Deteksi Penyakit pada Tanaman Padi Menggunakan Metode Fuzzy K-Nearest Neighbor

1st Diki Iskandar
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
diskandar@student.telkomuniversity.ac.id

2nd Tjokorda Agung Budi Wirayuda
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
cokagung@telkomuniversity.ac.id

3rd Prasti Eko Yunanto
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
gppras@telkomuniversity.ac.id

Abstrak — Penyakit pada tanaman padi merupakan salah satu faktor utama yang menyebabkan penurunan produktivitas hasil panen. Identifikasi penyakit yang masih dilakukan secara visual sering kali kurang akurat karena adanya kemiripan gejala antarjenis penyakit, sehingga diperlukan sistem klasifikasi berbasis citra digital yang mampu memberikan hasil secara lebih objektif dan akurat. Penelitian ini bertujuan mengembangkan sistem klasifikasi penyakit daun padi menggunakan metode Fuzzy K-Nearest Neighbor (F-kNN) dengan memanfaatkan kombinasi ekstraksi fitur citra. Fitur yang digunakan meliputi Gray Level Co-occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG), Histogram Warna HSV, dan Local Binary Pattern (LBP). Seluruh fitur melalui proses *standardization*, reduksi dimensi menggunakan *Principal Component Analysis* (PCA), dan *L2 normalization* sebelum diklasifikasikan. Optimasi parameter dilakukan menggunakan *GridSearchCV*, sedangkan evaluasi model menggunakan *Stratified K-Fold Cross Validation* dan data uji. Hasil penelitian menunjukkan bahwa kombinasi fitur HSV dan LBP menghasilkan representasi fitur terbaik, sedangkan konfigurasi parameter k = 3, kernel weighted, m = 2, dan cosine distance memberikan performa optimal. Model yang diusulkan memperoleh akurasi pengujian sebesar 97,98% dan menunjukkan kinerja yang lebih baik dibandingkan beberapa algoritma klasifikasi pembanding. Hasil tersebut menunjukkan bahwa pendekatan F-kNN efektif dalam mengklasifikasikan penyakit daun padi berbasis citra digital serta berpotensi mendukung proses identifikasi penyakit secara otomatis.

Kata kunci — fuzzy k-nearest neighbor, klasifikasi citra, penyakit padi, tanaman padi, ekstraksi fitur

I. PENDAHULUAN

Hasil panen padi pada tahun 2021 mengalami penurunan menurut Badan Pusat Statistik (BPS) [1]. Penurunan produksi padi tersebut mencapai 233.910 ton, yang merupakan angka cukup signifikan mengingat padi merupakan komoditas utama sebagai bahan makanan pokok masyarakat Indonesia. Apabila kondisi ini tidak segera ditanggulangi, maka penurunan hasil panen padi dapat terus berlanjut dan berpotensi menyebabkan kenaikan harga yang signifikan. Kondisi tersebut tentu akan berdampak pada sektor ekonomi, khususnya terhadap meningkatnya harga kebutuhan pokok lainnya.

Beberapa faktor yang menyebabkan penurunan hasil panen padi antara lain berkurangnya lahan sawah, kondisi cuaca yang tidak menentu, serta serangan hama dan penyakit pada tanaman padi. Penelitian ini berfokus pada faktor penurunan hasil panen yang disebabkan oleh penyakit tanaman padi. Menurut Herawati dalam penelitiannya pada tahun 2012, penyakit pada tanaman padi dikelompokkan menjadi tiga jenis utama, yaitu penyakit bercak daun cokelat (*Helminthosporium oryzae*), penyakit blast (*Pyricularia oryzae*), dan penyakit busuk pelepah daun (*Rhizoctonia* sp.) [2]. Masing-masing penyakit memiliki karakteristik gejala serta metode penanganan yang berbeda sehingga identifikasi penyakit secara tepat menjadi faktor penting dalam menentukan tindakan pengendalian yang sesuai. Kesalahan dalam proses identifikasi dapat menyebabkan penanganan yang kurang tepat sehingga berdampak pada penurunan kualitas maupun kuantitas hasil produksi padi.

Perkembangan teknologi *computer vision* memberikan peluang untuk membantu proses identifikasi penyakit tanaman secara otomatis melalui analisis citra digital. Dalam identifikasi penyakit tanaman padi, citra daun diproses melalui tahapan prapemrosesan dan ekstraksi fitur sehingga karakteristik warna, tekstur, dan bentuk dapat dimanfaatkan sebagai representasi citra dalam proses klasifikasi [3]. Pendekatan ini diharapkan mampu membantu petani maupun pihak terkait dalam melakukan identifikasi penyakit secara lebih cepat, objektif, dan konsisten dibandingkan dengan pengamatan secara manual.

Berbagai penelitian telah memanfaatkan teknik pengolahan citra dan *machine learning* untuk mendeteksi penyakit tanaman padi. Muslim dkk. mengembangkan sistem deteksi penyakit pada tanaman padi dan cabai berbasis klasifikasi citra menggunakan *Convolutional Neural Network* (CNN) yang diimplementasikan pada platform Android, serta menunjukkan bahwa pendekatan *deep learning* mampu mengenali penyakit tanaman secara otomatis [4]. Penelitian lain oleh Jinan dan Hayadi juga menerapkan metode CNN untuk mengklasifikasikan penyakit tanaman padi melalui citra daun dan memperoleh performa klasifikasi yang baik [5]. Selain itu, Saputra dkk. menggunakan ekstraksi fitur *Gray Level Co-occurrence Matrix* (GLCM) yang dikombinasikan dengan algoritma *K-Nearest Neighbor* (kNN) untuk klasifikasi citra penyakit daun padi, sehingga menunjukkan bahwa fitur tekstur berperan penting dalam membedakan karakteristik setiap jenis penyakit [6]. Hasil penelitian-penelitian tersebut menunjukkan bahwa deteksi penyakit tanaman padi berbasis citra telah berkembang cukup pesat melalui pendekatan *deep learning* maupun *machine learning* berbasis ekstraksi fitur.

Di sisi lain, metode *Fuzzy K-Nearest Neighbor* (F-kNN) telah banyak diterapkan pada berbagai permasalahan klasifikasi dan diagnosis. Rahayu dan Mendes menerapkan F-kNN untuk mendeteksi penyakit tanaman rambutan berdasarkan citra daun dan menunjukkan bahwa konsep derajat keanggotaan mampu meningkatkan proses pengambilan keputusan pada data yang memiliki kemiripan antar kelas [7]. Andromeda dkk. menggunakan F-kNN untuk diagnosis

penyakit tanaman kentang [8], sedangkan Akbari dkk. menerapkannya pada diagnosis penyakit tanaman cabai [9]. Selain pada bidang pertanian, metode F-kNN juga digunakan untuk klasifikasi penyakit demam berdarah [10] serta pengklasifikasian kualitas buah jeruk [11]. Berbagai penelitian tersebut menunjukkan bahwa F-kNN memiliki kemampuan yang baik dalam menangani data yang bersifat ambigu karena setiap data diberikan derajat keanggotaan terhadap seluruh kelas, bukan hanya ditentukan ke dalam satu kelas secara mutlak.

Berdasarkan penelitian-penelitian terdahulu, dapat diketahui bahwa klasifikasi penyakit daun padi berbasis citra masih didominasi oleh metode CNN maupun kNN. Sementara itu, penerapan metode F-kNN pada kasus penyakit daun padi masih relatif terbatas, meskipun metode tersebut telah menunjukkan hasil yang baik pada berbagai kasus klasifikasi penyakit tanaman lainnya [7]-[11]. Selain itu, pendekatan CNN umumnya memerlukan jumlah data latih yang lebih besar dibandingkan pendekatan *machine learning* berbasis ekstraksi fitur agar dapat menghasilkan performa yang optimal. Di sisi lain, salah satu kelemahan algoritma kNN adalah setiap data tetangga dianggap memiliki kontribusi yang sama dalam proses penentuan kelas sehingga berpotensi menghasilkan kesalahan klasifikasi ketika data berada pada batas antar kelas atau memiliki karakteristik yang saling menyerupai [12]. Kondisi tersebut menunjukkan adanya peluang untuk mengembangkan pendekatan klasifikasi yang mampu menangani ketidakpastian data dengan lebih baik pada proses identifikasi penyakit daun padi.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penerapan metode F-kNN dengan memanfaatkan kombinasi fitur warna menggunakan HSV, fitur tekstur menggunakan GLCM dan *Local Binary Pattern* (LBP), serta fitur bentuk menggunakan *Histogram of Oriented Gradients* (HOG). Integrasi konsep logika fuzzy pada algoritma kNN memungkinkan setiap citra daun memiliki derajat keanggotaan terhadap setiap kelas sehingga proses klasifikasi diharapkan menjadi lebih adaptif terhadap data yang memiliki karakteristik serupa [12]. Tujuan penelitian ini adalah merancang dan membangun model identifikasi penyakit daun padi menggunakan metode F-kNN serta mengevaluasi tingkat akurasi model yang dihasilkan. Hasil penelitian ini diharapkan dapat menjadi alternatif pendekatan klasifikasi penyakit daun padi berbasis citra dengan performa yang baik sehingga mampu mendukung identifikasi penyakit secara dini dan membantu pengambilan keputusan penanganan yang lebih tepat.

II. KAJIAN TEORI

Kajian teori merupakan landasan konseptual yang mendukung pengembangan metode pada penelitian ini. Teori-teori yang dibahas meliputi *computer vision*, pengolahan citra digital, metode ekstraksi fitur, *feature scaling* dan *dimensional reduction*, serta algoritma F-kNN. Seluruh konsep tersebut digunakan sebagai dasar dalam membangun model klasifikasi penyakit daun padi berbasis citra, mulai dari proses pengolahan citra hingga proses klasifikasi.

Computer Vision

Computer vision merupakan salah satu cabang kecerdasan buatan (*Artificial Intelligence/AI*) yang berfokus pada kemampuan komputer untuk memperoleh, memproses, menganalisis, dan memahami informasi visual dari citra maupun video digital [13]. Teknologi ini memungkinkan sistem mengenali karakteristik objek, seperti warna, tekstur, bentuk, dan pola, sehingga dapat dimanfaatkan untuk berbagai aplikasi, termasuk pengenalan objek, klasifikasi citra, deteksi objek, dan segmentasi citra [14]. Dalam implementasinya, sistem *computer vision* umumnya melibatkan beberapa tahapan, yaitu akuisisi citra, *preprocessing*, ekstraksi fitur, dan klasifikasi untuk menghasilkan informasi yang sesuai dengan tujuan analisis [15], [16].

Dalam bidang pertanian, *computer vision* telah banyak dimanfaatkan untuk mendukung konsep *smart farming*, seperti pemantauan pertumbuhan tanaman, identifikasi gulma, deteksi hama, klasifikasi kualitas hasil panen, serta identifikasi penyakit tanaman [14], [17]. Pada penelitian ini, *computer vision* digunakan sebagai landasan dalam membangun sistem klasifikasi penyakit daun padi berbasis citra. Citra daun padi diproses melalui tahapan *preprocessing* dan ekstraksi fitur untuk memperoleh representasi karakteristik visual, kemudian hasil ekstraksi tersebut digunakan sebagai masukan bagi algoritma F-kNN dalam proses klasifikasi [18], [19].

Pengolahan Citra Digital

Pengolahan citra digital merupakan proses memanipulasi dan menganalisis citra dalam bentuk digital untuk meningkatkan kualitas citra maupun mengekstraksi informasi yang relevan [20]. Dalam penelitian ini, pengolahan citra digital berperan sebagai tahapan awal sebelum proses ekstraksi fitur dan klasifikasi, sehingga karakteristik visual daun padi dapat direpresentasikan secara lebih baik. Citra digital sendiri merupakan representasi dua dimensi suatu objek yang tersusun atas kumpulan piksel, di mana setiap piksel menyimpan informasi intensitas atau warna yang membentuk keseluruhan citra [21].

Tahapan pengolahan citra pada umumnya meliputi *preprocessing* dan ekstraksi fitur. *Preprocessing* bertujuan untuk meningkatkan kualitas citra serta mengurangi pengaruh *noise* melalui proses seperti *resizing*, normalisasi, dan konversi citra berwarna (RGB) menjadi citra *grayscale* [22]. Konversi dari ruang warna RGB ke *grayscale* dilakukan untuk menyederhanakan informasi citra dengan menggabungkan komponen merah (*Red*), hijau (*Green*), dan biru (*Blue*) menjadi satu nilai intensitas menggunakan Persamaan (1).

$$Gray = 0{,}299R + 0{,}587G + 0{,}114B \quad (1)$$

Dengan R, G, dan B masing-masing menyatakan komponen warna merah (Red), hijau (Green), dan biru (Blue). Hasil pengolahan citra selanjutnya digunakan sebagai masukan pada proses ekstraksi fitur untuk memperoleh representasi warna, tekstur, dan bentuk yang mendukung proses klasifikasi penyakit daun padi.

Metode Ekstraksi Fitur

Ekstraksi fitur merupakan proses untuk memperoleh karakteristik penting dari suatu citra sehingga dapat digunakan sebagai representasi objek pada proses klasifikasi. Pada penelitian ini, representasi citra daun padi dibangun

melalui kombinasi fitur warna, tekstur, dan bentuk agar informasi visual yang diperoleh menjadi lebih komprehensif. Metode ekstraksi fitur yang digunakan meliputi GLCM, HOG, histogram warna HSV, dan LBP. Kombinasi keempat metode tersebut diharapkan mampu meningkatkan kemampuan model dalam membedakan karakteristik berbagai jenis penyakit daun padi.

GLCM merupakan metode ekstraksi fitur tekstur yang menganalisis hubungan spasial antar piksel berdasarkan tingkat keabuan (*gray level*) pada jarak dan arah tertentu [23]. Dari matriks GLCM dapat dihitung beberapa fitur statistik, yaitu *contrast*, *correlation*, *energy*, dan *homogeneity*, yang dirumuskan pada Persamaan (2)-(5).

$$Contrast = \sum_i \sum_j (i-j)^2 P(i,j) \quad (2)$$

$$Correlation = \frac{\sum_i \sum_j (i-\mu_i)\left(j-\mu_j\right)P(i,j)}{\sigma_i \sigma_j} \quad (3)$$

$$Energy = \sum_i \sum_j \left(P(i,j)\right)^2 \quad (4)$$

$$Homogeneity = \sum_i \sum_j \frac{P(i,j)}{1+|i-j|} \quad (5)$$

Dengan $P(i,j)$ menyatakan probabilitas kemunculan pasangan piksel pada tingkat keabuan i dan j, μ merupakan nilai rata-rata (*mean*), sedangkan σ merupakan simpangan baku (*standard deviation*) dari distribusi tingkat keabuan.

Selain tekstur, informasi bentuk direpresentasikan menggunakan HOG yang menghitung distribusi arah gradien pada citra sehingga mampu menggambarkan pola tepi dan struktur objek [24], [25]. Besar gradien (*magnitude*) dan arah gradien (*orientation*) dihitung menggunakan Persamaan (6) dan Persamaan (7).

$$G = \sqrt{G_x^2 + G_y^2} \quad (6)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (7)$$

Dengan G_x dan G_y masing-masing merupakan gradien citra pada arah horizontal dan vertikal, G adalah besar gradien, sedangkan θ menunjukkan arah gradien pada setiap piksel.

Karakteristik warna direpresentasikan menggunakan histogram warna pada ruang warna HSV (*Hue, Saturation,* dan *Value*) yang dinilai lebih sesuai dengan persepsi visual manusia dibandingkan ruang warna RGB [26], [27]. Komponen *value* dan *saturation* dihitung menggunakan Persamaan (8) dan Persamaan (9).

$$V = max(R,G,B) \quad (8)$$

$$S = \frac{V - \min(R,G,B)}{V}, \text{jika } V \neq 0 \quad (9)$$

Dengan R, G, dan B masing-masing menyatakan intensitas komponen warna merah (*Red*), hijau (*Green*), dan biru (*Blue*), sedangkan V dan S berturut-turut merupakan komponen *Value* dan *Saturation* pada ruang warna HSV.

Selain fitur warna dan bentuk, penelitian ini juga memanfaatkan LBP sebagai deskriptor tekstur lokal yang membandingkan intensitas piksel pusat dengan piksel-piksel di sekitarnya [28]-[31]. Nilai LBP dihitung menggunakan Persamaan (10).

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} s\left(g_p - g_c\right) \cdot 2^p \quad (10)$$

Dengan fungsi ambang (*threshold*) didefinisikan sebagai,

$$\begin{cases} s(x) = 1, & \text{jika } x \geq 0 \\ s(x) = 0, & \text{jika } x < 0 \end{cases} \quad (11)$$

di mana g_c merupakan intensitas piksel pusat, g_p adalah intensitas piksel tetangga, P menyatakan jumlah piksel tetangga, sedangkan $s(x)$ merupakan fungsi ambang yang menghasilkan nilai biner berdasarkan perbandingan intensitas.

Kombinasi fitur GLCM, HOG, HSV, dan LBP menghasilkan representasi citra yang mencakup informasi tekstur, bentuk, dan warna sehingga diharapkan mampu meningkatkan kemampuan model F-kNN dalam membedakan karakteristik berbagai penyakit daun padi.

Feature Scaling dan Dimensional Reduction

Feature scaling dan *dimensional reduction* merupakan tahapan prapemrosesan fitur yang bertujuan untuk menyeragamkan skala data, mengurangi dimensi fitur, serta meningkatkan efisiensi dan stabilitas proses klasifikasi [32]. Pada penelitian ini, tahapan tersebut dilakukan melalui tiga proses, yaitu *standardization*, *Principal Component Analysis* (PCA), dan *L2 normalization*. Urutan ini dipilih agar setiap fitur memiliki skala yang sebanding sebelum dilakukan reduksi dimensi dan normalisasi vektor.

Standardization digunakan untuk mentransformasikan setiap fitur sehingga memiliki rata-rata (*mean*) sebesar 0 dan simpangan baku (*standard deviation*) sebesar 1. Proses ini penting untuk mencegah dominasi fitur yang memiliki rentang nilai lebih besar, terutama pada algoritma berbasis jarak seperti F-kNN. Transformasi *standardization* dirumuskan pada Persamaan (12).

$$Z = \frac{X - \mu}{\sigma} \quad (12)$$

Dengan Z merupakan nilai hasil *standardization*, X adalah nilai fitur asli, μ menyatakan nilai rata-rata (*mean*), dan σ merupakan simpangan baku (*standard deviation*).

Setelah proses *standardization*, dilakukan reduksi dimensi menggunakan PCA. PCA mentransformasikan sekumpulan fitur yang saling berkorelasi menjadi sejumlah komponen utama (*principal components*) yang saling ortogonal dengan tetap mempertahankan sebagian besar variasi data. Transformasi PCA dinyatakan pada Persamaan (13).

$$Z = XW \quad (13)$$

Dengan X merupakan matriks data yang telah distandardisasi, W adalah matriks *eigenvector* yang diperoleh dari matriks kovarians, sedangkan Z merupakan data hasil transformasi pada ruang berdimensi lebih rendah.

Tahap terakhir adalah *L2 normalization*, yaitu proses normalisasi yang mengubah setiap vektor fitur menjadi memiliki panjang (*norm*) sebesar satu sehingga perhitungan jarak antar data menjadi lebih stabil. Transformasi *L2 normalization* dirumuskan pada Persamaan (14).

$$x' = \frac{x}{|x|_2} \tag{14}$$

Dengan x merupakan vektor fitur asli, x' adalah vektor hasil normalisasi, dan $|x|_2$ menyatakan norma *Euclidean* dari vektor x, yang dihitung menggunakan Persamaan (15).

$$|x|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{15}$$

Di mana x_i merupakan nilai fitur ke-i dan n menyatakan jumlah fitur. Melalui kombinasi *standardization*, PCA, dan *L2 normalization*, fitur yang dihasilkan memiliki skala yang seragam, dimensi yang lebih ringkas, serta lebih sesuai untuk digunakan pada proses klasifikasi berbasis jarak seperti F-kNN.

Fuzzy K-Nearest Neighbor

F-kNN merupakan pengembangan dari algoritma kNN yang mengintegrasikan konsep logika fuzzy ke dalam proses klasifikasi. Metode ini pertama kali diperkenalkan oleh Keller dkk. [33] sebagai solusi terhadap keterbatasan kNN konvensional yang hanya menentukan satu kelas berdasarkan mayoritas dari K tetangga terdekat. Pada kNN, setiap tetangga dianggap memiliki kontribusi yang sama dalam proses klasifikasi, sehingga performanya dapat menurun ketika data berada pada batas antar kelas atau memiliki karakteristik yang saling menyerupai.

F-kNN mengatasi permasalahan tersebut dengan memberikan derajat keanggotaan (*membership value*) terhadap setiap kelas. Dengan demikian, suatu data tidak hanya diklasifikasikan ke dalam satu kelas tertentu, tetapi juga memiliki tingkat keanggotaan pada seluruh kelas sehingga mampu merepresentasikan tingkat ketidakpastian data. Derajat keanggotaan suatu data uji x terhadap kelas ke-i dihitung menggunakan Persamaan (16).

$$u_i(x) = \frac{\sum_{j=1}^{K} \left(\frac{u_{ij}}{d(x, x_j)^{2/(m-1)}} \right)}{\sum_{j=1}^{K} \left(\frac{1}{d(x, x_j)^{2/(m-1)}} \right)} \tag{16}$$

Dengan $u_i(x)$ merupakan derajat keanggotaan data uji terhadap kelas ke-i, u_{ij} adalah derajat keanggotaan data latih ke-j, $d(x, x_j)$ menyatakan jarak antara data uji dan data latih, K merupakan jumlah tetangga terdekat, sedangkan mmm adalah parameter *fuzziness* dengan nilai $m > 1$.

Perhitungan jarak pada F-kNN umumnya menggunakan *Euclidean Distance*, yang mengukur kedekatan antara data uji dan data latih berdasarkan ruang fitur. Persamaan *Euclidean Distance* ditunjukkan pada Persamaan (17).

$$d(x, x_j) = \sqrt{\sum_{k=1}^{n} (x_k - x_{jk})^2} \tag{17}$$

Dengan x_k merupakan nilai fitur ke-k pada data uji, x_{jk} adalah nilai fitur ke-k pada data latih ke-j, dan n menyatakan jumlah fitur.

Selain menggunakan pembobotan berdasarkan jarak, performa F-kNN dapat ditingkatkan melalui fungsi pembobotan (*weighting*) atau *kernel* yang memberikan kontribusi berbeda terhadap setiap tetangga terdekat. Pada penelitian ini digunakan tiga pendekatan, yaitu *Inverse Distance Weighting*, *Radial Basis Function* (RBF), dan *Sigmoid Kernel*. Pembobotan berbasis *inverse distance* memberikan bobot yang berbanding terbalik terhadap jarak sehingga data yang lebih dekat memiliki pengaruh lebih besar. Bobot tersebut dirumuskan sebagai Persamaan (18).

$$W(x, x_j) = \frac{1}{d(x, x_j) + \epsilon} \tag{18}$$

Dengan ϵ merupakan konstanta bernilai kecil untuk menghindari pembagian dengan nol. Pendekatan berikutnya adalah RBF yang memodelkan tingkat kemiripan menggunakan fungsi Gaussian, sebagaimana ditunjukkan pada Persamaan (19).

$$K_{RBF}(x, x_j) = \exp\left(-\gamma \|x - x_j\|^2\right) \tag{19}$$

Dengan γ merupakan parameter yang mengontrol tingkat penyebaran (*spread*) fungsi kernel. Selain itu, penelitian ini juga menggunakan *Sigmoid Kernel* yang menghitung tingkat kemiripan berdasarkan hasil perkalian titik (*dot product*) antara dua vektor fitur. Persamaan *Sigmoid Kernel* ditunjukkan pada Persamaan (20).

$$K_{Sigmoid}(x, x_j) = \tanh(\alpha(x^T x_j) + c) \tag{20}$$

Dengan α merupakan parameter kemiringan (*slope*), c adalah konstanta *intercept*, dan $x^T x_j$ menyatakan hasil perkalian titik antara data uji dan data latih.

Melalui penerapan konsep derajat keanggotaan serta pendekatan pembobotan berbasis jarak dan *kernel*, F-kNN mampu memberikan representasi klasifikasi yang lebih adaptif terhadap data yang memiliki karakteristik serupa. Oleh karena itu, metode ini dipilih pada penelitian ini untuk mengklasifikasikan penyakit daun padi berdasarkan kombinasi fitur warna, tekstur, dan bentuk yang diperoleh dari proses ekstraksi fitur.

III. METODE

A. Dataset Penelitian

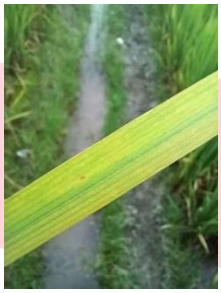
Dataset yang digunakan dalam penelitian ini merupakan dataset publik *Rice Leaf Diseases Detection* yang diperoleh melalui platform Kaggle [34]. Dataset ini berisi kumpulan citra daun tanaman padi yang telah diberi label berdasarkan kondisi kesehatan daun, sehingga sesuai digunakan untuk permasalahan klasifikasi citra (*image classification*). Penggunaan dataset publik memungkinkan proses penelitian

dilakukan menggunakan data yang telah tervalidasi serta memudahkan reproduksibilitas hasil penelitian oleh peneliti lain.

Dataset terdiri atas 18.445 citra daun padi yang dikelompokkan ke dalam sepuluh kelas, yaitu *Bacterial Leaf Blight*, *Brown Spot*, *Healthy*, *Leaf Blast*, *Leaf Scald*, *Narrow Brown Leaf Spot*, *Neck Blast*, *Rice Hispa*, *Sheath Blight*, dan *Tungro*. Setiap citra merepresentasikan kondisi daun dengan karakteristik visual yang berbeda, baik dari aspek warna, tekstur, maupun pola kerusakan akibat penyakit. Contoh citra dari masing-masing kelas ditunjukkan pada **Error! Unknown switch argument.** hingga **Error! Unknown switch argument.**.



GAMBAR 1
Bacterial leaf blight



GAMBAR 2
Brown spot



GAMBAR 3
Healty rice leaf



GAMBAR 4
Leaf blast

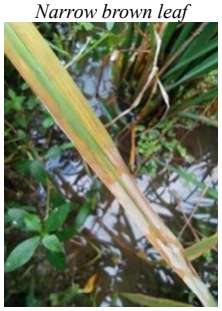




GAMBAR 5
Leaf scald



GAMBAR 6
Narrow brown leaf



GAMBAR 7
Rice hispa

GAMBAR 8
Sheath blight



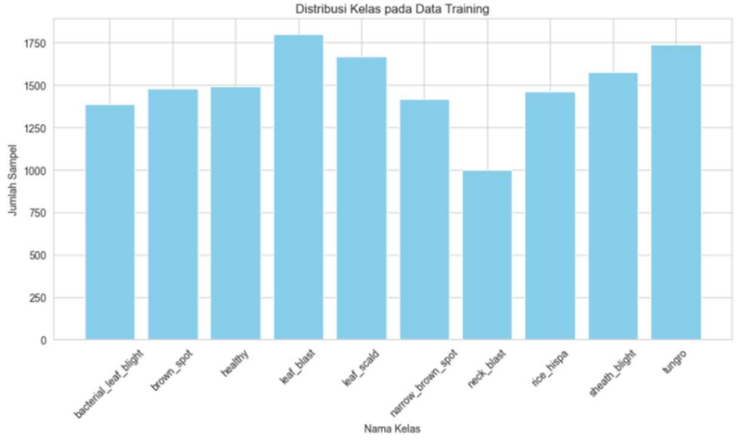
GAMBAR 9
Tungro



GAMBAR 10
Neck blast

Seluruh dataset telah dipisahkan menjadi data pelatihan (*training set*) dan data pengujian (*testing set*). Data pelatihan terdiri atas 15.023 citra, sedangkan data pengujian terdiri atas 3.422 citra. Pembagian ini digunakan untuk membangun model klasifikasi sekaligus mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah digunakan selama proses pelatihan.

Distribusi jumlah citra pada setiap kelas divisualisasikan menggunakan diagram batang untuk memberikan gambaran mengenai penyebaran data pada masing-masing kategori penyakit. Visualisasi distribusi data pelatihan dan data pengujian masing-masing ditunjukkan pada **Error! Unknown switch argument.** dan **Error! Unknown switch argument.**. Secara umum, distribusi data antar kelas relatif seimbang sehingga dapat membantu mengurangi potensi bias selama proses pelatihan dan evaluasi model.



GAMBAR 11
Distribusi Data Pelatihan

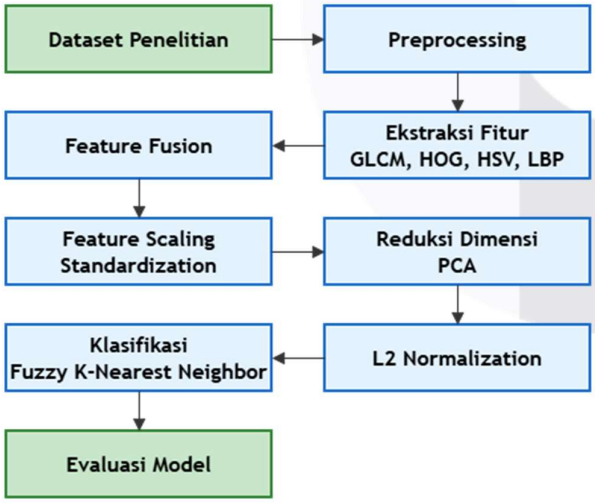


GAMBAR 12
Distribusi Data Pengujian

Seluruh citra pada dataset kemudian diproses melalui tahapan *preprocessing*, ekstraksi fitur, *feature scaling*, reduksi dimensi, dan klasifikasi menggunakan metode F-kNN sesuai dengan tahapan penelitian yang dijelaskan pada subbab berikutnya.

Alur Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang dirancang secara sistematis untuk menghasilkan model klasifikasi penyakit daun padi berbasis citra menggunakan metode F-kNN. Setiap tahapan saling berkaitan, dimulai dari penyediaan dataset, pengolahan citra, pembentukan representasi fitur, hingga proses klasifikasi dan evaluasi performa model. Alur penelitian secara keseluruhan ditunjukkan pada **Error! Unknown switch argument.**.



GAMBAR 13
Alur penelitian yang diusulkan

Berdasarkan **Error! Unknown switch argument.**, penelitian diawali dengan penggunaan dataset *Rice Leaf Diseases Detection* yang terdiri atas citra daun padi dari berbagai kategori penyakit. Seluruh citra terlebih dahulu melalui tahap *preprocessing* untuk menghasilkan data yang memiliki ukuran dan format yang seragam sehingga siap digunakan pada proses ekstraksi fitur.

Tahap berikutnya adalah ekstraksi fitur menggunakan empat metode yang saling melengkapi, yaitu GLCM untuk merepresentasikan karakteristik tekstur global, HOG untuk menangkap informasi bentuk dan tepi objek, Histogram warna HSV untuk menggambarkan distribusi warna, serta LBP untuk merepresentasikan tekstur lokal. Seluruh fitur yang dihasilkan kemudian digabungkan melalui proses *feature fusion* sehingga setiap citra direpresentasikan dalam sebuah vektor fitur yang lebih komprehensif.

Vektor fitur hasil penggabungan selanjutnya diproses melalui tahap *feature scaling* menggunakan standardisasi untuk menyamakan skala antarfitur. Setelah itu dilakukan reduksi dimensi menggunakan PCA guna mengurangi redundansi informasi sekaligus mempertahankan sebagian besar variasi data. Hasil reduksi dimensi kemudian dinormalisasi menggunakan *L2 Normalization* agar setiap vektor fitur memiliki panjang yang seragam sehingga proses perhitungan jarak menjadi lebih stabil.

Tahap klasifikasi dilakukan menggunakan metode F-kNN. Algoritma ini menentukan kelas suatu citra berdasarkan tingkat kedekatan terhadap sejumlah tetangga terdekat serta menghitung derajat keanggotaan (*membership value*) pada setiap kelas. Pendekatan tersebut memungkinkan sistem tidak hanya menghasilkan label prediksi, tetapi juga memberikan informasi mengenai tingkat keyakinan hasil klasifikasi.

Tahap terakhir adalah evaluasi performa model. Evaluasi dilakukan menggunakan *Stratified K-Fold Cross Validation* untuk memperoleh hasil yang lebih robust, disertai pengujian terhadap data uji menggunakan metrik evaluasi yang meliputi *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, *confidence*, *probability*, dan *entropy*. Selain itu, dilakukan pula pengujian terhadap berbagai kombinasi fitur serta konfigurasi parameter untuk memperoleh model dengan performa terbaik.

Preprocessing dan Ekstraksi Fitur

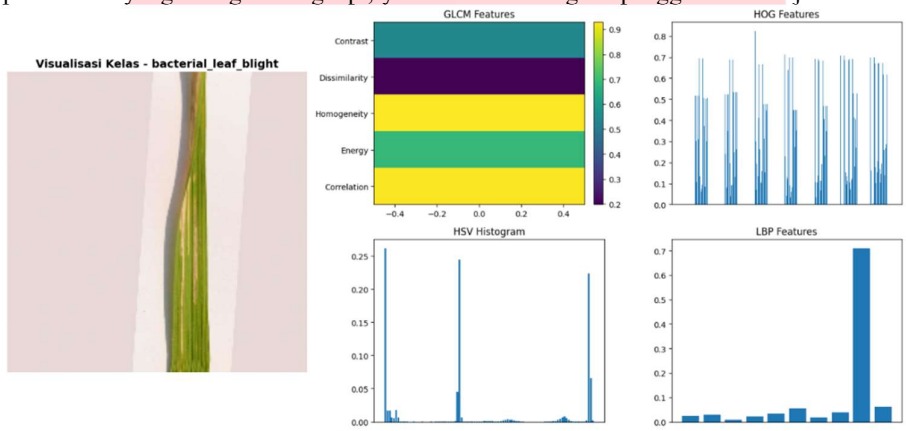
Sebelum dilakukan proses klasifikasi, seluruh citra pada dataset terlebih dahulu melalui beberapa tahapan pengolahan untuk menghasilkan representasi fitur yang mampu

menggambarkan karakteristik visual daun padi secara efektif. Tahapan tersebut meliputi *preprocessing*, ekstraksi fitur, *feature fusion*, *feature scaling*, reduksi dimensi menggunakan PCA dan *L2 normalization*. Rangkaian proses ini bertujuan untuk menghasilkan vektor fitur yang memiliki informasi diskriminatif, bebas dari redundansi, serta sesuai digunakan sebagai masukan bagi algoritma F-kNN.

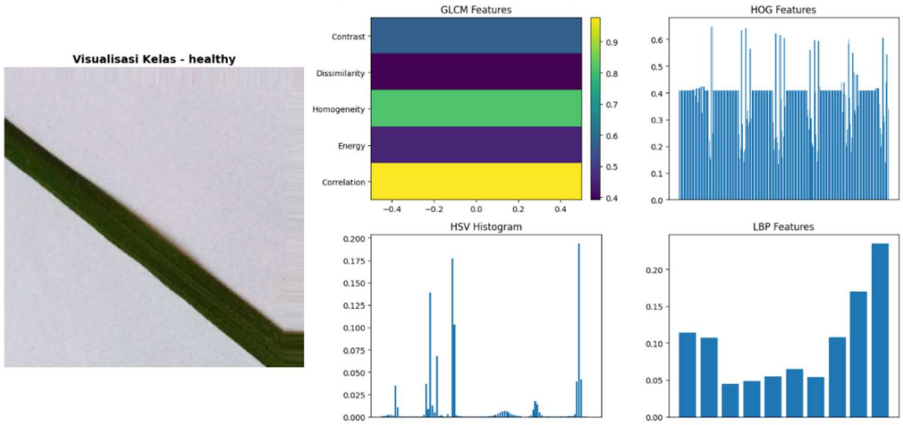
Pada tahap *preprocessing*, seluruh citra terlebih dahulu diseragamkan ukurannya melalui proses *resize* sehingga setiap citra memiliki dimensi yang konsisten. Selanjutnya dilakukan penyesuaian format warna sesuai dengan kebutuhan masing-masing metode ekstraksi fitur. Citra dikonversi ke ruang warna keabuan (*grayscale*) untuk proses ekstraksi menggunakan GLCM, HOG, dan LBP, sedangkan ekstraksi histogram warna dilakukan menggunakan ruang warna HSV. Tahap ini bertujuan untuk memastikan setiap metode memperoleh representasi citra yang sesuai dengan karakteristik fitur yang akan diekstraksi.

Setelah proses *preprocessing*, dilakukan ekstraksi fitur menggunakan empat metode yang saling melengkapi, yaitu GLCM, HOG, histogram warna HSV, dan LBP. GLCM digunakan untuk memperoleh informasi tekstur global berdasarkan hubungan spasial antar piksel, HOG digunakan untuk merepresentasikan pola bentuk dan tepi objek melalui distribusi orientasi gradien, histogram warna HSV digunakan untuk menangkap karakteristik distribusi warna daun, sedangkan LBP digunakan untuk menggambarkan pola tekstur lokal pada permukaan daun. Kombinasi keempat metode tersebut diharapkan mampu merepresentasikan karakteristik penyakit daun padi dari berbagai aspek visual sehingga meningkatkan kemampuan model dalam membedakan setiap kelas penyakit.

Fitur yang dihasilkan dari masing-masing metode kemudian digabungkan melalui proses *feature fusion* dengan menggabungkan seluruh vektor fitur menjadi satu representasi fitur terpadu. Proses ini menghasilkan vektor fitur berdimensi tinggi yang mengandung informasi tekstur, bentuk, warna, dan tekstur lokal secara bersamaan sehingga setiap citra memiliki representasi yang lebih komprehensif dibandingkan penggunaan satu jenis fitur saja.



GAMBAR 14
Visualisasi Ekstraksi Fitur Kelas *bacterial_leaf_blight*



GAMBAR 15
Visualisasi Ekstraksi Fitur Kelas *healthy*

Untuk memberikan gambaran mengenai representasi fitur yang dihasilkan, dilakukan visualisasi hasil ekstraksi pada salah satu citra daun padi dari kelas *Bacterial Leaf Blight* dan *Healthy* sebagaimana ditunjukkan pada **Error! Unknown switch argument.** dan **Error! Unknown switch argument.**. Visualisasi tersebut memperlihatkan bagaimana setiap metode ekstraksi menghasilkan representasi yang berbeda sesuai dengan karakteristik informasi yang diekstraksi.

Pada **Error! Unknown switch argument.** dan **Error! Unknown switch argument.**, fitur GLCM ditampilkan dalam bentuk *heatmap* yang merepresentasikan karakteristik tekstur berdasarkan beberapa ukuran statistik, sehingga pola

tekstur permukaan daun dapat diamati secara kuantitatif. Fitur HOG divisualisasikan sebagai grafik distribusi gradien yang menggambarkan struktur tepi serta bentuk bercak penyakit pada daun. Sementara itu, histogram warna HSV menunjukkan distribusi intensitas warna pada ruang warna *Hue*, *Saturation*, dan *Value*, sehingga perubahan warna akibat infeksi penyakit dapat direpresentasikan dengan lebih baik. Adapun LBP menghasilkan histogram pola tekstur lokal yang menggambarkan distribusi pola mikro pada permukaan daun.

Visualisasi tersebut menunjukkan bahwa setiap metode menangkap karakteristik visual yang berbeda. GLCM dan LBP lebih menekankan informasi tekstur, HOG merepresentasikan bentuk serta orientasi tepi, sedangkan histogram HSV menggambarkan karakteristik warna daun. Oleh karena itu, penelitian ini menggabungkan keempat jenis fitur tersebut melalui proses *feature fusion* sehingga diperoleh representasi citra yang lebih komprehensif sebelum memasuki tahap *feature scaling* dan reduksi dimensi.

Karena setiap metode ekstraksi menghasilkan rentang nilai yang berbeda, dilakukan *feature scaling* menggunakan StandardScaler agar seluruh fitur memiliki skala yang seragam dengan rata-rata nol dan simpangan baku satu. Standardisasi ini bertujuan untuk mencegah fitur dengan rentang nilai yang lebih besar mendominasi proses perhitungan jarak pada algoritma F-kNN.

Selanjutnya dilakukan reduksi dimensi menggunakan PCA dengan mempertahankan 95% total variansi data. Proses ini bertujuan untuk mengurangi jumlah dimensi fitur sekaligus mempertahankan sebagian besar informasi penting yang terkandung pada data. Selain meningkatkan efisiensi komputasi, reduksi dimensi juga membantu mengurangi redundansi antarfitur sehingga representasi data menjadi lebih ringkas.

Tahap terakhir adalah *L2 normalization*, yaitu proses normalisasi panjang setiap vektor fitur sehingga memiliki norma sebesar satu. Normalisasi ini dilakukan untuk menghasilkan representasi vektor yang lebih konsisten dan meningkatkan stabilitas perhitungan jarak antar sampel pada proses klasifikasi menggunakan F-kNN.

Untuk memberikan gambaran mengenai perubahan jumlah dimensi setelah proses reduksi menggunakan PCA, penelitian ini mengevaluasi lima belas skenario kombinasi fitur yang terdiri atas fitur tunggal, kombinasi dua fitur, kombinasi tiga fitur, serta kombinasi keempat fitur. Perubahan dimensi sebelum dan sesudah proses reduksi disajikan pada **Error! Unknown switch argument.**.

TABEL 1
Dimensi fitur sebelum dan sesudah reduksi

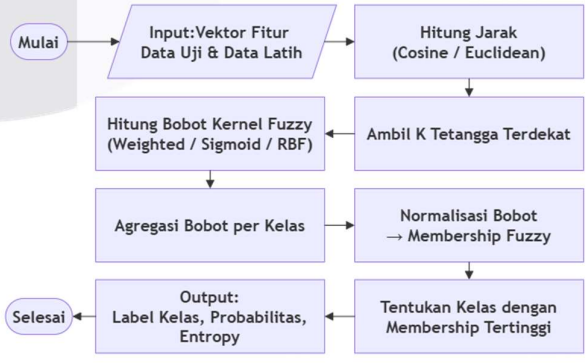
Skenario	Kombinasi Fitur	Dimensi Awal	Dimensi Akhir
S1	GLCM	20	4
S2	HOG	294	142
S3	HSV	24	13
S4	LBP	10	3
S5	GLCM + HOG	314	143
S6	GLCM + HSV	44	16
S7	GLCM + LBP	30	6
S8	HOG + HSV	318	153
S9	HOG + LBP	304	142
S10	HSV + LBP	34	15
S11	GLCM + HOG + HSV	338	153
S12	GLCM + HOG + LBP	324	142
S13	GLCM + HSV + LBP	54	17
S14	HOG + HSV + LBP	328	153
S15	GLCM + HOG + HSV + LBP	348	153

Berdasarkan **Error! Unknown switch argument.**, terlihat bahwa penerapan PCA mampu mengurangi jumlah dimensi fitur secara signifikan pada seluruh skenario tanpa menghilangkan sebagian besar informasi penting yang terkandung dalam data. Sebagai contoh, kombinasi seluruh fitur menghasilkan reduksi dimensi dari 348 menjadi 153 fitur, sehingga proses klasifikasi dapat dilakukan dengan lebih efisien tanpa mengorbankan representasi karakteristik citra secara signifikan. Hasil akhir dari tahapan ini berupa vektor fitur yang telah melalui proses standardisasi, reduksi dimensi, dan normalisasi sehingga siap digunakan sebagai masukan pada proses klasifikasi menggunakan metode F-kNN.

Implementasi Fuzzy K-Nearest Neighbor

Setelah diperoleh representasi fitur hasil proses *preprocessing*, ekstraksi fitur, *feature scaling*, dan reduksi dimensi, tahapan berikutnya adalah proses klasifikasi menggunakan metode F-kNN. Pada penelitian ini, algoritma F-kNN diimplementasikan sebagai metode klasifikasi berbasis *instance-based learning*, di mana proses pembelajaran dilakukan dengan memanfaatkan seluruh data latih sebagai referensi tanpa membangun model parametrik secara eksplisit.

Implementasi metode dimulai dengan memasukkan vektor fitur hasil ekstraksi sebagai data masukan. Untuk setiap data uji, sistem menghitung tingkat kemiripan terhadap seluruh data latih menggunakan metrik jarak yang telah ditentukan. Berdasarkan nilai jarak tersebut, sistem memilih sejumlah *K* tetangga terdekat yang memiliki tingkat kemiripan paling tinggi terhadap data uji. Alur implementasi metode F-kNN pada penelitian ini ditunjukkan pada Error! Unknown switch argument..



GAMBAR 16
Alur implementasi metode Fuzzy K-Nearest Neighbor

Setelah diperoleh *K* tetangga terdekat, sistem menghitung derajat keanggotaan (*membership value*) setiap kelas dengan

mempertimbangkan kontribusi seluruh tetangga berdasarkan kedekatan jaraknya. Berbeda dengan metode kNN konvensional yang menggunakan mekanisme pemungutan suara (*majority voting*), F-kNN memberikan bobot kontribusi pada setiap tetangga sehingga menghasilkan nilai keanggotaan untuk seluruh kelas yang tersedia. Nilai tersebut kemudian dinormalisasi sehingga dapat diinterpretasikan sebagai distribusi probabilitas terhadap masing-masing kelas. Tahap berikutnya adalah proses pengambilan keputusan (*decision making*), yaitu dengan memilih kelas yang memiliki nilai keanggotaan tertinggi sebagai hasil klasifikasi. Selain menghasilkan label prediksi, sistem juga menghasilkan nilai *confidence*, yang menunjukkan tingkat keyakinan model terhadap kelas yang dipilih, serta nilai *entropy* untuk mengukur tingkat ketidakpastian prediksi. Nilai entropy yang rendah menunjukkan bahwa distribusi keanggotaan terkonsentrasi pada satu kelas sehingga model memiliki tingkat keyakinan yang tinggi, sedangkan nilai entropy yang tinggi menunjukkan bahwa distribusi keanggotaan tersebar pada beberapa kelas sehingga keputusan klasifikasi menjadi lebih ambigu.

Pada penelitian ini, implementasi F-kNN dilakukan dengan mengevaluasi beberapa konfigurasi parameter yang meliputi jumlah tetangga (*K*), fungsi kernel, serta metrik jarak. Penentuan konfigurasi terbaik dilakukan melalui proses eksperimen menggunakan *GridSearchCV* yang dikombinasikan dengan *Stratified K-Fold Cross Validation*. Konfigurasi dengan performa terbaik selanjutnya digunakan sebagai model akhir untuk melakukan klasifikasi pada data pengujian.

Dengan implementasi tersebut, metode F-kNN tidak hanya menghasilkan prediksi kelas penyakit daun padi, tetapi juga menyediakan informasi mengenai tingkat keyakinan dan ketidakpastian prediksi. Informasi tambahan ini memberikan dasar yang lebih kuat dalam proses interpretasi hasil klasifikasi, terutama pada citra yang memiliki karakteristik visual yang saling menyerupai.

Skenario Eksperimen dan Evaluasi

Skenario eksperimen pada penelitian ini disusun untuk mengevaluasi pengaruh kombinasi fitur, konfigurasi parameter, serta metode klasifikasi terhadap performa sistem dalam mendeteksi penyakit daun padi. Seluruh eksperimen dilakukan menggunakan data yang telah melalui tahapan preprocessing, ekstraksi fitur, feature scaling, dan reduksi dimensi sebagaimana dijelaskan pada subbab sebelumnya. Pengujian terhadap kombinasi fitur dilakukan menggunakan lima belas skenario yang telah disajikan pada Error! Unknown switch argument., kemudian kombinasi terbaik digunakan pada tahap optimasi parameter dan perbandingan model.

Untuk memperoleh hasil evaluasi yang objektif dan mengurangi bias akibat pembagian data, penelitian ini menggunakan *Stratified K-Fold Cross Validation* dengan konfigurasi sebagaimana ditunjukkan pada **Error! Unknown switch argument.**.

TABEL 2
Konfigurasi Stratified K-Fold Cross Validation

Parameter	Nilai
Jumlah Fold	5
Shuffle	True
Random State	42

Setelah kombinasi fitur terbaik diperoleh, dilakukan proses optimasi parameter menggunakan *GridSearchCV*. Parameter yang dievaluasi pada penelitian ini ditunjukkan pada **Error! Unknown switch argument.**.

TABEL 3
Parameter Pengujian Fuzzy K-Nearest Neighbor

Parameter	Nilai yang Diuji
Jumlah Tetangga (*k*)	3, 5, 7, 10
Kernel	*Weighted*, *Sigmoid*, RBF
Distance Metric	*Euclidean*, *Cosine*
m (Weighted Kernel)	2, 3, 4, 5
alpha (Sigmoid Kernel)	5, 10
gamma (RBF Kernel)	0.1, 1.0

Konfigurasi terbaik yang diperoleh kemudian dibandingkan dengan beberapa algoritma klasifikasi sebagai model pembanding. Daftar model yang digunakan ditunjukkan pada **Error! Unknown switch argument.**.

TABEL 4
Model Pembanding

No	Model
1	Fuzzy K-Nearest Neighbor
2	K-Nearest Neighbor
3	Naïve Bayes
4	Support Vector Machine
5	Multi-Layer Perceptron

Performa model dievaluasi menggunakan metrik klasifikasi serta metrik evaluasi berbasis fuzzy untuk mengukur kemampuan model dalam melakukan klasifikasi dan tingkat keyakinan hasil prediksi. Ringkasan metrik evaluasi yang digunakan disajikan pada **Error! Unknown switch argument.**.

TABEL 5
Metrik Evaluasi

Kelompok	Metrik
Metrik Klasifikasi	*Accuracy, Precision, Recall, F1-Score*
Analisis Klasifikasi	*Confusion Matrix*
Evaluasi Fuzzy	*Confidence, Probability (Membership Value), Entropy*

Berdasarkan skenario tersebut, penelitian ini mengevaluasi pengaruh kombinasi fitur, konfigurasi parameter, serta metode klasifikasi terhadap performa sistem dalam mengidentifikasi penyakit daun padi. Hasil dari setiap tahapan eksperimen disajikan dan dibahas secara rinci pada Bab IV.

IV. HASIL DAN PEMBAHASAN

A. Hasil Pengujian Kombinasi Fitur

Pengujian pertama pada penelitian ini bertujuan untuk mengevaluasi pengaruh kombinasi fitur terhadap kinerja metode F-kNN dalam mengklasifikasikan penyakit daun padi. Sebanyak 15 skenario kombinasi fitur diuji menggunakan empat metode ekstraksi fitur, yaitu GLCM, HOG, histogram warna HSV, dan LBP. Setiap skenario menggunakan konfigurasi parameter F-kNN yang sama sehingga perbedaan performa yang diperoleh hanya

dipengaruhi oleh representasi fitur yang digunakan. Evaluasi dilakukan menggunakan *Stratified 5-Fold Cross Validation*, dan hasil pengujian dinyatakan sebagai nilai rata-rata akurasi beserta standar deviasinya.

TABEL 6
Hasil Pengujian Kombinasi Fitur Menggunakan F-kNN

Skenario	Kombinasi Fitur	Cross-Validation Accuracy (Mean ± Std)
S1	GLCM	0.680224 ± 0.010060
S2	HOG	0.610532 ± 0.008250
S3	HSV	0.913932 ± 0.004413
S4	LBP	0.503362 ± 0.006716
S5	GLCM + HOG	0.665714 ± 0.008604
S6	GLCM + HSV	0.918392 ± 0.007357
S7	GLCM + LBP	0.834987 ± 0.006008
S8	HOG + HSV	0.731147 ± 0.010280
S9	HOG + LBP	0.641417 ± 0.006654
S10	**HSV + LBP**	**0.928377 ± 0.002668**
S11	GLCM + HOG + HSV	0.753246 ± 0.010869
S12	GLCM + HOG + LBP	0.681889 ± 0.008677
S13	GLCM + HSV + LBP	0.927778 ± 0.003882
S14	HOG + HSV + LBP	0.737936 ± 0.009425
S15	GLCM + HOG + HSV + LBP	0.760035 ± 0.010686

Berdasarkan **Error! Unknown switch argument.**, kombinasi fitur HSV + LBP (S10) menghasilkan performa terbaik dengan rata-rata akurasi 0.9284 ± 0.0027, diikuti oleh kombinasi GLCM + HSV + LBP (S13) sebesar 0.9278 ± 0.0039 dan GLCM + HSV (S6) sebesar 0.9184 ± 0.0074. Hasil tersebut menunjukkan bahwa fitur warna memiliki peran dominan dalam membedakan berbagai jenis penyakit daun padi, sedangkan penambahan informasi tekstur lokal melalui LBP mampu meningkatkan kemampuan model dalam mengenali pola visual yang lebih spesifik.

Sebaliknya, penggunaan fitur tunggal selain HSV menghasilkan performa yang relatif rendah. Fitur LBP (S4) memperoleh akurasi terendah sebesar 0.5034, diikuti oleh HOG (S2) sebesar 0.6105 dan GLCM (S1) sebesar 0.6802. Temuan ini mengindikasikan bahwa informasi tekstur maupun bentuk secara individual belum cukup untuk merepresentasikan karakteristik penyakit daun padi secara optimal. Perubahan warna yang muncul akibat infeksi penyakit tampaknya menjadi karakteristik visual yang lebih diskriminatif dibandingkan tekstur atau bentuk daun.

Menariknya, penambahan seluruh jenis fitur tidak selalu menghasilkan peningkatan performa. Kombinasi lengkap GLCM + HOG + HSV + LBP (S15) hanya mencapai akurasi 0.7600, jauh lebih rendah dibandingkan kombinasi HSV + LBP (S10) maupun GLCM + HSV + LBP (S13). Hasil ini menunjukkan bahwa semakin banyak fitur yang digunakan tidak selalu memberikan informasi yang lebih baik. Kehadiran fitur HOG pada beberapa kombinasi diduga menambah informasi yang kurang relevan sehingga meningkatkan redundansi fitur meskipun proses reduksi dimensi menggunakan PCA telah diterapkan.

Secara keseluruhan, hasil pengujian menunjukkan bahwa kombinasi HSV + LBP memberikan representasi fitur yang paling efektif untuk klasifikasi penyakit daun padi. Oleh karena itu, kombinasi fitur tersebut dipilih sebagai skenario terbaik dan digunakan pada tahap optimasi parameter F-kNN yang dibahas pada subbab berikutnya.

Hasil Optimasi Parameter

Proses optimasi parameter dilakukan menggunakan *GridSearchCV* untuk memperoleh konfigurasi terbaik pada model F-kNN berdasarkan nilai rata-rata akurasi *Stratified 5-Fold Cross Validation*. Parameter yang dioptimasi meliputi jumlah tetangga (*k*), jenis kernel, metrik jarak, serta parameter kernel sesuai dengan jenis fungsi keanggotaan yang digunakan. Ringkasan lima konfigurasi terbaik hasil proses optimasi disajikan pada **Error! Unknown switch argument.**.

TABEL 7
Hasil Optimasi Parameter Fuzzy K-Nearest Neighbor (Top-5)

Peringkat	k	Kernel	Distance Metric	Parameter Kernel	CV Accuracy (Mean ± Std)
1	3	*Weighted*	*Cosine*	m = 2	**0.9795 ± 0.0016**
2	5	*Weighted*	*Cosine*	m = 2	0.9778 ± 0.0016
3	3	*Weighted*	*Euclidean*	m = 2	0.9767 ± 0.0011
4	3	*Weighted*	*Cosine*	m = 3	0.9767 ± 0.0011
5	7	*Weighted*	*Cosine*	m = 2	0.9766 ± 0.0017

Berdasarkan **Error! Unknown switch argument.**, konfigurasi terbaik diperoleh pada *k* = 3, *kernel weighted*, *cosine distance metric*, dan parameter *fuzziness* m = 2, dengan rata-rata akurasi *cross-validation* sebesar 97,95% ± 0,16%. Nilai simpangan baku yang relatif kecil menunjukkan bahwa konfigurasi tersebut menghasilkan performa yang konsisten pada setiap lipatan (*fold*) selama proses validasi silang.

Hasil optimasi juga memperlihatkan bahwa *kernel weighted* mendominasi seluruh lima peringkat terbaik. Temuan ini menunjukkan bahwa mekanisme pembobotan berdasarkan jarak memberikan representasi derajat keanggotaan yang lebih efektif dibandingkan pendekatan kernel lainnya. Selain itu, penggunaan *cosine distance* memberikan performa yang sedikit lebih baik dibandingkan *euclidean distance*. Kondisi tersebut mengindikasikan bahwa setelah melalui proses standardisasi, reduksi dimensi menggunakan PCA, dan *L2 normalization*, kesamaan arah antar vektor fitur menjadi lebih representatif daripada besarnya jarak absolut.

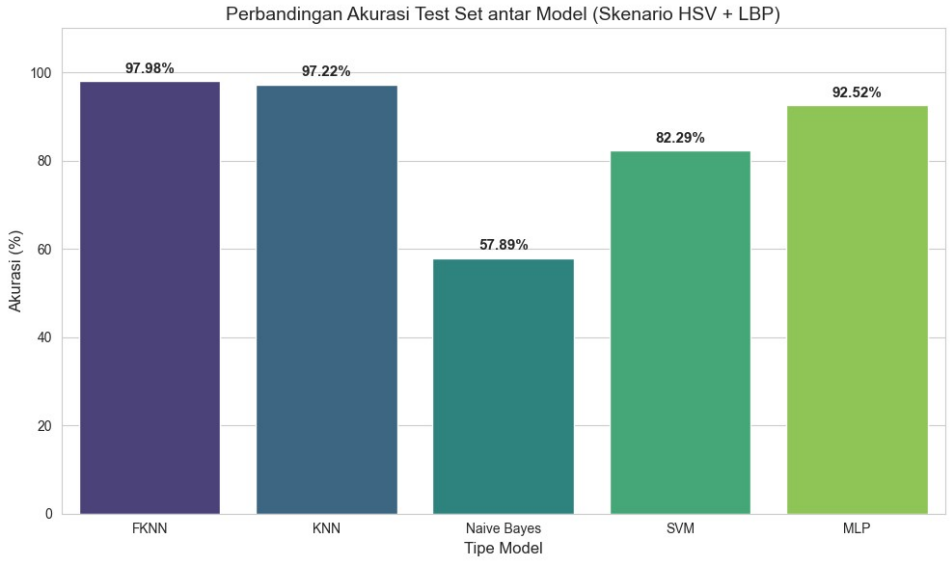
Dari sisi jumlah tetangga, konfigurasi dengan *k* = 3 menghasilkan performa terbaik, sedangkan peningkatan nilai *k* menjadi 5 dan 7 menunjukkan kecenderungan penurunan akurasi meskipun relatif kecil. Hal ini mengindikasikan bahwa penggunaan jumlah tetangga yang lebih sedikit mampu mempertahankan pengaruh sampel-sampel terdekat yang paling relevan, sementara jumlah tetangga yang lebih besar berpotensi memasukkan sampel dari kelas lain sehingga menurunkan ketepatan klasifikasi.

Berdasarkan hasil tersebut, konfigurasi *k* = 3, *kernel weighted*, m = 2, dan *cosine distance* dipilih sebagai konfigurasi akhir model F-kNN dan digunakan pada seluruh pengujian selanjutnya, termasuk perbandingan dengan algoritma klasifikasi lainnya.

Hasil Perbandingan Model

Setelah diperoleh kombinasi fitur dan konfigurasi parameter terbaik, dilakukan perbandingan performa antara model F-kNN dengan beberapa algoritma klasifikasi pembanding, yaitu kNN, *Support Vector Machine* (SVM), *Naïve Bayes* (NB), dan *Multi-Layer Perceptron* (MLP). Seluruh model dievaluasi menggunakan dataset, skema pembagian data, dan

prosedur evaluasi yang sama sehingga hasil yang diperoleh dapat dibandingkan secara objektif.

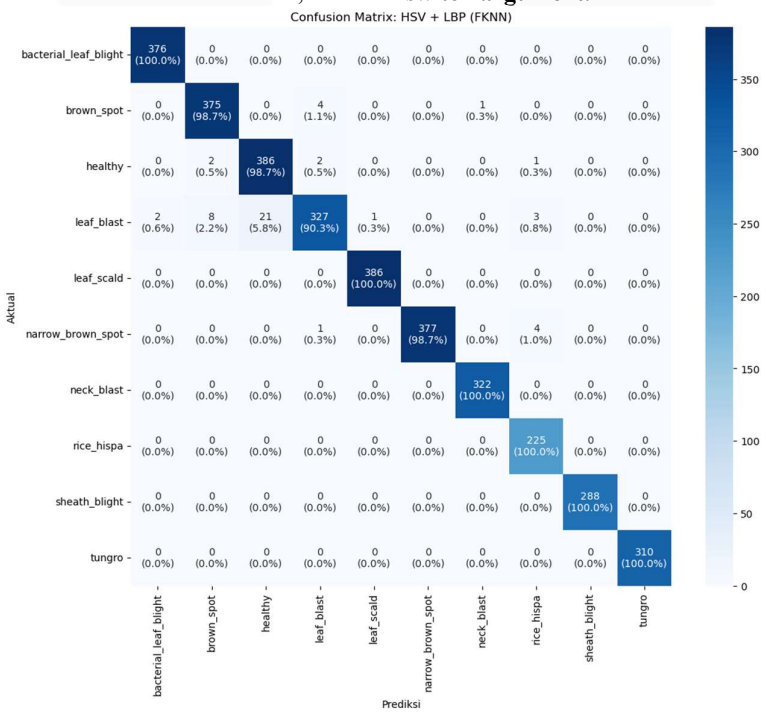


GAMBAR 17
Perbandingan Performa Model Klasifikasi

Berdasarkan **Error! Unknown switch argument.**, model F-kNN menunjukkan performa terbaik dibandingkan seluruh model pembanding. Hasil tersebut menunjukkan bahwa penerapan konsep *fuzzy membership* mampu meningkatkan kemampuan model dalam menangani data yang memiliki batas antarkelas yang tidak tegas (*overlapping classes*). Berbeda dengan kNN konvensional yang menentukan kelas berdasarkan *majority voting*, F-kNN mempertimbangkan derajat keanggotaan setiap tetangga sehingga keputusan klasifikasi menjadi lebih adaptif terhadap karakteristik data.

Model kNN konvensional menempati posisi kedua dengan performa yang relatif mendekati F-kNN. Sementara itu, MLP menghasilkan performa yang cukup baik, namun masih berada di bawah kedua metode berbasis tetangga. Di sisi lain, SVM dan Naïve Bayes memperoleh akurasi yang lebih rendah, yang menunjukkan bahwa kedua metode tersebut kurang mampu merepresentasikan kompleksitas karakteristik fitur citra daun padi dibandingkan pendekatan berbasis kemiripan.

Untuk memberikan gambaran yang lebih rinci mengenai kemampuan model terbaik dalam mengklasifikasikan setiap kelas penyakit, hasil prediksi F-kNN divisualisasikan menggunakan *confusion matrix* pada **Error! Unknown switch argument.**.



GAMBAR 18
Confusion Matrix Model Fuzzy K-Nearest Neighbor

Berdasarkan **Error! Unknown switch argument.**, sebagian besar sampel berhasil diklasifikasikan dengan benar, yang ditunjukkan oleh dominasi nilai pada diagonal utama *confusion matrix*. Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan setiap kelas penyakit daun padi. Beberapa kesalahan klasifikasi masih ditemukan pada kelas-kelas yang memiliki karakteristik visual yang serupa, terutama pada pola warna dan tekstur daun. Meskipun demikian, jumlah kesalahan tersebut relatif kecil dibandingkan jumlah prediksi yang benar, sehingga secara keseluruhan model menunjukkan kemampuan klasifikasi yang tinggi dan konsisten.

Pembahasan

Hasil penelitian menunjukkan bahwa performa sistem klasifikasi penyakit daun padi dipengaruhi oleh tiga aspek utama, yaitu pemilihan kombinasi fitur, konfigurasi parameter model, dan mekanisme klasifikasi yang digunakan. Ketiga aspek tersebut saling berkaitan dalam membentuk representasi data dan menentukan kemampuan model dalam membedakan setiap kelas penyakit.

Pada tahap ekstraksi fitur, kombinasi HSV dan LBP menghasilkan performa terbaik dibandingkan kombinasi fitur lainnya. Temuan ini menunjukkan bahwa karakteristik visual penyakit daun padi lebih banyak direpresentasikan melalui informasi warna dan tekstur lokal dibandingkan informasi bentuk maupun tekstur global. Perubahan warna pada permukaan daun, seperti munculnya bercak berwarna cokelat, kuning, atau keabu-abuan, merupakan salah satu indikator utama yang membedakan berbagai jenis penyakit. Sementara itu, fitur LBP mampu melengkapi informasi tersebut dengan menangkap pola tekstur lokal pada area yang terinfeksi. Sebaliknya, penambahan seluruh fitur secara bersamaan tidak menghasilkan peningkatan performa, yang mengindikasikan adanya redundansi informasi serta masuknya fitur yang kurang relevan terhadap proses klasifikasi.

Optimasi parameter menggunakan *GridSearchCV* menunjukkan bahwa konfigurasi $k = 3$, *kernel weighted*, *cosine distance*, dan $m = 2$ merupakan kombinasi yang paling sesuai untuk data pada penelitian ini. Jumlah tetangga yang relatif kecil memungkinkan proses klasifikasi lebih dipengaruhi oleh sampel-sampel yang memiliki tingkat kemiripan paling tinggi. Selain itu, penggunaan *weighted kernel* memberikan kontribusi yang lebih besar kepada tetangga yang berjarak lebih dekat sehingga proses penentuan derajat keanggotaan menjadi lebih representatif. Penggunaan *cosine distance* juga memberikan hasil yang lebih baik dibandingkan *euclidean distance*, yang menunjukkan bahwa kesamaan arah antar vektor fitur lebih efektif digunakan setelah proses standardisasi, reduksi dimensi menggunakan PCA, dan *L2 normalization*.

Perbandingan dengan beberapa algoritma klasifikasi menunjukkan bahwa F-kNN memberikan performa terbaik pada dataset yang digunakan. Keunggulan tersebut menunjukkan bahwa pendekatan berbasis derajat keanggotaan (*membership degree*) mampu menangani ketidakpastian pada batas antarkelas dengan lebih baik dibandingkan metode klasifikasi konvensional. Pada data citra penyakit daun padi, beberapa kelas memiliki karakteristik visual yang saling menyerupai sehingga pendekatan klasifikasi yang menghasilkan tingkat keyakinan terhadap setiap kelas menjadi lebih sesuai dibandingkan pendekatan yang langsung menetapkan satu kelas secara tegas.

Hasil *confusion matrix* juga menunjukkan bahwa sebagian besar sampel berhasil diklasifikasikan dengan benar, meskipun masih terdapat sejumlah kecil kesalahan pada kelas-kelas yang memiliki karakteristik visual serupa. Kesalahan tersebut menunjukkan bahwa beberapa jenis penyakit memiliki distribusi warna dan tekstur yang saling tumpang tindih sehingga sulit dipisahkan secara sempurna hanya berdasarkan informasi visual. Meskipun demikian, jumlah kesalahan yang relatif kecil menunjukkan bahwa kombinasi fitur yang digunakan telah mampu merepresentasikan karakteristik citra secara efektif.

Secara keseluruhan, penelitian ini menunjukkan bahwa integrasi ekstraksi fitur yang tepat, proses *feature scaling* dan reduksi dimensi, serta penerapan metode F-kNN mampu menghasilkan sistem klasifikasi penyakit daun padi dengan performa yang tinggi dan stabil. Hasil tersebut mengindikasikan bahwa pendekatan berbasis *feature fusion* dan klasifikasi fuzzy merupakan alternatif yang efektif untuk mendukung deteksi penyakit tanaman padi secara otomatis berbasis citra digital.

V. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode F-kNN yang dipadukan dengan kombinasi ekstraksi fitur Histogram Warna HSV dan LBP mampu menghasilkan sistem klasifikasi penyakit daun padi berbasis citra digital dengan performa yang tinggi dan stabil. Kombinasi fitur HSV dan LBP terbukti memberikan representasi yang paling efektif terhadap karakteristik visual penyakit daun padi dibandingkan kombinasi fitur lainnya, sedangkan proses optimasi parameter berhasil memperoleh konfigurasi terbaik pada $k = 3$, kernel weighted, cosine distance, dan $m = 2$, sehingga menghasilkan performa klasifikasi yang optimal. Hasil perbandingan dengan beberapa algoritma klasifikasi menunjukkan bahwa F-kNN memberikan kinerja yang lebih baik dibandingkan metode pembanding, yang menunjukkan bahwa pendekatan berbasis derajat keanggotaan (*fuzzy membership*) mampu menangani ketidakpastian dan kemiripan karakteristik antarkelas secara lebih efektif. Dengan demikian, penelitian ini menunjukkan bahwa integrasi ekstraksi fitur yang tepat, proses *feature scaling* dan reduksi dimensi, serta penerapan metode F-kNN merupakan pendekatan yang efektif untuk mendukung deteksi penyakit daun padi secara otomatis berbasis citra digital. Untuk pengembangan selanjutnya, penelitian ini dapat diperluas dengan memanfaatkan dataset yang lebih beragam, mengeksplorasi metode ekstraksi fitur maupun klasifikasi yang lebih mutakhir, serta mengimplementasikan sistem dalam bentuk aplikasi yang dapat digunakan secara langsung untuk mendukung identifikasi penyakit tanaman padi di lapangan.

REFERENSI

[1] Badan Pusat Statistik Indonesia, "Luas Panen dan Produksi Padi di Indonesia 2021," Mar. 2022. [Online]. Available: https://www.bps.go.id/id/pressrelease/2022/03/01/1909/produksi-padi-tahun-2021-turun-0-43-persen--angka-tetap-.html. [Accessed: May 14, 2026].

[2] W. D. Herawati, *Budidaya Padi*. Jakarta, Indonesia: PT. Buku Kita, 2012.

[3] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York, NY, USA: Springer, 2011.

[4] R. Muslim, Zaeniah, A. Akbar, B. Imran, and Zaenudin, "Disease Detection of Rice and Chili Based on Image Classification Using Convolutional Neural Network Android-Based," *PILAR Nusa Mandiri*, vol. 19, no. 2, pp. 85-96, 2023, doi: 10.33480/pilar.v19i2.4669

[5] A. Jinan and B. H. Hayadi, "Klasifikasi Penyakit Tanaman Padi Menggunakan Metode Convolutional Neural Network Melalui Citra Daun (Multilayer Perceptron)," *Journal of Computer and Engineering Science*, vol. 1, no. 2, pp. 37-44, 2022.

[6] R. A. Saputra, Suharyanto, S. Wasiyanti, D. F. Saefudin, A. Supriyatna, and A. Wibowo, "Rice Leaf Disease Image Classifications Using kNN Based On GLCM Feature Extraction," *Journal of Physics: Conference Series*, vol. 1641, no. 1, p. 012080, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012080.

[7] T. K. Rahayu and J. H. Mendes, "Deteksi Penyakit Tanaman Rambutan Berdasarkan Citra Daun Menggunakan Fuzzy K-Nearest Neighbour," *Mustek Anim Ha*, vol. 10, no. 2, pp. 41-46, 2021, doi: 10.35724/mustek.v10i02.4143

[8] J. Andromeda, N. Hidayat, and R. K. Dewi, "Implementasi Metode FkNN (Fuzzy K-Nearest Neighbor) untuk Diagnosis Penyakit Tanaman Kentang," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 3, no. 12, pp. 10900-10905, 2019.

[9] W. G. Akbari, N. Hidayat, and N. Santoso, "Diagnosis Penyakit Cabai Menggunakan Metode Fuzzy K-Nearest Neighbor (FkNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 1, pp. 1070-1074, 2019.

[10] F. Ramadhani, A. Satria, and I. P. Sari, "Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah," *Hello World: Jurnal Ilmu Komputer*, vol. 2, no. 2, pp. 58-62, 2023, doi: 10.56211/helloworld.v2i2.253

[11] I. Syaputra, R. S. Lubis, and H. Cipta, "Pengklasifikasian Kualitas Buah Jeruk dengan Menerapkan Metode Fuzzy K-Nearest Neighbor," *Journal of Maritime and Education*, vol. 4, no. 1, pp. 368-371, 2022, doi: 10.54196/jme.v4i1.72

[12] J. M. Keller and M. R. Gray, "A fuzzy K-nearest neighbor algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 4, pp. 580-585, 1985, doi: 10.1109/TSMC.1985.6313426.

[13] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, no. 1, Art. no. 22, pp. 1-18, Feb. 2021, doi: 10.1186/s13007-021-00722-9.

[14] M. Ouhami, A. Hafiane, Y. Es-Saady, M. El Hajji, and R. Canals, "Computer Vision, IoT and Data Fusion for Crop Disease Detection Using Machine Learning: A Survey and Ongoing Research," *Remote Sensing*, vol. 13, no. 13, Art. no. 2486, pp. 1-29, Jun. 2021, doi: 10.3390/rs13132486.

[15] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Hoboken, NJ, USA: Pearson, 2018.

[16] A. Upadhyay, N. S. Chandel, K. P. Singh, S. K. Chakraborty, B. M. Nandede, M. Kumar, and A. Elbeltagi, "Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture," *Artificial Intelligence Review*, vol. 58, no. 3, Art. no. 92, pp. 1-64, 2025, doi: 10.1007/s10462-024-11100-x.

[17] L. Li, S. Zhang, and B. Wang, "Plant Disease Detection and Classification by Deep Learning A Review," *IEEE Access*, vol. 9, pp. 56683-56698, 2021, doi: 10.1109/ACCESS.2021.3069646.

[18] J. Lu, L. Tan, and H. Jiang, "Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification," *Agriculture*, vol. 11, no. 8, Art. no. 707, pp. 1-18, Jul. 2021, doi: 10.3390/agriculture11080707.

[19] A. S. Abade, P. A. Ferreira, and F. B. Vidal, "Plant diseases recognition on images using convolutional neural networks: A systematic review," *Computers and Electronics in Agriculture*, vol. 185, Art. no. 106125, pp. 1-16, Jun. 2021, doi: 10.1016/j.compag.2021.106125.

[20] G. Dhingra, V. Kumar, and H. D. Joshi, "Study of digital image processing techniques for leaf disease detection and classification," *Multimedia Tools and Applications*, vol. 77, no. 15, pp. 19951-20000, Aug. 2018, doi: 10.1007/s11042-017-5445-8.

[21] J. Yoon, "Digital Images, Resolution, and Color Models," in *Hands-On Image Processing with Python*. New York, NY, USA: Apress, 2022, pp. 245-258, doi: 10.1007/978-1-4842-8963-1_9.

[22] Y. Satuluri, K. V. R. Prabha, J. S. Polavarapu, H. V. Battina, V. Kota, and P. Ramesh, "Pre-Processing Algorithms for Accurate Analysis of an Image: A Review," in *Proc. 2023 International Conference on Smart Systems for Intelligent Technology (ICSSIT)*, 2023, pp. 1535-1540, doi: 10.1109/ICSSIT55814.2023.10060937.

[23] M. Chen, X. Yu, and S. Su, "Image Texture Analysis Based on Gray Level Co-Occurrence Matrix," in *Proc. 2023 IEEE International Conference on Information Technology and Mechatronics Engineering (ITME)*, Nov. 2023, pp. 451-455, doi: 10.1109/ITME60234.2023.00096.

[24] W. Liu and A. Ralescu, "Advanced HOG research: Multi-scale partitioning HOG algorithm," *IEEE Access*, pp. 1-1, Jan. 2025, doi: 10.1109/access.2025.3609976.

[25] J. Diaz-Escobar and V. Kober, "A robust HOG-based descriptor for pattern recognition," in *Proc. SPIE*, vol. 9971, p. 7, Sep. 2016, doi: 10.1117/12.2237963.

[26] J. Chaki and N. Dey, "Histogram-Based Image Color Features," in *Computer Vision and Image Processing in Intelligent Applications. Singapore: Springer*, 2021, pp. 29-41, doi: 10.1007/978-981-15-5761-3_2.

[27] J. Jos and K. A. Venkatesh, "Pseudo Color Region Features for Plant Disease Detection," in *Proc. 2020 International Conference on Inventive Computation Technologies (INOCON)*, Nov. 2020, doi: 10.1109/INOCON50539.2020.9298337.

[28] T. Ojala, M. Pietikäinen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distribution," *Pattern Recognition*, vol. 19, no. 3, pp. 51-59, 1996, doi: 10.1016/0031-3203(95)00067-4

[29] N. Kaur, N. Nazir, and Manik, "A Review of Local Binary Pattern Based Texture Feature Extraction," in *Proc. 2021 IEEE International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*, Sep. 2021, doi: 10.1109/ICRITO51393.2021.9596485.

[30] G. Gilanie, "Local Binary Pattern (LBP)," Jan. 2025, doi: 10.13140/RG.2.2.33246.55369.

[31] A. K. Bedi, R. K. Sunkaria, and S. K. Randhawa, "Local Binary Pattern Variants: A Review," in *Proc. 2018 International Conference on Smart Computing and Communication (ICSCCC)*, Dec. 2018, doi: 10.1109/ICSCCC.2018.8703326.

[32] A. A. Amer, S. D. Ravana, and R. A. A. Habeeb, "On enhancing data classification using local mean-based fuzzy K-nearest neighbor algorithms," *Advances in Data Analysis and Classification*, Sep. 2025, doi: 10.1007/s11634-025-00653-6.

[33] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 4, pp. 580-585, Jul. 1985, doi: 10.1109/TSMC.1985.6313426.

[34] Y. Itadori, "Rice Leaf Diseases Detection," *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/loki4514/rice-

leaf-diseases-detection. [Accessed: May 14, 2026]