

Perbandingan Random Forest dan Support Vector Machine untuk Klasifikasi Penyakit Daun Tomat Berbasis Citra Digital

Pelagia Melati^{1*}, Windi Irmayani², Muhammad Ifan Rifani Ihsan³

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

Email: ^{1*}melatipelagia@gmail.com, ²windi.wnr@bsi.ac.id, ³ifan.mii@bsi.ac.id

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

Received: 29 Juni 2026 | Revision: 4 Juli 2026 | Accepted: 8 Juli 2026

Abstrak

Penyakit pada daun tomat merupakan salah satu penyebab utama menurunnya kualitas dan produktivitas hasil panen, sehingga diperlukan metode deteksi yang cepat dan akurat. Perkembangan teknologi pengolahan citra digital dan *machine learning* dapat dimanfaatkan untuk membantu proses identifikasi penyakit daun tomat secara otomatis. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* dan *Support Vector Machine* (SVM) dalam mengklasifikasikan penyakit daun tomat berbasis citra digital. Data dikumpulkan melalui studi pustaka dan penggunaan *dataset Tomato Leaf Disease Detection* yang diperoleh dari Kaggle. Tahapan penelitian meliputi pengumpulan *dataset*, pra-pemrosesan citra, ekstraksi fitur menggunakan Histogram HSV, *Histogram of Oriented Gradients* (HOG), dan *Gray Level Co-occurrence Matrix* (GLCM), pembagian data latih dan data uji dengan rasio 80:20, pelatihan model menggunakan algoritma *Random Forest* dan SVM, serta evaluasi menggunakan *confusion matrix*, akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memiliki performa lebih baik dibandingkan SVM, dengan akurasi sebesar 86,33%, presisi 87,00%, *recall* 86,00%, dan *F1-score* 86,00%, sedangkan SVM memperoleh akurasi 74,67%, presisi 75,00%, *recall* 75,00%, dan *F1-score* 74,00%. Berdasarkan hasil tersebut, *Random Forest* terbukti lebih efektif dalam mengklasifikasikan penyakit daun tomat berbasis citra digital. Penelitian ini diharapkan dapat menjadi referensi dalam pengembangan sistem deteksi penyakit tanaman berbasis *machine learning*.

Kata Kunci: : Klasifikasi Penyakit Daun Tomat, Citra Digital, *Random Forest*, *Support Vector Machine*, *Machine Learning*.

Abstract

Tomato leaf disease is one of the main causes of declining harvest quality and productivity, so a fast and accurate detection method is required. Advances in digital image processing and machine learning can be utilized to support the automatic identification of tomato leaf diseases. This study aims to compare the performance of the *Random Forest* and *Support Vector Machine* (SVM) algorithms in classifying tomato leaf diseases based on digital images. Data were collected through a literature study and the use of the *Tomato Leaf Disease Detection* dataset obtained from Kaggle. The research stages include dataset collection, image preprocessing, feature extraction using HSV Histogram, *Histogram of Oriented Gradients* (HOG), and *Gray Level Co-occurrence Matrix* (GLCM), data splitting into training and testing sets with an 80:20 ratio, model training using *Random Forest* and SVM, and evaluation using *confusion matrix*, accuracy, precision, recall, and *F1-score*. The results show that *Random Forest* performs better than SVM, achieving an accuracy of 86.33%, precision of 87.00%, recall of 86.00%, and *F1-score* of 86.00%, while SVM obtained an accuracy of 74.67%, precision of 75.00%, recall of 75.00%, and *F1-score* of 74.00%. Based on these results, *Random Forest* is proven to be more effective in classifying tomato leaf diseases based on digital images. This research is expected to serve as a reference for the development of machine learning-based plant disease detection systems.

Keywords: *Tomato Leaf Disease Classification*, *Digital Image*, *Random Forest*, *Support Vector Machine*, *Machine Learning*.

1. PENDAHULUAN

Tomat (*Solanum lycopersicum* L.) merupakan tanaman hortikultura dari keluarga Solanaceae yang dibudidayakan secara luas di berbagai negara, termasuk Indonesia. Tomat mengandung beragam nutrisi penting seperti vitamin C, vitamin A, dan likopen yang bermanfaat bagi kesehatan manusia. Berdasarkan data Badan Pusat Statistik, hasil panen

tomat di Indonesia mencapai 1,12 juta ton, sehingga tomat menjadi salah satu komoditas pertanian dengan nilai ekonomi yang tinggi [1].

Meskipun memiliki nilai ekonomi tinggi, produktivitas tanaman tomat masih menghadapi berbagai kendala, salah satunya adalah serangan penyakit pada daun. Serangan penyakit dapat menurunkan hasil panen hingga 20-80% apabila tidak dikenali dan ditangani sejak dini sehingga berpotensi menimbulkan kerugian besar bagi petani. Beberapa penyakit yang umum menyerang daun tomat meliputi Bercak Bakteri (*Bacterial Spot*), Hawar Awal (*Early Blight*), Hawar Akhir (*Late Blight*), Jamur Daun (*Leaf Mold*), Bercak Daun Septoria (*Septoria Leaf Spot*), Tungau Laba-laba (*Spider Mites*), Bercak Target (*Target Spot*), Virus Keriting Kuning (*Yellow Leaf Curl Virus*), dan Virus Mosaik Tomat (*Tomato Mosaic Virus*) [2]. Selama ini, identifikasi penyakit pada daun tomat masih banyak dilakukan secara manual melalui pengamatan visual terhadap gejala yang tampak. Metode ini bersifat subjektif, membutuhkan pengetahuan dan pengalaman yang mendalam, serta memakan waktu yang cukup lama, terutama bagi petani di lokasi terpencil yang sulit mengakses tenaga ahli pertanian [3]. Kemajuan teknologi kecerdasan buatan, khususnya di bidang pengolahan citra digital dan pembelajaran mesin, memberikan peluang besar untuk mengembangkan sistem klasifikasi penyakit tanaman secara otomatis. Metode berbasis citra digital mengambil foto daun tanaman yang selanjutnya diolah oleh komputer untuk mengekstraksi ciri visual seperti warna, tekstur, dan bentuk, yang kemudian dijadikan dasar untuk mengklasifikasikan berbagai jenis penyakit [4]. *Dataset* yang digunakan dalam penelitian ini adalah *Tomato Leaf Disease Detection* yang merupakan versi terbaru dari *dataset* PlantVillage dan telah banyak digunakan dalam riset klasifikasi penyakit tanaman [7]. *Dataset* ini berisi ribuan gambar daun tomat yang diberi label ke dalam 10 kategori, terdiri dari 9 kategori penyakit dan 1 kategori daun sehat. Adanya kesamaan visual antar beberapa jenis penyakit menghadirkan tantangan tersendiri dalam proses klasifikasi, sehingga diperlukan algoritma yang mampu membedakan pola visual tersebut dengan baik.

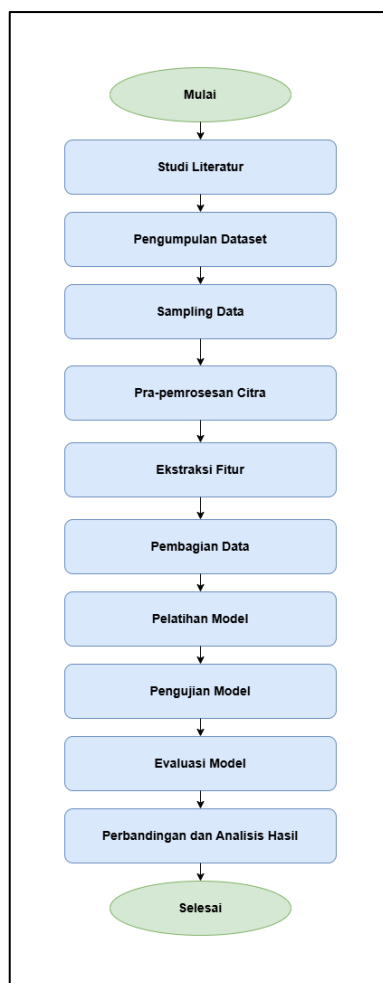
Dua algoritma pembelajaran mesin yang sering digunakan dan terbukti efektif untuk klasifikasi berbasis citra adalah *Random Forest* (RF) dan *Support Vector Machine* (SVM). *Random Forest* merupakan metode *ensemble* yang membangun banyak pohon keputusan secara bersamaan dan menggabungkan hasilnya melalui voting mayoritas untuk memperoleh prediksi yang lebih akurat dan stabil. Sementara itu, SVM bekerja dengan mencari *hyperplane* terbaik yang memisahkan kelas data dengan margin maksimum, sehingga mampu menangani data berdimensi tinggi seperti citra digital dengan baik. Kedua algoritma memiliki karakteristik, kelebihan, dan kelemahan masing-masing, sehingga diperlukan penelitian perbandingan untuk mengetahui algoritma mana yang memberikan hasil optimal dalam klasifikasi penyakit daun tomat.

Beberapa penelitian terkait telah dilakukan sebelumnya. Rahman dkk. mengembangkan sistem klasifikasi penyakit daun tomat berbasis pembelajaran mendalam dengan pendekatan segmentasi pada *dataset* PlantVillage yang menghasilkan akurasi tinggi, namun membutuhkan sumber daya komputasi yang besar [5]. Khan dkk. menggabungkan fitur *Gray Level Co-occurrence Matrix* (GLCM) dan *Scale-Invariant Feature Transform* (SIFT) dengan algoritma SVM pada 2.700 gambar daun tomat sembilan kategori penyakit dan memperoleh akurasi yang tinggi, tetapi tidak membandingkan performanya dengan algoritma lain seperti *Random Forest* [6]. B. Khan dkk. mengembangkan tujuh model hibrida yang mengombinasikan *Convolutional Neural Network* (CNN) dengan algoritma *machine learning* tradisional, termasuk *Random Forest* dan SVM, pada *dataset* PlantVillage menggunakan optimasi Bayesian dan memperoleh kinerja yang sangat baik [10]. Pandiyaraju dkk. mengembangkan model *ensemble* adaptif berbasis *exponential moving average* untuk klasifikasi penyakit daun tomat dan membandingkannya dengan *Random Forest* serta SVM sebagai baseline independen dengan fitur yang dirancang secara manual [12]. Nyasulu dkk. melakukan studi komparatif klasifikasi penyakit fungal pada tomat menggunakan fitur tekstur GLCM dengan beberapa algoritma *machine learning* [9]. Berdasarkan tinjauan terhadap penelitian-penelitian tersebut, dapat diidentifikasi adanya gap penelitian (*research gap*). Sebagian besar penelitian terbaru berfokus pada pendekatan *deep learning*, khususnya CNN, karena mampu menghasilkan akurasi tinggi [10], [12]. Akan tetapi, metode *deep learning* umumnya memerlukan sumber daya komputasi yang besar, waktu pelatihan yang lama, serta jumlah data yang relatif banyak. Di sisi lain, penelitian yang secara langsung membandingkan kinerja *Random Forest* dan SVM pada *dataset* *Tomato Leaf Disease Detection* yang mencakup sembilan jenis penyakit dan satu kelas daun sehat, dengan kombinasi fitur warna (Histogram HSV), tekstur (GLCM), dan bentuk (HOG), masih relatif terbatas. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma *Random Forest* dan *Support Vector Machine* dalam mengklasifikasikan 10 kategori penyakit daun tomat berbasis citra digital, menggunakan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian ini diharapkan dapat memberikan sumbangan ilmiah berupa saran algoritma yang paling sesuai untuk diterapkan dalam sistem klasifikasi penyakit pada daun tomat, serta menjadi dasar pengembangan sistem deteksi penyakit tanaman berbasis *machine learning* yang efisien namun tetap memberikan akurasi yang memadai, sehingga dapat membantu petani mengenali penyakit pada daun tomat dengan lebih cepat, mudah, dan tepat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui serangkaian langkah terencana untuk menilai kemampuan algoritma *Random Forest* dan *Support Vector Machine* (SVM) dalam mengidentifikasi penyakit pada daun tomat menggunakan citra digital. Tahapan penelitian dimulai dari studi literatur, pengumpulan *dataset*, pra-pemrosesan citra, ekstraksi fitur, pembagian data, pelatihan model, pengujian dan evaluasi, hingga analisis perbandingan hasil. Alur tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Kerangka Metodologi Penelitian

Berdasarkan Gambar 1, penelitian dimulai dengan tinjauan pustaka untuk memahami teori terkait pengolahan citra digital, ekstraksi fitur, algoritma *Random Forest*, algoritma *Support Vector Machine*, serta berbagai penyakit yang menyerang daun tomat. *Dataset* yang digunakan berasal dari *Tomato Leaf Disease Detection Dataset* yang diambil dari Kaggle. Proses sampling dilakukan dengan mengambil 300 gambar untuk setiap kelas sehingga jumlah total gambar yang digunakan mencapai 3.000 gambar yang terbagi ke dalam 10 kategori, seperti disajikan pada Tabel 1.

Tabel 1. Distribusi Kelas Dataset Tomato Leaf Disease Detection

No.	Kelas Penyakit	Jumlah Citra
1	Bacterial Spot	300
2	Early Blight	300
3	Late Blight	300
4	Leaf Mold	300
5	Septoria Leaf Spot	300
6	Spider Mites	300
7	Target Spot	300

8	Yellow Leaf Curl Virus	300
9	Mosaic Virus	300
10	Healthy (Sehat)	300
Total		3.000

Sumber: Hasil Penelitian (2026)

2.2 Pra-pemrosesan dan Ekstraksi Fitur

Setelah *dataset* terkumpul, dilakukan pra-pemrosesan citra menggunakan pustaka OpenCV (cv2) pada Python. Setiap citra diubah ukurannya menjadi 64×64 piksel menggunakan fungsi cv2.resize() dengan interpolasi INTER_AREA agar kualitas citra tetap terjaga. Citra kemudian dikonversi ke ruang warna grayscale untuk kebutuhan ekstraksi fitur HOG dan GLCM, serta ke ruang warna HSV (*Hue, Saturation, Value*) untuk ekstraksi fitur warna.

Ekstraksi fitur dilakukan dengan menggabungkan tiga metode. Fitur warna diperoleh menggunakan histogram warna pada ruang warna HSV yang lebih konsisten terhadap perubahan pencahayaan dibandingkan RGB. Fitur tekstur dan bentuk diperoleh melalui metode *Histogram of Oriented Gradients* (HOG) dengan parameter 8 arah orientasi gradien, ukuran sel 8×8 piksel, dan ukuran blok 2×2 sel, sehingga dapat menangkap informasi pola tepi dan struktur daun. Fitur tekstur statistik tambahan diperoleh melalui *Gray Level Co-occurrence Matrix* (GLCM) yang menghitung kontras, dissimilarity, homogenitas, energi, dan korelasi pada sudut 0°, 45°, 90°, dan 135°. Ketiga jenis fitur tersebut digabungkan menjadi satu vektor fitur berdimensi tinggi yang selanjutnya distandarisasi menggunakan StandardScaler dari Scikit-learn agar memiliki rata-rata mendekati nol dan deviasi standar satu.

2.3 Pembagian Data dan Pelatihan Model

Data yang telah diekstraksi fiturnya dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan fungsi train_test_split dari Scikit-learn dengan metode stratified sampling, sehingga proporsi setiap kelas tetap seimbang pada kedua subset. Model *Random Forest* dilatih dengan membangun banyak pohon keputusan yang bekerja secara *ensemble* untuk menghasilkan prediksi yang konsisten dan akurat. Model SVM dilatih menggunakan fungsi kernel *Radial Basis Function* (RBF) untuk menangani data yang tidak dapat dipisahkan secara linear, dengan parameter C (regularisasi) dan gamma yang disesuaikan melalui pengujian untuk mencapai performa optimal.

2.4 Evaluasi Model

Kinerja kedua model dievaluasi menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang dihitung dengan fungsi classification_report dan confusion_matrix dari Scikit-learn menggunakan metode rata-rata makro. Selain pengujian train-test split, dilakukan pula validasi silang 5-fold (*5-fold cross validation*) untuk menilai konsistensi dan kemampuan generalisasi model pada data yang berbeda. Rumus keempat metrik evaluasi tersebut adalah sebagai berikut.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Precision = TP / (TP + FP) \quad (2)$$

$$Recall = TP / (TP + FN) \quad (3)$$

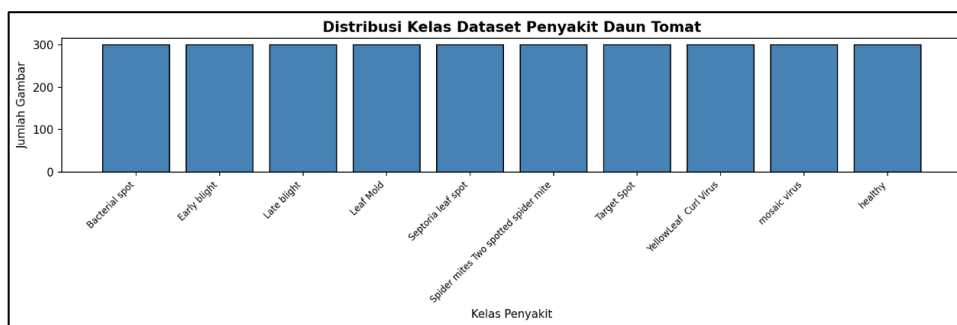
$$F1-Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

dengan TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative) diperoleh dari confusion matrix pada setiap kelas.

3. HASIL DAN PEMBAHASAN

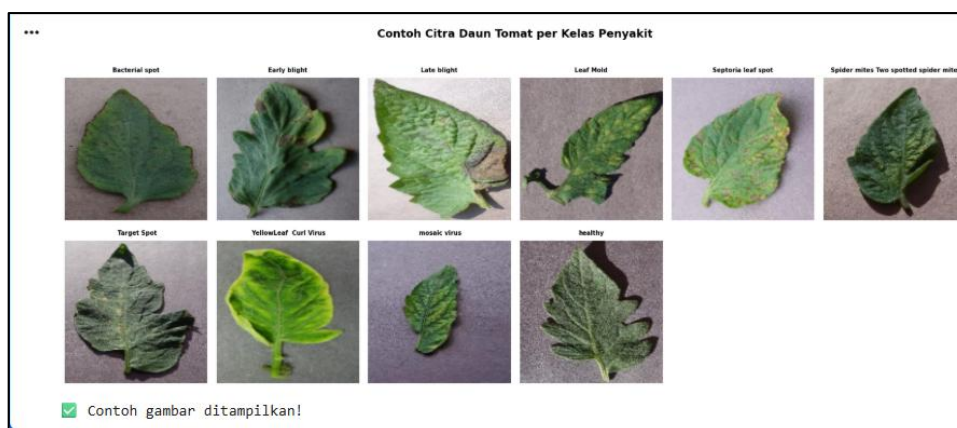
3.1 Dataset dan Distribusi Kelas

Penelitian ini menggunakan *dataset* yang terdiri dari 3.000 gambar daun tomat yang terbagi rata ke dalam 10 kategori, yaitu 9 kategori penyakit dan 1 kategori daun sehat, dengan masing-masing kategori berjumlah 300 gambar sehingga distribusi data tergolong seimbang (*balanced dataset*). Kategori yang digunakan meliputi *Bacterial Spot*, *Early Blight*, *Late Blight*, *Leaf Mold*, *Septoria Leaf Spot*, *Spider Mites*, *Target Spot*, *Yellow Leaf Curl Virus*, *Mosaic Virus*, dan *Healthy*. Distribusi data yang seimbang sangat penting dalam pelatihan model *machine learning* karena dapat meminimalkan risiko bias terhadap kategori tertentu. Gambar 2 memperlihatkan distribusi jumlah citra pada setiap kelas yang digunakan dalam penelitian ini.



Gambar 2. Distribusi Kelas Dataset Penyakit Daun Tomat

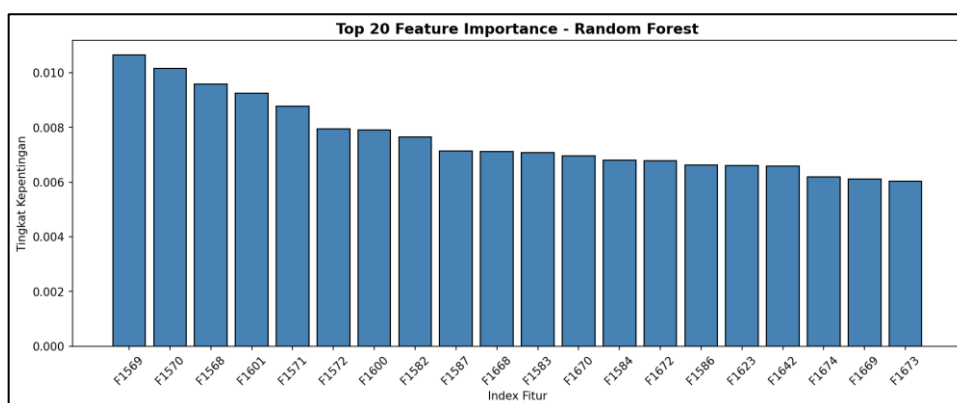
Berdasarkan Gambar 2, seluruh kelas memiliki jumlah data yang seragam, yaitu 300 citra per kelas. Kondisi ini menunjukkan bahwa *dataset* yang digunakan adalah *balanced dataset*, sehingga dapat mendukung model klasifikasi dalam mempelajari ciri-ciri masing-masing kelas secara seimbang dan mengurangi risiko bias selama pelatihan. Gambar 3 menunjukkan contoh citra daun tomat dari setiap kelas yang digunakan dalam penelitian ini.



Gambar 3. Contoh Citra Daun Tomat per Kelas Penyakit

3.2 Hasil Ekstraksi Fitur

proses ekstraksi fitur dilakukan untuk mengubah citra digital menjadi data numerik yang dapat diproses oleh algoritma pembelajaran mesin. Kombinasi fitur HOG dan histogram warna HSV menghasilkan representasi vektor fitur berdimensi tinggi untuk setiap gambar. Analisis *Feature importance* Random Forest dilakukan untuk mengetahui fitur-fitur yang paling berkontribusi terhadap proses klasifikasi. Gambar 4 menunjukkan 20 fitur teratas berdasarkan tingkat kepentingannya.



Gambar 4. Top 20 Feature Importance pada Random Forest

Berdasarkan Gambar 4, fitur dengan indeks F1569 berada pada posisi paling atas dengan nilai kepentingan sekitar 0,011, diikuti oleh fitur F1570 dengan nilai sekitar 0,010. Seluruh 20 fitur utama memiliki nilai kepentingan pada kisaran 0,007

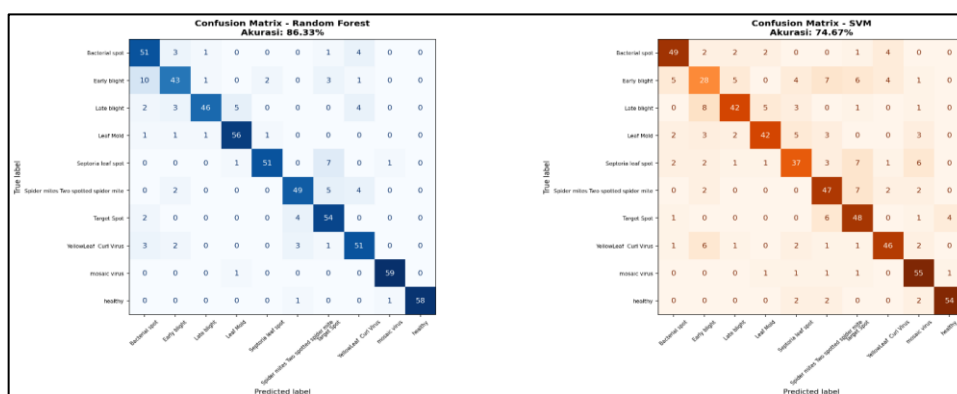
hingga 0,011, yang menunjukkan tidak adanya satu fitur yang sangat dominan, melainkan terdapat distribusi kepentingan yang relatif merata. Hal ini menandakan bahwa model *Random Forest* menggunakan kombinasi berbagai fitur, baik dari HOG maupun histogram warna, untuk menghasilkan prediksi yang akurat.

3.3 Hasil Pengujian Model

Pengujian model dilaksanakan dengan metode pembagian data train-test pada rasio 80:20, di mana 80% data digunakan untuk pelatihan dan 20% untuk pengujian. Selain itu, dilakukan pula validasi silang 5-fold untuk memastikan keandalan serta konsistensi kinerja model. Metrik yang digunakan untuk evaluasi meliputi akurasi, presisi, *recall*, dan *F1-score*.

3.3.1 Confusion Matrix

Confusion matrix digunakan untuk menampilkan rincian jumlah prediksi yang akurat dan tidak akurat pada setiap kategori. Gambar 5 memperlihatkan *confusion matrix* dari kedua model yang diuji.



Gambar 5. *Confusion Matrix Random Forest (kiri) dan SVM (kanan)*

Berdasarkan Gambar 5, model *Random Forest* menunjukkan performa klasifikasi yang lebih baik dibandingkan SVM. *Random Forest* memperoleh akurasi sebesar 86,33%, sedangkan SVM memperoleh akurasi sebesar 74,67%. Pada model *Random Forest*, beberapa kelas berhasil diklasifikasikan dengan sangat baik, seperti *Mosaic Virus* (59 dari 60 citra), *Healthy* (58 dari 60 citra), *Leaf Mold* (56 dari 60 citra), dan *Target Spot* (54 dari 60 citra). Namun, masih terdapat kesalahan klasifikasi, terutama pada kelas *Early Blight* yang sering diprediksi sebagai *Bacterial Spot*, serta kelas *Late Blight* yang masih tertukar dengan *Early Blight* karena memiliki karakteristik visual yang hampir serupa. Pada model SVM, kesalahan klasifikasi lebih banyak ditemukan, misalnya kelas *Early Blight* hanya berhasil dikenali sebanyak 28 dari 60 citra uji, sedangkan kelas *Septoria Leaf Spot* hanya berhasil diprediksi dengan benar sebanyak 37 citra. Hal ini menunjukkan bahwa kemampuan SVM dalam membedakan penyakit dengan pola bercak yang mirip masih lebih rendah dibandingkan *Random Forest*.

3.3.2 Classification Report

Classification report digunakan untuk menilai kinerja model dengan beberapa ukuran, yaitu *precision*, *recall*, *f1-score*, dan *support*. *Precision* mengindikasikan sejauh mana prediksi model akurat untuk suatu kelas, *recall* mengukur kemampuan model dalam mengenali seluruh data pada kelas tertentu, dan *f1-score* merupakan rata-rata harmonis antara *precision* dan *recall*. Tabel 2 dan Tabel 3 menyajikan *classification report* untuk model *Random Forest* dan SVM.

Tabel 2. Classification Report Random Forest

Kelas Penyakit	Precision	Recall	F1-Score	Support
Bacterial Spot	0.74	0.85	0.79	60
Early Blight	0.80	0.72	0.75	60
Late Blight	0.94	0.77	0.84	60
Leaf Mold	0.89	0.93	0.91	60
Septoria Leaf Spot	0.94	0.85	0.89	60
Spider Mites	0.86	0.82	0.84	60
Target Spot	0.76	0.90	0.82	60
Yellow Leaf Curl	0.80	0.85	0.82	60
Virus				

Mosaic Virus	0.97	0.98	0.98	60
Healthy	0.95	0.97	0.98	60
Accuracy			0.86	600
Macro Avg	0.87	0.86	0.86	600
Weighted Avg	0.87	0.86	0.86	600

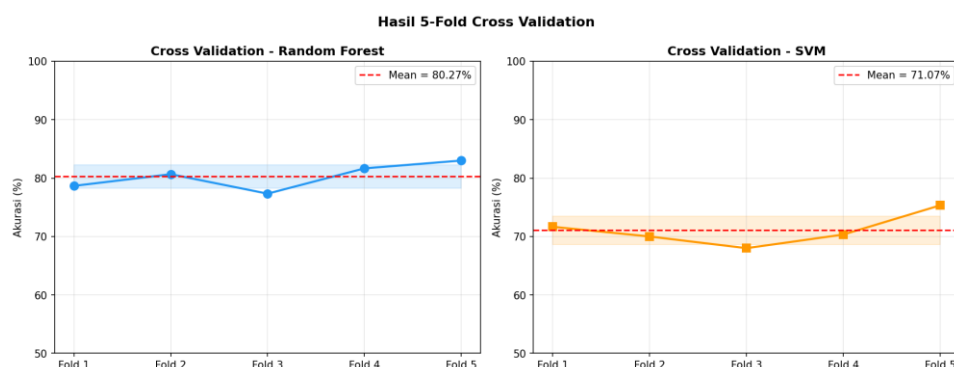
Tabel 3. Classification Report SVM

Kelas Penyakit	Precision	Recall	F1-Score	Support
Bacterial Spot	0.82	0.82	0.82	60
Early Blight	0.55	0.47	0.50	60
Late Blight	0.79	0.70	0.74	60
Leaf Mold	0.82	0.70	0.76	60
Septoria Leaf Spot	0.69	0.62	0.65	60
Spider Mites	0.67	0.78	0.72	60
Target Spot	0.67	0.80	0.73	60
Yellow Leaf Curl	0.81	0.77	0.79	60
Virus				
Mosaic Virus	0.75	0.92	0.83	60
Healthy	0.92	0.90	0.91	60
Accuracy			0.75	600
Macro Avg	0.75	0.75	0.74	600
Weighted Avg	0.75	0.75	0.74	600

Berdasarkan hasil *classification report*, algoritma *Random Forest* memiliki performa yang lebih baik dibandingkan SVM. Hal ini ditunjukkan oleh nilai akurasi *Random Forest* sebesar 86%, sedangkan SVM memperoleh akurasi sebesar 75%. Nilai *macro average* dan *weighted average* pada *Random Forest* juga lebih tinggi, yaitu *precision* sebesar 0,87, *recall* sebesar 0,86, dan *F1-score* sebesar 0,86, sedangkan pada SVM masing-masing sebesar *precision* 0,75, *recall* 0,75, dan *F1-score* 0,74. Pada model *Random Forest*, kelas *Healthy* dan *Mosaic Virus* memperoleh nilai *F1-score* tertinggi sebesar 0,98, sedangkan kelas *Leaf Mold* memperoleh *F1-score* sebesar 0,91. Sementara itu, pada model SVM masih terdapat kelas dengan performa relatif rendah, terutama *Early Blight* dengan *recall* 0,47 dan *F1-score* 0,50, serta *Septoria Leaf Spot* dengan *recall* 0,62 dan *F1-score* 0,65. Hasil ini menunjukkan bahwa *Random Forest* lebih konsisten dan memiliki kemampuan klasifikasi yang lebih baik dibandingkan SVM dalam mengidentifikasi penyakit pada daun tomat.

3.3.3 Hasil 5-Fold Cross Validation

Penggunaan metode *5-fold cross validation* bertujuan untuk mengevaluasi konsistensi performa model dan menguji kemampuan generalisasi model pada data yang beragam. Dalam teknik ini, *dataset* dibagi menjadi lima bagian yang setara, dan model dilatih serta diuji secara bergantian pada setiap bagian selama lima kali iterasi. Gambar 6 menyajikan hasil *5-fold cross validation* untuk kedua model.

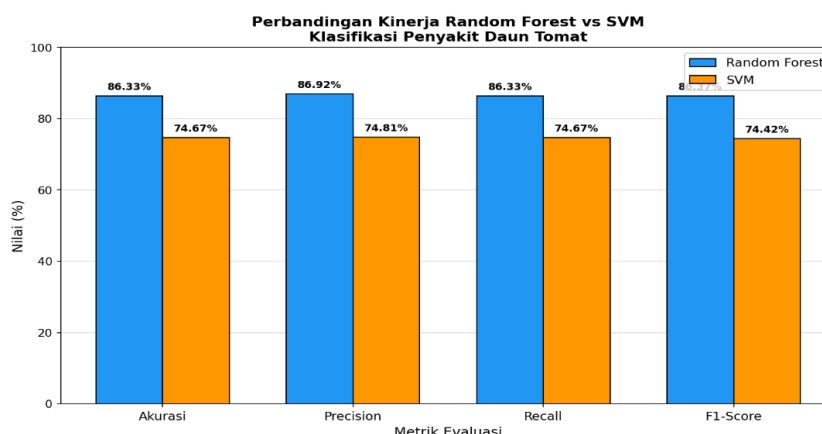


Gambar 6. Hasil 5-Fold Cross Validation Random Forest dan SVM

Berdasarkan Gambar 6, model *Random Forest* menunjukkan kinerja yang relatif stabil pada setiap fold dengan tingkat akurasi berkisar antara 77,33% hingga 83,00% dan rata-rata akurasi sebesar 80,27%. Meskipun terdapat sedikit variasi antar fold, perbedaannya tidak terlalu besar, sehingga menunjukkan bahwa model *Random Forest* memiliki kemampuan generalisasi yang baik. Di sisi lain, model SVM memperoleh akurasi yang berkisar antara 68,00% hingga 75,33% dengan rata-rata akurasi sebesar 71,07%. Hasil ini menunjukkan bahwa performa SVM masih berada di bawah *Random Forest* pada setiap *fold*, dengan variasi nilai akurasi yang lebih besar, yang mengindikasikan bahwa kinerja SVM cenderung kurang stabil dibandingkan *Random Forest* pada dataset yang digunakan.

3.1 Perbandingan Kinerja Random Forest dan SVM

Perbandingan kinerja kedua model dilakukan berdasarkan empat metrik evaluasi utama, yaitu akurasi, presisi, *recall*, dan *F1-score*. Gambar 7 menyajikan perbandingan nilai keempat metrik tersebut untuk model *Random Forest* dan SVM.



Gambar 7. Perbandingan Kinerja Random Forest vs SVM

Berdasarkan Gambar 7, model *Random Forest* selalu lebih baik pada seluruh ukuran evaluasi dibandingkan SVM. Tabel 4 memberikan ringkasan hasil perbandingan kinerja kedua algoritma secara keseluruhan.

Tabel 4. Perbandingan Kinerja Random Forest dan SVM

Metrik Evaluasi	Random Forest	SVM	Selisih
Akurasi	86,33%	74,67%	11,66%
Presisi	87,00%	75,00%	12,00%
Recall	86,00%	75,00%	11,00%
F1-Score	86,00%	74,00%	12,00%

Berdasarkan Tabel 4, algoritma *Random Forest* memperoleh akurasi sebesar 86,33%, presisi 87,00%, *recall* 86,00%, dan *F1-score* sebesar 86,00%. Sementara itu, algoritma SVM memperoleh akurasi sebesar 74,67%, presisi 75,00%, *recall* 75,00%, dan *F1-score* sebesar 74,00%. Hasil tersebut menunjukkan bahwa *Random Forest* memberikan performa yang lebih baik dibandingkan SVM pada seluruh metrik evaluasi, dengan selisih kinerja berkisar antara 11,00% hingga 12,00%.

3.2 Pembahasan

Hasil pengujian menunjukkan bahwa algoritma *Random Forest* memiliki performa yang lebih unggul dibandingkan SVM dalam mengklasifikasikan penyakit pada daun tomat berdasarkan citra digital. Perbedaan kinerja yang cukup mencolok ini dapat dijelaskan dari beberapa sudut pandang. Pertama, *Random Forest* sebagai metode ensemble memanfaatkan kombinasi dari banyak pohon keputusan yang beroperasi secara bersamaan. Setiap pohon dilatih menggunakan subset data dan fitur yang berbeda, sehingga menghasilkan prediksi yang lebih kuat dan tahan terhadap gangguan serta variasi dalam data citra. Mekanisme bagging yang digunakan dalam *Random Forest* secara efisien mengurangi varians model tanpa mengorbankan bias secara signifikan.

Kedua, dataset citra daun tomat memiliki ukuran fitur yang sangat besar akibat penggabungan fitur HOG, histogram HSV, dan GLCM. Dalam kondisi dengan ukuran fitur yang besar, *Random Forest* umumnya lebih efisien karena hanya memanfaatkan bagian acak dari fitur pada setiap pemisahan pohon, sehingga proses komputasi menjadi lebih efisien dan model dapat lebih baik dalam menemukan fitur-fitur penting yang tersebar dalam ruang fitur. Ketiga, meskipun SVM secara teori lebih baik pada ruang fitur berdimensi tinggi, kinerjanya sangat bergantung pada pemilihan kernel dan

parameter yang tepat. Pada penelitian ini, kemungkinan besar pengaturan parameter SVM belum optimal untuk karakteristik dataset yang digunakan, yang berakibat pada akurasi yang lebih rendah. Hal ini terlihat dari variasi yang lebih besar pada hasil validasi silang SVM dibandingkan Random Forest.

Hasil evaluasi menggunakan confusion matrix memperlihatkan bahwa beberapa kelas penyakit dengan karakteristik visual yang hampir sama masih sering tertukar oleh kedua model, terutama pada kelas Late Blight, Early Blight, dan Septoria Leaf Spot. Kesalahan klasifikasi tersebut disebabkan oleh kemiripan pola bercak, warna daun, serta tekstur permukaan daun pada citra digital, sehingga model mengalami kesulitan dalam mengenali perbedaan antar kelas secara tepat. Kondisi ini merupakan salah satu tantangan umum dalam klasifikasi penyakit tanaman berbasis pengolahan citra digital.

Selain itu, hasil analisis feature importance pada algoritma Random Forest menunjukkan bahwa fitur-fitur dengan kontribusi terbesar berasal dari hasil ekstraksi fitur HOG, yang menandakan bahwa informasi tekstur, pola tepi, dan arah gradien pada citra memiliki peran penting dalam klasifikasi penyakit daun tomat. Distribusi nilai feature importance yang relatif merata juga menunjukkan bahwa model memanfaatkan kombinasi berbagai fitur secara keseluruhan, tidak hanya bergantung pada satu fitur tertentu. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan keunggulan metode ensemble berbasis pohon keputusan dalam tugas klasifikasi citra pertanian [10], [12], serta memperkuat hasil penelitian yang menekankan pentingnya kombinasi fitur warna dan tekstur dalam klasifikasi penyakit daun tomat [6], [9]. Pemanfaatan Random Forest di bidang klasifikasi penyakit tanaman terbukti dapat menghasilkan performa yang kompetitif, bahkan bila dibandingkan dengan beberapa model deep learning pada ukuran dataset yang sedang, tanpa memerlukan sumber daya komputasi yang besar seperti pada pendekatan CNN [5], [10]. Penelitian lain yang menerapkan fitur tekstur berbasis polynomial dengan algoritma SVM bahkan mampu mencapai akurasi hingga 98,80% pada dataset PlantVillage [15], yang mengindikasikan bahwa performa SVM pada penelitian ini masih dapat ditingkatkan melalui pemilihan metode ekstraksi fitur dan penyetelan parameter yang lebih optimal.

Berdasarkan seluruh hasil pengujian dan analisis yang telah dilakukan, dapat disimpulkan bahwa hipotesis penelitian terbukti, yaitu terdapat perbedaan signifikan dalam kinerja klasifikasi antara algoritma Random Forest dan Support Vector Machine pada dataset Tomato Leaf Disease Detection, dengan Random Forest menunjukkan kinerja yang lebih baik dibandingkan SVM dalam klasifikasi penyakit daun tomat berbasis citra digital pada dataset yang digunakan dalam penelitian ini.

4. KESIMPULAN

Berdasarkan studi yang telah dilakukan mengenai perbandingan efektivitas algoritma *Random Forest* dan *Support Vector Machine* (SVM) dalam mengklasifikasikan penyakit pada daun tomat menggunakan citra digital, dapat disimpulkan bahwa klasifikasi penyakit daun tomat dapat dilakukan melalui beberapa tahapan, yaitu pra-pemrosesan gambar, ekstraksi fitur warna dan tekstur menggunakan Histogram HSV, HOG, dan GLCM, pembagian data ke dalam data latih dan data uji, pelatihan model, serta evaluasi menggunakan confusion matrix dan classification report. Proses ini memungkinkan sistem untuk mengidentifikasi dan membedakan ciri-ciri dari masing-masing kategori penyakit daun tomat. Berdasarkan hasil pengujian terhadap 10 kondisi daun tomat (9 jenis penyakit dan 1 daun sehat), algoritma *Random Forest* dan SVM mampu melakukan klasifikasi dengan baik. Namun, *Random Forest* menunjukkan kinerja yang lebih unggul dengan akurasi sebesar 86,33%, presisi 87,00%, *recall* 86,00%, dan *F1-score* 86,00%, sedangkan SVM memperoleh akurasi 74,67%, presisi 75,00%, *recall* 75,00%, dan *F1-score* 74,00%. Nilai *precision*, *recall*, dan *F1-score* yang dihasilkan *Random Forest* juga lebih baik pada sebagian besar kategori, sehingga *Random Forest* lebih efektif untuk klasifikasi penyakit daun tomat pada penelitian ini. Hasil validasi silang 5-fold juga menunjukkan bahwa *Random Forest* memiliki kinerja yang lebih stabil dengan rata-rata akurasi 80,27%, dibandingkan SVM yang memperoleh rata-rata akurasi 71,07%. Berdasarkan hasil evaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*, algoritma *Random Forest* menunjukkan performa yang lebih unggul dibandingkan SVM, sehingga *Random Forest* adalah algoritma yang paling sesuai untuk mengklasifikasikan penyakit daun tomat pada dataset yang digunakan dalam penelitian ini. Untuk penelitian selanjutnya, disarankan untuk meningkatkan jumlah dan variasi data gambar, melakukan optimasi hyperparameter menggunakan *Grid Search* atau *Random Search*, menerapkan metode augmentasi data untuk menurunkan risiko *overfitting*, serta membandingkan *Random Forest* dan SVM dengan algoritma pembelajaran mesin dan pembelajaran mendalam lainnya seperti *K-Nearest Neighbor*, *XGBoost*, dan *Convolutional Neural Network*, agar dapat diperoleh model dengan kinerja yang lebih baik dan dapat diimplementasikan dalam bentuk aplikasi berbasis web atau perangkat *mobile* yang dapat digunakan langsung oleh petani.

REFERENCES

- [1] G. Henry, A. Panjaitan, and F. Simatupang, "Pemodelan Klasifikasi Penyakit Daun Tanaman Tomat dengan Convolutional Neural Network Algorithm," vol. 4, no. 5, pp. 2667–2675, 2024, doi: 10.30865/klik.v4i5.1646.
- [2] G. Zhu, C. Ma, S. Yu, X. Zhang, and J. Jiang, "Transcriptome Analyses Reveal the Key Regulators of Tomato Compound Leaf Development," 2023.
- [3] S. Kaur, N. Khanal, R. Dearth, and R. Kariyat, "Morphological characterization of intraspecific variation for trichome traits in tomato (*Solanum lycopersicum*)," *Botanical Studies*, 2023, doi: 10.1186/s40529-023-00370-3.
- [4] M. Shafay, T. Hassan, M. Owais, I. Hussain, S. G. Khawaja, and L. Seneviratne, "Recent advances in plant disease detection: challenges and opportunities," 2025.
- [5] T. Rahman, S. D. Dipto, I. J. June, A. Momin, and R. Al. Mamun, "Machine Learning-based Disease Classification in Tomato (*Solanum lycopersicum*) Plants," vol. 12, no. 3, pp. 151–159, 2024.
- [6] R. Khan, N. U. Din, A. Zaman, and B. Huang, "Processing: An SVM-Based Approach with GLCM and SIFT Features," vol. 2024, 2024, doi: 10.1155/2024/9918296.
- [7] A. Chelladurai, D. P. M. Kumar, and S. S. Askar, "Classification of tomato leaf disease using Transductive Long Short-Term Memory with an attention mechanism," pp. 1–15, 2025, doi: 10.3389/fpls.2024.1467811.
- [8] O. Rainio, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, pp. 1–14, 2024, doi: 10.1038/s41598-024-56706-x.
- [9] C. Nyasulu, A. Diattara, A. Traore, and C. Ba, "A comparative study of Machine Learning-based classification of Tomato fungal diseases: Application of GLCM texture features," *Heliyon*, vol. 9, no. 11, p. e21697, 2023, doi: 10.1016/j.heliyon.2023.e21697.
- [10] B. Khan et al., "Bayesian optimized multimodal deep hybrid learning approach for tomato leaf disease classification," *Scientific Reports*, vol. 0123456789, pp. 1–30, 2024, doi: 10.1038/s41598-024-72237-x.
- [11] H. Riggs and M. Tariq, "Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms," 2023.
- [12] V. Pandiyaraju et al., "Improved tomato leaf disease classification through adaptive ensemble models with exponential moving average fusion and enhanced weighted gradient optimization," pp. 1–26, 2024, doi: 10.3389/fpls.2024.1382416.
- [13] S. Ahmed, M. B. Hasan, T. Ahmed, M. R. K. Sony, and M. H. Kabir, "Less is more: Lighter and faster deep neural architecture for tomato leaf disease classification," *IEEE Access*, vol. 10, pp. 68868–68884, 2022, doi: 10.1109/ACCESS.2022.3187203.
- [14] A. Guerrero-Ibañez and A. Reyes-Muñoz, "Monitoring Tomato Leaf Disease through Convolutional Neural Networks," *Electronics*, vol. 12, no. 1, p. 229, Jan. 2023, doi: 10.3390/electronics12010229.
- [15] A. R. Al-Shamasneh and R. W. Ibrahim, "Classification of tomato leaf images for detection of plant disease using conformable polynomials image features," *MethodsX*, 2024, doi: 10.1016/j.mex.2024.102844.