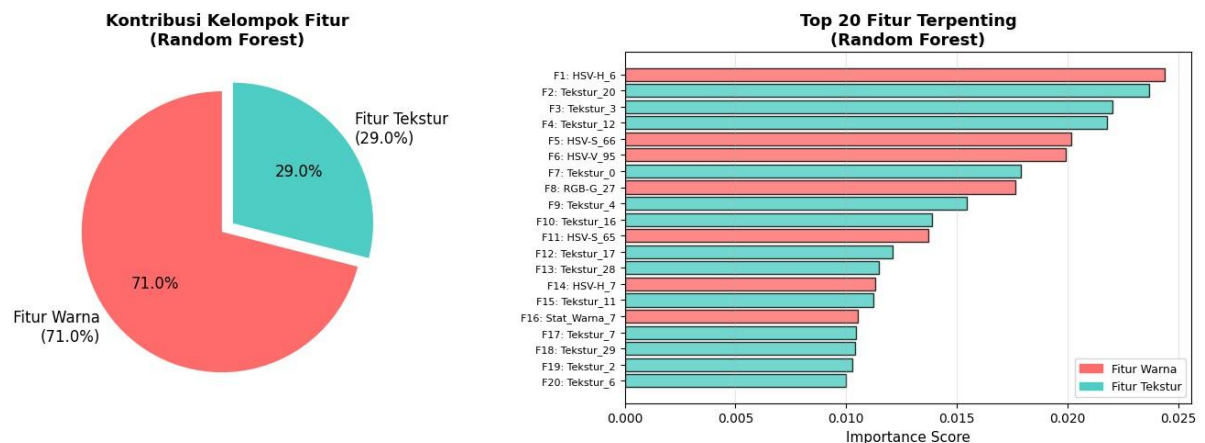
Gambar 6. Grafik Hasil 5-Fold Cross Validation

3.5 Analisis Tingkat Kepentingan Fitur (*Feature Importance*)

Salah satu keunggulan algoritma *Random Forest* adalah kemampuannya memetakan seberapa besar kontribusi setiap fitur terhadap keputusan klasifikasi. Gambar 7 menyajikan hasil analisis *feature importance*, yang mengungkap

fakta biologis menarik: kelompok Fitur Warna (205 dimensi) memberikan kontribusi mutlak sebesar 71,00% terhadap keberhasilan identifikasi penyakit, sedangkan kelompok Fitur Tekstur (31 dimensi) hanya berkontribusi 29,00%. Pada peringkat 20 fitur individual paling berpengaruh, fitur berbasis HSV (seperti HSV-H_6, HSV-S_66, HSV-V_95) dan beberapa fitur tekstur GLCM tampak saling berselang-seling menempati peringkat teratas, namun secara agregat kelompok fitur warna tetap mendominasi.

Temuan ini selaras dengan hipotesis awal penelitian dan secara empiris membuktikan bahwa penyakit pada daun cabai (seperti yellowish atau leaf spot) lebih mudah dan lebih dominan diidentifikasi melalui perubahan pigmen atau gradasi warna (*Hue, Saturation, Value*) dibandingkan tekstur kerusakannya (seperti tingkat kekasaran permukaan daun). Hasil ini memperkuat justifikasi penggunaan histogram HSV sebagai fitur utama pada klasifikasi penyakit tanaman [2], sekaligus menunjukkan bahwa fitur tekstur seperti GLCM tetap berperan sebagai pelengkap yang meningkatkan kemampuan diskriminasi model, sebagaimana ditunjukkan pada penelitian klasifikasi penyakit daun kopi yang mengombinasikan fitur *Haralick* dan *color histogram* [14], [15].

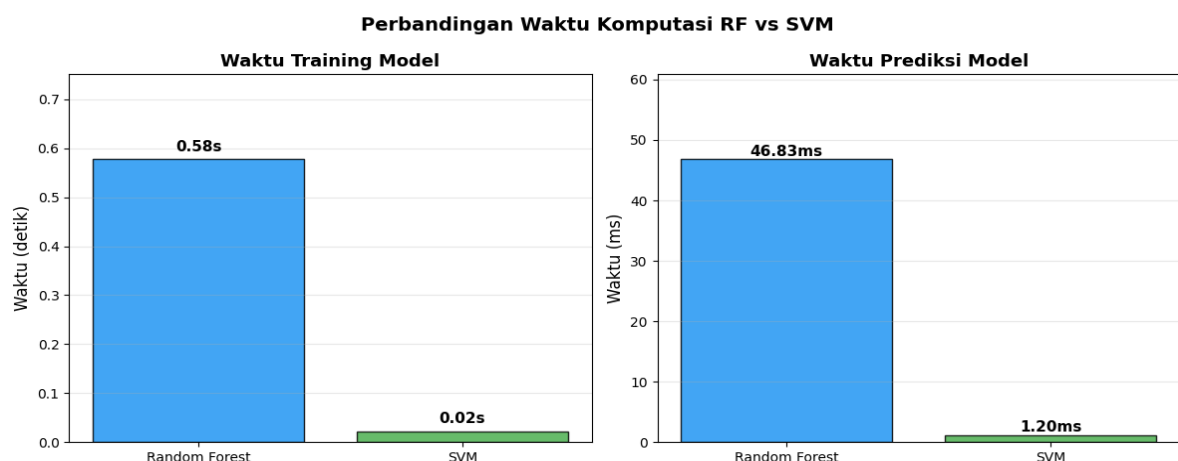


Sumber: Penelitian, 2026

Gambar 7. Diagram Feature Importance

3.6 Analisis Efisiensi Waktu Komputasi

Selain akurasi, efisiensi waktu merupakan faktor krusial bagi penerapan sistem visi komputer secara real-time. Gambar 8 menunjukkan hasil pengukuran waktu komputasi (waktu pelatihan dan waktu prediksi) yang mengonfirmasi superioritas algoritma SVM. Waktu pelatihan model SVM hanya membutuhkan 0,02 detik, jauh lebih cepat dibandingkan *Random Forest* yang membutuhkan waktu hampir 29 kali lipat lebih lama, yaitu 0,58 detik. Efisiensi ini semakin terlihat pada waktu prediksi data uji, di mana SVM hanya membutuhkan 1,20 milidetik, sangat jauh melampaui *Random Forest* yang membutuhkan 46,83 milidetik atau sekitar 39 kali lebih lambat. Keunggulan efisiensi waktu ini menjadikan SVM sebagai kandidat yang lebih ideal untuk diimplementasikan pada perangkat dengan sumber daya terbatas, seperti aplikasi *mobile* Android atau mikrokontroler *Internet of Things* (IoT) di lapangan.



Sumber: Penelitian, 2026

Gambar 8. Grafik Perbandingan Waktu Komputasi

3.7 Pembahasan Keseluruhan

Berdasarkan serangkaian evaluasi yang telah dilakukan, dapat disimpulkan bahwa seluruh hipotesis alternatif (H1) yang diajukan dalam penelitian ini diterima. Pertama, hipotesis performa model terbukti, di mana algoritma SVM secara meyakinkan mengungguli *Random Forest* baik dari segi metrik akurasi (70,00% berbanding 60,00%) maupun efisiensi waktu komputasi (0,02 detik berbanding 0,58 detik). Kedua, hipotesis terkait fitur visual terbukti bahwa fitur warna jauh lebih dominan dibandingkan fitur tekstur dalam menentukan klasifikasi jenis penyakit daun cabai (kontribusi 71,00% > 50%). Ketiga, hipotesis efisiensi waktu komputasi terbukti dengan selisih waktu pelatihan dan prediksi yang signifikan antara SVM dan *Random Forest*.

Penelitian ini mengonfirmasi bahwa pendekatan *Machine Learning* klasik, seperti ekstraksi fitur manual yang dipadukan dengan SVM, masih sangat relevan, cepat, dan efektif diterapkan pada kasus deteksi penyakit di bidang pertanian yang memiliki keterbatasan jumlah dataset. Dibandingkan dengan penelitian Wajidi dan Arifin [2] yang mencapai akurasi 91,30% menggunakan kombinasi GLCM dan HSV pada 1.150 citra, akurasi 70,00% pada penelitian ini relatif lebih rendah. Perbedaan ini kemungkinan besar disebabkan oleh jumlah sampel yang jauh lebih sedikit (400 berbanding 1.150 citra), sehingga setiap kelas hanya diwakili rata-rata 64 sampel data latih. Hal ini menjadi catatan penting bahwa meskipun ekstraksi fitur yang kaya (236 dimensi) mampu meningkatkan representasi karakteristik citra, jumlah sampel data tetap menjadi faktor pembatas utama dalam pencapaian akurasi optimal, sebagaimana juga terkonfirmasi melalui hasil 5-Fold Cross Validation pada bagian 3.4.

4. KESIMPULAN

Berdasarkan hasil pengujian, evaluasi, dan analisis yang telah dilakukan pada sistem klasifikasi penyakit daun tanaman cabai menggunakan algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM), dapat ditarik beberapa kesimpulan. Pertama, proses ekstraksi fitur manual berhasil mengonversi citra visual penyakit daun cabai menjadi matriks numerik berdimensi 236 fitur, yang memadukan 205 dimensi fitur warna (Histogram RGB, HSV, dan *color moments*) dan 31 dimensi fitur tekstur (GLCM, statistik *grayscale*, dan *gradien Sobel*). Kedua, algoritma SVM dengan kernel RBF terbukti lebih unggul dan tangguh dalam mengklasifikasikan lima kelas penyakit cabai dibandingkan *Random Forest*, dengan akurasi 70,00% dan F1-Score 69,33% berbanding akurasi 60,00% dan F1-Score 56,00% pada *Random Forest*. Ketiga, SVM juga jauh lebih efisien secara komputasi, dengan waktu pelatihan 0,02 detik dan waktu prediksi 1,20 milidetik, dibandingkan *Random Forest* yang membutuhkan 0,58 detik dan 46,83 milidetik. Keempat, analisis *feature importance* mengonfirmasi bahwa kelompok fitur warna mendominasi kontribusi identifikasi penyakit sebesar 71,00%, jauh melampaui kelompok fitur tekstur yang hanya berkontribusi 29,00%, membuktikan bahwa perubahan warna daun menjadi indikator paling dominan dalam pengenalan pola penyakit cabai. Kelima, pengujian 5-Fold Cross Validation memperlihatkan penurunan rata-rata akurasi (menjadi 37,50% untuk RF dan 32,50% untuk SVM), yang mengindikasikan bahwa ukuran dataset yang digunakan saat ini (400 citra) masih terlalu kecil untuk menjamin stabilitas akurasi pada skenario validasi berlapis. Penelitian selanjutnya disarankan untuk memperbesar volume dataset guna meningkatkan stabilitas generalisasi model, menerapkan reduksi dimensi fitur seperti *Principal Component Analysis* (PCA) untuk mengurangi beban komputasi tanpa mengorbankan akurasi, mengeksplorasi pendekatan *Deep Learning* seperti *Convolutional Neural Network* apabila volume data yang lebih besar tersedia, serta mengimplementasikan model SVM yang telah dilatih ke dalam aplikasi mobile atau perangkat IoT untuk mendukung deteksi dini penyakit tanaman cabai secara *real-time* di lapangan.

REFERENCES

- [1] C. H. Fathan Fathurrohman, "165-185+Catur+Herison," Prosiding Seminar Nasional Pertanian Pesisir (SENATASI), 2023.
- [2] F. Wajidi and N. Arifin, "Leaf Disease Detection in Chili Plants Using GLCM and HSV Combination with SVM Classification," *Insect (Informatics and Security): Jurnal Teknik Informatika*, vol. 11, no. 2, 2025, doi: 10.33506/insect.v11i2.4820.
- [3] S. F. Nazila, Y. Arman, D. Wahyuni, N. Nurhasanah, and Y. S. Putra, "Deteksi Dini Serangan Hama Penyakit pada Cabai Rawit Menggunakan Metode Image Recognition," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 2, 2023, doi: 10.28932/jutisi.v9i2.6342.
- [4] R. Suhendra and I. Juliwardi, "Identifikasi dan Klasifikasi Penyakit Daun Jagung Menggunakan Support Vector Machine," *Jurnal Teknologi Informasi*, vol. 1, no. 1, pp. 29–35, 2022.

- [5] N. Apriani and Supatman, "Analisis Ekstraksi Citra GLCM dan Wavelet untuk Klasifikasi Penyakit Cabai dengan Metode SVM," *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, vol. 7, no. 2, pp. 895–903, 2025, doi: 10.51401/jinteks.v7i2.5860.
- [6] M. Faqihuddin and R. A. Purnama, "Komparasi SVM dan Random Forest Berbasis Histogram Warna untuk Deteksi Penyakit Anggur," *Jurnal Teknologi dan Sistem Informasi Bisnis*, vol. 8, no. 1, pp. 34–39, 2026, doi: 10.47233/jteksis.v8i1.2340.
- [7] R. K. Dewi and R. V. H. Ginardi, "Identifikasi Penyakit pada Daun Tebu dengan Gray Level Co-Occurrence Matrix dan Color Moments," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 11, no. 2, 2024.
- [8] S. Agustiani, Y. Tajul Arifin, A. Junaidi, S. Khotimatul Wildah, and A. Mustopa, "Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram," *Jurnal Komputasi*, vol. 10, no. 1, pp. 65–74, 2022, doi: 10.23960/komputasi.v10i1.2961.
- [9] S. Yuliany and A. Nur Rachman, "Implementasi Deep Learning pada Sistem Klasifikasi Hama Tanaman Padi Menggunakan Metode Convolutional Neural Network (CNN)," *Jurnal Buana Informatika*, vol. 13, no. 1, 2022.
- [10] U. Khultsum and A. Subekti, "Penerapan Algoritma Random Forest dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat," *Jurnal Media Informatika Budidarma*, vol. 5, no. 1, p. 186, 2021, doi: 10.30865/mib.v5i1.2624.
- [11] J. Kusuma, "Lisensi Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0," *Journal of Applied Computer Science and Technology (JACOST)*, vol. 4, 2022, doi: 10.52158/jacost.484.
- [12] H. P. Hadi and E. H. Rachmawanto, "Ekstraksi Fitur Warna dan GLCM pada Algoritma KNN untuk Klasifikasi Kematangan Rambutan," *Jurnal Informatika Polinema*, vol. 8, no. 3, pp. 63–68, 2022, doi: 10.33795/jip.v8i3.949.
- [13] A. Purnamawati, W. Nugroho, D. Putri, and W. F. Hidayat, "Deteksi Penyakit Daun pada Tanaman Padi Menggunakan Algoritma Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN," *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, no. 1, 2020, doi: 10.30743/infotekjar.v5i1.2934.
- [14] S. K. Wildah, A. Latif, A. Mustopa, Suharyanto, M. S. Maulana, and A. Sasongko, "Klasifikasi Penyakit Daun Kopi Menggunakan Kombinasi Haralick, Color Histogram dan Random Forest," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 11, no. 1, pp. 35–39, 2023, doi: 10.26418/justin.v11i1.60985.
- [15] L. X. Boa Sorte, C. T. Ferraz, F. Fambrini, R. D. R. Goulart, and J. H. Saito, "Coffee leaf disease recognition based on deep learning and texture attributes," *Procedia Computer Science*, vol. 159, pp. 135–144, 2019, doi: 10.1016/j.procs.2019.09.168.