

Identifikasi Penyakit Daun pada Tanaman Padi Menggunakan Ekstraksi Fitur *Gray Level Co-occurrence Matrix* (GLCM) dan Metode *K-Nearest Neighbour* (KNN)

Diki Wahyudi Rustanto^{a1}, Febri Liantoni^{a2}, Nurcahya Pradana Taufik Prakisy^{a3}

^aPendidikan Teknik Informatika dan Komputer, Universitas Sebelas Maret Surakarta

¹dikiwahyudir40@student.uns.ac.id

²febri.liantoni@gmail.com

³nurcahya.pradana@gmail.com

Abstrak

Negara Indonesia merupakan salah satu negara agrikultur, di mana bidang pertanian berperan penting dalam menjaga keberlangsungan hidup. Hal ini dikarenakan, sebagian besar masyarakat Indonesia menggunakan beras sebagai bahan pangan pokok mereka. Sedangkan ketersediaan bahan pangan pokok masyarakat sudah berkurang, karena adanya alih fungsi lahan-lahan pertanian menjadi perumahan, industri, dan lain-lain. Bukan hanya itu saja, permasalahan lain yang dapat menurunkan ketersediaan bahan pangan yaitu seperti, kondisi iklim atau cuaca, sistem pengairan, serangan hama dan masih banyak lagi permasalahan yang dapat mengakibatkan panen menjadi kurang maksimal. Penggunaan teknologi dalam bidang pertanian seharusnya menjadi lebih mudah dan membantu para petani dalam mendeteksi penyakit yang menyerang daun padi. Karena itu, deteksi dan klasifikasi hama pada daun padi perlu dilakukan untuk mengevaluasi akurasi, presisi, dan recall menggunakan perhitungan matriks kebingungan (confusion matrix) dengan menerapkan algoritma K-Nearest Neighbors (KNN). Dari hasil klasifikasi tersebut menghasilkan nilai akhir akurasi paling tinggi yaitu sebesar 73% pada jarak piksel (d) yaitu 5 dan nilai tetangga (k) yaitu 3 pada offset 0°. Hal ini menunjukkan bahwa algoritma KNN cukup baik dalam melakukan klasifikasi.

Kata kunci: Penyakit Daun Padi, K-Nearest Neighbor, Gray Level Co-occurrence Matrix, Image Classification

Identification of Leaf Diseases in Rice Using Feature Extraction of the *Gray Level Co-occurrence Matrix* (GLCM) and the *K-Nearest Neighbor* (KNN) Method

Abstract

Indonesia is an agrarian country, where the agricultural sector plays a crucial role in sustaining life. This is because the majority of Indonesians use rice as their staple food. However, the availability of staple food has decreased due to the conversion of agricultural lands into housing, industries, and other purposes. Not only that, other issues such as climate or weather conditions, irrigation systems, pest attacks, and many more can lead to suboptimal harvests. The use of technology in agriculture should be made easier and more helpful for farmers to detect diseases that affect rice leaves. Therefore, the detection and classification of pests on rice leaves need to be performed to evaluate accuracy, precision, and recall using the confusion matrix calculation by implementing the K-Nearest Neighbors (KNN) algorithm. From the classification results, the highest accuracy value obtained was 73% with a pixel distance (d) of 5 and a number of neighbors (k) of 3 at an offset of 0°. This indicates that the KNN algorithm performs reasonably well in the classification process.

Keywords: Rice Leaf Diseases, K-Nearest Neighbor, Gray Level Co-occurrence Matrix, Image Classification

I. PENDAHULUAN

Indonesia merupakan salah satu negara agrikultur, yang mana bidang pertanian berperan penting dalam menjaga keberlangsungan hidup. Hal ini dikarenakan, sebagian besar masyarakat Indonesia menggunakan beras sebagai bahan pangan pokok mereka. Oleh karena itu, meningkatnya jumlah penduduk akan menyebabkan peningkatan dalam konsumsi pangan [1], [2]. Namun,

ketersediaan bahan pangan pokok masyarakat sudah berkurang, karena adanya alih fungsi lahan-lahan pertanian menjadi perumahan, industri dan lain-lain. Hal ini memberikan gambaran bahwa semakin besar alih fungsi lahan sawah yang terjadi, maka akan semakin besar nilai produksi padi yang hilang akibat kegiatan alih fungsi lahan sawah tersebut [3]. Bukan hanya itu saja, permasalahan lain yang dapat menurunkan ketersediaan bahan pangan yaitu

seperti, kondisi iklim atau cuaca, system pengairan, serangan hama dan masih banyak lagi permasalahan yang dapat mengakibatkan hasil panen menjadi tidak optimal.

Serangan hama salah satu dari beberapa beberapa penyebab penurunan ketersediaan bahan pangan dan merupakan musuh utama para petani. Tanaman padi yang mengalami serangan penyakit menunjukkan tanda-tanda berupa bercak dan pola warna khusus. Gejala penyakit tersebut terlihat hampir di seluruh bagian tanaman padi. Bahkan akan sangat tampak jelas pada bagian daun. Karena bagian daun merupakan bagian yang terpampang dengan jelas bahkan dari jarak beberapa meter masih bisa terlihat gejala penyakit yang menyerang daun pada tanaman padi. Di samping itu, penampang daun padi memiliki lebar dan luas yang lebih besar daripada bagian-bagian lain dari tanaman padi. Dengan demikian, daun padi bisa dipakai sebagai titik awal penelitian serta sebagai landasan untuk menerapkan teknologi pertanian guna meningkatkan kualitas hasil panen.

Pada saat ini, pemanfaatan teknologi telah menyebar ke berbagai sektor, termasuk pendidikan, ekonomi, dan termasuk juga pertanian. Teknologi sendiri diciptakan untuk membantu dan mempermudah tugas atau pekerjaan manusia. Sejalan dengan tujuan diciptakan teknologi, maka penggunaan teknologi dalam bidang pertanian seharusnya menjadi lebih mudah dan membantu para petani dalam mendeteksi penyakit yang menyerang daun padi. Sehingga petani akan lebih cepat dalam mendapatkan data mengenai penyakit yang menyerang tanaman padi. Hal ini dapat dilakukan dengan menggunakan data mining dengan berbagai algoritma untuk mengetahui keakuratan proses pendeteksian penyakit tanaman. Data mining atau penambangan data adalah suatu pendekatan yang berguna untuk mendapatkan informasi berharga dari kumpulan data melalui penerapan pengetahuan seperti statistik, matematika, dan pengenalan pola [4]. Data mining sendiri mempunyai banyak jenis algoritma metode di dalamnya, seperti SVM, KNN dan algoritma metode lainnya.

Menurut hasil penelitian yang dilakukan oleh Agustiani dan rekannya, mereka melakukan klasifikasi dengan memanfaatkan citra yang telah melalui proses augmentasi, kemudian menerapkan metode ekstraksi fitur Color Histogram, dan akhirnya menggunakan algoritma Random Forest untuk melakukan klasifikasi [2], [4]. Disamping itu, dalam penelitian ini juga dilakukan perbandingan, termasuk perbandingan dalam ekstraksi fitur dan algoritma, untuk mencari solusi terbaik. Hasil tertinggi yang dicapai adalah tingkat akurasi sekitar 99.65% dari metode yang diajukan [2], [4]. Penelitian berikutnya memperkenalkan sistem deteksi yang menggunakan machine learning dengan beragam algoritma, termasuk Decision Tree, Random Forest, Naive Bayes, SVM, dan KNN. Dari perbandingan kelima jenis algoritma tersebut, dapat ditemukan tiga jenis model: Model Overfit (Random Forest, Decision Tree, dan Naive Bayes), Model Underfit (SVM), dan Model Berkualitas (KNN) [5]. Jadi metode terbaik diantara kelima tersebut yaitu metode KNN dengan nilai akurasi 87%, karna model ini konsisten baik pada kedua evaluasi [5]. KNN tidak terbukti memiliki masalah

overfitting karena secara konsisten berkinerja baik pada data train dan data test [5].

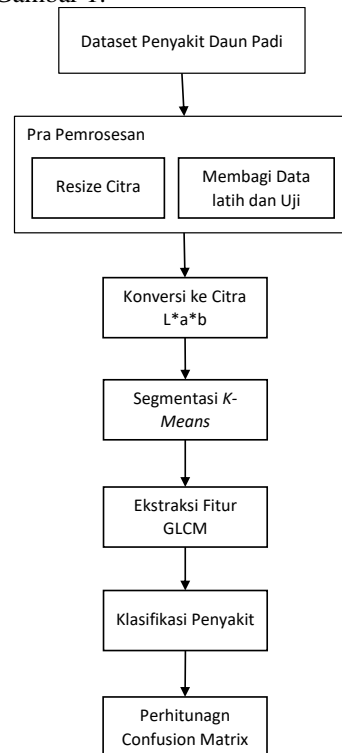
Selain dari beberapa penelitian diatas, Huda dkk. [6], mengatakan bahwa secara penggunaan keseluruhan GLCM dan KNN dalam sistem identifikasi penyakit daun padi menghasilkan tingkat akurasi maksimum sekitar 93,3%. Dalam penelitian ini, metode k-nearest neighbor digunakan untuk melakukan klasifikasi pada penyakit tanaman kopi arabika. Hasil penelitian menunjukkan bahwa nilai k terbaik adalah k=11 dengan menggunakan jarak Euclidean, dan tingkat akurasi yang dicapai mencapai 94% [7].

Penelitian ini melibatkan pengukuran kinerja metode KNN dengan berbagai nilai K pada 1000 objek data produksi jenis kaca yang berasal dari pusat dataset. Hasil pengujian menunjukkan bahwa performa terbaik ditemukan saat K=3, dengan tingkat akurasi mencapai 64%, presisi sekitar 63%, recall sekitar 71%, dan F-Measure sekitar 67% [8].

Berdasarkan beberapa penelitian yang telah dilakukan seperti pada pemaparan diatas, pada penelitian ini pengidentifikasian penyakit daun pada tanaman padi akan melalui pengolahan citra gambar dengan menggunakan algoritma K-Nearest Neighbor. Hal ini dikarenakan menggunakan algoritma K-Nearest Neighbor memiliki nilai akurasi yang tinggi dan konsisten dengan tingkat kegagalan dan masalah yang sangat minimalis.

II. METODOLOGI

Pada tahapan metodologi penelitian ini dilakukan beberapa alur atau tahapan penelitian. Alur penelitian tersebut digunakan untuk mendapatkan hasil klasifikasi dengan nilai performa dari metode yang baik. Tahapan-tahapan tersebut terdapat dalam penelitian yang ditunjukkan Gambar 1.



Gambar 1. Alur Penelitian

A. Dataset Penelitian

Dataset penelitian ini diperoleh dari *Kaggle Repository*. Pada tahapan pengumpulan dataset penelitian, bisa diperoleh dengan mengunjungi dan mengunduh dataset pada laman kaggle yang telah disumbangkan para peneliti lain dengan kata kunci *leaf disease*.

Dataset tersebut berupa citra daun berpenyakit. Daun adalah salah satu karakteristik yang khas dari tumbuhan, dapat diamati dengan mudah, dan cukup mewakili sehingga dapat digunakan sebagai objek untuk mengekstraksi fitur tumbuhan [9]. Citra daun tersebut terbagi menjadi tiga kategori, yaitu *bacterial leaf blight* (hawar daun bakteri), *leaf smut*, dan *brown spot* (bercak coklat).

Setiap kelas tersebut mempunyai ciri-ciri masing. Hawar daun bakteri mempunyai perkembangan penyakit mengakibatkan keringnya tanaman dan lipatan pada helaian daun sepanjang tulang daun utama [10], [11]. Bercak coklat (*brown spot*) memiliki ciri-ciri pada daun dikenali dari adanya tanda-tanda seperti bercak kecil berwarna coklat tua, perkembangan bercak menjadi coklat di pinggir dan warna kuning pucat, putih kotor, coklat, atau kelabu di bagian tengahnya [10]. *Leaf smut* memiliki ciri-ciri adanya bercak-bercak kecil berwarna hitam pada daun. Lesi berbentuk lingkaran hingga oval atau tidak seragam berbentuk dengan tekstur kasar [12].

B. Pra Pemrosesan

Tahapan pra pemrosesan ini, fitur dari dataset dilakukan proses *resize* ukuran *pixel* menjadi ukuran 400×400 *pixel*. Proses *resize* dilakukan dengan menggunakan code matlab. Format yang dihasilkan dari proses tersebut berbentuk .jpg.

Kemudian setelah melakukan proses *resize* ukuran piksel citra sebanyak dataset yang diambil, membagi dataset menjadi data latih dan data uji.

C. Konversi ke Citra L^*a^*b

Setelah melakukan *resize* citra dengan ukuran piksel 400×400 , kemudian mengkonversikan citra tersebut ke dalam citra L^*a^*b . Citra L^*a^*b terdiri dari tiga komponen utama, yaitu Luminance (kecerahan) dan dua komponen berwarna, A dan B. Tahap ini perlu dilaksanakan karena langkah segmentasi menggunakan metode K-Means harus diterapkan pada gambar yang berada dalam format skema warna L^*a^*b .

D. Segmentasi

Tahap berikutnya adalah segmentasi. Tujuan dari proses pemisahan ini adalah untuk mengidentifikasi dan memisahkan wilayah objek dari latar belakangnya agar mempermudah analisis lebih lanjut. Segmentasi dilakukan melalui code yang dibuat di software Matlab. Metode segmentasi ini adalah *k-means clustering*.

Lalu citra yang sudah melakukan proses segmentasi kemudian diubah ke citra *grayscale*. Tujuan dari langkah ini adalah untuk memfasilitasi proses ekstraksi fitur GLCM selanjutnya. Proses ekstraksi tersebut hanya dapat dilakukan apabila citra telah dikonversi ke dalam format grayscale.

E. Ekstraksi Gray Level Co-occurrence Matrix (GLCM)

Dalam penelitian ini, fitur ekstraksi dilakukan dengan menggunakan *Gray Level Co-occurrence Matrix* (GLCM), suatu metode yang memiliki tujuan yaitu untuk mengungkap informasi tekstur dari gambar tersebut. Konstruksi GLCM sangat mudah diimplementasikan untuk kuantifikasi pola permukaan heterogen dan kekasaran yang ditunjukkan dalam citra digital [13]. GLCM merupakan sebuah matriks ko-okurensi di mana elemen-elemennya mewakili jumlah kemunculan piksel-piksel dengan tingkat keabuan tertentu, yang terletak pada jarak (d) dan sudut (Θ) tertentu satu sama lain [14]. GLCM sering digunakan karena mampu memberikan tingkat kognisi yang tinggi dengan mempertimbangkan aspek-aspek seperti kontras, korelasi, homogenitas, dan energi [13], [15]. Mengenai sifat tekstur menggunakan metode GLCM dengan parameter kontras, korelasi, energi dan homogenitas.

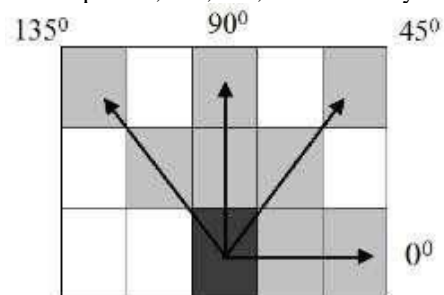
$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{i,j} (i - j)^2 \quad (1)$$

$$\text{Correlation} = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad (2)$$

$$\text{Energy} = \sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (3)$$

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2} \quad (4)$$

GLCM dibentuk dengan menghitung intensitas pada satu garis warna abu-abu yang berada di dalam gambar, dengan memperbandingkan hubungan spasial linier antara dua piksel [13], [16]. Pasangan parameter ($0, d$) digunakan untuk menyatakan hubungan spasial, dimana 0 mengacu pada orientasi dan d menyatakan jarak diantara dua piksel. Dalam ekstraksi fitur tekstur ini memiliki sudut – sudut yang digunakan dalam menghitung jarak pikselnya atau biasa dilambangkan dengan notasi (d). Sudut dalam ekstraksi ini meliputi 0° , 45° , 90° , dan seterusnya.



Gambar 2. Sudut dalam GLCM

F. Klasifikasi

Penelitian ini menggunakan metode klasifikasi *K-Nearest Neighbor* (K-NN). *K-Nearest Neighbor* merupakan salah satu metode yang melakukan klasifikasi terhadap objek dengan jarak terdekat. K-NN merupakan suatu teknik yang digunakan untuk mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki jarak

terdekat dengan objek tersebut [17]. Berdasarkan jarak Euclidean, biasanya dekat atau jauhnya neighbor dihitung.

$$dis = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots} \quad (5)$$

Dalam metode KNN ini terdapat dua buah fase didalamnya. Proses KNN terbagi menjadi dua tahap, yaitu tahap pelatihan (training) dan tahap klasifikasi atau pengujian (testing) [5]. Fase pertama yaitu pembelajaran, algoritma hanya menyimpan data vektor – vektor fitur dan klasifikasi. Fase kedua yaitu pengujian, fitur yang sama akan dilakukan penghitungan untuk melakukan uji coba. Dalam proses ini, kami menghitung jarak dari vektor yang baru terhadap seluruh vektor data pembelajaran, dan kemudian kita memilih sejumlah k tetangga yang memiliki jarak terdekat [5].

G. Confusion Matrix

Metode ini merepresentasikan hasil klasifikasi menggunakan matriks [8]. Berikut adalah tabel *confusion matrix* pada Tabel I.

TABEL I
TABEL CONFUSION MATRIX

	Positive	Negative
Positive	<i>True Positive</i>	<i>False Positive</i>
Negative	<i>False Negative</i>	<i>True Negative</i>

True Positive adalah jumlah data positif yang benar-benar diklasifikasikan sebagai positif, sementara *False Positive* ialah banyaknya data negatif yang salah diklasifikasikan sebagai positif. *False Negative* adalah jumlah data positif yang salah diklasifikasikan sebagai negatif, dan *True Negative* adalah jumlah data negatif yang benar-benar diklasifikasikan sebagai negative. Metode pengujian *confusion matrix* dapat menghasilkan perhitungan dengan 4 output, di antaranya yaitu:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (8)$$

Presisi (*Precision*) mengukur sejauh mana klasifikasi yang positif (positif secara prediksi) benar-benar adalah positif. Di sini, TP (*True Positive*) adalah jumlah sampel yang diklasifikasikan dengan benar sebagai positif, dan FP (*False Positive*) adalah jumlah sampel yang salah diklasifikasikan sebagai positif. Dalam kata lain, presisi menghitung berapa banyak hasil positif yang benar dari semua hasil positif yang diklasifikasikan oleh model.

Recall (*Recall*) juga dikenal sebagai sensitivitas atau *true positive rate* mengukur sejauh mana model mampu mengidentifikasi semua sampel positif yang sebenarnya. Di sini, TP (*True Positive*) adalah jumlah sampel yang diklasifikasikan dengan benar sebagai positif, dan FN (*False Negative*) adalah jumlah sampel yang salah diklasifikasikan sebagai negatif. Recall menghitung berapa banyak sampel positif yang berhasil diidentifikasi oleh model dari semua sampel positif yang ada.

Akurasi (*Accuracy*) mengukur persentase keseluruhan klasifikasi yang benar dari semua sampel. Di sini, TP (*True Positive*) adalah jumlah sampel yang diklasifikasikan dengan benar sebagai positif, TN (*True Negative*) adalah jumlah sampel yang diklasifikasikan dengan benar sebagai negatif, FP (*False Positive*) adalah jumlah sampel yang salah diklasifikasikan sebagai positif, dan FN (*False Negative*) adalah jumlah sampel yang salah diklasifikasikan sebagai negatif. Akurasi menghitung berapa banyak sampel yang diklasifikasikan dengan benar (positif dan negatif) dari semua sampel yang ada.




III. HASIL DAN PEMBAHASAN

Penelitian yang dilakukan berfokus pada cara mengidentifikasi dan membedakan citra penyakit pada daun tanaman padi dengan melakukan penerapan algoritma K-Nearest Neighbor (K-NN) untuk mengklasifikasikan citra-citra tersebut ke dalam tiga kelas klasifikasi. Selain itu, penelitian ini bertujuan untuk mengevaluasi ketepatan akurasi dari algoritma K-Nearest Neighbor (K-NN) dalam mengidentifikasi citra penyakit pada daun tanaman padi (Hawar Bakteri, Bercak coklat, dan Jamur daun), yang setiap jenis penyakit daunnya terdiri dari 40 citra..

A. Dataset Penelitian

Dataset penelitian ini mempunyai jumlah sebanyak 120 citra. Citra-citra tersebut terbagi menjadi tiga kategori yaitu, *Bacterial leaf blight* atau Hawar daun, *Brown spot*, dan *Leaf smut* yang setiap kelasnya terdiri dari 40 citra. Berikut adalah contoh citra berpenyakit pada Tabel II.

TABEL II
CITRA DAUN BERPENYAKIT

		
<i>Bacterial leaf blight</i>	<i>Brown spot</i>	<i>Leaf smut</i>

B. Pra Pemrosesan

Sesuai dengan alur penelitian tadi citra yang telah dikelompokkan tersebut, semua citra dilakukan *resize* pixel dengan ukuran 400 × 400 piksel. Berikut tampilan gambar yang telah melakukan proses *resize* pada Gambar 3.



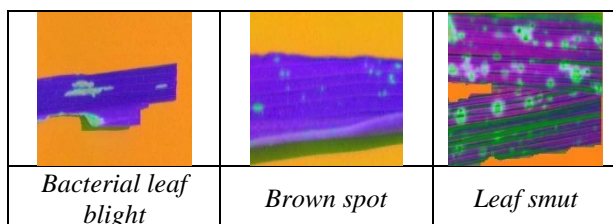
Gambar 3. Gambar Citra Ukuran 400 × 400

Persentase pembagian data latih dan data uji yaitu 75% : 25%. Jadi data latih yang di dapat dengan banyak 90 citra, Sementara itu, terdapat 30 citra sebagai data uji yang digunakan dalam pengujian.

C. Konversi ke Citra L^*a^*b

Tahapan pertama pada pengolahan citra. Sebelumnya, dilakukan transformasi warna dari citra dalam format RGB dilakukan untuk mengubahnya menjadi gambar dengan menggunakan skema warna L^*a^*b . Proses tersebut menjadi penting karena tahap segmentasi menggunakan algoritma K-Means diterapkan dalam citra yang menggunakan skema warna L^*a^*b sebagai dasar. Di bawah ini adalah hasil dari citra daun berpenyakit setelah dikonversi dari skema warna RGB menjadi format gambar L^*a^*b terdapat pada Tabel III.

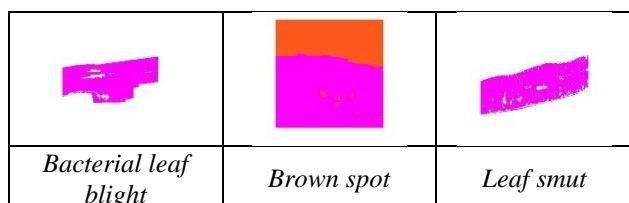
TABEL III
CITRA L^*a^*b



D. Segmentasi

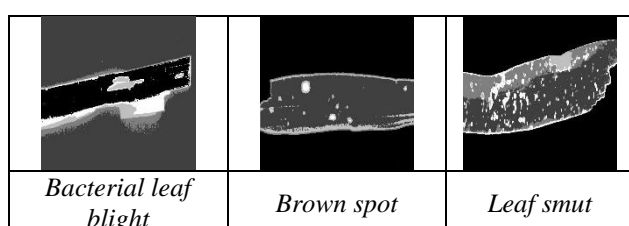
Tahap berikutnya adalah segmentasi. Tujuan dari proses segmentasi ini adalah untuk memisahkan area atau bagian objek dengan area latar belakang pada gambar agar memudahkan dalam melakukan analisis lebih lanjut. Segmentasi dilakukan melalui code yang dibuat di perangkat lunak yaitu Matlab. Pada proses segmentasi, algoritma yang dipakai dalam penelitian ini ialah segmentasi *k-means clustering*. Nilai parameter yang dipakai dalam clustering ini ialah 5. Hasil tahap segmentasi terdapat pada Tabel IV.

TABEL IV
CITRA SEGMENTASI



Setelah mendapatkan objek bercak dari hasil segmentasi, tindakan selanjutnya adalah mengonversi citra objek setelah melalui proses segmentasi diubah menjadi citra dalam format grayscale. Tindakan ini menjadi penting karena ketika melakukan ekstraksi fitur tekstur menggunakan GLCM, proses tersebut hanya bisa dilakukan ketika gambar sudah dikonversi menjadi skema grayscale. Berikut contoh dari proses konversi *RGB* ke *Grayscale* terdapat pada Tabel V.

TABEL V
CITRA GRAYSCALE



E. Ekstraksi GLCM

Tujuan utama dari ekstraksi fitur ini adalah untuk mengidentifikasi dan mendeskripsikan karakteristik atau fitur dari objek yang ingin Anda deteksi pada gambar. Fitur yang membedakan adalah fitur tekstur daun. Mengenai sifat tekstur menggunakan metode GLCM dengan parameter kontras, korelasi, energi dan homogenitas. Materi visual yang akan dipakai dalam proses pelatihan akan disiapkan kemudian, diikuti oleh pembuatan database gambar pelatihan yang akan digunakan dalam proses pengujian. Berikut adalah hasil dari ekstraksi fitur pada data pelatihan menggunakan *offset* 0° terdapat pada Tabel VI.

TABEL VI
HASIL EKSTRAKSI DATA LATIH

Citra	Kontras	Korelasi	Energi	Homogeneity
1	0.44839	0.94708	0.54652	0.98511
2	0.28765	0.96702	0.56928	0.99201
...
45	0.18825	0.98682	0.56873	0.98854
46	0.63328	0.92614	0.36973	0.96736
...
89	1.15211	0.90179	0.30776	0.97191
90	0.61814	0.92587	0.43300	0.97161

Pada Tabel VI tersebut untuk citra 1 – 30 termasuk dalam kelas *Bacterial leaf blight*, kemudian citra 31 – 60 termasuk dalam citra *Brown spot*, dan terakhir yaitu citra 61 – 90 merupakan kelas *Leaf smut*.

Kemudian hasil ekstraksi pada data uji dengan *offset* 0° terdapat pada Tabel VII.

TABEL VII
HASIL EKSTRAKSI DATA UJI

Citra	Kontras	Korelasi	Energi	Homogeneity
1	0.53377	0.95496	0.38732	0.98137
2	0.08624	0.99164	0.37124	0.99521
...
15	0.35579	0.96881	0.24947	0.97625
16	0.68503	0.96145	0.33173	0.97499
...
29	0.32075	0.94438	0.58889	0.98873
30	0.47531	0.97162	0.48510	0.98318

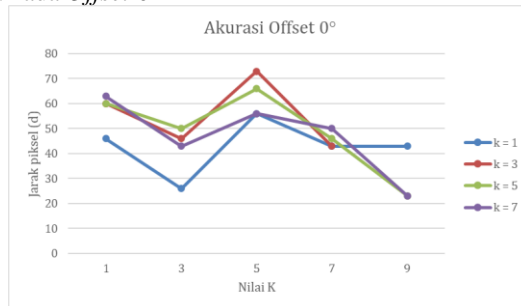
F. Klasifikasi K-NN

Setelah menerima database citra dari hasil training, proses pengelompokan data dapat dijalankan dengan memanfaatkan algoritma *K-nearest neighbor*. *K-Nearest Neighbor* ialah teknik algoritma yang dipakai untuk memperoleh informasi dengan melakukan perhitungan jarak atau kedekatan.

Berdasarkan prinsip ini, data yang baru dapat ditempatkan dalam kelompok yang paling banyak muncul di antara sejumlah tetangga terdekat yang ditentukan.

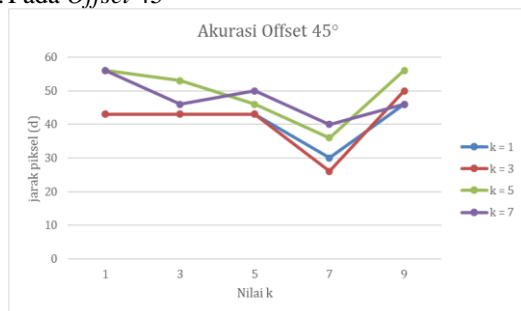
Biasanya, nilai k ini merupakan bilangan bulat positif dan cenderung memiliki nilai yang kecil. Setelah itu, hasil klasifikasi dapat ditentukan. Perhitungan dilakukan menggunakan algoritma K-NN, di mana jarak antara data uji dan data pelatihan dihitung untuk mencari tetangga terdekat. Pada penelitian ini, parameter k memiliki nilai 1, 3, 5, dan 7. Kemudian jarak piksel atau nilai d yang dipakai pada penelitian ini ialah 1, 3, 5, 7, dan 9 dengan nilai sudut yang berbeda-beda. Berikut merupakan hasil nilai akurasi dari klasifikasi citra uji menggunakan ekstraksi fitur *GLCM* dan *K-NN* :

1. Pada Offset 0°



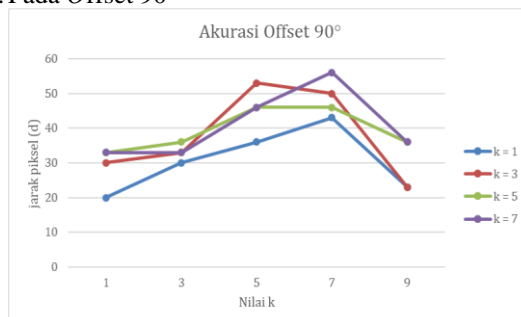
Gambar 4. Hasil Akurasi pada Offset 0°

2. Pada Offset 45°



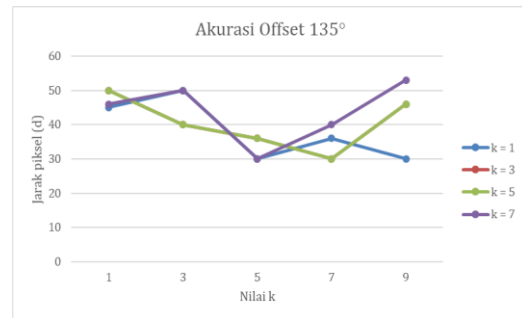
Gambar 5. Hasil Akurasi pada Offset 45°

3. Pada Offset 90°



Gambar 6. Hasil Akurasi pada Offset 90°

4. Pada Offset 135°



Gambar 7. Hasil Akurasi pada Offset 135°

Dari beberapa data gambar hasil akurasi diatas mulai dari perhitungan offset 0° sampai dengan offset 135° didapatkan nilai akurasi maksimal, terdapat pada $k = 3$ dan $d = 5$ dengan nilai sebesar 73%.

G. Confusion Matrix

Hasil pengujian pada proses klasifikasi daun berpenyakit menggunakan algoritma KNN menunjukkan bahwa tingkat akurasi tertinggi diperoleh berdasarkan analisis tekstur pada $k = 3$ dengan jarak piksel (d) sebesar 5 pada offset 0° dengan hasil sebesar 73%. Oleh karena itu, algoritma KNN telah terbukti efektif untuk dimanfaatkan dalam tahap klasifikasi. Di bawah ini merupakan tabel matriks kebingungan (confusion matrix) dari hasil percobaan yang dilakukan dengan menggunakan k sebagai parameter yaitu 3 dan jarak piksel dengan nilai 5.

TABEL VIII
PERHITUNGAN CONFUSION MATRIX

Prediksi	<i>Bacterial leaf blight</i>	<i>Brown spot</i>	<i>Leaf smut</i>
<i>Bacterial leaf blight</i>	10	0	0
<i>Brown spot</i>	2	6	2
<i>Leaf smut</i>	3	1	6

Tabel VIII menampilkan hasil dari proses klasifikasi yang menggunakan algoritma KNN dengan nilai parameter $k = 3$. Nilai $k = 3$ menghasilkan akurasi tertinggi sebesar 73%.

Kemudian setelah diperoleh tabel perhitungan *confusion matrix*, dilanjutkan dengan melakukan perhitungan untuk mengukur tingkat akurasi, presisi, dan recall. Untuk nilai akurasi sudah kita dapatkan dari proses klasifikasi dengan nilai akurasi sebesar 73%. Berikut adalah hasil perhitungan presisi dan recall setiap kelas pada Tabel IX.

TABEL IX
PERHITUNGAN PRESISI DAN RECALL

Kelas	Presisi	Recall
<i>Bacterial leaf blight</i>	67%	100%
<i>Brown spot</i>	85%	60%
<i>Leaf smut</i>	75%	60%

IV. KESIMPULAN

Algoritma K-NN terbukti cukup efektif dalam mengidentifikasi penyakit daun pada tanaman padi. Dalam penelitian ini, K-NN digunakan untuk mengklasifikasikan

penyakit daun tanaman padi menjadi beberapa kategori penyakit berdasarkan fitur-fitur yang diambil dari daun tersebut. Hasil dari klasifikasi mempunyai tingkat akurasi sebesar 73% yang didapat dari $k = 3$ dengan d (jarak piksel) = 5.

kinerja model klasifikasi bervariasi untuk setiap kelas. Model mencapai recall yang sempurna (100%) untuk kelas *Bacterial leaf blight*, tetapi recall yang lebih rendah (60%) untuk kelas *Brown spot* dan *Leaf smut*. Hal ini menunjukkan bahwa model cenderung memiliki tingkat kepekaan yang lebih rendah dalam mendeteksi sampel-sampel *Brown spot* dan *Leaf smut* yang sebenarnya.

Namun, presisi secara relatif cukup tinggi untuk semua kelas, dengan tingkat tertinggi terlihat pada kelas *Brown spot* (85%). Ini menunjukkan bahwa model cenderung memberikan hasil positif yang benar untuk kelas-kelas tersebut, tetapi ada kemungkinan adanya beberapa *false positive*. Dalam keseluruhan, model mungkin memberikan hasil yang lebih akurat untuk kelas *Bacterial leaf blight* dibandingkan dengan kelas *Brown spot* dan kelas *Leaf smut*.

DAFTAR PUSTAKA

- [1] T. Purwa, "Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017)," *Jurnal Matematika, Statistika dan Komputasi*, vol. 16, no. 1, p. 58, Jun. 2019, doi: 10.20956/jmsk.v16i1.6494.
- [2] S. Agustiani, Y. Tajul Arifin, A. Junaidi, S. Khotimatul Wildah, and A. Mustopa, "Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram 1,*)," [Online]. Available: <https://www.kaggle.com/vbookshelf/rice-leaf->
- [3] J. Sosial and E. Pertanian, "The Impact Of Wetland Conversion On Food Security In Daerah Istimewa Yogyakarta," 2018.
- [4] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *Jurnal Riset Komputer*, vol. 8, no. 6, pp. 2407–389, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [5] A. Purnamawati, W. Nugroho, D. Putri, and W. F. Hidayat, "InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan Attribution-NonCommercial 4.0 International. Some rights reserved Deteksi Penyakit Daun pada Tanaman Padi Menggunakan Algoritma Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN," vol. 5, no. 1, 2020, doi: 10.30743/infotekjar.v5i1.2934.
- [6] 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
- [7] J. Khatib Sulaiman, G. A. Pongdatu, J. Rusman, and I. Artikel Abstrak, "Klasifikasi Penyakit pada Tanaman Kopi Arabika Menggunakan Metode K-Nearest Neighbor (KNN) Berbasis Citra," *Indonesian Journal of Computer Science*, vol. 12, Apr. 2023.
- [8] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca," *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, Dec. 2019, doi: 10.33096/ilkom.v11i3.489.269-274.
- [9] F. Liantoni, "Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor," *ULTIMATICS*, vol. VII, no. 2, 2015.
- [10] V. Susanti, S. Dwi Nurcahyanti, and D. R. Masnilah, "Perkembangan Penyakit dan Pertumbuhan Lima Varietas Padi (*Oryza sativa* L.) dengan Sistem Tanam Blok The Development of a Disease and Growth Five Varieties of Rice (*Oryza sativa* L.) with a System of Planting Block," 2018.
- [11] R. Masnilah, W. Sri Wahyuni, S. Dwi N, A. Majid, H. Susilo, and A. Wafa, "Insiden dan Keparahan Penyakit Penting Tanaman Padi di Kabupaten Jember," *Agritrop*, vol. 18, Jul. 2020.
- [12] M. A. Azim, M. K. Islam, M. M. Rahman, and F. Jahan, "An effective feature extraction method for rice leaf disease classification," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 2, pp. 463–470, Apr. 2021, doi: 10.12928/TELKOMNIKA.v19i2.16488.
- [13] M. Muhathir, M. H. Santoso, and D. A. Larasati, "Wayang Image Classification Using SVM Method and GLCM Feature Extraction," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 4, no. 2, pp. 373–382, Jan. 2021, doi: 10.31289/jite.v4i2.4524.
- [14] D. P. Pamungkas, "Ekstraksi Citra menggunakan Metode GLCM dan KNN untuk Identifikasi Jenis Anggrek (Orchidaceae)," *Innovation in Research of Informatics (INNOVATICS)*, vol. 1, no. 2, pp. 51–56, 2019.
- [15] A. Pariyandani, E. P. Wanti, and Muhathir, "Analysis of the Naïve Bayes Method in Classifying Formalized Fish Images Using GLCM Feature Extraction," *Journal of Computer Science, Information Technology and Telecommunication Engineering*, Sep. 2020, doi: 10.30596/jcositte.v1i2.5171.
- [16] K. Lloyd, P. L. Rosin, D. Marshall, and S. C. Moore, "Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures," *Mach Vis Appl*, vol. 28, no. 3–4, pp. 361–371, May 2017, doi: 10.1007/s00138-017-0830-x.
- [17] F. Liantoni and H. Nugroho, "Klasifikasi Daun Herbal Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor," vol. 5, no. 1, 2015.