

IDENTIFIKASI PENYAKIT DAUN SELADA MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM) BERBASIS EKSTRAKSI FITUR VISUAL

Lettuce Leaf Disease Identification Using Support Vector Machine (SVM) Based on Visual Feature Extraction

Muh. Ulil Amri, 105841106221@student.unismuh.ac.id^{(1)*}, Chyquitha Danuputri,
S.Kom., M.Kom, chyquithadanuputri@unismuh.ac.id⁽²⁾, Rizki Yusliana Bakti, S.T., M.T,
rizkiyusliana@unismuh.ac.id⁽³⁾, Ir. Muhammad Syafaat S. Kuba, S.T., M.T,
syafaat_skuba@unismuh.ac.id⁽⁴⁾, dan Muhyiddin A M Hayat S.Kom., M.T,
muhyiddin@unismuh.ac.id⁽⁵⁾

¹⁾²⁾³⁾⁵⁾ Informatika/Fakultas Teknik, Universitas Muhammadiyah Makassar

⁴⁾ Pengairan/Fakultas Teknik, Universitas Muhammadiyah Makassar

Diterima 11 Agustus 2025 / Disetujui 20 Februari 2026

ABSTRACT

Lettuce (Lactuca sativa L.) is a horticultural commodity with high economic value, but its productivity is often hampered by disease attacks. Manual identification by farmers is often inaccurate due to similar visual symptoms between diseases. This study aims to develop an automated disease identification system using a Support Vector Machine (SVM) based on visual feature extraction. The research focuses on classifying four conditions of lettuce leaves: Healthy, Cercospora Leaf Spot (caused by Cercospora longissima), Tip Burn, and Etiolation. The methodology involves extracting color features from the HSV color space and texture features using the Gray-Level Co-occurrence Matrix (GLCM). The effectiveness of data augmentation and hyperparameter optimization using Particle Swarm Optimization (PSO) was also evaluated through three comparative scenarios. The results showed that data augmentation significantly increased the model's accuracy from a baseline of 69.57% to 92.28%. Further optimization with PSO improved the performance to a final accuracy of 93.63%. The best model demonstrated a balanced F1-Score above 0.91 for all classes, proving that the combination of HSV and GLCM feature extraction, data augmentation, and PSO-based SVM optimization is a robust and effective method for lettuce disease identification, offering a promising tool for precision agriculture..

Keywords: Disease Identification, Lettuce, Support Vector Machine, Feature Extraction, Particle Swarm Optimization.

ABSTRAK

Selada (*Lactuca sativa L.*) merupakan komoditas hortikultura bernilai ekonomi tinggi, namun produktivitasnya sering terhambat oleh serangan penyakit. Identifikasi manual oleh petani seringkali tidak akurat karena kemiripan gejala visual antar penyakit. Penelitian ini bertujuan untuk membangun sistem identifikasi penyakit otomatis menggunakan Support Vector Machine (SVM) berbasis ekstraksi fitur visual. Penelitian berfokus pada klasifikasi empat kondisi daun selada: Sehat, Bercak Daun Cercospora (disebabkan oleh *Cercospora longissima*), Tipburn, dan Etiolasi. Metodologi yang digunakan meliputi ekstraksi fitur warna dari ruang warna HSV dan fitur tekstur menggunakan Gray-Level Co-occurrence Matrix (GLCM). Efektivitas augmentasi data dan optimasi hyperparameter menggunakan Particle Swarm Optimization (PSO) juga dievaluasi melalui tiga skenario perbandingan. Hasil penelitian menunjukkan bahwa augmentasi data secara signifikan meningkatkan akurasi model dari baseline 69,57% menjadi 92,28%. Optimasi lebih lanjut dengan PSO berhasil meningkatkan performa hingga mencapai akurasi final sebesar 93,63%. Model terbaik menunjukkan F1-Score yang seimbang di atas 0,91 untuk semua kelas, membuktikan bahwa kombinasi metode ekstraksi fitur HSV dan GLCM, augmentasi data, dan optimasi

*Korespondensi Penulis:

E-mail: 105841106221@student.unismuh.ac.id

SVM menggunakan PSO merupakan pendekatan yang andal dan efektif untuk identifikasi penyakit daun selada, serta menawarkan alat bantu yang prospektif untuk pertanian presisi.

Kata Kunci: Identifikasi Penyakit, Selada, Support Vector Machine, Ekstraksi Fitur, Particle Swarm Optimization.

PENDAHULUAN

Sektor pertanian memegang peranan strategis dalam perekonomian nasional dan upaya pemenuhan ketahanan pangan. Di antara berbagai komoditas hortikultura, selada (*Lactuca sativa* L.) memiliki kontribusi penting sebagai sumber pangan segar bagi masyarakat. Namun, upaya peningkatan produktivitas selada seringkali dihadapkan pada tantangan signifikan berupa gangguan kesehatan tanaman, baik yang bersifat patogenik maupun fisiologis. Kegagalan dalam membedakan penyebab gangguan ini dapat berujung pada kerugian ekonomi yang signifikan bagi petani [1]. Identifikasi penyakit secara manual oleh petani seringkali bersifat subjektif dan tidak akurat karena kemiripan gejala visual antar penyakit, terutama pada tahap awal, sehingga menyulitkan proses pembedaan [2]

Perkembangan teknologi *Computer Vision* dan pembelajaran mesin (*Machine Learning*) menawarkan solusi yang prospektif untuk mengatasi permasalahan ini [3]. Berbagai pendekatan telah dieksplorasi untuk melakukan identifikasi penyakit tanaman secara otomatis. Pendekatan *state-of-the-art* saat ini didominasi oleh metode *deep learning*, khususnya *Convolutional Neural Network* (CNN). Sebagai contoh, penelitian oleh Bhujel dkk. berhasil mengembangkan arsitektur CNN ringan yang dilengkapi mekanisme atensi (*attention-based*) untuk klasifikasi penyakit daun tomat dengan akurasi yang sangat tinggi [4]. Meskipun sangat kuat, pendekatan CNN umumnya memerlukan sumber daya komputasi yang besar dan dataset pelatihan yang ekstensif [5].

Di sisi lain, pendekatan *machine learning* klasik seperti Support Vector Machine (SVM), yang dikombinasikan dengan rekayasa fitur manual, tetap menjadi alternatif yang relevan dan lebih ringan secara komputasi. Kombinasi SVM dengan ekstraksi fitur *Gray-Level Co-occurrence Matrix* (GLCM) telah terbukti efektif pada berbagai tanaman hortikultura, seperti pada deteksi penyakit cabai [6]. Penelitian yang lebih spesifik oleh Marlita dan Mustofa juga berhasil menerapkan kerangka kerja serupa untuk klasifikasi penyakit pada daun kentang [7]. Namun, studi-studi ini, meskipun berhasil, cenderung berfokus pada penerapan dasar dan belum secara spesifik mengeksplorasi dampak dari optimasi *hyperparameter* secara sistematis, yang merupakan faktor krusial untuk memaksimalkan potensi performa model SVM.

Dari tinjauan tersebut, teridentifikasi sebuah celah penelitian (*research gap*) yang jelas. Terdapat kebutuhan untuk mengevaluasi apakah kinerja model SVM klasik dapat ditingkatkan secara signifikan melalui proses optimasi yang sistematis, sehingga mampu menjadi alternatif yang lebih efisien dibandingkan model *deep learning* yang kompleks. Penelitian ini memosisikan diri untuk mengisi celah tersebut. Dengan mengadopsi kerangka kerja SVM dan ekstraksi fitur visual (HSV dan GLCM) yang telah terbukti efektif, penelitian ini menambahkan sebuah lapisan investigasi yang sering terlewatkan.

Oleh karena itu, kontribusi utama dan kebaruan dari penelitian ini adalah melakukan analisis perbandingan secara spesifik antara model SVM dengan parameter standar dan model SVM yang *hyperparameter* nya dioptimalkan menggunakan algoritma *Particle Swarm Optimization* (PSO). Dengan demikian, penelitian ini tidak hanya bertujuan untuk membangun sebuah sistem deteksi, tetapi juga untuk secara kuantitatif membuktikan dan mengukur sejauh mana proses optimasi dapat meningkatkan akurasi dan keandalan sebuah model klasifikasi klasik dalam domain identifikasi penyakit daun selada.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk mengembangkan dan mengevaluasi model klasifikasi penyakit daun selada. Metodologi penelitian dirancang untuk memastikan transparansi dan reproduktibilitas.

Dataset dan Pembagian Data

Dataset yang digunakan merupakan data primer yang terdiri dari 230 citra digital daun selada, yang diambil menggunakan kamera ponsel cerdas di lokasi budidaya DEEDAD HIDROPONIK, Makassar.

Komposisi Kelas

Tabel 1. Komposisi Dataset Induk

No	Kelas Kategori	Jumlah Gambar
1.	Sehat	49
2.	<i>Cercospora longissima</i>	79
3.	Tip Burn	54
4.	Etiolasi	48
Total		230

Dapat diamati bahwa terdapat ketidakseimbangan ringan (*mild imbalance*) pada dataset, di mana kelas *Cercospora longissima* memiliki jumlah sampel yang lebih banyak. Fakta ini menjadi pertimbangan penting dalam evaluasi model, di mana metrik seperti F1-Score menjadi lebih relevan daripada akurasi saja.

Dataset induk dibagi menjadi 80% data latih (*training set*) dan 20% data uji (*testing set*). Pembagian ini dilakukan secara acak namun tetap menjaga proporsi setiap kelas (*stratified split*). Proses optimasi *hyperparameter* menggunakan validasi silang 5-lipat (*5-fold cross-validation*) yang diterapkan hanya pada data latih untuk mencegah kebocoran informasi dari data uji.

Mengingat jumlah dataset asli yang terbatas, teknik augmentasi data diterapkan secara dinamis pada data latih untuk Skenario 2 dan 3. Jenis augmentasi yang diterapkan meliputi

- Rotasi. Rotasi acak gambar dalam rentang ± 20 derajat.
- Flipping. Pembalikan gambar secara horizontal dan vertikal.
- Zoom Pembesaran acak pada gambar hingga 20%.
- Pergeseran (Shift) Pergeseran horizontal dan vertikal acak hingga 10% dari lebar/tinggi gambar.
- Penyesuaian Kecerahan (Brightness): Perubahan kecerahan acak dalam rentang [0.8, 1.2].

Dua jenis fitur diekstraksi dari setiap citra yang telah di-resize menjadi 224x224 piksel. Fitur Warna (HSV): Citra dikonversi ke ruang warna HSV. Fitur warna diekstrak dengan menghitung tiga momen statistik mean, standar deviasi, dan skewness dari histogram untuk setiap kanal Hue (H), Saturation (S), dan Value (V), menghasilkan total 9 fitur warna. Pendekatan ini relevan dengan studi analisis color space yang telah dilakukan sebelumnya [8].

Fitur Tekstur (GLCM): Citra dikonversi ke format grayscale. Fitur tekstur diekstraksi menggunakan Gray-Level Co-occurrence Matrix (GLCM), yang dihitung dengan jarak piksel (d)

= 1 pada empat orientasi sudut: 0°, 45°, 90°, dan 135°. Nilai dari keempat sudut ini kemudian dirata-ratakan untuk menghasilkan satu nilai tunggal untuk setiap parameter tekstur. Efektivitas GLCM telah divalidasi dalam berbagai penelitian, seperti pada identifikasi penyakit cabai [6]

Model Klasifikasi dan Optimasi sebagai Model utama yang digunakan adalah Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF). Formulasi SVM SVM bekerja dengan mencari hyperplane optimal yang memaksimalkan margin antar kelas. Untuk data non-linear, fungsi kernel RBF digunakan rumus (1)

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \dots\dots\dots(1)$$

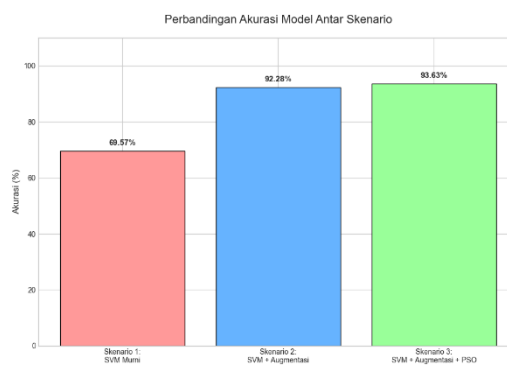
Optimasi Hyperparameter dengan PSO. Untuk Skenario 3, hyperparameter SVM yaitu parameter regularisasi (C) dan parameter kernel (gamma) dioptimalkan menggunakan *Particle Swarm Optimization* (PSO). PSO bekerja dengan "kawanan" partikel yang menjelajahi ruang pencarian [C, gamma] untuk menemukan kombinasi yang memberikan nilai fitness tertinggi. Dalam penelitian ini, fungsi fitness didefinisikan sebagai akurasi rata-rata dari validasi silang 5-lipat pada data latih. Pendekatan optimasi SVM-PSO ini telah terbukti berhasil pada kasus serupa [9]

Kinerja model dievaluasi menggunakan Akurasi, Presisi, Recall, dan F1-Score yang diturunkan dari *confusion matrix*. [10]

$$\begin{aligned} \text{Akurasi} &= (TP + TN) / (TP + TN + FP + FN) \\ \text{Presisi} &= TP / (TP + FP) \\ \text{Recall} &= TP / (TP + FN) \\ \text{F1-Score} &= 2 * (\text{Presisi} * \text{Recall}) / (\text{Presisi} + \text{Recall}) \end{aligned}$$

HASIL DAN PEMBAHASAN

Analisis perbandingan akurasi dari ketiga skenario (Gambar 1) menunjukkan dampak signifikan dari setiap teknik.



Gambar 1. Grafik Perbandingan Akurasi Antar Skenario

Hasil menunjukkan lompatan kinerja signifikan dari Skenario 1 (akurasi 69.57%) ke Skenario 2 (akurasi 92.28%). Peningkatan lebih lanjut ke Skenario 3 (akurasi 93.63%) menunjukkan keberhasilan optimasi PSO.

Analisis dan Interpretasi Hasil

Efektivitas Augmentasi Data. Peningkatan akurasi yang drastis setelah augmentasi data dapat dijelaskan oleh ukuran dataset awal yang terbatas. Augmentasi secara artifisial menciptakan variasi baru, memaksa model untuk belajar fitur yang lebih robust dan mencegah *overfitting*[11]. Peran PSO dalam Fine-Tuning. Peningkatan lebih lanjut oleh PSO menunjukkan peran *fine-tuning*. PSO secara cerdas menjelajahi ruang parameter [C, gamma] untuk menemukan "sweet spot" yang paling sesuai dengan struktur data, menyempurnakan *decision boundary* SVM lebih baik daripada parameter default.

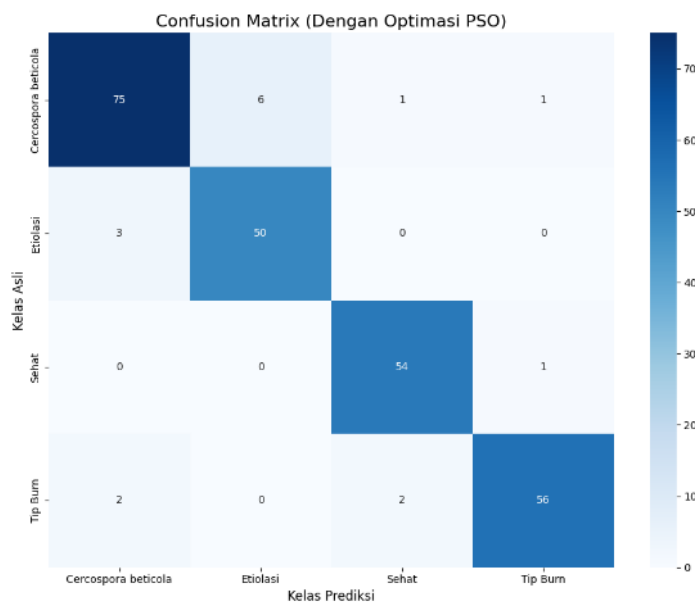
Analisis Model Terbaik (Skenario 3)

Model terbaik (Skenario 3) dievaluasi secara mendalam. Tabel 2 menunjukkan laporan klasifikasi untuk setiap kelas.

Tabel 2. Laporan Klasifikasi Model Terbaik (Skenario 3)

Kelas	Presisi	Recall	F1-Score
Cercospora beticola	0.94	0.90	0.92
Etiolasi	0.89	0.94	0.92
Sehat	0.95	0.98	0.96
Tip Burn	0.97	0.93	0.95

F1-Score untuk semua kelas berada di atas 0.91, menandakan kinerja model sangat baik dan seimbang. *Confusion matrix* (Gambar 2) mengonfirmasi bahwa sebagian besar prediksi benar.



Gambar 2. Confusion Matrix Model Terbaik (Skenario 3)

Berdasarkan hasil penelitian, dapat disimpulkan bahwa implementasi model SVM dengan fitur gabungan HSV dan GLCM berhasil dilakukan. Augmentasi data terbukti menjadi faktor kunci yang secara drastis meningkatkan akurasi model dari 69.57% menjadi 92.28%. Proses optimasi *hyperparameter* menggunakan PSO terbukti mampu memberikan peningkatan kinerja lebih lanjut hingga mencapai akurasi final 93.63%. Kombinasi dari rekayasa fitur yang relevan, augmentasi data, dan optimasi *hyperparameter* terbukti mampu menghasilkan sebuah model klasifikasi yang andal dan akurat untuk identifikasi penyakit pada daun selada.

SIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa implementasi model SVM dengan fitur gabungan HSV dan GLCM berhasil dilakukan. Augmentasi data terbukti menjadi faktor kunci yang secara drastis meningkatkan akurasi model dari 69.57% menjadi 92.28%. Proses optimasi *hyperparameter* menggunakan PSO terbukti mampu memberikan peningkatan kinerja lebih lanjut hingga mencapai akurasi final 93.63%. Kombinasi dari rekayasa fitur yang relevan, augmentasi data, dan optimasi *hyperparameter* terbukti mampu menghasilkan sebuah model klasifikasi yang andal dan akurat untuk identifikasi penyakit pada daun selada.

DAFTAR PUSTAKA

- [1] M. Indah Kusuma, J. Orbit, W. I. Safira, N. M. Zuhri, N. Maulida, and S. Ayomi, "Kerusakan Tanaman Selada (*Lactuca sativa* L) oleh OPT pada Budidaya Hidroponik di Kota Semarang," 2024.
- [2] A. Nursanti, E. T. Suciando, and A. Mumpuni, "Identifikasi Jamur Patogen dan Tingkat Persentase Penyakit pada Tanaman Selada (*Lactuca sativa* L.) di Sentra Tanaman Sayur Desa Serang, Kecamatan Karangreja, Kabupaten Purbalingga," *Jurnal Ilmiah Biologi Unsoed*, vol. 3, no. 1, pp. 9–19, 2021.
- [3] T. M. Surya Mulyana, A. Fritz Wijaya, J. A. Ginting, I. Gusti Ngurah Suryantara, and R. T. Kwa, "Boundary Identification using Backtracking Algorithm in Depth First Search Approach," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, Dec. 2025, pp. 390–395. doi: 10.1109/ICIMCIS68501.2025.11327050.
- [4] A. Bhujel, N. E. Kim, E. Arulmozhi, J. K. Basak, and H. T. Kim, "A Lightweight Attention-Based Convolutional Neural Networks for Tomato Leaf Disease Classification," *Agriculture (Switzerland)*, vol. 12, no. 2, Feb. 2022, doi: 10.3390/agriculture12020228.
- [5] D. C. Sulaiman and T. M. S. Mulyana, "Web-Based Writing Learning Application of Basic Hanacaraka Using Convolutional Neural Network Method," *Ultimatics : Jurnal Teknik Informatika*, pp. 28–34, Jun. 2023, doi: 10.31937/ti.v15i1.2993.
- [6] F. Zikra, K. Usman, and R. Patmasari, "Deteksi Penyakit Cabai Berdasarkan Citra Daun Menggunakan Metode Gray Level Co-Occurrence Matrix Dan Support Vector Machine," 2021.
- [7] R. Marlita and D. Mustofa, "Implementasi Support Vector Machine Pada Klasifikasi," *JIKA : Jurnal Ilmu Komputer dan Aplikasinya*, vol. 1, no. 1, pp. 6–11, 2025, doi: 10.5281/zenodo.14885356.

- [8] M. Fatham, M. Akbar, H. Fitriyah, and S. R. Akbar, "Analisis Color Space untuk Spesifikasi Perancangan Perangkat Lunak pada Embedded System Deteksi Penyakit Busuk Selada," 2023. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] W. A. Saputro, P. N. Andono, and M. A. Soeleman, "Klasifikasi SVM Menggunakan Optimasi PSO Untuk Kelayakan Biji Kopi Dengan Level Medium Roast," *Techno.Com*, vol. 24, no. 2, pp. 543–555, May 2025, doi: 10.62411/tc.v24i2.12657.
- [10] T. M. S. Mulyana, D. Widyaningrum, and H. Herlina, "OCR HURUF JAWA DENGAN FITUR KODE RANTAI DAN LEVENSHTTEIN DISTANCE," *Network Engineering Research Operation*, vol. 6, no. 1, p. 67, Apr. 2021, doi: 10.21107/nero.v6i1.217.
- [11] J. Jahja, H. Agung, T. M. S. Mulyana, L. Hakim, and E. Sipayung, "Classification of heart disease trigger factors using Naive Bayes method to predict the risk of heart disease using IoT-based heart rate sensors," 2023, p. 020014. doi: 10.1063/5.0119083.