

Klasifikasi Daun Mangga Yang Terkena Hama Dengan Metode Gray Level Co-occurrence Matrix Menggunakan Support Vector Machine Dan K-Nearest Neighbor Berbasis Data Kaggle

Reva Nursyawalni^{*1}, Jamaludin Indra², Tatang Rohana³, Deden Wahiddin⁴

^{1,2,3,4}Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang, Indonesia

Email: ¹if21.revanursyawalni@mhs.ubpkarawang.ac.id, ²jamaludin.indra@ubpkarawang.ac.id,
³tatang.rohana@ubpkarawang.ac.id, ⁴deden.wahiddin@ubpkarawang.ac.id

Abstrak

Penurunan produksi buah mangga di sebabkan oleh kerusakan atau serangan hama pada daun mangga ada beberapa jenis hama pada daun mangga yang umum menyerang antara lain kutu daun (*Aphis gossypii*), bercak daun *alternaria*, *anthracnose*, penggerek batang dan lain-lain. Untuk memperoleh hasil klasifikasi yang lebih akurat dan performa model yang optimal, dibutuhkan sistem yang mampu menghasilkan tingkat akurasi terbaik. Sebagai respons terhadap urgensi tersebut, penelitian ini bertujuan untuk mengklasifikasikan daun mangga yang terkena hama dengan memanfaatkan algoritma *Support Vector Machine* dan *K-Nearest Neighbor*, serta penggunaan *Gray Level Co-occurrence Matrix* sebagai metode untuk mengekstraksi tekstur gambar. Rangkaian tahapan dalam penelitian ini meliputi *pre-processing*, augmentasi data, ekstraksi fitur, proses klasifikasi oleh kedua algoritma, dan dievaluasi menggunakan akurasi. Hasilnya, algoritma *Support Vector Machine* dengan kernel *Radial Basis Function* mencapai 78% untuk algoritma *K-Nearest Neighbor* mencapai akurasi 80% dengan ketanggaan $k=3$

Kata kunci: Daun Mangga, Klasifikasi, Support Vector Machine, K-Nearest Neighbor, Gray Level Co-occurrence Matrix

Classification of Pest-Infected Mango Leaves Using Gray Level Co-occurrence Matrix Method with Support Vector Machine and K-Nearest Neighbor Algorithms Based on Kaggle Dataset

Abstract

The decline in mango fruit production is primarily caused by damage or pest infestations on mango leaves. Several common pests that typically attack mango leaves include aphids (*Aphis gossypii*), *Alternaria* leaf spots, *anthracnose*, stem borers, and others. To achieve more accurate classification results and optimal model performance, a system capable of producing the highest possible accuracy is required. In response to this urgency, this study aims to classify pest-infected mango leaves using the *Support Vector Machine (SVM)* and *K-Nearest Neighbor (K-NN)* algorithms, with the *Gray Level Co-occurrence Matrix (GLCM)* method employed for texture feature extraction. The stages of this research include *pre-processing*, data augmentation, feature extraction, classification using both algorithms, and evaluation based on accuracy. The results show that the SVM algorithm with a *Radial Basis Function (RBF)* kernel achieved an accuracy of 78%, while the K-NN algorithm achieved 80% accuracy with $k = 3$.

Keywords: Mango Leaves, Classification, Support Vector Machine, K-Nearest Neighbor, Gray Level Co-occurrence Matrix

1. PENDAHULUAN

Tanaman mangga merupakan komoditas hortikultura yang memiliki potensi produksi signifikan dan dapat dikembangkan berkat tingginya tingkat keanekaragaman genetik [1]. Pohon mangga ditanam di seluruh Indonesia karena tanaman hortikultura khas daerah tropis [2]. Hortikultura adalah penggabungan antara bahasa latin, 'hortus' dan 'culture', merujuk pada kegiatan budidaya tanaman di sekitar pemukiman [3]. Buahnya mempunyai banyak manfaat, yaitu salah satu jenis antioksidan yang membantu melindungi tubuh dari kerusakan sel. Mangga mengandung antioksidan dalam jumlah tinggi yang efektif untuk mencegah serta mengurangi efek kerusakan kulit yang disebabkan oleh sinar UV dan proses penuaan [4].

Menurut data BPS Produksi buah mangga di Indonesia, memperoleh total 33.088.950 kuintal pada tahun 2022. Provinsi Jawa Timur adalah wilayah dengan produksi tertinggi, diikuti oleh provinsi Jawa Tengah, Jawa Barat, Nusa Tenggara Barat, dan Sulawesi Selatan merupakan sentra produksi utama, dengan akumulasi hasil panen sebesar 28.180.450 kuintal [5]. Namun, pada tahun 2024 menurut Badan Pusat Statistik, Indonesia menghasilkan buah mangga sekitar 33.000.000 kuintal per tahunnya [6]. Penurunan produksi buah mangga disebabkan oleh adanya kerusakan pada daun mangga. Daun adalah bagian penting dari tumbuhan karena berfungsi sebagai tempat terjadinya fotosintesis dan respirasi, yang menghasilkan berbagai nutrisi bagi tanaman. Oleh karena itu, kerusakan atau serangan hama pada daun dapat mempengaruhi pertumbuhan buah mangga, bahkan berpotensi menyebabkan kematian pada tanaman tersebut [1].

Beberapa jenis hama yang umum menyerang daun mangga antara lain kutu daun (*Aphis gossypii*), bercak daun *Alternaria*, *anthracnose*, penggerek batang, *deformitas* mangga, serangan *webber*, dan infeksi *gall* adalah beberapa hama umum pada daun mangga yang disebabkan oleh patogen seperti mikroorganisme, virus, dan kondisi iklim yang tidak diinginkan [7]. Hama ini dapat menyebabkan kerugian ekonomi bagi petani dan pemilik, tanpa penanganan yang tepat, kerusakan pada daun ini akan mempengaruhi pertumbuhan tanaman dan mengurangi produktivitas tanaman mangga [8].

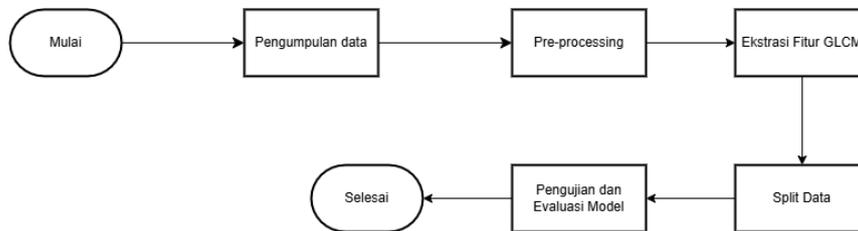
Untuk mengatasi tantangan ini, teknologi klasifikasi oleh *machine learning* dapat menjadi solusi untuk mengklasifikasikan daun mangga yang terkena hama menggunakan algoritma SVM dan KNN dengan metode GLCM. Kedua algoritma ini adalah data dari *supervised learning* [9]. Pembelajaran terawasi (*supervised learning*) merupakan pendekatan dalam machine learning di mana proses pelatihan dilakukan dengan bantuan data berlabel. Label ini berperan sebagai panduan atau pengawas yang membantu model belajar untuk menghasilkan prediksi dengan tingkat akurasi tertentu [10]. Sebagai algoritma *supervised learning*, SVM banyak digunakan dalam berbagai aplikasi klasifikasi karena efektivitasnya dalam menangani baik data linear maupun *non-linear* [11]. Algoritma ini juga dapat menyelesaikan masalah klasifikasi dan regresi. Performa dari klasifikasi ini diambil dan dibandingkan mana yang menjadi klasifikasi terbagus [12]. Algoritma KNN mengklasifikasikan objek baru dengan menggunakan sampel pelatihan, algoritma ini mengidentifikasi jarak terdekat antara data yang sedang dianalisis dengan K tetangga terdekat yang ada dalam dataset pelatihan [13]. Kedua algoritma ini dibantu oleh ekstraksi fitur untuk mengambil ciri atau fitur dari suatu bentuk atau tekstur [14]. Dalam bidang pengolahan citra, GLCM digunakan untuk mengekstraksi tekstur gambar dengan mempelajari pola hubungan antar piksel yang berdekatan [15]. Metode perhitungan ini bekerja efektif dalam menghilangkan gangguan atau artefak pada citra. Di samping itu, teknik ini dapat menampilkan perbedaan tekstur secara tegas. Dengan memanfaatkan GLCM, pola kemunculan piksel tertentu di berbagai wilayah gambar dapat terdeteksi secara tepat [16].

Adapun penelitian dari Sari dkk (2021) [17], dalam penelitian yang berjudul "*Classification of Chili Leaf Disease Using the Gray Level Co-occurrence Matrix (GLCM) and the Support Vector Machine (SVM) Methods*", diperoleh akurasi sebesar 88% saat klasifikasi dilakukan terhadap 5 kelas penyakit daun cabai. Jumlah data uji yang digunakan adalah 25 gambar, dengan 5 gambar untuk setiap penyakit daun cabai yang diklasifikasikan. Penelitian lainnya oleh Supriyanto dkk (2023) [18], penelitian tersebut berjudul "*Robusta Coffee Leaf Disease Classifications Using SVM Method and GLCM Feature Extraction*" memperoleh 320 data latih dan 80 data uji. Jumlah sampel daun kopi dengan bercak daun dan karat adalah 200 daun, di mana setiap jenis penyakit pada daun terdiri dari seratus enam puluh data latih dan empat puluh data uji dengan menggunakan kernel Gaussian RBF dan pendekatan *Gray Level Co-occurrence* yang dihasilkan, seperti ASM atau energi, kontras, korelasi, IDM atau homogenitas, dan entropi dengan sudut 0°, 45°, 90°, 135°, serta sudut rata-rata. Akurasi tertinggi mencapai 97,5%, presisi 95,24%, recall 100%, dan F1-score 97,56%. Penelitian yang menggunakan algoritma KNN oleh A.A.Huda dkk (2022) [19], penelitian ini berjudul "Implementasi *Gray Level Co-occurrence Matrix (GLCM)* untuk Klasifikasi Penyakit Daun Padi," hasil klasifikasi menunjukkan akurasi maksimum sebesar 93,3% berhasil dicapai pada proses klasifikasi tiga kelas daun, dengan komposisi data latih sebanyak 210 dan data uji sebanyak 30 sampel. Sementara itu, dalam penelitian berjudul "Klasifikasi Citra Daun dengan GLCM (*Gray Level Co-Occurrence*) dan K-NN (*K-Nearest Neighbor*)", dilakukan analisis klasifikasi untuk delapan jenis citra daun dengan memanfaatkan empat ratus data latih dan delapan puluh data uji. Hasilnya, akurasi tertinggi yang diperoleh dalam proses klasifikasi mencapai 98% oleh Rosiva dkk (2022) [20].

Penelitian ini memiliki perbedaan dengan yang sebelumnya, pada penelitian ini melakukan klasifikasi terhadap daun mangga yang terkena hama menggunakan 2 algoritma yaitu, SVM dan K-NN dengan memanfaatkan metode ekstraksi fitur GLCM. Penelitian ini diharapkan mampu mengklasifikasikan daun mangga yang terkena hama dan memungkinkan akan meningkatkan akurasi yang baik.

2. METODE PENELITIAN

2.1. Tahap Penelitian



Gambar 1. Tahap Penelitian

Prosedur penelitian disusun melalui beberapa tahapan yang sistematis dan berurutan, dimulai dari pengumpulan data, *pre-processing*, ekstrasi fitur, *split* data, pengujian dan evaluasi model. Diatas merupakan gambaran flowchart penelitian. Berikut ini merupakan tahap penelitian

2.1.1. Pengumpulan Data

Sebuah teknik yang memperoleh informasi dari sekumpulan sumber yang relevan dengan subjek atau tujuan penelitian merupakan pengertian dari pengumpulan dataset [21]. Dataset ini berisi jenis-jenis daun mangga yang terkena hama dengan total 4000 gambar dengan memiliki 8 folder dan sudah terlabel dari platform kaggle

2.1.2. Pre-Processing

Pre-processing adalah langkah persiapan data untuk memastikan kelancaran dan keteraturan dalam tahap pemrosesan selanjutnya [21]. *Pre-processing* ini memuat tahapan *resize* dan menggunakan augmentasi *flip* horizontal

a. *Resize*

Pada tahap ini gambar akan di *resize* menjadi 100x100 piksel karena gambar tersebut beragam maka dibutuhkannya sebuah *resize* dalam mempersiapkan gambar sebelum diuji

b. *Augmentasi*

Gambar akan di *augmentasi* menggunakan sebuah *flip* dengan menggunakan fungsi pemanggil *cv2* dan *horizontal* yang dibalikkan ke kiri serta ke kanan yang nilainya adalah 1

2.1.3. Ekstrasi Fitur GLCM

Untuk mengolah data-data yang ada perlu dilakukan proses ekstrasi fitur yang memiliki tujuan yaitu menemukan data-data yang paling baik dan berguna dalam proses *learning* yang dilakukan [22]. Metode GLCM digunakan untuk mengambil ciri tekstur dari gambar dengan melihat pola hubungan antara piksel-piksel yang saling berdekatan [23]. Fitur tekstur yang digunakan dalam studi ini meliputi rata-rata, varians, kontras, dissimilarity, homogenitas, energi, korelasi, dan momen kedua sudut. Fitur-fitur ini dapat digunakan untuk pelatihan lebih lanjut dari algoritma machine learning seperti *Random Forest* atau *ExtraTrees Classifier*, yang membangun beberapa pengklasifikasi pohon keputusan selama pelatihan dan menggabungkan prediksi mereka untuk membuat keputusan akhir [24]. *Contrast* merupakan ukuran distribusi elemen matriks yang mempresentasikan perbedaan warna atau skala keabuan, dimana nilai yang semakin tinggi menunjukkan perbedaan warna atau skala keabuan yang semakin besar. Sementara itu, *energy* mengukur keseragaman dan disebut sebagai *angular second moment*, dengan nilai tinggi menunjukkan kemiripan piksel dan nilai rendah menunjukkan heterogenitas [25]. Berikut ini merupakan rumus yang selalu dipakai oleh ekstrasi fitur GLCM

$$Contrast = \sum_{(i,j)}(i - j)^2 P(i, j) \tag{1}$$

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

$$IDM = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i, j) \tag{3}$$

2.1.4. Split Data

Pembagian dataset merupakan langkah krusial yang dapat memengaruhi performa algoritma klasifikasi [25]. Dalam studi ini, dataset dipisahkan menjadi dua subset, yakni 80% *training* dan 20% *testing*. Pemisahan tersebut dimaksudkan agar model dapat dilatih pada sebagian data, Selanjutnya, model diuji dengan menggunakan data yang tidak digunakan dalam fase pelatihan. Proses ini dilakukan secara sistematis untuk mempertahankan proporsi kelas yang seimbang di kedua subset.

2.1.5. Pengujian dan Evaluasi Model

Pengujian kali ini menggunakan model SVM kernel RBF dan K-NN dengan $k=3$ untuk mengklasifikasikan daun mangga yang terkena hama berdasarkan tekstur daun

a. Support Vector Machine

Model dimana metode sering diterapkan dalam proses klasifikasi data adalah model SVM (*Support Vector Machine*), dengan cara memisahkan data ke dalam dua kelompok melalui pencarian *hyperplane* yang optimal [26]. Pada tahap ini digunakannya sebuah kernel RBF. Dalam kernel RBF, fungsi Gaussian digunakan sebagai dasar untuk menilai seberapa mirip dua vektor input dalam domain fitur [27]

b. K-Nearest Neighbor

K-NN yaitu proses klasifikasi sederhana [28] dimana model bekerja dengan memperhitungkan jarak antara data baru dan data *training*, memilih k tetangga terdekat, dan menetapkan kelas berdasarkan mayoritas dari tetangga tersebut [28]. Syarat nilai k adalah tidak boleh lebih besar dari jumlah data latih, dan nilai k tidak boleh harus ganjil [29]. Nilai k pada K-NN harus menggunakan nilai ganjil jika digunakan untuk proses klasifikasi [30], nilai k yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antar setiap klasifikasi menjadi lebih kabur [29]. Memilih K-NN dengan $K=3$ dikatakan memenuhi syarat untuk proses klasifikasi. Keunggulan algoritma K-NN bekerja dengan sangat baik untuk *dataset* kecil dan sedang, tetapi kurang efektif saat digunakan untuk *dataset* besar atau dengan banyak fitur [31]. Algoritma ini akan bekerja dengan cara mencari K data terdekat dengan data yang ingin diprediksi. Berdasarkan kelas mayoritas dari data tetangga terdekat tersebut, prediksi akan dibuat.

Setelah selesai proses pengujian dilanjutkan dengan tahap evaluasi yaitu menggunakan *Confusion Matrix*. Evaluasi ini akan dipakai untuk mengukur seberapa baik model yang diterapkan setelah tahap *modelling* selesai. *Confusion Matrix* menentukan performa model dengan nilai-nilai yang terkandung didalamnya. Yaitu, akan menggunakan akurasi presisi, *recall*, dan *f1-score*. Data yang telah di pisahkan kemudian di *testing* untuk menghitung akurasi, presisi, *recall*, dan *f1-score*. Untuk menghitung nilai tersebut *Confusion Matrix* yang mencakup empat kategori evaluasi: TP (*True Positive*), TN (*True Negatif*), FP (*False Negatif*), dan FN (*False Negatif*). Berikut untuk menghitung evaluasi:

a. Akurasi (*Accuracy*): Mengukur proporsi keseluruhan prediksi yang benar, baik untuk kelas positif maupun negatif.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (4)$$

b. Presisi (*Precision*): Mengukur seberapa tepat model dalam memprediksi kasus positif.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

c. Recall (*Sensitivity*): Mengukur sejauh mana model mampu mengidentifikasi kasus positif.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

d. F1-Score: Merupakan rata-rata harmonis antara presisi dan recall, yang memberikan gambaran keseimbangan antara keduanya.

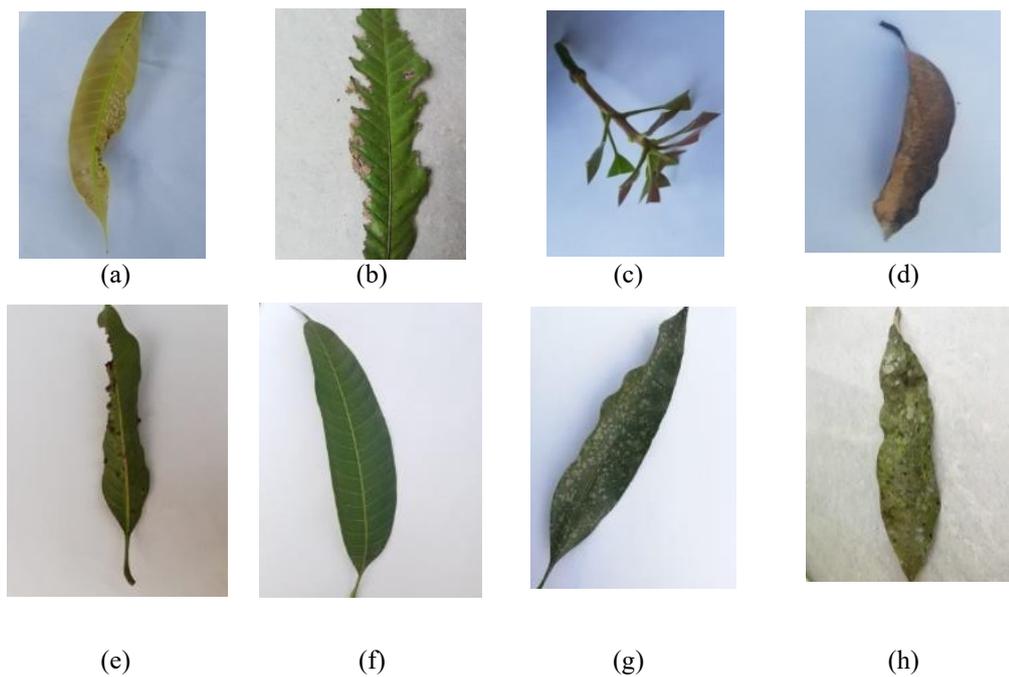
$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Setelah melewati rancangan tahapan pengumpulan data, *preprocessing*, ekstraksi fitur, split data, pengujian dan evaluasi model, maka hasil pengujian klasifikasi daun mangga yang terkena hama akan dijelaskan dan dibahas lebih lanjut.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Teknik pertama yang sangat penting adalah pengumpulan data. Dalam dataset ini adalah sebuah gambar daun mangga yang terkena berbagai jenis hama. Dataset yang digunakan terdiri dari gambar-gambar yang dilabeli secara jelas. Gambar-gambar ini berasal dari *platform Kaggle* yang telah dipilih untuk kualitas dan keberagaman gambar yang dimilikinya. Dataset ini disediakan dalam beberapa folder setiap folder bervariasi. Total keseluruhannya yaitu 4000 gambar dengan ukuran 240×320 piksel berwarna RGB berbentuk format JPG . Gambar daun mangga yang terkena hama dapat ditunjukkan pada gambar 2.

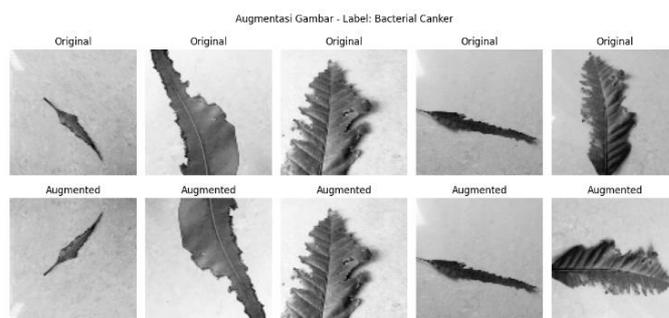


Gambar 1. (a). *Anthracnose*, (b). *Bacterial Canker*, (c). *Cutting Weevil*, (d). *Die Back*, (e). *Gall Midge*, (f). *Healthy*, (g). *Powdery Mildew*, dan (h). *Sooty Mould*

3.2.

Pre-processing

Proses *pre-processing* melibatkan beberapa tahap, mulai dari membaca citra daun, mengubah ukurannya menjadi 100×100 piksel, hingga mengonversi citra ke dalam format grayscale, total data augmentasi adalah 7.810. Hasil dari tahap ini kemudian digunakan sebagai input untuk proses berikutnya, yakni ekstraksi fitur tekstur dengan pendekatan GLCM. Seluruh data, baik yang digunakan untuk pelatihan maupun pengujian, akan melalui tahapan ini. Berikut ini merupakan salah satu hasil *pre-processing* dari 8 kelas yang digunakan tertera pada Gambar 3.



Gambar 2. Hasil pre-processing

3.3. Ekstraksi Fitur GLCM

Ekstraksi fitur ini akan diambil melalui tekstur. Pada proses ini digunakan metode GLCM. Setelah melewati tahap *pre-processing*, gambar akan diproses lebih lanjut untuk mengekstraksi fitur-fiturnya. Proses ekstraksi ini dimulai dengan membentuk matriks ko-occurrence dari masing-masing gambar, yang kemudian digunakan untuk menghitung parameter-parameter tekstur GLCM seperti kontras, dissimilarity, homogenitas, energi, korelasi, dan entropi, dengan orientasi sudut yang digunakan sebesar 0°. Tabel 1 menyajikan hasil ekstraksi fitur dari GLCM yang dihitung berdasarkan parameter kontras, dissimilarity, homogeneity, energi, korelasi, serta entropi.

Tabel 1. Hasil ekstraksi fitur GLCM

	Contrast	Dissimilarity	Homogeneity	Energy	Correlation	Entropy
0	29.6865	1.6147	0.7049	0.0951	0.9900	8.5051
1	49.5932	3.2054	0.3987	0.0542	0.9767	9.8676
2	77.4887	2.2116	0.7577	0.1052	0.9700	8.2231
3	30.9371	1.3648	0.7302	0.0887	0.9899	8.4730
4	33.5502	1.8641	0.6611	0.0987	0.9939	8.6919
5	52.6877	2.0285	0.6823	0.1003	0.9927	8.5960
6	63.0992	3.3222	0.5541	0.0722	0.9920	9.8926
7	65.2948	3.8222	0.4358	0.0501	0.9881	10.5376

Hasil ekstraksi fitur GLCM ini adalah hasil dari nilai rata-rata kelas tersebut

3.4. Split Data

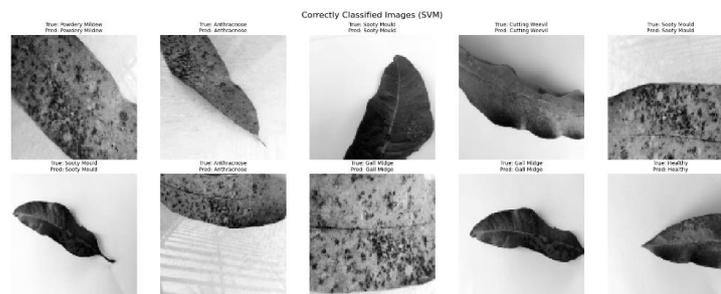
Setelah proses ekstraksi fitur selesai dilakukan, tahap selanjutnya adalah membagi dataset menjadi 2 dengan proporsi masing-masing sebesar 80% untuk *training* dan 20% untuk *testing*. Hasil pemisahan tertera pada tabel 2.

Tabel 2. Hasil Pemisahan Data

Proporsi	Hasil Pemisahan
80:20	6.248:1.562

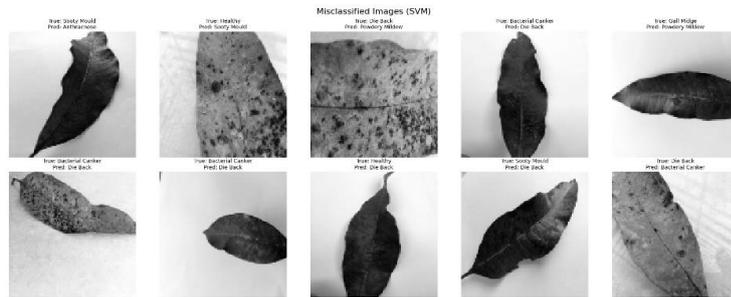
3.5. Pengujian dan Evaluasi Model

Pada pengujian dan evaluasi model ini akan menampilkan hasil sebuah prediksi dan menjelaskan sebuah *confusion matrix* dari daun mangga yang terkena hama. Kebenaran hasil prediksi dari model algoritma SVM (*Support Vector Machine*) menggunakan kernel RBF tertera dalam gambar 4



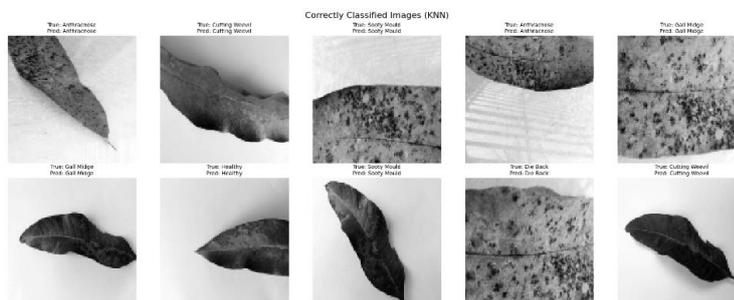
Gambar 4. Kebenaran hasil prediksi model SVM

Gambar 5 merupakan kesalahan hasil prediksi oleh model algoritma SVM (*Support Vector Machine*) menggunakan kernel RBF

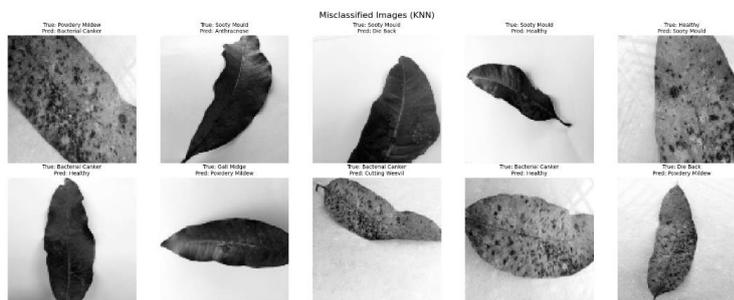


Gambar 5. Kesalahan hasil prediksi model SVM

Hasil kebenaran dan kesalahan prediksi oleh model algoritma K-NN dengan ketanggan yang terdekat yaitu k=3 tertera pada gambar 6 serta gambar 7



Gambar 6. Hasil kebenaran prediksi K-NN



Gambar 7. Hasil kesalahan prediksi K-NN

Untuk klasifikasi SVM digunakan pengklasifikasian *Support Vector Machine Radial Basis Function (RBF)*. Hasil dari klasifikasi menggunakan RBF dengan nilai akurasi 78% dan K-NN dengan nilai akurasi sebesar 80% dapat tertera di tabel 3 serta 4.

Tabel 1. Report Klasifikasi SVM

SVM Classification Report				
Class	Precision	Recall	F1-Score	Accuracy
<i>Anthracnose</i>	70%	71%	71%	71%
<i>Bacterial Canker</i>	90%	87%	88%	87%
<i>Cutting Weevil</i>	100%	98%	99%	98%
<i>Die Back</i>	79%	74%	77%	74%
<i>Gall Midge</i>	62%	70%	66%	70%
<i>Healthy</i>	70%	66%	68%	66%
<i>Powdery Mildew</i>	87%	84%	86%	84%
<i>Sooty Mould</i>	72%	75%	74%	75%
Average				78%

Tabel 2. Report Klasifikasi K-NN

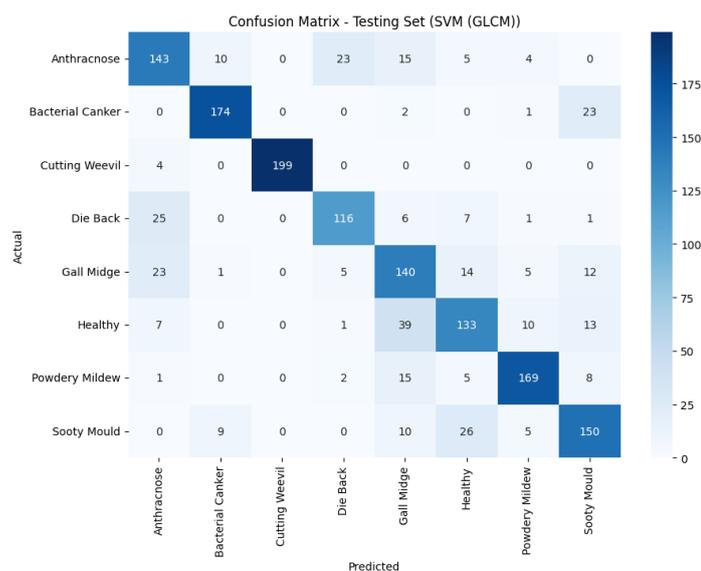
<i>K-Nearest Neighbor Classification Report</i>				
<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>
<i>Anthracnose</i>	73%	78%	75%	78%
<i>Bacterial Canker</i>	92%	93%	92%	93%
<i>Cutting Weevil</i>	99%	100%	100%	100%
<i>Die Back</i>	79%	85%	82%	85%
<i>Gall Midge</i>	60%	61%	61%	61%
<i>Healthy</i>	75%	66%	70%	66%
<i>Powdery Mildew</i>	85%	83%	84%	83%
<i>Sooty Mould</i>	73%	78%	78%	78%
<i>Average</i>				80%

Berdasarkan *confusion matriks* yang dihasilkan oleh model klasifikasi menggunakan algoritma SVM, didapatkan hasil analisa yang tertera di tabel 5.

Tabel 3. Hasil analisa SVM

Label	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>	<i>False Negatif (FN)</i>	<i>True Negative (TN)</i>
<i>Anthracnose</i>	143	1222	60	57
<i>Bacterial Canker</i>	174	1262	20	26
<i>Cutiing Weevil</i>	199	1278	1	4
<i>Die Back</i>	116	1295	31	40
<i>Gall Midge</i>	140	1207	75	60
<i>Healthy</i>	133	1212	67	70
<i>Powdery Mildew</i>	169	1233	49	31
<i>Sooty Mould</i>	150	1221	57	54

Confusion Matrix model SVM tertera pada gambar 8



Gambar 8. Model SVM

Performa *confusion matrix* model SVM dalam mengklasifikasikan delapan kategori. Setiap baris pada matriks menunjukkan jumlah kasus aktual, sedangkan setiap kolom menunjukkan prediksi yang diberikan oleh model.

Kelas *anthracnose* Model berhasil mengklasifikasikan dengan benar sebanyak **143 sampel**. Kesalahan klasifikasi terjadi ke kelas *die back* **23 sampel**, *gall midge* **15 sampel**, dan *bacterial canker* **10 sampel**

Kelas *bacterial canker* Sebanyak **174 sampel** diklasifikasikan dengan benar. Kesalahan terbanyak ke kelas *sooty mould* **23 sampel** dan **1 sampel** ke *powdery mildew*

Kelas *cutting weevil* klasifikasi benar sebanyak **199 sampel**, hanya terdapat **4 sampel** yang salah klasifikasi ke *anthracnose*.

Kelas *die back* model mengklasifikasikan benar sebanyak **116 sampel**, dengan kesalahan klasifikasi ke *anthracnose* **25 sampel**, serta **1 sampel** ke *gall midge*, *healthy*, dan *powdery mildew*.

Kelas *gall midge* Klasifikasi benar sebanyak **140 sampel**. Kesalahan klasifikasi terjadi ke *healthy* **14 sampel**, *sooty mould* **12 sampel**.

Kelas *healthy* Sebanyak **133 sampel** diklasifikasikan dengan benar, namun **39 sampel** salah ke *gall midge*.

Kelas *powdery mildew* Klasifikasi benar sebanyak **169 sampel**. Kesalahan klasifikasi ke *die back* **15 sampel**.

Kelas *sooty mould* Sebanyak **150 sampel** diklasifikasikan dengan benar. Kesalahan klasifikasi cukup banyak ke *healthy* **26 sampel**, *gall midge* **10 sampel**.

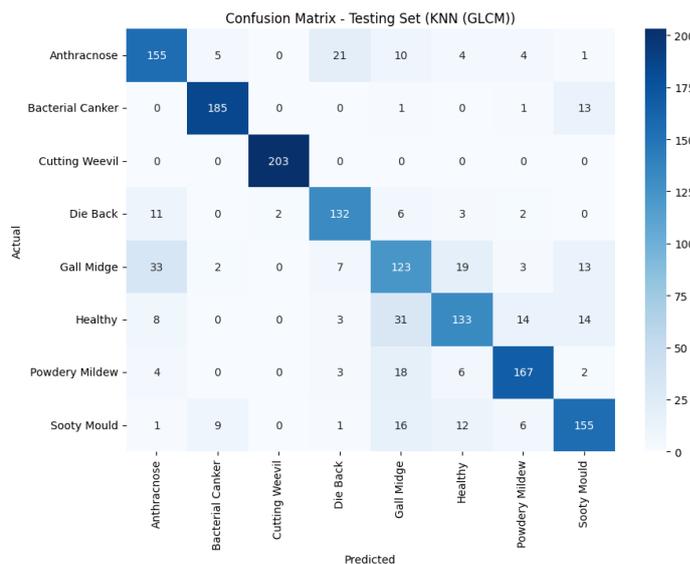
Dari hasil *confusion matrix* tersebut, diketahui bahwa model mampu mengklasifikasikan beberapa kelas dengan akurasi yang tinggi, namun masih terdapat beberapa kelas yang sering mengalami kesalahan prediksi.

Terdapat tabel 6 yang merupakan hasil analisa dari *confusion matrix* dari model K-NN

Tabel 4. Hasil analisa K-NN

Label	True Positive (TP)	False Positive (FP)	False Negatif (FN)	True Negative (TN)
<i>Anthracnose</i>	155	1696	57	41
<i>Bacterial Canker</i>	185	1733	16	15
<i>Cutting Weevil</i>	203	1742	2	2
<i>Die Back</i>	132	1759	36	22
<i>Gall Midge</i>	123	1666	83	77
<i>Healthy</i>	133	1687	59	70
<i>Powdery Mildew</i>	167	1706	43	33
<i>Sooty Mould</i>	155	1718	43	33

Setelah hasil analisa berikut gambar 9 merupakan *confusion matrix* dari model K-NN



Gambar 9. Confusion Matrix model K-NN

Performa model *K-Nearest Neighbors* (K-NN) dalam mengklasifikasikan delapan kategori. Baris pada matriks menunjukkan kelas sebenarnya, sedangkan kolom menunjukkan prediksi model.

Pada kelas *anthracnose* model berhasil mengklasifikasikan dengan benar sebanyak **155 sampel**. Kesalahan klasifikasi terjadi pada kelas *die back* sebanyak **21 sampel**, kelas *gall midge* sebesar **10 sampel**, kelas *bacterial canker* memiliki **5 sampel** kelas *healthy* dan *powdery mildew* masing-masing memiliki **4 sampel** dan kelas *sooty mould* memiliki **1 sampel**

Kelas *bacterial canker* Sebanyak **185 sampel** diklasifikasikan dengan benar. Kesalahan klasifikasi terjadi pada kelas *sooty mould* sebanyak **13 sampel**, *anthracnose* dan *powdery mildew* masing-masing **1 sampel** serta kelas *cutting weevil* klasifikasi benar sebanyak **203 sampel**.

Kelas *die back* model mengklasifikasikan benar sebanyak **132 sampel**. Kesalahan klasifikasi terjadi di kelas *anthracnose* sebanyak 11 sampel, *gall midge* memiliki 6 sampel, *cutting weevil* memiliki **2 sampel**, *healthy* memiliki kesalahan klasifikasi sebesar **3 sampel**, *powdery mildew* memiliki **2 sampel**

Kelas *gall midge* klasifikasi benar sebanyak **123 sampel**. Kesalahan klasifikasi terjadi pada kelas *healthy* **19 sampel**, *anthracnose* sebanyak **33 sampel**, *sooty mould* memiliki **13 sampel**, *die back* memiliki 7 sampel, *bacterial canker* **2 sampel** dan *powdery mildew* **3 sampel**.

Kelas *healthy* sebanyak **133 sampel** diklasifikasikan dengan benar. Kesalahan klasifikasi terjadi di kelas *gall midge* sebanyak **31 sampel**, *sooty mould* dan *powdery mildew* masing-masing 14 sampel, *die back* dan *anthracnose* masing-masing 3–8 sampel

Kelas *powdery mildew* klasifikasi benar sebanyak **167 sampel**. Kesalahan klasifikasi terjadi pada kelas *gall midge* sebanyak **18 sampel**, kelas *healthy* memiliki **6 sampel**, kelas *die back* dan *anthracnose*: masing-masing memiliki 3 dan 4 sampel, dan kelas *sooty mould* memiliki 2 sampel

Kelas *sooty mould* sebanyak **155 sampel** diklasifikasikan dengan benar. Kesalahan klasifikasi terjadi pada kelas *gall midge* sebanyak **16 sampel**, kelas *healthy* sebesar **12 sampel**, kelas *Bacterial Canker* **9 sampel** dan kelas *anthracnose* **1 sampel**.

4. KESIMPULAN

Algoritma *K-Nearest Neighbor* terbukti lebih unggul daripada *Support Vector* dalam tugas klasifikasi gambar daun mangga yang terkena hama sebanyak 4000 gambar dengan metode ekstraksi fitur GLCM, SVM tersebut menggunakan kernel RBF melalui proses pre-processing dengan di resize 100x100 piksel dan menggunakan augmentasi flipping mencapai akurasi sebesar 78% , sedangkan K-NN mencatatkan akurasi 80% dengan ketanggaan $k=3$. Temuan ini menegaskan bahwa K-NN masih unggul dibandingkan algoritma SVM selain itu keunggulan K-NN dapat bekerja dengan baik mengklasifikasikan dataset dengan ukuran kecil ataupun sedang, dalam prosesnya GLCM kurang begitu efektif untuk mengambil ekstraksi fitur tekstur dan jika augmentasi ditambah maka hasil testing kedua algoritma tersebut akan mengalami *overfitting* maupun *underfitting*. Dengan demikian, hasil penelitian ini disarankan untuk memperluas sumber dataset, meningkatkan ukuran dataset, dan menguji algoritma yang berbeda atau *Deep Learning*, agar memberikan hasil akurasi lebih akurat dan lebih komprehensif.

DAFTAR PUSTAKA

- [1] H. Pebriola Br Manik, K. Ibnutama, S. Yakub, S. Informasi, and S. Triguna Dharma, "Penerapan Metode Sobel Dalam Mendeteksi Tepi Citra Daun Mangga Untuk Mendeteksi Serangan Hama Tungau," vol. 3, no. 2, pp. 293–303, 2024, [Online]. Available: <https://ojs.trigunadharna.ac.id/index.php/jsi>
- [2] H. Oktafa, D. G. Pratita, H. Y. Riskiawan, and R. Firgiyanto, "Peningkatan Kapasitas Petani melalui Transfer Teknologi dan Pengetahuan Penanganan Pasca Panen, Diversifikasi, dan Pemasaran Mangga di Kecamatan Wuluhan," *Jurnal Ilmiah Pengembangan dan Penerapan IPTEKS*, vol. Vol. 22, No. 02, 2024, Accessed: Apr. 27, 2025. [Online]. Available: <https://ejournal.unib.ac.id/dharmaraflesia/article/view/36763>
- [3] A. N. A. Lukman *et al.*, *Mengenal Budidaya Tanaman Holtikultura: Pendidikan Lingkungan Hidup*. 2022.
- [4] M. A. Siahaan, E. Gultom, I. S. Tafonao, A. Laia, and W. Gea, "MANFAAT KULIT MANGGA SEBAGAI MASKER ORGANIK," 2021. Accessed: Aug. 13, 2025. [Online]. Available: <https://ejournal.sari-mutiara.ac.id/index.php/JAM/article/view/4618>
- [5] A. H. Mardiah, Y. S. Wulandari, and A. F. Syahputra, "Analisis Faktor-Faktor yang Memengaruhi Pendapatan Petani Mangga (*Mangifera indica* L.) di Desa Sumberjaya Kecamatan Tempuran Kabupaten Karawang," *Paspalum: Jurnal Ilmiah Pertanian*, vol. 12, no. 2, Nov. 2024, doi: 10.35138/paspalum.v12i2.760.

- [6] D. R. Pratama, A. S. Berutu, F. Husin, M. Y. Panjaitan, and Taslim, "Inovasi Edible Coating Buah Mangga Berbasis Kitosan Kulit Udang dengan Aditif Ekstrak Daun Asam Jawa sebagai Antimikroba," *Jurnal Teknik Kimia USU*, vol. 14, no. 1, pp. 53–61, Mar. 2025, doi: 10.32734/jtk.v14i1.19093.
- [7] U. Muthaiah and S. Chitra, "Mango Pest Detection Using Entropy-ELM with Whale Optimization Algorithm," *Intelligent Automation and Soft Computing*, vol. 35, no. 3, pp. 3447–3458, 2023, doi: 10.32604/iasc.2023.028869.
- [8] A. Selvakumar and A. Balasundaram, "Machine Learning based Classification of Diseased Mango Leaves," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 7, pp. 38–44, 2022, doi: 10.17762/ijritcc.v10i7.5563.
- [9] A. Wijoyo, A. Y. Saputra, S. Ristanti, R. Sya'ban, M. Amalia, and R. Febriansyah, "Pembelajaran Machine Learning," *OKTAL: Jurnal Ilmu Komputer dan Science*, vol. 3, 2024, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal>
- [10] K. Kristiawan and A. Widjaja, "Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3182.
- [11] A. M. Siregar, "Analisis Sentimen Pindah Ibu Kota Negara (IKN) Baru pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *Faktor Exacta*, vol. 16, no. 3, Oct. 2023, doi: 10.30998/faktorexacta.v16i3.16703.
- [12] T. A. Putra and I. Permana, "Klasifikasi Penerima Bantuan Program Indonesia Pintar (PIP) Pada Siswa SMK Menggunakan Algoritma KNN, NBC dan C4.5," *Technology and Science (BITS)*, vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6395.
- [13] W. R. Hasibuan, I. P. Sari, and M. Basri, "Klasifikasi Kerusakan (Cacat) pada Biji Kopi Arabika Menggunakan Algoritma KNN (K-Nearest Neighbor)," *Blend Sains Jurnal Teknik*, vol. 3, no. 4, pp. 452–459, May 2025, doi: 10.56211/blendsains.v3i4.781.
- [14] F. Mahrus Fathoni, C. Aji Putra, and A. Lina Nurlaili, "KLASIFIKASI PENYAKIT DAUN ANGGUR MENGGUNAKAN METODE K-NEAREST NEIGHBOR BERDASARKAN GRAY LEVEL CO-OCCURRENCE MATRIX," vol. 3, no. 1, 2024, [Online]. Available: <https://ojs.unsiq.ac.id/index.php/biner>
- [15] R. A. Saputra, D. Puspitasari, and T. Baidawi, "Deteksi Kematangan Buah Melon dengan Algoritma Support Vector Machine Berbasis Ekstraksi Fitur GLCM," 2022. [Online]. Available: <http://ejournal.bsi.ac.id/ejournal/index.php/infortech200>
- [16] D. Zahirah, P. Purnawansyah, N. Kurniati, and H. Darwis, "DIGITAL IMAGE CLASSIFICATION OF HERBAL LEAVES USING KNN AND CNN WITH GLCM FEATURES," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 1, pp. 61–67, Jan. 2024, doi: 10.52436/1.jutif.2024.5.1.1162.
- [17] Y. Sari, A. R. Baskara, and R. Wahyuni, "Classification of Chili Leaf Disease Using the Gray Level Co-occurrence Matrix (GLCM) and the Support Vector Machine (SVM) Methods," in *2021 6th International Conference on Informatics and Computing, ICIC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ICIC54025.2021.9632920.
- [18] A. Supriyanto, R. Rizal Isnanto, and O. D. Nurhayati, "Robusta Coffee Leaf Disease Classifications Using SVM Method and GLCM Feature Extraction," *JURNAL NASIONAL TEKNIK ELEKTRO DAN TEKNOLOGI INFORMASI*, vol. 12, 2023.
- [19] A. A. Huda, B. Setiaji, and F. R. Hidayat, "IMPLEMENTASI GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) UNTUK KLASIFIKASI PENYAKIT DAUN PADI," 2022. [Online]. Available: www.ejournal.unib.ac.id/index.php/pseudocode
- [20] S. A. Rosiva Srg, M. Zarlis, and W. Wanayumini, "Identifikasi Citra Daun dengan GLCM (Gray Level Co-Occurrence) dan K-NN (K-Nearest Neighbor)," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 477–488, Mar. 2022, doi: 10.30812/matrik.v21i2.1572.
- [21] A. Mustika Rani and N. Hendrastuty, "Perbandingan Algoritma NBC Dan SVM Untuk Melakukan Analisis Sentimen Terhadap PP NO.82 Tahun 2021," *Technology and Science (BITS)*, vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.6496.
- [22] M. Sc. , Ph. D. Prof. Drs. Ec. Ir Riyanarto Sarno, S. K. Dr. Shoffi Izza Sabilla, S. Kom. , M. K. Malikhah, S. Kom. , M. K. Doni Putra Purbawa, and S. Kom. , M. K. M. Syauqi Hanif Ardani, *Machine Learning Deep Learning:Konsep dan Pemrograman Python*, I, Cetakan I: 2023. 2023.

-
- [23] Y. Chen, "Covid-19 Classification Based on Gray-Level Co-occurrence Matrix and Support Vector Machine," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 60, Springer Science and Business Media Deutschland GmbH, 2021, pp. 47–55. doi: 10.1007/978-981-15-9682-7_6.
- [24] Y. Hao *et al.*, "Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix," *PLoS One*, vol. 17, no. 5 May, May 2022, doi: 10.1371/journal.pone.0267955.
- [25] Y. F. Achmad, A. Yulfitri, and P. Maharani, "Penerapan Algoritma GLCM dan KNN dalam Pengenalan Jenis Jerawat," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 6, no. 2, pp. 74–82, Nov. 2022, doi: 10.31603/komtika.v6i2.8078.
- [26] R. S. T. Gultom, D. Widiyanto, and A. B. Pangaribuan, *KLASIFIKASI PENYAKIT PENGOROK TANAMAN DAUN PADA TANAMAN MANGGIS MENGGUNAKAN METODE GRAY LEVEL CO-OCCURRENCE MATRIX DAN SUPPORT VECTOR MACHINE*. 2020.
- [27] R. S. Wijaya, A. Qur'ania, and I. Anggraeni, "Klasifikasi Penyakit Cacar Monyet Menggunakan Support Vector Machine (SVM)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1253–1260, Jul. 2024, doi: 10.57152/malcom.v4i4.1417.
- [28] S. A. Rosiva Srg, M. Zarlis, and W. Wanayumini, "Identifikasi Citra Daun dengan GLCM (Gray Level Co-Occurrence) dan K-NN (K-Nearest Neighbor)," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 477–488, Mar. 2022, doi: 10.30812/matrik.v21i2.1572.
- [29] M. Nur Salam, Bijanto, and E. Proborini, "IMPLEMENTATION OF K-NN ALGORITHM BASED EUCLIDEAN DISTANCE TO DETERMINE STUDENT'S PERCEPTION OF PRODUCTIVE LEARNING AT SMK TUNAS HARAPAN PATI IMPLEMENTASI ALGORITMA K-NN BERBASIS EUCLIDEAN DISTANCE UNTUK MENENTUKAN PERSEPSI SISWA TERHADAP PEMBELAJARAN PRODUKTIF DI SMK TUNAS HARAPAN PATI," *Jurnal EDUKASI ELEKTROMATIKA (JEE)2022*, vol. 3, no. 2, 2022.
- [30] A. Khairi, A. Fais Ghozali, and A. Darul Nur Hidayah, "IMPLEMENTASI K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI MASYARAKAT PRA SEJAHTERA DESA SAPIKEREP KECAMATAN SUKAPURA," *TRILOGI Ilmu Teknologi, Kesehatan, dan Humaniora*, vol. 2, no. 3, pp. 319–323, 2021, Accessed: May 22, 2025. [Online]. Available: https://risbang.unuja.ac.id/media/arsip/berkas_penelitian/15_JFPwQpZ.pdf
- [31] T. Brian, E. N. Sholikhah, A. Nur, A. Maulidhia, and S. Wibowo, "Application of K-Nearest Neighbor (KNN) Algorithm to Predict Drinking Water Quality Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Kualitas Air Minum," *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem & Komputer*, vol. 5, 2025, doi: 10.32503/jtecs.v5i1.6715.