

Implementasi Metode K-NN Untuk Klasifikasi Penyakit Daun Teh Berdasarkan Fitur Tekstur GLCM

Rifqi Alvian Ardiansyah^{*1}, Fetty Tri Anggraeny², Hendra Maulana³

^{1,2,3}Informatika, Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur, 60294

E-mail: ¹rifqialvian46@gmail.com, ²fettyanggraeny.if@upnjatim.ac.id,

³hendra.maulana.if@upnjatim.ac.id.

Abstrak

Tanaman teh mengalami penurunan produktivitas yang cukup signifikan akibat serangan penyakit yang menyerang pada daun teh, dan berdampak langsung terhadap kualitas dan kuantitas hasil panen. Penelitian ini bertujuan untuk melakukan klasifikasi penyakit daun teh menggunakan metode algoritma K-Nearest Neighbor (K-NN) dan ekstraksi fitur tekstur Gray Level Co-occurrence Matrix (GLCM). Dataset yang digunakan terdiri dari 1.800 citra daun teh yang terbagi ke dalam enam kategori. Pada proses preprocessing meliputi konversi citra ke grayscale, resize, dan normalisasi nilai piksel. Fitur tekstur yang diekstraksi mencakup contrast, correlation, energy, dan homogeneity. Model K-NN diuji dengan variasi nilai K dari 1 hingga 10 dan tiga skenario pembagian data latih dan uji, yaitu 90:10, 80:20, dan 70:30. Evaluasi dilakukan menggunakan confusion matrix. Hasil terbaik diperoleh pada skenario 80:20 dengan nilai K=1, menghasilkan akurasi sebesar 95,83%. Hasil ini menunjukkan bahwa penggunaan ekstraksi fitur GLCM dan algoritma K-NN efektif dalam memberikan kinerja klasifikasi yang tinggi, sehingga berpotensi besar untuk diterapkan dalam sistem deteksi otomatis penyakit tanaman berbasis citra digital.

Kata Kunci— Penyakit Daun Teh, Gray Level Co-occurrence Matrix, K-Nearest Neighbor, Klasifikasi, Ekstraksi Fitur

Abstract

Tea plants experience a significant decline in productivity due to disease attacks on tea leaves, and have a direct impact on the quality and quantity of the harvest. This study aims to classify tea leaf diseases using the K-Nearest Neighbor (K-NN) algorithm method and Gray Level Co-occurrence Matrix (GLCM) texture feature extraction. The dataset used consists of 1,800 tea leaf images divided into six categories. The preprocessing process includes image conversion to grayscale, resizing, and normalizing pixel values. The extracted texture features include contrast, correlation, energy, and homogeneity. The K-NN model was tested with variations in K values from 1 to 10 and three training and test data division scenarios, namely 90:10, 80:20, and 70:30. The evaluation was carried out using a confusion matrix. The best results were obtained in the 80:20 scenario with a K value = 1, resulting in an accuracy of 95.83%. These results indicate that the use of GLCM feature extraction and the K-NN algorithm is effective in providing high classification performance, so it has great potential to be applied in an automatic detection system for plant diseases based on digital images.

Keywords— Tea Leaf Disease, Gray Level Co-occurrence Matrix, K-Nearest Neighbor, Classification, Feature Extraction

1. PENDAHULUAN

Tanaman teh (*Camellia Sinensis*) merupakan salah satu komoditas perkebunan yang memiliki peranan penting dalam perekonomian Indonesia. Teh, selain sebagai minuman populer secara global, juga menjadi salah satu sumber devisa negara yang signifikan [1]. Indonesia tercatat sebagai eksportir teh terbesar keenam di dunia, mencerminkan tingginya permintaan global terhadap produk ini yang tidak hanya karena rasanya, tetapi juga memiliki manfaat untuk kesehatan [2].

Namun demikian, dalam satu dekade terakhir, produktivitas teh nasional mengalami penurunan yang cukup signifikan. Salah satu penyebab utama dari penurunan tersebut adalah serangan penyakit yang menyerang daun teh dan berdampak langsung terhadap kualitas dan kuantitas hasil panen [3]. Berdasarkan data dari Badan Pusat Statistik (BPS), total produksi teh Indonesia pada tahun 2022 tercatat sebesar 136.800 ton, mengalami penurunan sebesar 5,72% dibandingkan dengan produksi tahun sebelumnya yang mencapai 145.100 ton [3]. Kondisi ini menunjukkan pentingnya upaya identifikasi dan penanganan penyakit secara lebih sistematis dan efisien guna menjaga stabilitas produksi.

Klasifikasi penyakit pada daun teh merupakan pendekatan yang relevan dalam konteks ini karena memungkinkan proses identifikasi penyakit dilakukan secara cepat dan tepat. Klasifikasi merupakan metode untuk mengelompokkan data berdasarkan karakteristik tertentu, yang telah ditentukan sebelumnya [4]. Dalam pengolahan citra daun teh, fitur visual seperti tekstur merupakan elemen kunci yang berperan penting dalam proses klasifikasi. Oleh karena itu, diperlukan metode ekstraksi fitur yang mampu menangkap informasi tekstur secara akurat dan efektif [5].

Grey Level Co-occurrence Matrix (GLCM) merupakan salah satu teknik ekstraksi fitur tekstur yang telah banyak digunakan dalam pengolahan citra. GLCM merepresentasikan hubungan spasial antar piksel pada tingkat keabuan tertentu, sehingga memungkinkan analisis tekstur citra berdasarkan sudut dan jarak antar piksel [6]. Ciri tekstur yang diperoleh dari GLCM dapat digunakan sebagai masukan dalam algoritma klasifikasi untuk membedakan antara berbagai jenis penyakit pada daun teh.

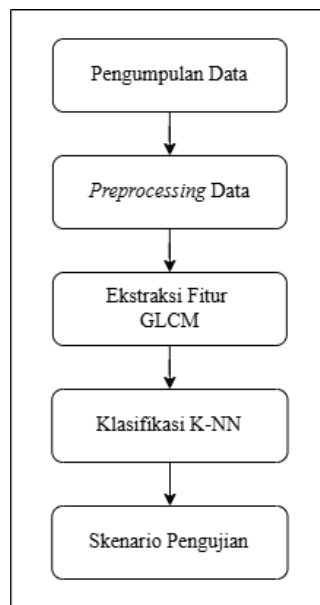
Dalam proses klasifikasi, algoritma *K-Nearest Neighbor* (K-NN) sering digunakan karena sifatnya yang sederhana namun efektif. K-NN menentukan kelas suatu data berdasarkan sejumlah data tetangga terdekat dengan jarak terkecil pada ruang fitur [7]. Metode ini terdiri dari dua tahap utama, yaitu tahap pelatihan (*training*) dan tahap pengujian (*testing*). Kombinasi antara ekstraksi fitur tekstur GLCM dan klasifikasi menggunakan K-NN telah terbukti memberikan akurasi yang tinggi dalam beberapa studi terdahulu [8].

Beberapa penelitian terdahulu menunjukkan bahwa penggunaan GLCM dan K-NN dalam tugas klasifikasi citra menghasilkan performa yang menjanjikan. Misalnya, penelitian yang dilakukan oleh [7] berhasil mengklasifikasikan jenis buah durian dengan akurasi sebesar 93% menggunakan kombinasi GLCM dan K-NN pada dataset berjumlah 1602 citra. Penelitian lain oleh [9] dalam klasifikasi buah anggur menunjukkan akurasi tertinggi sebesar 99,54% dengan nilai $K=2$ menggunakan dataset sebanyak 2624 citra. Hasil ini mengindikasikan potensi metode ini dalam menangani permasalahan klasifikasi citra berbasis tekstur.

Penelitian ini bertujuan untuk mengevaluasi sejauh mana efektivitas penerapan metode ekstraksi fitur tekstur menggunakan *Grey Level Co-occurrence Matrix* (GLCM) serta algoritma klasifikasi *K-Nearest Neighbor* (K-NN) dalam mengidentifikasi penyakit pada daun teh. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem klasifikasi penyakit tanaman yang lebih efisien dan akurat. Selain itu, temuan yang diperoleh juga diharapkan dapat membantu para petani teh dalam melakukan deteksi dini serta penanganan penyakit secara tepat, sehingga dapat meningkatkan produktivitas dan kualitas hasil panen secara berkelanjutan [10].

2. METODE PENELITIAN

Penelitian ini dilakukan untuk mengevaluasi model klasifikasi penyakit daun teh menggunakan ekstraksi fitur tekstur *Gray Level Co-occurrence Matrix* (GLCM) dan algoritma klasifikasi *K-Nearest Neighbor* (K-NN). Tahapan penelitian terdiri dari serangkaian prosedur sistematis yang dirancang untuk mengarahkan proses penelitian secara menyeluruh, mulai dari tahap awal hingga tahap akhir. Tahapan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Pada tahap pengumpulan data, dilakukan pencarian dataset yang relevan dengan penelitian ini. Dataset yang akan dikumpulkan merupakan citra dari beberapa penyakit daun teh. Pada penelitian ini, Dataset diambil dari website *Open Source Kaggle* dengan judul “*Plant Disease*”, yang dibuat oleh Venum1612. Dataset yang diambil berjumlah 1800 citra, yang terbagi dalam 6 kategori, yaitu *Red Leaf Spot*, *Algal Leaf Spot*, *Anthracnose*, *Brown Blight*, *Bird Eye Spot*, dan *Healthy*. Jumlah dataset yang akan digunakan bisa dilihat pada Tabel 1.

Tabel 1. Dataset Penyakit Daun Teh

No.	Kelas	Jumlah Dataset
1	Healthy	300
2	Red Leaf Spot	300
3	Algal Leaf Spot	300
4	Bird Eye Spot	300
5	Brown Blight	300
6	Anthracnose	300
Total		1800

2.2. Preprocessing Data

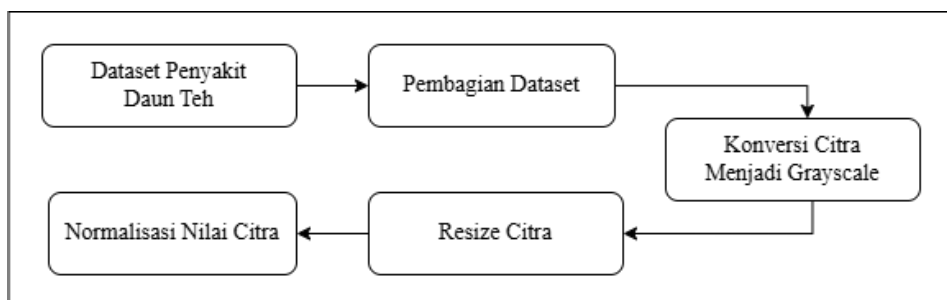
Preprocessing data merupakan langkah untuk meningkatkan kualitas gambar dan meningkatkan akurasi. Tujuan tahap *preprocessing* data adalah untuk memastikan bahwa data yang digunakan untuk pelatihan model merupakan data yang berkualitas [11]. Langkah awal dimulai dari proses *resize* citra, dalam penelitian ini citra yang digunakan memiliki resolusi awal yang beragam. Untuk memastikan konsistensi dan kompatibilitas dalam pemrosesan, seluruh gambar diubah ukurannya menjadi 256x256 piksel. Ukuran ini dipilih karena merupakan ukuran yang cukup umum dalam berbagai visi komputer dan mampu mempertahankan detail tekstur penting yang dibutuhkan untuk proses ekstraksi fitur. Langkah selanjutnya adalah konversi citra dari format RGB ke *grayscale*, langkah ini dilakukan untuk menyederhanakan data dengan hanya mempertahankan informasi intensitas cahaya, sehingga proses ekstraksi fitur tekstur GLCM dapat dilakukan lebih efisien. Konversi *grayscale* membantu mengurangi dimensi data tanpa kehilangan informasi penting terkait pola tekstur daun teh. Setelah proses konversi, dilakukan normalisasi nilai intensitas piksel dari rentang [0, 255] ke [0, 1] dengan cara membagi setiap nilai piksel dengan 255. Proses ini bertujuan untuk menyamakan skala antar fitur, mempercepat proses pelatihan, dan menghindari dominasi nilai piksel tinggi dalam perhitungan jarak pada algoritma K-NN. Selanjutnya, dilakukan pembagian dataset menjadi data latih dan data uji. Proses ini bertujuan untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya dan menghindari *overfitting*. Dalam penelitian ini, digunakan tiga skenario pembagian data, yaitu:

Skenario 1: 90% data latih (1620 citra) dan 10% data uji (180 citra)

Skenario 2: 80% data latih (1440 citra) dan 20% data uji (360 citra)

Skenario 3: 70% data latih (1260 citra) dan 30% data uji (540 citra)

Masing-masing skenario ini diterapkan untuk menguji sejauh mana perubahan proporsi data mempengaruhi performa klasifikasi model. Dengan melakukan evaluasi terhadap ketiga skenario tersebut, dapat diidentifikasi kondisi optimal yang memberikan akurasi dan kemampuan generalisasi terbaik. Secara keseluruhan, tahap *preprocessing* ini memastikan bahwa data input memiliki format yang seragam, terstandarisasi, dan siap digunakan dalam proses ekstraksi fitur dan klasifikasi. Dengan kualitas data yang baik, model diharapkan dapat belajar secara lebih efektif dan memberikan hasil klasifikasi yang akurat. Tahapan *preprocessing* dapat dilihat pada Gambar 2.

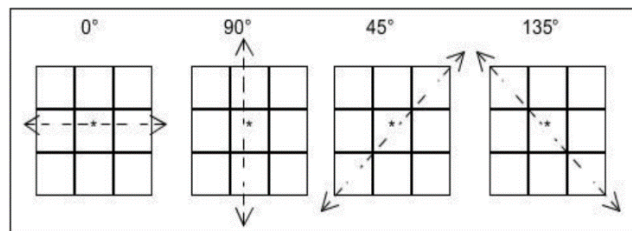


Gambar 2. *Preprocessing Data*

2.3. Gray Level Co-Occurrence Matrix (GLCM)

GLCM merupakan metode fitur tekstur dalam citra digital untuk mengevaluasi hubungan spasial antara nilai intensitas piksel. GLCM mempunyai 28 fitur untuk menggambarkan pola spesial yang diperkenalkan oleh Haralick pada tahun 1973 [7]. GLCM bekerja dengan menghitung frekuensi kemunculan pasangan nilai intensitas piksel dalam jarak dan sudut

tertentu. sudut yang umum digunakan dalam perhitungan GLCM meliputi 0° , 45° , 90° , dan 135° [12]. Dengan menganalisis kedekatan antar piksel dalam berbagai sudut tersebut, GLCM mampu mengungkapkan informasi pola tekstur yang kompleks dalam citra, yang sangat berguna dalam proses identifikasi atau klasifikasi objek berbasis citra.



Gambar 3. Sudut GLCM

Perhitungan menggunakan metode *Gray Level Co-Occurrence Matrix* (GLCM) dilakukan melalui beberapa tahapan. Tahapan pertama dimulai dengan pembentukan matriks awal GLCM yang diperoleh dari pasangan dua piksel yang berjajar dalam citra berdasarkan orientasi sudut yang telah ditentukan sebelumnya, seperti 0° , 45° , 90° , atau 135° [13]. Proses pembentukan matriks ini memperhitungkan hubungan spasial antara piksel dengan mempertimbangkan intensitas nilai *grayscale* (keabuan) dari citra. Dengan kata lain, matriks ini merepresentasikan frekuensi kemunculan pasangan nilai piksel dalam citra, yang kemudian digunakan sebagai dasar dalam ekstraksi fitur tekstur. Dan fitur yang digunakan meliputi *energy*, *correlation*, *contrast*, dan *homogeneity*. Proses ini sangat penting dalam analisis tekstur karena memungkinkan model untuk memahami pola distribusi intensitas dalam citra secara lebih mendalam [14].

2.4. Klasifikasi K-Nearest Neighbor (K-NN)

Algoritma *K-Nearest Neighbor* (K-NN) merupakan metode yang banyak digunakan pada data *mining* dan *machine learning* dikarenakan kemudahan implementasi dan memiliki performa yang baik [15]. Algoritma *K-Nearest Neighbor* (KNN) sering digunakan pada data yang memiliki label dalam klasifikasi dan melakukan identifikasi berdasarkan jarak terdekat. *K-Nearest Neighbor* dalam pengolahan citra digital dapat digunakan sebagai perhitungan jarak terhadap suatu objek atau citra untuk melakukan proses identifikasi atau klasifikasi. Tujuan dari metode KNN dalam pengolahan citra adalah mampu mengidentifikasi objek baru berdasarkan data latih dan data uji dengan prinsip kerja menentukan jarak terdekat dari data yang telah dievaluasi dengan data latih [7]. Metode K-NN bergantung pada pencarian jarak terdekat antara sejumlah k tetangga terdekat dalam data latih dan data uji [9].

2.5. Skenario Pengujian

Pada *Skenario Pengujian*, dijelaskan bahwa proses pengujian model klasifikasi penyakit daun teh dilakukan dengan ekstraksi fitur tekstur GLCM dan algoritma klasifikasi K-NN. Untuk memastikan performa model secara menyeluruh, penelitian ini menerapkan tiga skenario pembagian dataset, yakni data latih 90% dan data uji 10%, data latih 80% dan data uji 20%, serta data latih 70% dan data uji 30%. Skenario ini bertujuan untuk menganalisis pembagian data terhadap akurasi dan efektivitas model, serta untuk mendeteksi kemungkinan *overfitting* atau *underfitting* pada model klasifikasi. Pada ekstraksi fitur tekstur GLCM, digunakan beberapa parameter utama, yaitu *contrast*, *correlation*, *energy*, dan *homogeneity*. Masing-masing parameter tersebut merepresentasikan karakteristik tekstur citra berdasarkan hubungan spasial antar piksel. Hasil ekstraksi fitur ini digunakan sebagai input untuk proses klasifikasi dengan algoritma *K-Nearest Neighbor* (K-NN).

Dalam proses klasifikasi, dilakukan pengujian menggunakan variasi nilai K dari K=1 hingga K=10, untuk menentukan nilai K yang paling optimal dan menghasilkan performa terbaik. Untuk setiap kombinasi rasio data dan nilai K, dilakukan klasifikasi terhadap data uji berdasarkan kedekatan nilai fitur terhadap data latih, menggunakan perhitungan *Euclidean Distance* sebagai metode pengukuran jarak antar data. Evaluasi dari hasil klasifikasi dilakukan dengan menggunakan *confusion matrix*, yang menghasilkan empat metrik utama, yaitu akurasi, presisi, *recall*, dan *F1-score*. Metrik ini digunakan untuk mengukur sejauh mana model mampu mengklasifikasikan data secara akurat dan konsisten. Melalui skenario pengujian ini, penelitian bertujuan menemukan kombinasi antara nilai K dan proporsi pembagian data yang memberikan hasil klasifikasi terbaik.

3. HASIL DAN PEMBAHASAN

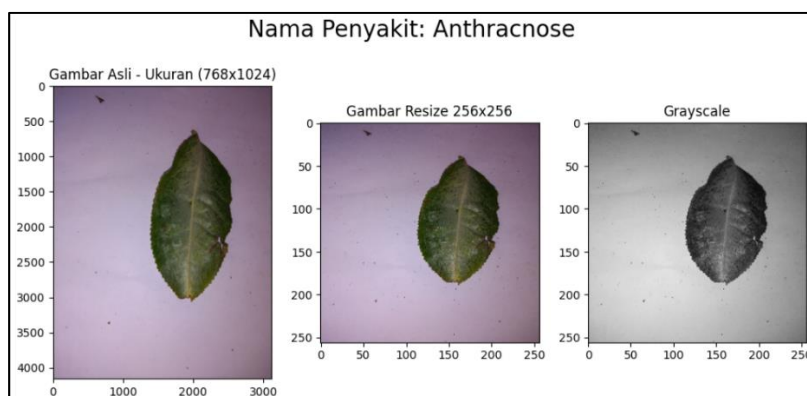
Pada hasil dan pembahasan, akan dijelaskan penerapan model algoritma K-NN berdasarkan fitur tekstur GLCM dalam melakukan klasifikasi penyakit pada daun teh. Proses ini mencakup seluruh tahapan, mulai dari persiapan data hingga evaluasi performa model, yang akan diuraikan secara rinci pada subbab berikutnya. Tahapan tersebut meliputi *preprocessing* data, ekstraksi fitur menggunakan GLCM, pembangunan model klasifikasi K-NN, serta analisis kinerja model berdasarkan beberapa skenario pengujian.

3.1. Preprocessing Data

Preprocessing data citra, merupakan sebuah fase penting yang bertujuan untuk memastikan kesiapan dataset sebelum memasuki tahap ekstraksi fitur dan klasifikasi. *Preprocessing* ini dilakukan secara sistematis dan hati-hati, mengikuti alur kerja yang ditetapkan untuk meningkatkan konsistensi, kualitas, dan keseragaman data. Gambar 2 memberikan representasi visual dari langkah-langkah ini, menggambarkan proses transformasi data mentah menjadi bentuk yang lebih terstruktur dan dapat dianalisis.

Langkah pertama dalam *preprocessing* adalah konversi citra berwarna ke *grayscale*, langkah ini merupakan transformasi penting untuk menyederhanakan representasi visual tanpa kehilangan karakteristik tekstur yang penting bagi analisis GLCM. Selanjutnya, setiap citra diubah ukurannya menjadi 256x256 supaya dimensi lebih seragam, memastikan bahwa seluruh dataset memiliki struktur spasial yang identik, hal ini penting untuk memastikan bahwa parameter tekstur yang dihitung tidak terpengaruh oleh variasi dimensi citra. Setelah proses *resizing*, tahap normalisasi dilakukan dengan merubah rentang intensitas piksel dari 0-255 menjadi rentang 0-1. Proses ini dilakukan untuk menyelaraskan skala fitur yang akan diekstrak, sehingga setiap input berada dalam skala numerik yang seragam, serta mengurangi kompleksitas komputasi dalam penghitungan jarak selama proses klasifikasi K-NN.

Preprocessing tidak hanya menyelaraskan dan membersihkan data, tetapi juga menetapkan dasar penting yang memungkinkan algoritma untuk fokus pada pola-pola tekstur yang relevan tanpa terganggu oleh noise atau ketidakteraturan data. Dengan menstandarkan bentuk dan intensitas citra, model K-NN yang akan digunakan di tahap selanjutnya dapat bekerja secara optimal dalam mengklasifikasikan penyakit daun teh berdasarkan fitur GLCM yang telah diekstrak.

Gambar 4. Hasil *Preprocessing*

Gambar 4 merupakan proses citra daun teh yang mengidap penyakit *Anthracnose* melalui serangkaian tahap *preprocessing* yang sistematis. Di sisi kiri ditampilkan citra asli beresolusi tinggi (768x1024), yang masih mengandung berbagai variasi ukuran, *noise* latar belakang, serta pencahayaan yang tidak seragam. Gambar tengah menunjukkan hasil setelah dilakukan *resize* ke dimensi 256x256 piksel, sebuah tahap penting untuk memastikan bahwa seluruh citra dalam dataset memiliki skala yang konsisten dan dapat diolah secara efisien oleh model klasifikasi.

Sementara itu, gambar paling kanan menampilkan hasil konversi citra menjadi format *grayscale* yang telah dinormalisasi. Proses konversi ini menghapus informasi warna yang tidak diperlukan dan menyederhanakan struktur data, sementara normalisasi dilakukan dengan mereduksi rentang intensitas piksel menjadi skala 0 hingga 1. Tujuannya adalah untuk menyelaraskan nilai-nilai piksel antar citra sehingga ekstraksi fitur berbasis tekstur menggunakan metode GLCM dapat dilakukan secara konsisten, tanpa dipengaruhi perbedaan pencahayaan atau intensitas asli.

Perbandingan visual dari ketiga tahapan ini merupakan tahap penting *preprocessing* dalam menyaring dan menyiapkan data mentah menjadi input yang lebih bersih, seragam, dan siap digunakan dalam proses pelatihan model pembelajaran mesin. Proses ini tidak hanya meningkatkan efisiensi komputasi, tetapi juga secara langsung berkontribusi terhadap peningkatan akurasi dan stabilitas model klasifikasi yang dibangun.

3.2. Gray Level Co-Occurrence Matrix (GLCM)

Setelah tahap *preprocessing* selesai. Tahapan dimulai dengan ekstraksi fitur GLCM. Dimana proses pada ekstraksi fitur menggunakan fitur utama, yaitu *energy*, *contrast*, *homogeneity*, *correlation*. Fitur *energy* berfungsi untuk mengukur konsistensi intensitas dalam tekstur, *contrast* untuk mengukur intensitas kontras antar piksel, *homogeneity* berfungsi untuk mengukur seberapa lama nilai piksel serupa dalam GLCM, dan *correlation* untuk mengukur keterkaitan piksel dalam jarak tertentu.

	Kategori	Fitur	degree 0°	degree 45°	degree 90°	degree 135°
0	Red Leaf Spot	Kontras	0.120865	0.17691	0.12192	0.176378
1	Red Leaf Spot	Korelasi	0.903578	0.853095	0.900725	0.862096
2	Red Leaf Spot	Energi	0.195353	0.173631	0.193183	0.173554
3	Red Leaf Spot	Homogenitas	0.504278	0.509654	0.509374	0.507222
4	Healthy	Kontras	0.0682524	0.0886086	0.0551468	0.0909838
5	Healthy	Korelasi	0.88806	0.855713	0.917136	0.862201
6	Healthy	Energi	0.0898954	0.0787689	0.0929729	0.0787918
7	Healthy	Homogenitas	0.440313	0.436205	0.468172	0.435251
8	Brown Blight	Kontras	0.0550235	0.0835698	0.0628324	0.0882055
9	Brown Blight	Korelasi	0.938725	0.904227	0.930797	0.907517
10	Brown Blight	Energi	0.184766	0.163414	0.180067	0.163252
11	Brown Blight	Homogenitas	0.731647	0.755125	0.72709	0.750685
12	Bird Eye Spot	Kontras	0.329313	0.3744	0.353841	0.372838
13	Bird Eye Spot	Korelasi	0.629057	0.575439	0.598634	0.583113
14	Bird Eye Spot	Energi	0.047863	0.0433837	0.0542171	0.0434278
15	Bird Eye Spot	Homogenitas	0.564499	0.589288	0.611548	0.589468
16	Anthracoese	Kontras	0.33039	0.377077	0.34334	0.37817
17	Anthracoese	Korelasi	0.598306	0.53758	0.5753	0.541651
18	Anthracoese	Energi	0.052909	0.0464581	0.0575865	0.0464488
19	Anthracoese	Homogenitas	0.578935	0.588897	0.604536	0.587733
20	Algal Leaf Spot	Kontras	0.127831	0.171185	0.126539	0.177797
21	Algal Leaf Spot	Korelasi	0.884549	0.843324	0.886236	0.844395
22	Algal Leaf Spot	Energi	0.151949	0.130795	0.142735	0.130855
23	Algal Leaf Spot	Homogenitas	0.713584	0.721156	0.676855	0.720964

Gambar 5. Hasil Fitur GLCM

Gambar 5 merupakan tampilan hasil proses ekstraksi fitur GLCM dari 6 kategori citra. Setiap baris menunjukkan nilai rata-rata fitur GLCM untuk satu citra berdasarkan sudut 0, 45, 90, dan 135 derajat, dengan fitur yang dihitung yaitu, *contrast*, *correlation*, *energy*, *homogeneity*. Dapat dilihat bahwa masing-masing kategori penyakit memiliki pola nilai fitur GLCM yang berbeda-beda. Misalnya, pada kategori *Red Leaf Spot*, fitur *correlation* menunjukkan nilai yang cukup tinggi, menunjukkan adanya pola hubungan antar piksel yang kuat, sedangkan untuk nilai *contrast* cenderung rendah, menunjukkan perbedaan lokal antar piksel yang minimal. Sebaliknya untuk kategori *Healthy* memiliki nilai korelasi dan homogenitas yang konsisten tinggi di semua sudut, menunjukkan permukaan daun yang relatif halus dan seragam, sesuai dengan karakteristik daun sehat.

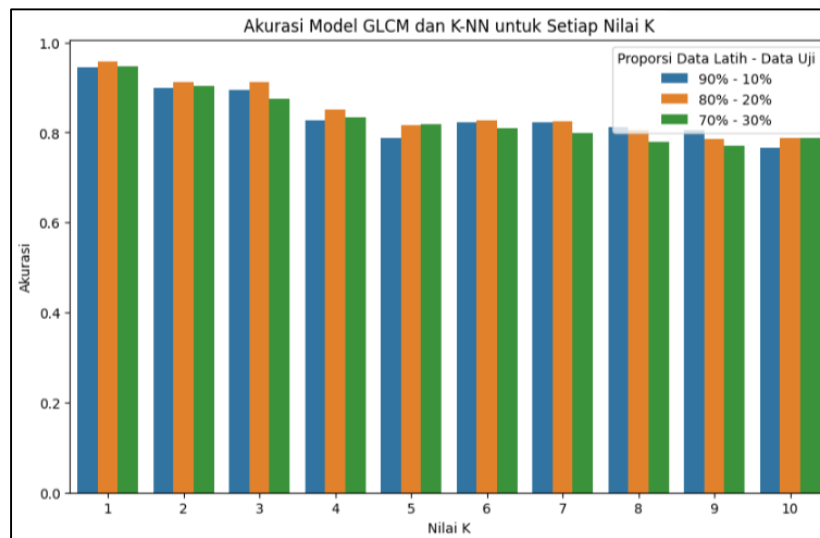
3.3. Skenario Pengujian

Model GLCM dan K-NN menunjukkan performa optimalnya dalam skenario pengujian dengan tiga proporsi data, yaitu 70:30, 80:20, dan 90:10 antara data pelatihan dan pengujian. Dalam setiap skenario, nilai K = 1 secara konsisten memberikan hasil yang paling optimal. Skenario 80:20 menghasilkan performa tertinggi, dengan akurasi sebesar 95,83%, *precision* 96,19%, *recall* dan *F1-score* masing-masing 95,83%. Proporsi ini memberikan keseimbangan ideal antara kapasitas pembelajaran model dan kemampuan generalisasi terhadap data baru. Sementara itu, skenario 70:30 dan 90:10 masing-masing menghasilkan akurasi sebesar 94,62% dan 94,44%, menunjukkan stabilitas performa model meskipun terdapat variasi dalam jumlah data pelatihan dan pengujian. Hasil ini menegaskan bahwa kombinasi GLCM dan K-NN dengan K = 1 efektif dalam klasifikasi citra penyakit daun teh berbasis tekstur, dengan skenario 80:20 sebagai proporsi yang paling optimal, seperti yang dirinci pada Tabel 2, merupakan kinerja dari model GLCM dan K-NN.

Tabel 2. Hasil Pengujian Keseluruhan Proporsi Data

Model	Proporsi Data	Nilai K Terbaik	Matrix			
			Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)

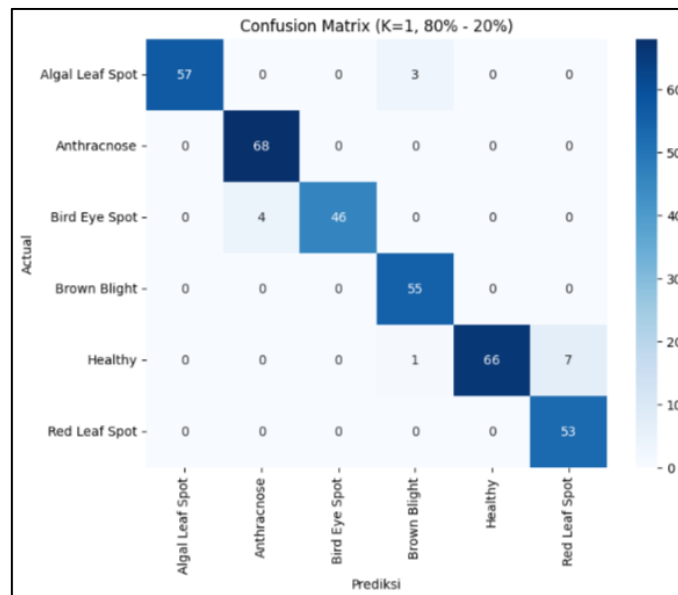
	70:30	1	94.62	94.62	94.79	94.61
GLCM dan K-NN	80:20	1	95.83	95.83	96.19	95.83
	90:10	1	94.44	94.44	95.16	94.46



Gambar 6. Grafik Akurasi Berdasarkan Nilai K

Gambar 6 memperlihatkan grafik akurasi model GLCM dan K-NN berdasarkan variasi nilai K untuk tiga skenario proporsi pembagian data: 90:10, 80:20, dan 70:30. Dari grafik ini dapat disimpulkan bahwa nilai K=1 memberikan akurasi tertinggi pada ketiga proporsi data, dengan performa paling optimal terjadi pada skenario 80% data latih dan 20% data uji. Hal ini menunjukkan bahwa semakin besar porsi data latih yang tersedia, semakin baik model dalam mengenali pola dan melakukan klasifikasi dengan tepat. Pola naik turun akurasi mulai terlihat saat nilai K meningkat, terutama dari K=4 hingga K=10, di mana akurasi mengalami proses naik turun secara bertahap. Hal ini mengindikasikan bahwa mempertimbangkan terlalu banyak tetangga dalam proses klasifikasi justru dapat menyebabkan dampak negatif terhadap ketepatan prediksi, karena model mulai kehilangan sensitivitas terhadap karakteristik lokal yang penting. Secara keseluruhan, hasil ini menunjukkan bahwa K=1 merupakan nilai optimal untuk digunakan dalam kombinasi GLCM dan K-NN pada klasifikasi citra penyakit daun teh.

Hasil *confusion matrix* untuk skenario pengujian paling optimal yang diperoleh dari model GLCM dan K-NN, seperti yang diilustrasikan dalam Gambar 7. Dataset yang digunakan dalam pengujian ini terdiri dari 1.800 citra, yang dibagi dengan rasio 80:20, menghasilkan 1.440 gambar untuk pelatihan dan 360 gambar untuk pengujian. Data pelatihan digunakan untuk membangun model klasifikasi berdasarkan fitur tekstur yang diekstraksi menggunakan metode GLCM, sementara data pengujian berfungsi sebagai set evaluasi untuk menilai kemampuan model dalam melakukan generalisasi terhadap data baru.



Gambar 7. Hasil *Confussion Matrix* Rasio 80:20 dengan K=1

Gambar 7 adalah *confussion matrix* yang dihasilkan setelah evaluasi model terhadap data uji. *Confussion matrix* ini menjadi alat penting dalam mengukur performa klasifikasi model dengan membandingkan antara actual dengan prediksi yang dihasilkan oleh model. Melalui analisis ini, dapat diketahui bahwa dari 360 gambar uji, sebanyak 346 gambar berhasil diklasifikasikan dengan benar, dan 14 gambar lainnya mengalami kesalahan klasifikasi. Kesalahan ini diklasifikasikan ke dalam kategori *false positive* dan *false negative*, yang memberikan gambaran menyeluruh mengenai kekuatan serta kelemahan prediksi model. *Confussion matrix* memberikan wawasan tentang seberapa efektif pendekatan GLCM dan K-NN dalam membedakan kategori penyakit daun teh yang memiliki kemiripan tekstur. Misalnya, akurasi tinggi pada kelas seperti *Anthracnose* dan *Brown Blight* menunjukkan bahwa fitur tekstur GLCM cukup representatif untuk membedakan keduanya. Namun, kesalahan yang terjadi menunjukkan adanya kemiripan fitur pada kategori tertentu.

```

Classification Report untuk skenario 80% - 20%, K=1:
precision  recall  f1-score  support
Algal Leaf Spot      1.00    0.95    0.97      60
Anthracnose          0.94    1.00    0.97      68
Bird Eye Spot        1.00    0.92    0.96      50
Brown Blight         0.93    1.00    0.96      55
Healthy              1.00    0.89    0.94      74
Red Leaf Spot        0.88    1.00    0.94      53

accuracy              0.96              360
macro avg             0.96    0.96    0.96      360
weighted avg          0.96    0.96    0.96      360
    
```

Gambar 8. Hasil *Classification Report* Rasio 80:20 dengan K=1

Gambar 8 menunjukkan *classification report* untuk model optimal GLCM dan K-NN pada skenario pengujian 80% data latih dan 20% data uji dengan nilai K = 1. Evaluasi ini mencakup kinerja model dalam mengklasifikasikan enam kelas penyakit daun teh, dengan total 360 gambar uji yang didistribusikan secara proporsional ke setiap kelas. Nilai akurasi keseluruhan didapatkan sebesar 96%, mencerminkan kestabilan dan efektivitas model dalam melakukan klasifikasi citra daun teh berdasarkan fitur tekstur GLCM. *Classification report* ini memberikan ringkasan menyeluruh atas performa model dari GLCM dan K-NN dengan K = 1

pada proporsi data 80:20, merupakan pendekatan yang sangat andal untuk sistem klasifikasi berbasis citra.

4. KESIMPULAN

Pada penelitian yang telah dilakukan, model klasifikasi penyakit daun teh menggunakan metode algoritma K-NN dan ekstraksi fitur tekstur GLCM, menunjukkan performa yang sangat baik dengan akurasi tertinggi sebesar 95.83% pada skenario proporsi 80:20 dengan nilai $K=1$. Hasil ini menunjukkan bahwa fitur tekstur yang dihasilkan dari GLCM mampu menangkap pola spasial antar piksel yang relevan dalam membedakan jenis penyakit daun teh. Performa model tetap stabil pada skenario lain dengan proporsi 70:30 dan 90:10, masing-masing menghasilkan akurasi sebesar 94.62% dan 94.44%. Konsistensi ini menunjukkan bahwa penggunaan ekstraksi fitur GLCM dan algoritma K-NN terbukti efektif dan efisien untuk diterapkan dalam klasifikasi citra berbasis tekstur.

5. SARAN

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan lebih lanjut agar model klasifikasi penyakit daun teh dapat menjadi lebih optimal. Pertama, disarankan untuk menggunakan dataset yang lebih banyak dan lebih beragam agar model dapat belajar lebih banyak variasi citra dan meningkatkan kemampuan generalisasi. Kedua penggunaan algoritma klasifikasi lain seperti SVM atau Random Forest dapat dijadikan pembandingan untuk melihat kelebihan dan kekurangan masing-masing metode.

DAFTAR PUSTAKA

- [1] Wildah, S. K., and Latif, A, Kombinasi Metode Fitur Ekstraksi untuk Identifikasi Penyakit pada Daun Teh, *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, Vol. 11, No. 3, pp. 447-452, 2023.
- [2] Diah, A. K., Puspita, A. Nilogiri, H. Oktavianto, and M. Jember, Deteksi Penyakit Daun Teh Menggunakan Metode Convolutional Neural Network (CNN), Vol. 5, No.1, pp. 45-50, 2023.
- [3] Mustikasari, A. R. Ramli, and A. K. Gibran, Analisis Kinerja Algoritma Machine Learning Untuk Deteksi Penyakit Daun Teh Dengan Particle Swarm Optimization, 2023. [Online]. Available: <https://www.kaggle.com/code/arafathhmm/tea-leaf->
- [4] Hardi, N, Komparasi Algoritma MobileNet Dan Nasnet Mobile Pada Klasifikasi Penyakit Daun Teh, *Jurnal Rekayasa Perangkat Lunak*, vol. 3, no. 1, pp. 50–55, 2022.
- [5] Kurniati, F. T, Model Ekstraksi Fitur Berbasis Tekstur untuk Identifikasi Keaslian Objek, vol. 18, pp. 184–189, 2024.
- [6] Azizah, A. P, and P. S. W. Gede, Analisis Pengenalan Pola Daun Berdasarkan Fitur Canny Edge Detection dan Fitur GLCM Menggunakan Metode Klasifikasi k-Nearest Neighbor (kNN), 2021. [Online]. Available: <http://jcosine.if.unram.ac.id/>
- [7] Sarimole, F. M, and A. Syaeful, CLASSIFICATION OF DURIAN TYPES USING FEATURES EXTRACTION GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) AND K-NEAREST NEIGHBORS (KNN), Vol. 4, No. 1, pp. 111-121, 2022.

- [8] Mustofa, M. I., M. T. Furqon, and D. E. Ratnawati, Penggunaan Metode Ekstraksi Fitur Tekstur Gray Level Co-occurrence Matrix dan K-Nearest Neighbor untuk Identifikasi Jenis Penyakit Tanaman Apel, 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] Pulung, N. A. and S. H. Nugraini, Texture Feature Extraction in Grape Image Classification Using K-Nearest Neighbor, *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 5, pp. 768–775, Oct. 2022, doi: 10.29207/resti.v6i5.4137.
- [10] Suherman, A. H. N. Ibrahim, H. Syahrian, V. P. Rahadi, and M. K. Prayoga, KLASIFIKASI DAUN TEH GAMBUNG VARIETAS ASSAMICA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK DENGAN ARSITEKTUR LENET-5, *JOURNAL OF ELECTRICAL AND SYSTEM CONTROL ENGINEERING*, vol. 4, no. 2, pp. 63–71, Feb. 2021, doi: 10.31289/jesce.v4i2.4136.
- [11] Ramadhan, G. S. P, M. Maimunah, and S. Nugroho, Optimasi Data Preprocessing dan Hyperparameter Tuning pada Klasifikasi Penyakit Daun Apel menggunakan DenseNet169, *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 1352–1362, Dec. 2024, doi: 10.47065/bits.v6i3.6134.
- [12] Fathoni, F. M., Klasifikasi Penyakit Daun Tomat Menggunakan Algoritma K-NN Berdasarkan Ekstraksi Fitur GLCM dan LBP, *Jurnal Teknik Informatika dan Teknologi Informasi*, vol. 4, no. 1, pp. 39–50, Jan. 2024, doi: 10.55606/jutiti.v4i1.3417.
- [13] Ikhlasul, R, A. Wijaya, N. David, M. Veronica, and R. Toyib, Klasifikasi Kualitas Buah Sawit Menggunakan Metode Gray Level Co-occurrence Matrix Dengan Variasi Arah Obyek, Vol 20, No. 1, pp. 256-263, 2024.
- [14] Prasaja, Y. A., S. Agustin, U. Chotijah, and F. Mar'i, PERBANDINGAN METODE GLCM DAN LBP DALAM KLASIFIKASI JENIS KAYU (id 2), *INDEXIA : Informatic and Computational Intelligent Journal*, vol. 04, no. 2, pp. 60–84, 2022.
- [15] Yudhana, A, and Agus Jaka Sri Hartanta, ALGORITMA K-NN DENGAN EUCLIDEAN DISTANCE UNTUK PREDIKSI HASIL PENGGERGAJIAN KAYU SENGON, *TRANSMISI*, vol. 22, no. 4, 2020, doi: 10.14710/transmisi.22.4.107-141.