

Classification of Rice Leaf Diseases Using Support Vector Machine with HSV and GLCM-Based Feature Extraction

Rizal Ramli *¹, Evanita *², Aditya Akbar Riadi *³

* Teknik Informatika, Universitas Muria Kudus

202151160@std.umk.ac.id¹, evanita@umk.ac.id², aditya.akbar@umk.ac.id³

Article Info

Article history:

Received 2025-07-22

Revised 2025-09-11

Accepted 2025-09-20

Keyword:

GLCM,

HSV,

Rice Leaf Disease,

SVM,

Image Classification.

ABSTRACT

This study aims to classify rice leaf diseases using the Support Vector Machine (SVM) algorithm based on image processing and feature extraction. A total of 600 rice leaf images were collected, each representing one of five disease types: bacterial blight, leaf smut, leaf blast, brown spot, and hispa. The images underwent preprocessing, including resizing, background removal, and feature extraction using HSV and GLCM methods. Extracted features were then used to train and test an SVM classification model. The evaluation using confusion matrix showed an overall accuracy of 83%, with class-specific F1-scores ranging from 0.72 to 0.90. These results indicate that SVM is effective in classifying rice leaf diseases and can potentially assist farmers in early disease detection to reduce crop loss.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Beras merupakan salah satu komoditas pangan pokok yang sangat penting bagi masyarakat Indonesia. Berdasarkan data Badan Pusat Statistik Nasional, produksi beras pada tahun 2024 untuk konsumsi pangan penduduk mencapai angka 30,62 juta ton. Angka tersebut mengalami penurunan sebesar 1,54 persen atau sekitar 480,04 ribu ton dibandingkan tahun sebelumnya [1]. Di sisi lain, jumlah penduduk Indonesia terus mengalami kenaikan setiap tahunnya [2]. Hal ini mengindikasikan adanya peningkatan kebutuhan pangan yang signifikan, sementara ketersediaan produksi justru mengalami penurunan.

Penurunan produksi beras disebabkan oleh berbagai faktor, seperti berkurangnya lahan pertanian padi, ketidakstabilan cuaca, serta serangan hama dan penyakit. Penyakit pada tanaman padi, khususnya yang menyerang bagian daun, umumnya disebabkan oleh bakteri dan jamur [3]. Serangan penyakit tersebut pada masa pertumbuhan dapat mengakibatkan penurunan produktivitas panen secara drastis. Oleh karena itu, diperlukan upaya deteksi dini penyakit pada daun padi guna mencegah penyebaran lebih lanjut dan meminimalisasi kerugian hasil panen.

Melalui perkembangan teknologi yang pesat, pemanfaatan *machine learning* untuk pengolahan citra

menjadi salah satu solusi potensial. Algoritma *Support Vector Machine* (SVM) merupakan salah satu algoritma yang banyak digunakan untuk klasifikasi citra berbasis fitur, sehingga penting untuk menguji sejauh mana efektivitasnya dalam mendeteksi penyakit daun padi. SVM dikenal cukup efektif dalam mengklasifikasikan citra karena mampu menemukan *hyperplane* terbaik yang memisahkan antar kelas secara optimal [4][5].

Dibandingkan dengan algoritma *machine learning* tradisional lainnya seperti *K-Nearest Neighbor* (KNN), *Naïve Bayes*, atau *Decision Tree*, SVM sering menunjukkan kinerja lebih stabil pada data dengan dimensi tinggi seperti citra digital [6]. KNN misalnya, meskipun sederhana, sering kali sensitif terhadap jumlah data dan pemilihan parameter k [7]. *Naïve Bayes* dapat bekerja baik pada data teksual, tetapi performanya menurun jika data citra memiliki distribusi fitur yang kompleks [8]. *Decision Tree* mudah diinterpretasikan, namun cenderung mengalami *overfitting* [9]. Oleh karena itu, SVM dipandang sebagai *baseline* yang kuat karena kemampuannya menangani data non-linear dengan menggunakan kernel *trick* dan menjaga margin pemisahan antar kelas yang maksimal.

Meskipun saat ini *deep learning* seperti algoritma *Convolutional Neural Network* (CNN) banyak digunakan

dalam klasifikasi citra karena mampu mengekstraksi fitur otomatis, pendekatan ini membutuhkan dataset besar serta daya komputasi tinggi[10]. Dengan demikian, penggunaan SVM masih relevan, khususnya dalam konteks penelitian akademik dengan keterbatasan data maupun komputasi. Hal ini menjadikan penelitian ini penting sebagai eksplorasi efektivitas SVM dibanding algoritma ML tradisional maupun pendekatan *deep learning*.

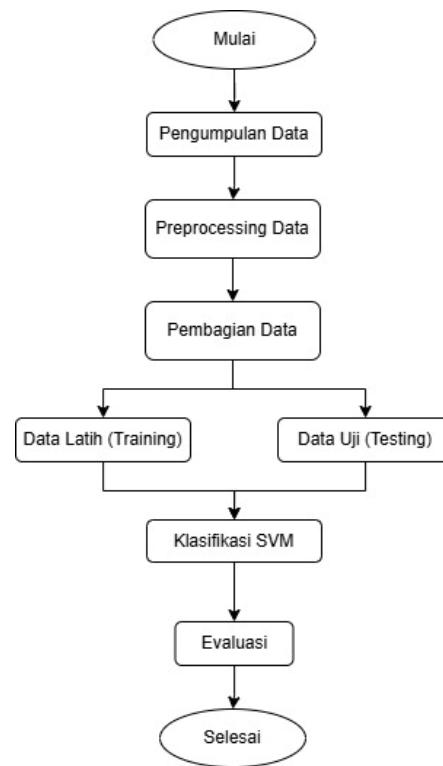
Beberapa penelitian terdahulu telah menunjukkan efektivitas SVM dalam klasifikasi citra. Kurniadi *et al.* [5] mengimplementasikan SVM untuk klasifikasi citra buah dan membandingkannya dengan KNN, dengan dataset sebanyak 11.219 citra pada 19 kelas menggunakan fitur RGB, menghasilkan akurasi 93,09%. Penelitian lain juga dilakukan oleh Meiriyama [11] yang juga memanfaatkan SVM untuk klasifikasi citra buah dengan 100 dataset pada 6 kelas menggunakan fitur HSV, menghasilkan akurasi 94%. Penelitian lain dilakukan Neneng [12] menggunakan SVM untuk klasifikasi jenis daging dengan 1.800 citra pada 4 kelas menggunakan fitur GLCM, menghasilkan akurasi 87,5%. Selanjutnya, penelitian dilakukan oleh arifin *et all.* [13] melakukan klasifikasi citra tanaman obat herbal dengan SVM, menggunakan 1.720 citra pada 4 kelas dengan fitur GLCM, menghasilkan akurasi 87%.

Berdasarkan penelitian terdahulu, dapat dilihat bahwa pemilihan metode ekstraksi fitur memengaruhi kinerja klasifikasi. Pada penelitian ini digunakan kombinasi HSV (*Hue, Saturation, Value*) dan GLCM (*Gray Level Co-occurrence Matrix*). HSV lebih representatif dalam mengekstraksi informasi warna dibandingkan model RGB karena memisahkan komponen warna dan intensitas cahaya [14], sehingga cocok digunakan untuk klasifikasi yang fokus pada perbedaan warna objek, seperti penyakit pada daun padi. Sementara itu, GLCM unggul dalam menangkap pola tekstur pada citra yang sering menjadi ciri khas penyakit pada daun [15]. Kombinasi keduanya diharapkan mampu meningkatkan akurasi klasifikasi dibandingkan penggunaan ekstraksi fitur lainnya seperti LBP dan HOG yang hanya fokus pada tekstur lokal serta bentuk tepian objek yang kurang cocok untuk ekstraksi fitur pada penyakit daun padi.

Dengan demikian, penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma SVM dalam klasifikasi citra daun padi berdasarkan jenis penyakitnya. Aspek kebaruan dalam penelitian ini terletak pada penggunaan kombinasi metode ekstraksi fitur HSV dan GLCM yang jarang digunakan secara bersamaan dalam penelitian klasifikasi penyakit padi sebelumnya. Selain itu, penelitian ini juga menggunakan jumlah dataset yang lebih optimal serta variasi jenis penyakit yang lebih luas, sehingga diharapkan dapat memberikan kontribusi nyata dalam upaya deteksi dini penyakit padi. Hasil penelitian ini diharapkan dapat membantu petani dalam mengidentifikasi penyakit sejak dini guna mengurangi potensi penurunan hasil panen di kemudian hari.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma *Support Vector Machine* (SVM) untuk melakukan klasifikasi penyakit pada daun padi. SVM bekerja dengan cara menemukan *hyperplane* terbaik yang memisahkan antar kelas secara optimal. Evaluasi dilakukan menggunakan *Confusion Metric* untuk menilai akurasi model prediktif terhadap data actual. Diagram alur penelitian ditunjukkan pada Gambar 1



Gambar 1. Diagram Alir

Uraian mendalam terkait masing-masing tahap diatas dijabarkan pada bagian berikut:

1) Pengumpulan Data

Penelitian ini diawali dengan pengumpulan data, data yang digunakan merupakan citra digital daun padi yang berpenyakit, citra berasal dari dokumentasi pribadi penulis. Citra daun terdiri dari 5 jenis penyakit, diantaranya *bacterial blight*, *leaf smut*, *leaf blast*, *brown spot* dan *hispa*. Citra daun diambil dengan kamera *smartphone* resolusi 48MP, diambil dari jarak ketinggian 10-15cm dari objek citra dengan latar belakang putih. Jumlah keseluruhan data adalah 600 dataset dengan masing-masing kelas sejumlah 120 citra berbentuk JPG. Total data secara keseluruhan adalah 600 data.

2) Preprocessing Data

Setelah pengumpulan data, dilakukan preprocessing data yang mencakup *resize* dan ekstraksi fitur numerik HSV-GLCM. Proses *resize* dilakukan untuk menyeragamkan ukuran citra. Dengan melakukan *resize* ke ukuran seragam 224×224 piksel, setiap citra akan memiliki jumlah piksel yang sama sehingga perhitungan fitur tekstur dan warna menjadi lebih konsisten. Langkah ini penting untuk meningkatkan akurasi model dan mengurangi beban komputasi karena ukuran citra yang lebih kecil mempercepat proses ekstraksi fitur dan pelatihan model.

TABEL I
CONTOH CITRA YANG DIAMBIL

Bacterial Leaf Blight	Blast
Bworn Spot	Hispa
Leaf Smut	

Setelah *resize* data, dilakukan ekstraksi fitur guna menghasilkan data numerik. Hasil ekstraksi menghasilkan fitur 7 fitur, dengan 4 fitur GLCM, 3 fitur HSV dan yang terakhir adalah label kelas.

a.) Ekstraksi Fitur HSV

- 1) *Hue*: Hue merepresentasikan nunsa warna yang sebenarnya, dengan atribut kualitatif pada

warna, seperti hijau, biru, merah, serta warna lainnya [16] dan digunakan untuk menentukan tingkat kemerahan [17].

$$H = \tan \frac{3(G - B)}{(R - R) + (R - B)}$$

- 2) *Saturation*: *Saturation* merepresentasikan tingkat kemurnian pada warna, semakin tinggi *saturation*, semakin kaya juga warna dan tingkat kejemuhananya [16].

$$S = \tan \frac{V - \min(R, G, B)}{V}$$

- 3) *Value*: *Value* merepresentasikan tingkat kecerahan atau intensitas warna, semakin tinggi nilai *value*, warna yang dihasilkan juga semakin terang [16].

$$V = \frac{R + G + B}{3}$$

TABEL II
HASIL EKSTRAKSI FITUR HSV

Hasil Masking HSV	HSV_Mean
	31, 174, 131
	39, 180, 187
	36, 180, 187

	36, 181, 88
	30, 148, 138

b.) *Fitur GLCM*

Fitur GLCM adalah matriks yang mampu menggambarkan secara rinci setiap unsur pada citra [18], termasuk tekstur pada citra. Tekstur merupakan salah satu fitur penting dalam melakukan identifikasi citra karena memberikan detail mengenai struktur permukaan, perubahan intensitas, dan kecerahan warna [19]. GLCM memiliki 4 unsur, yang terdiri dari Kontras, Homogenitas, Energi, dan Korelasi.

- 1) *Kontras:* Kontras memiliki fungsi mengukur frekuensi spasial pada citra serta perbedaan momen pada GLCM, perhitungan kontras adalah sebagai berikut:

$$Contrast = \sum_{n=1}^L n^2 \left\{ \sum_{|i-j|=n} GLCM(i,j) \right\}$$

- 2) *Homogenitas:* Homogenitas memiliki fungsi untuk kesamaan variasi intensitas pada citra, perhitungan kontras adalah sebagai berikut:

$$ASM = \sum_{i=1}^L \sum_{j=1}^L (GLCM(i,j))^2$$

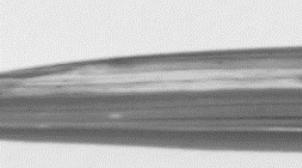
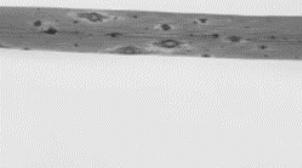
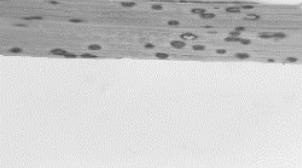
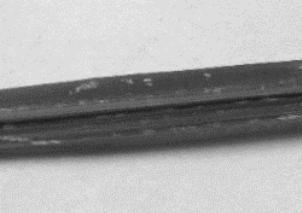
- 3) *Energi:* Energi memiliki fungsi untuk ukuran konsentrasi pasangan dengan intensitas keabuan. Perhitungan energi adalah sebagai berikut:

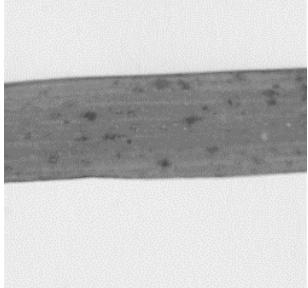
$$Energy = \sum_{i,j} (P(i,j)^2)$$

- 4) *Korelasi:* Korelasi memiliki fungsi untuk mengukur ketergantungan linear antar nilai aras

keabuan dalam sebuah citra. perhitungan korelasi adalah sebagai berikut:

$$Correlation = \frac{\sum_{i=1}^L \sum_{j=1}^L (i - \mu_i)(j - \mu_j)(GLCM(i,j))}{\sqrt{\sigma_i^2 \sigma_j^2}}$$

Hasil Grayscale	Kontras-Homogenitas-Energi-Korelasi
	11.14, 0.522, 0.155, 0.998
	40.66, 0.737, 0.376, 0.985
	82.71, 0.564, 0.126, 0.953
	57.26, 0.240, 0.045, 0.993

	25.93, 0.489, 0.202, 0.995
---	-----------------------------------

3) Pembagian Data

Sebelum masuk dalam proses pelatihan model klasifikasi, dataset dibagi secara merata dalam penelitian ini. Seperti yang ditunjukkan pada Tabel IV, dataset dibagi menjadi dua bagian, yaitu bagian pelatihan dan bagian pengujian, dengan rasio 80% bagian untuk data pelatihan, 20% bagian untuk data pengujian. Seluruh dataset terdiri dari 600 gambar, yang mencakup 5 penyakit pada daun padi dengan jumlah gambar yang seimbang di setiap kelas. Dataset pelatihan terdiri dari 420 gambar dengan 96 gambar per kelas dan dataset pengujian terdiri dari 120 gambar dengan 24 gambar per kelas.

TABEL IV
DISTRIBUSI DATA TRAIN DAN TESTING

Kategori	Training	Testing	Jumlah Gambar
Bacterial Blight	96	24	120
Leaf Smut	96	24	120
Leaf Blast	96	24	120
Brown Spot	96	24	120
Hispa	96	24	120
Jumlah Gambar keseluruhan			600

4) Support Vector Machine

Support Vector Machine (SVM) mulai diperkenalkan pada 1992 oleh Vladimir Vapnik, Bernhard Boser dan Isabelle Guyon. Cara kerja SVM adalah menemukan *hyperplane* terbaik dengan mencari jarak terjauh antar kelas. *Hyperplane* memiliki fungsi untuk memisahkan satu kelas dengan kelas lainnya [13].

Dalam penggunaannya, SVM memiliki parameter penting yang mempengaruhi kinerja klasifikasi. Pemilihan parameter yang tepat sangat berpengaruh terhadap kemampuan model untuk menghasilkan *hyperplane* optimal yang memisahkan data secara akurat. Parameter-parameter utama pada SVM antara lain kernel yang berfungsi memetakan data ke ruang berdimensi lebih tinggi sehingga memungkinkan pemisahan data yang tidak linear. Beberapa jenis kernel yang biasa digunakan antara lain *linear*, *Radial Basis Function* (RBF), dan *polynomial*.

Parameter kedua adalah gamma (γ), yang mengontrol seberapa jauh pengaruh satu titik data terhadap pembentukan *hyperplane*. Nilai gamma yang besar membuat model lebih fokus pada data di sekitar margin, sedangkan nilai gamma kecil membuat model lebih general dengan pengaruh yang lebih luas.

Parameter terakhir adalah C, yaitu *cost* parameter yang mengatur keseimbangan antara margin yang besar dan kesalahan klasifikasi. Nilai C yang tinggi akan membuat model meminimalkan kesalahan dengan margin yang sempit, sedangkan nilai C yang rendah memberikan toleransi kesalahan lebih besar dengan margin yang lebih lebar.

Pada penelitian ini, untuk menentukan parameter yang paling optimal, pemilihan parameter dibantu dengan *GridSearchCV*, yaitu metode pencarian sistematis terhadap kombinasi parameter yang telah ditentukan sebelumnya. *GridSearchCV* akan mengevaluasi setiap kombinasi nilai kernel, C, dan gamma menggunakan k-fold *cross-validation*, sehingga diperoleh parameter dengan performa tertinggi berdasarkan metrik evaluasi seperti *precision*, *recall*, *accuracy*, dan *F1 score*.

5) Evaluasi Model

Evaluasi model merupakan salah satu tahapan yang penting dalam membangun model *machine learning*. Evaluasi model memiliki tujuan untuk menilai seberapa efektif model yang telah dibuat. Pada penelitian ini, evaluasi model dilakukan dengan menggunakan *confusion matrix* [20]. Pada hasil *confusion matrix*, baris merepresentasikan hasil kelas yang sebenarnya, sedangkan kolom merepresentasikan kelas pada hasil prediksi [21].

Confusion matrix dipilih karena memiliki matriks penilaian yang cukup lengkap, diantaranya *precision*, *recall*, *accuracy*, dan *F1 Score*. *Precision* menghitung seberapa tepat prediksi positif, *recall* menghitung seberapa baik model menemukan data positif, *accuracy* menghitung seberapa banyak prediksi yang benar, dan *F1 Score* merupakan gabungan dari persisi dan *recall*[13].

Persamaan *precision*, *recall*, *accuracy*, dan *F1 score* sebagai berikut:

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+FN} \times 100\%$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari dokumentasi pribadi yang dilakukan di area persawahan dengan kondisi lingkungan yang bervariasi. Proses pengambilan gambar dilakukan menggunakan kamera *smartphone* dengan resolusi tinggi agar kualitas citra memadai untuk keperluan analisis. Dataset yang diperoleh terdiri dari 600 citra daun padi yang terbagi ke dalam 5 kelas, yaitu: *bacterial blight*, *leaf smut*, *leaf blast*, *brown spot* dan *hispa*.

Setiap kelas memiliki jumlah citra yang relatif seimbang agar tidak terjadi bias pada model. Proses dokumentasi dilakukan dengan memperhatikan pencahaayaan alami dan sudut pengambilan gambar yang konsisten untuk meminimalkan perbedaan intensitas cahaya dan bayangan.

B. Preprocessing Data

Tahap preprocessing dilakukan untuk mempersiapkan data agar siap digunakan dalam proses klasifikasi. Data yang telah dikumpulkan melalui dokumentasi pribadi sebelumnya diolah agar memiliki format seragam dan fitur yang relevan. Proses preprocessing ini mencakup beberapa langkah berikut:

1). Resize Citra

Resize citra dilakukan menyeragamkan ukuran input pada tahap ekstraksi fitur dan klasifikasi, seluruh citra *resize* menjadi ukuran 256×256 piksel menggunakan *function cv2.resize* dari pustaka *OpenCV*. Ukuran ini dipilih karena cukup besar untuk mempertahankan detail penting pada daun namun tidak terlalu besar sehingga mempengaruhi efisiensi komputasi.

2). Background Removal

Citra daun padi yang diperoleh memiliki latar belakang yang beragam. Kehadiran latar belakang dapat memengaruhi hasil ekstraksi fitur. Oleh karena itu, dilakukan penghapusan latar belakang secara otomatis menggunakan pendekatan *thresholding* pada ruang warna HSV (*Hue*, *Saturation*, *Value*). Proses ini dilakukan dengan cara mengubah citra dari ruang warna RGB ke HSV menggunakan *cv2.cvtColor*. Kemudian menentukan rentang nilai HSV yang sesuai untuk warna daun hijau menggunakan fungsi *cv2.inRange* untuk membuat masker berdasarkan rentang HSV tersebut dan yang terakhir menggabungkan masker dengan citra asli untuk memperoleh citra daun tanpa latar belakang. Hasil dari tahap ini adalah citra daun yang bersih dan siap diekstraksi fiturnya.

3). Ekstraksi GLCM

Ekstraksi fitur tekstur pada penelitian ini dilakukan menggunakan *Gray Level Co-occurrence Matrix* (GLCM). GLCM merupakan metode statistik yang menganalisis tekstur dengan menghitung seberapa sering

pasangan nilai intensitas piksel (*gray level*) muncul pada jarak dan orientasi tertentu dalam citra. Nilai tersebut kemudian digunakan untuk menghasilkan fitur tekstur seperti *Contrast*, *Homogeneity*, *Energy*, dan *Correlation*. Pada penelitian ini, parameter yang digunakan untuk perhitungan GLCM adalah sebagai berikut:

TABEL V
PARAMETER SERTA PENJELASAN GLCM

Parameter	Penjelasan
<i>Gray</i>	Citra dalam skala keabuan (<i>grayscale</i>), input utama untuk membentuk GLCM
<i>Distances = [5]</i>	Jarak antar piksel yang dibandingkan adalah 5 piksel
<i>Angles = [0]</i>	Arah hubungan piksel adalah horizontal (0 derajat atau ke kanan).
<i>Levels = 256</i>	Banyaknya level intensitas piksel. Jika citra 8-bit, maka 256 level (0–255).
<i>Symmetric = True</i>	Matriks akan disimetriskan. Jadi (i,j) dan (j,i) dihitung bersama.
<i>Normed = True</i>	Matriks akan dinormalisasi menjadi proporsi (nilai antara 0–1).

Fitur yang diambil meliputi:

- 1). *Contrast*: Menggambarkan perbedaan intensitas antara piksel.
- 2). *Homogeneity*: Menunjukkan tingkat keseragaman tekstur.
- 3). *Energy*: Menunjukkan tingkat keteraturan pola citra.
- 4). *Correlation*: Menggambarkan hubungan linear antar piksel.

Ekstraksi fitur dilakukan menggunakan fungsi *graycomatrix* dan *graycoprops* dari pustaka *skimage.feature*.

4). Ekstraksi HSV

Selain tekstur, informasi warna juga berperan penting dalam membedakan citra daun yang sehat dan yang terkena penyakit. Pada penelitian ini, ekstraksi fitur warna dilakukan dengan menggunakan model ruang warna HSV (*Hue*, *Saturation*, *Value*). HSV dipilih karena mampu merepresentasikan warna dengan lebih mendekati persepsi manusia dibandingkan ruang warna RGB.

- 1). *Hue* (H) merepresentasikan jenis warna (misalnya merah, hijau, biru) dengan rentang 0° – 360° .
- 2). *Saturation* (S) menunjukkan tingkat kejemuhan warna, dari 0 (abu-abu) hingga 255 (warna jenius).
- 3). *Value* (V) menggambarkan tingkat kecerahan warna, dari 0 (hitam) hingga 255 (putih).

Proses ekstraksi dilakukan dengan cara mengonversi citra dari ruang warna RGB ke HSV menggunakan fungsi konversi, kemudian menerapkan *masking* dengan menggunakan nilai *threshold* HSV untuk menyeleksi bagian citra yang sesuai dengan warna daun. Dalam penelitian ini, digunakan batas bawah (*lower*) [25, 40, 40] dan batas atas (*upper*) [90, 255, 255]. Rentang ini dipilih karena mewakili warna hijau pada daun dengan variasi saturasi dan kecerahan tertentu. Setelah masking, dilakukan perhitungan rata-rata nilai *Hue*, *Saturation*, dan *Value* pada area yang tersegmentasi.

Penerapan *threshold* ini bertujuan untuk menghilangkan bagian citra yang tidak relevan, seperti latar belakang, sehingga hanya area daun yang dianalisis. Dengan demikian, fitur warna yang dihasilkan lebih representatif terhadap karakteristik daun padi yang sebenarnya.

C. Pembagian Data

Setelah proses ekstraksi fitur selesai, data yang telah diperoleh dibagi menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*). Pembagian ini bertujuan agar model *Support Vector Machine* (SVM) dapat dilatih menggunakan sebagian data, kemudian dievaluasi performanya menggunakan data yang tidak pernah dilihat sebelumnya.

Pada penelitian ini, pembagian dilakukan dengan proporsi 80% untuk data latih dan 20% untuk data uji. Proporsi ini dipilih karena umum digunakan dalam penelitian klasifikasi citra dan mampu memberikan keseimbangan antara data yang digunakan untuk pelatihan dan pengujian.

Pembagian data dilakukan secara acak (*random*) untuk menghindari bias yang dapat terjadi apabila data dikelompokkan secara berurutan. Dengan cara ini, setiap kelas penyakit memiliki representasi yang seimbang baik pada data latih maupun data uji.

Langkah ini sangat penting karena kinerja model akan sangat bergantung pada kualitas dan keseimbangan data yang digunakan dalam pelatihan dan pengujian.

D. Klasifikasi Support Vector Machine

Pada penelitian ini, algoritma *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan penyakit pada daun padi. Proses pelatihan model dilakukan dengan menerapkan Grid Search yang dikombinasikan dengan *K-Fold Cross Validation* ($K = 5$) untuk memperoleh parameter terbaik.

Parameter yang diuji meliputi C , γ , dan kernel dengan rentang nilai yang ditentukan. Daftar parameter yang diuji ditunjukkan pada Tabel VI.

TABEL VI
PARAMETER SERTA PENJELASAN GLCM

Parameter	Nilai yang Diuji
C	[0.1, 1, 10]
γ	[0.001, 0.01, 0.1]
kernel	['linear', 'rbf']

Selanjutnya, validasi silang (*cross-validation*) dilakukan untuk setiap kombinasi parameter. Hasil rata-rata akurasi pada proses validasi silang ditunjukkan pada Tabel VII.

TABEL VII
PARAMETER SERTA PENJELASAN GLCM

Parameter			
C	Γ	Kernel	Accuracy
100	0.100	rbf	0.83
100	0.100	rbf	0.82
100	0.10	rbf	0.80
1	0.100	rbf	0.78
10	0.10	rbf	0.74
0.1	0.100	rbf	0.73
100	0.001	rbf	0.73
10	0.001	linear	0.74
10	0.100	linear	0.74
10	0.010	linear	0.74

Berdasarkan Tabel VII kombinasi parameter terbaik yang diperoleh adalah $\text{kernel} = \text{RBF}$, $C = 100$, dan $\gamma = 0.100$ dengan nilai akurasi rata-rata validasi sebesar 83%. Nilai ini menunjukkan bahwa model memiliki performa yang cukup baik selama proses pelatihan.

E. Evaluasi

Setelah mendapatkan parameter terbaik, model SVM dilatih ulang menggunakan data latih, kemudian diuji menggunakan data uji. Hasil evaluasi model pada data uji ditunjukkan pada Tabel VIII, yang berisi nilai *precision*, *recall*, dan *F1-score* untuk setiap kelas. Secara keseluruhan, model mampu mencapai tingkat akurasi sebesar 83%, yang menunjukkan performa prediksi yang konsisten.

TABEL VIII
HASIL EVALUASI

Kelas	Precision	Recall	F1-Score	Support
1	0.88	0.92	0.90	24
2	0.77	1.00	0.87	24
3	0.95	0.83	0.89	24
4	0.71	0.83	0.77	24
5	0.93	0.58	0.72	24
Akurasi			0.83	120

Berdasarkan hasil evaluasi, secara keseluruhan model SVM berhasil mencapai akurasi sebesar 83% pada data uji. Secara detail, nilai *precision* tertinggi diperoleh pada kelas 3 dengan skor 0,95, yang menunjukkan kemampuan model

dalam memprediksi kelas tersebut dengan tingkat kesalahan yang sangat rendah. Kelas ini juga memiliki *F1-score* sebesar 0,89, yang menandakan keseimbangan antara *precision* dan *recall* yang baik.

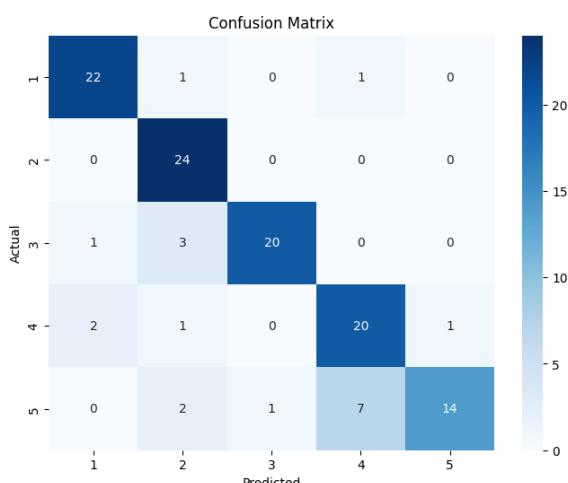
Sebaliknya, kelas 4 dan kelas 5 menunjukkan performa yang relatif lebih rendah dibandingkan kelas lainnya. Pada kelas 4, *precision* sebesar 0,71 dan *F1-score* 0,77 mengindikasikan adanya beberapa kesalahan prediksi terhadap kelas ini. Hal serupa terjadi pada kelas 5, dengan *precision* 0,93 namun *recall* hanya 0,58, menunjukkan bahwa model cukup sering salah dalam mendeteksi kelas ini sehingga banyak sampel yang seharusnya masuk kelas 5 tidak teridentifikasi dengan benar.

Rata-rata keseluruhan (*macro average*) untuk *precision*, *recall*, dan *F1-score* masing-masing adalah 0,85, 0,83, dan 0,83. Hal ini menunjukkan performa yang cukup baik untuk model yang diterapkan pada lima kelas yang berbeda. *Weighted average* juga memiliki nilai yang sama (0,85; 0,83; 0,83), menandakan konsistensi hasil pada keseluruhan data uji.

Secara umum, hasil ini membuktikan bahwa SVM dengan parameter yang telah ditentukan mampu mengklasifikasikan penyakit daun padi dengan tingkat akurasi yang cukup baik, meskipun terdapat beberapa kelas yang masih perlu dioptimalkan, khususnya pada kelas yang memiliki *recall* rendah.

F. Confusion Matrix

Untuk mengetahui distribusi prediksi model terhadap data uji, digunakan *Confusion Matrix* yang ditampilkan pada Gambar 2.



Gambar 2. Confusion Matrix

Confusion matrix ini menggambarkan distribusi prediksi model terhadap label aktual dari lima kelas penyakit daun padi.

- Kelas 1 (Aktual 1) diprediksi dengan benar sebanyak 22 sampel, sedangkan terdapat 1 sampel yang salah diklasifikaikan sebagai kelas 2 dan 1 sampel

diklasifikasi sebagai kelas 4. Hal ini menunjukkan bahwa model memiliki tingkat ketepatan yang tinggi untuk kelas ini.

- Kelas 2 diprediksi sempurna dengan 24 sampel benar dan tidak ada kesalahan prediksi ke kelas lain. Ini mengindikasikan bahwa ciri-ciri pada kelas 2 dapat dikenali dengan sangat baik oleh model.
- Kelas 3 memiliki 20 prediksi benar, namun terdapat 1 sampel yang salah diklasifikasi ke kelas 1 dan 3 sampel ke kelas 2. Kesalahan prediksi ini kemungkinan terjadi karena adanya kemiripan fitur antara kelas 3 dengan kelas 1 dan 2.
- Kelas 4 diprediksi dengan benar sebanyak 20 sampel, tetapi terdapat 2 sampel yang diklasifikasi sebagai kelas 1, 1 sampel sebagai kelas 2, dan 1 sampel sebagai kelas 5. Hal ini menunjukkan bahwa model masih mengalami kebingungan dalam membedakan kelas 4 dari kelas lain.
- Kelas 5 memiliki performa paling rendah, dengan hanya 14 prediksi benar dari total data uji kelas ini. Terdapat 2 sampel yang diklasifikasi sebagai kelas 2, 1 sampel sebagai kelas 3, dan 7 sampel sebagai kelas 4. Tingginya kesalahan pada kelas ini menunjukkan bahwa ciri visual kelas 5 cenderung mirip dengan kelas 4, sehingga model sulit membedakannya.

Secara keseluruhan, *confusion matrix* menunjukkan bahwa model memiliki performa yang baik pada kelas 1, 2, dan 3, namun mengalami kesulitan pada kelas 4 dan terutama kelas 5. Kesalahan prediksi yang dominan antara kelas 4 dan 5 mengindikasikan bahwa fitur-fitur kedua kelas ini perlu dipisahkan lebih jelas, misalnya dengan menambahkan teknik *feature extraction* yang lebih mendalam atau menggunakan metode augmentasi data untuk menyeimbangkan distribusi dan variasi samp

G. Perbandingan dengan Algoritma lain

Untuk mengevaluasi keunggulan algoritma *Support Vector Machine* (SVM), dilakukan perbandingan kinerja dengan beberapa algoritma lain yang umum digunakan pada permasalahan klasifikasi, yaitu *K-Nearest Neighbor* (KNN), *Naive Bayes* (NB), dan *Decision Tree* (DT). Proses pelatihan dan pengujian dilakukan dengan dataset yang sama, serta metrik evaluasi yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil perbandingan kinerja ditunjukkan pada Tabel IX.

Algoritma	HASIL EVALUASI			
	Akurasi	Precision	Recall	F1-score
SVM	0.83	0.84	0.83	0.83
KNN (k=5)	0.81	0.82	0.81	0.81
Naive Bayes	0.68	0.72	0.68	0.67
DecisionTree	0.82	0.83	0.82	0.82

Berdasarkan Tabel VIII dapat dilihat bahwa SVM dengan kernel RBF menunjukkan performa terbaik

dibandingkan algoritma lainnya. Model SVM memperoleh akurasi sebesar 0.83%, lebih tinggi dibandingkan KNN (81%), Decision Tree (82%), dan Naive Bayes (68%). Selain itu, nilai *precision*, *recall*, dan *F1-score* pada SVM juga konsisten lebih tinggi, yang mengindikasikan bahwa model ini tidak hanya akurat, tetapi juga seimbang dalam mendeteksi kedua kelas (sehat dan terinfeksi).

Kinerja *Decision Tree* menduduki peringkat kedua dengan akurasi 82%, diikuti oleh KNN dengan akurasi 81%, sedangkan *Naive Bayes* memiliki performa terendah, yaitu 68%. Rendahnya kinerja *Naive Bayes* kemungkinan disebabkan oleh asumsi independensi antar fitur yang tidak sepenuhnya terpenuhi pada dataset ini.

Secara keseluruhan, hasil ini mengindikasikan bahwa algoritma SVM dengan kernel RBF merupakan metode yang paling tepat untuk mengklasifikasikan penyakit pada daun padi dalam penelitian ini.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Support Vector Machine* (SVM) dapat digunakan secara efektif untuk melakukan klasifikasi jenis penyakit pada daun padi. Penelitian ini menggunakan dataset sebanyak 600 citra daun padi yang terdiri dari lima jenis penyakit, yaitu *bacterial blight*, *leaf smut*, *leaf blast*, *brown spot*, dan *hispa*. Proses klasifikasi diawali dengan tahapan *preprocessing* citra, termasuk *resize* dan segmentasi warna, kemudian dilanjutkan dengan ekstraksi fitur warna (HSV) dan tekstur (GLCM). Hasil ekstraksi kemudian digunakan untuk melatih dan menguji model SVM menggunakan parameter kernel RBF, C=100, dan gamma=0,1. Evaluasi model menggunakan *confusion matrix* menunjukkan akurasi keseluruhan sebesar 83%, dengan nilai rata-rata *precision*, *recall*, dan *F1-score* yang relatif tinggi. Meskipun demikian, performa klasifikasi pada kelas 5 (*hispa*) masih perlu ditingkatkan terutama pada aspek *recall*. Secara keseluruhan, pendekatan ini menunjukkan potensi yang baik untuk membantu proses deteksi dini penyakit pada tanaman padi, yang pada akhirnya dapat membantu petani dalam mencegah penurunan hasil panen secara signifikan.

DAFTAR PUSTAKA

- [1] BPS, “Produksi Gabah Kering Giling.” [Online]. Available: <https://www.bps.go.id/pressrelease/2025/02/03/2414/pada-2024--luas-panen-padi-mencapai-sekitar-10-05-juta-hektare-dengan-produksi-padi-sebanyak-53-14-juta-ton-gabah-kering-giling--gkg--.html>
- [2] BPS, “Laju Pertumbuhan Penduduk.” Accessed: Jun. 11, 2025. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/MTk3NiMy/laju-pertumbuhan-penduduk.html>
- [3] Y. Defitri, “Identifikasi jamur patogen penyebab penyakit pada tanaman padi (*Oryza sativa*) di Lubuk Ruso kecamatan Pemayung kabupaten Batanghari Jambi,” *J. Ilm. Univ. Batanghari Jambi*, vol. 13, no. 4, pp. 113–117, 2018.
- [4] Y. Amrozi, D. Yuliati, A. Susilo, N. Novianto, and R. Ramadhan, “Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 394–399, 2022, doi: 10.32736/sisfokom.v11i3.1502.
- [5] B. W. Kurniadi, H. Prasetyo, G. L. Ahmad, B. Aditya Wibisono, and D. Sandya Prasvita, “Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 1–11, 2021.
- [6] I. Irma, M. Muchtar, R. Adawiyah, and S. Sarimuddin, “Klasifikasi Tingkat Kematangan Cabai Merah Keriting Menggunakan Svm Multiclass Berdasarkan Ekstraksi Fitur Warna,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 1747–1755, 2024, doi: 10.23960/jitet.v12i3.4430.
- [7] M. Lestari, “Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) untuk Mendeteksi Penyakit Jantung,” *Fakt. Exacta*, vol. 7, no. September 2010, pp. 366–371, 2014.
- [8] S. D. Kamil, D. Widiyanto, and N. Chamidah, “Perbandingan Metode Decision Tree dengan Naive Bayes Dalam Klasifikasi Tumor Otak Citra MRI,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 539–550, 2020.
- [9] A. T. Cahyani, “Algoritma Random Forest , Xgboost Dan Decision Tree Dengan Ekstraksi Fitur Algoritma Random Forest , Xgboost Dan Decision Tree Dengan Ekstraksi Fitur,” 2024.
- [10] M. N. M. Hakim, A. B. Nugroho, and A. E. Minarno, “Prediksi Tumor Otak Menggunakan Metode Convolutional Neural Network,” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 17, no. 1, p. 48, 2023, doi: 10.30872/jim.v17i1.5246.
- [11] Meiriyama, “Klasifikasi Citra Buah berbasis fitur warna HSV dengan klasifikatorSVM,” *J. Komput. Terap.*, vol. 4, no. 1, pp. 50–61, 2018, [Online]. Available: <http://jurnal.pcr.ac.id>
- [12] N. Neneng, K. Adi, and R. Isnanto, “Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM),” *J. Sist. Inf. Bisnis*, vol. 6, no. 1, p. 1, 2016, doi: 10.21456/vol6iss1pp1-10.
- [13] A. Arifin, J. Hendyli, and D. E. Herwindiati, “Klasifikasi Tanaman Obat Herbal Menggunakan Metode Support Vector Machine,” *Comput. J. Comput. Sci. Inf. Syst.*, vol. 5, no. 1, p. 25, 2021, doi: 10.24912/computatio.v1i1.12811.
- [14] Ridho Surya Pangestu, Hari Purwadi, and Agusma Wajiansyah, “Ekstraksi Ciri Image Wajah Berdasarkan Ciri Warna Hue Saturation Value (HSV) dan Geometri,” *J. Komputer, Inf. dan Teknol.*, vol. 5, no. 1, p. 11, 2025, doi: 10.53697/jkomitek.v5i1.2472.
- [15] Adela Regita Azzahra, Purnawansyah, H. Darwis, and D. Widayati, “Klasifikasi Daun Herbal Menggunakan Metode CNN dan Naïve Bayes dengan Fitur GLCM,” *Indones. J. Comput. Sci.*, vol. 12, no. 4, 2023, doi: 10.33022/ijcs.v12i4.3362.
- [16] Pramudiya, Cerwyn Asyraq, Aldo Kadafi, and Ricky Putra Sardika, “Analisis Gambar Menggunakan Metode Grayscale Dan Hsv (Hue, Saturation, Value),” *Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 14, no. 3, pp. 174–180, 2024.
- [17] R. P. Putra, J. Jumadi, and D. Lianda, “Pengolahan Citra Digital Untuk Mengidentifikasi Tingkat Kematangan Buah Kelapa Sawit Berdasarkan Warna Rgb Dan Hsv Dengan Menggunakan Metode Self Organizing Map (SOM),” *J. Media Infotama*, vol. 20, no. 1, p. 341149, 2024.
- [18] J. Sofian and R. H. Laluma, “Jenis Tumor Otak Dengan Metode Image Threshold Dan Glcm Menggunakan Algoritma K-Nn (Nearest Neighbor) Classifier Berbasis Web,” *J. Infotronik*, vol. 4, no. 2, pp. 51–56, 2019.
- [19] F. Siqueira, W. Schwartz, and H. Pedrini, “Multi-Scale Gray Level Co-Occurrence Matrices for Texture Description,” *Neurocomputing*, vol. 120, pp. 1–10, 2013, doi: 10.1016/j.neucom.2012.09.042.
- [20] M. A. Ramadhan and R. Andarsyah, *Klasifikasi Text Spam Menggunakan Metode Support Vector Machine Dan Naïve Bayes*. Bandung: Penerbit Buku Pedia, 2022.
- [21] J. Xu, Y. Zhang, and D. Miao, “Three-way confusion matrix for classification: A measure driven view,” *Inf. Sci. (New York)*, vol. 507, pp. 772–794, 2020, doi: 10.1016/j.ins.2019.06.064.