



# Deteksi Penyakit Tanaman Padi (*Oryza Sativa L.*) Menggunakan Support Vector Machine (SVM) Dan Random Forest Pada Citra Daun

Bintang Karmila Gulo\*, Agustinus Rudatyo Himamunanto, Jatmika

Fakultas Fisika dan Komputer, Program Studi Informatika, Universitas Kristen Immanuel, Yogyakarta, Indonesia

Email: <sup>1,\*</sup>bintanggulo2020@gmail.com, <sup>2</sup>rudatyo@gmail.com, <sup>3</sup>jatmika@ukrimuniversity.ac.id

Email Penulis Korespondensi: bintanggulo2020@gmail.com

**Abstrak**—Padi (*Oryza sativa L.*) merupakan tanaman pangan utama yang rentan terhadap serangan penyakit, yang dapat menurunkan produktivitas dan hasil panen petani. Penelitian ini bertujuan mengembangkan sistem klasifikasi penyakit daun padi berbasis citra digital menggunakan algoritma *Support Vector Machine (SVM)* dan *Random Forest*. Dataset terdiri dari tiga kelas penyakit (*Blast*, *Blight*, dan *Tungro*), yang diproses melalui tahap pra-pemrosesan seperti *resize*, normalisasi, dan augmentasi. Ekstraksi fitur dilakukan dengan menggunakan histogram HSV, nilai rata-rata RGB, dan matriks *Gray Level Co-occurrence Matrix (GLCM)* untuk memperoleh ciri warna dan tekstur. Data kemudian dibagi dengan rasio 80:20 untuk pelatihan dan pengujian model. Hasil evaluasi menunjukkan bahwa *Random Forest* memberikan kinerja terbaik dengan akurasi 97,73%, nilai presisi dan *recall* di atas 0,94, serta skor F1 rata-rata sebesar 0,98. Penelitian ini menunjukkan bahwa pendekatan klasifikasi citra berbasis *machine learning* dapat menjadi solusi efektif untuk mendeteksi penyakit pada tanaman padi secara dini.

**Kata Kunci:** Daun Padi, Klasifikasi Citra, SVM, Random Forest, GLCM

**Abstract**—Rice (*Oryza sativa L.*) is a major food crop that is susceptible to disease attacks, which can reduce farmers' productivity and yields. This study aims to develop a digital image-based rice leaf disease classification system using the Support Vector Machine (SVM) and Random Forest algorithms. The dataset consists of three disease classes (Blast, Blight, and Tungro), which are processed through pre-processing stages such as resizing, normalization, and augmentation. Feature extraction is performed using HSV histograms, RGB average values, and Gray Level Co-occurrence Matrix (GLCM) to obtain color and texture characteristics. The data is then divided with a ratio of 80:20 for model training and testing. The evaluation results show that Random Forest provides the best performance with an accuracy of 97.73%, precision and recall values above 0.94, and an average F1 score of 0.98. This study shows that a machine learning-based image classification approach can be an effective solution for early detection of diseases in rice plants.

**Keywords:** Rice Leaves, Image Classification, SVM, Random Forest, GLCM

## 1. PENDAHULUAN

Padi merupakan salah satu komoditas pangan utama di Indonesia yang memiliki peran strategis dalam ketahanan pangan nasional. Namun, produktivitas dan kualitas hasil panen padi seringkali terancam oleh berbagai penyakit tanaman yang dapat menyebabkan kerugian signifikan bagi petani. Dalam konteks ini, identifikasi dan deteksi dini penyakit pada tanaman padi menjadi hal yang sangat krusial untuk menjaga keberlanjutan dan efisiensi produksi pertanian [1]. Keterlambatan dalam mendeteksi dan menangani penyakit dapat berakibat pada penurunan hasil panen yang drastis serta kerugian ekonomi yang besar bagi petani. Oleh karena itu, pengembangan metode deteksi yang cepat, akurat, dan dapat diandalkan menjadi prioritas utama dalam upaya peningkatan produktivitas pertanian padi [2].

Seiring dengan kemajuan teknologi, berbagai metode inovatif telah dikembangkan untuk mengidentifikasi penyakit pada tanaman padi. Salah satu pendekatan yang menjanjikan adalah penggunaan teknologi pengolahan citra digital. Metode ini memungkinkan analisis visual yang lebih presisi dan objektif terhadap gejala penyakit yang muncul pada daun tanaman padi dibandingkan dengan metode konvensional yang mengandalkan pengamatan manual. Penelitian terkait penggunaan citra digital untuk deteksi penyakit tanaman telah menunjukkan hasil yang menjanjikan dalam berbagai bidang pertanian. Sebagai contoh, penelitian mengembangkan sistem pengolahan citra untuk mengidentifikasi kesegaran ikan, menunjukkan potensi teknologi ini dalam penilaian kualitas produk pertanian. Sementara itu, penelitian berhasil menerapkan model *Convolutional Neural Network (CNN)* untuk mendiagnosis penyakit pada daun mangga, membuktikan efektivitas pendekatan *machine learning* dalam deteksi penyakit tanaman [3].

Dalam konteks yang lebih spesifik, penelitian lain telah membahas segmentasi citra penyakit pada tanaman bawang merah menggunakan metode *K-Means* dan *Otsu*, yang menunjukkan bahwa teknik segmentasi citra dapat digunakan untuk mengidentifikasi area yang terinfeksi dengan tingkat akurasi yang tinggi [4]. Meskipun penelitian-penelitian tersebut telah memberikan kontribusi signifikan dalam pengembangan metode deteksi penyakit tanaman berbasis citra digital, aplikasi serupa untuk tanaman padi masih relatif terbatas. Mengingat padi adalah komoditas pangan utama, terdapat kebutuhan mendesak untuk mengembangkan sistem deteksi penyakit yang khusus untuk tanaman ini [5].

Berdasarkan latar belakang ini, tujuan penelitian ini adalah untuk mengembangkan sistem deteksi penyakit tanaman padi berdasarkan pengolahan citra digital yang menggunakan metode *Support Vector Machine (SVM)* dan *Random Forest*. Metode kedua ini dipilih karena kemampuan mereka untuk menangani data yang kompleks dan menghasilkan hasil klasifikasi yang akurat. Pengembangan sistem ini diharapkan dapat memberikan solusi praktis bagi petani untuk mengidentifikasi penyakit pada tanaman padi secara dini dan akurat, sehingga petani dapat melakukan tindakan pencegahan dan pengendalian yang tepat waktu untuk mengurangi kerugian akibat penyakit dan meningkatkan produktivitas tanaman padi secara keseluruhan.

Selain itu, diharapkan penelitian ini akan memberikan kontribusi yang signifikan bagi kemajuan teknologi pertanian presisi di Indonesia. Sistem yang dikembangkan dapat meningkatkan manajemen penyakit tanaman, membantu



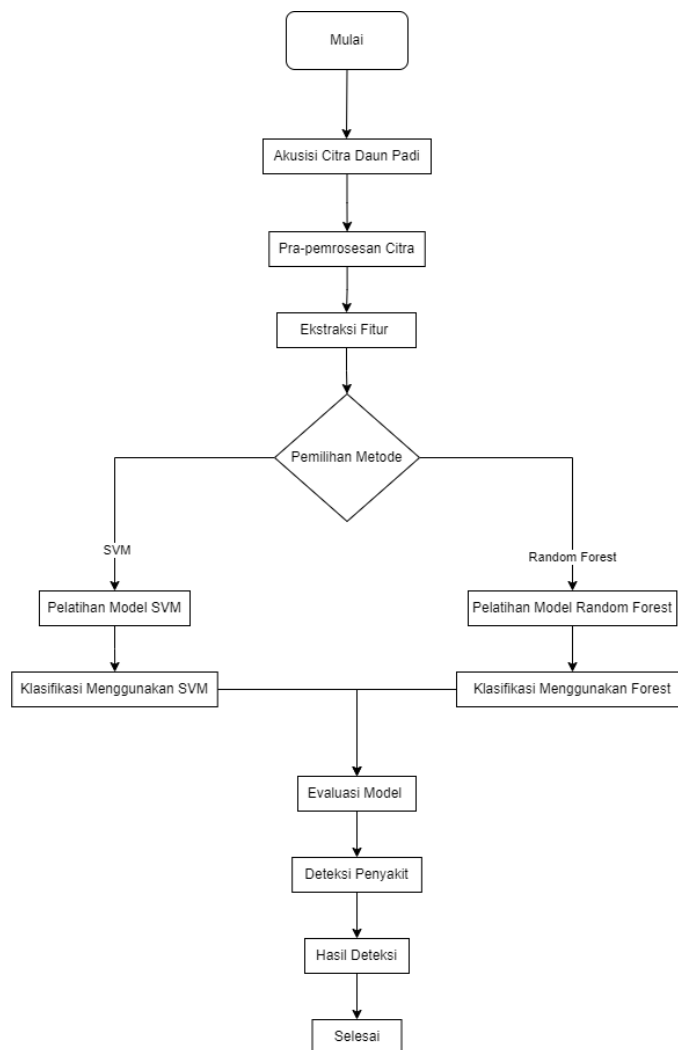
proses pengambilan keputusan pertanian berbasis data, dan meningkatkan ketahanan pangan nasional dengan memanfaatkan teknologi pemrosesan gambar digital dan mesin pembelajaran. Secara lebih luas, penelitian ini sejalan dengan upaya global untuk memasukkan AI dan TI ke dalam sektor pertanian. Di masa depan, hasil penelitian ini diharapkan akan menjadi dasar untuk pengembangan sistem deteksi penyakit tanaman yang lebih canggih. Sistem ini dapat diterapkan pada berbagai jenis tanaman pangan.

## 2. METODOLOGI PENELITIAN

Karena penelitian ini menggunakan pendekatan kuantitatif, fokusnya adalah mengukur dan menganalisis hasil deteksi penyakit pada daun padi secara numerik. Data citra daun padi akan diproses menggunakan algoritma *Support Vector Machine* dan *Random Forest*, dan keluarannya adalah tingkat akurasi klasifikasi penyakit, yang akan dianalisis secara statistik. Gambar berikut menunjukkan proses penelitian.

### 2.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif, dengan fokus pada pengukuran dan analisis hasil deteksi penyakit pada daun padi secara numerik. Data citra daun padi akan diproses menggunakan algoritma *Support Vector Machine* dan *Random Forest*. Keluaran dari pemrosesan ini adalah tingkat akurasi klasifikasi penyakit yang akan dianalisis secara statistik. Gambar berikut menunjukkan proses penelitian. Gambar 1 berikut menunjukkan alur proses penelitian mulai dari akuisisi citra daun padi hingga evaluasi model klasifikasi menggunakan algoritma SVM dan *Random Forest*.



**Gambar 1.** Flowchart Sistem Deteksi Penyakit Daun Padi Menggunakan *SVM* dan *Random Forest*

Gambar 1 memperlihatkan alur proses dalam sistem deteksi penyakit daun padi yang dikembangkan dalam penelitian ini. Proses dimulai dari pengumpulan data, yaitu citra daun padi yang diambil dari dataset publik *Kaggle*. Setelah itu, data citra mengalami *preprocessing* yang mencakup *resize*, normalisasi, dan augmentasi untuk meningkatkan kualitas serta keseimbangan data antar kelas penyakit. Langkah berikutnya adalah ekstraksi fitur, yang bertujuan mengubah citra daun menjadi representasi numerik menggunakan karakteristik warna (HSV dan RGB) serta tekstur



(GLCM). Fitur yang diperoleh disimpan dalam bentuk file .csv untuk digunakan pada tahap pelatihan. Proses selanjutnya adalah pelatihan model klasifikasi, di mana algoritma *Support Vector Machine (SVM)* dan *Random Forest* digunakan untuk mempelajari pola dari data yang telah diekstraksi. Model kemudian diuji menggunakan data uji yang dipisahkan dari data latih dengan rasio 80:20. Langkah terakhir adalah evaluasi model, yang dilakukan menggunakan *Confusion Matrix* dan metrik evaluasi seperti akurasi, presisi, *recall*, dan *f1-score*. Tujuan akhir dari sistem ini adalah mengidentifikasi penyakit daun padi secara otomatis dengan akurasi yang tinggi

## 2.2 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari *Kaggle Repository* pada tautan <https://www.kaggle.com/datasets/rahmi21/rice-datasets> yang terdiri dari 500 citra daun padi dengan tiga kategori penyakit, yaitu *Blast* (200 gambar), *Blight* (220 gambar), dan *Tungro* (80 gambar). Setiap gambar berformat RGB dengan resolusi 256x256 piksel. Distribusi kelas yang tidak seimbang perlu diperhatikan pada tahap pelatihan untuk menghindari bias terhadap kelas mayoritas.

## 2.3 Preprocessing Data

Tahap ini bertujuan menyiapkan citra agar optimal untuk diproses algoritma machine learning, meliputi:

- Resize*: Semua citra diubah ukurannya menjadi  $128 \times 128$  piksel agar dimensi seragam.
- Normalisasi: Nilai RGB setiap piksel dibagi 255 sehingga berada pada rentang 0–1.
- Augmentasi: Untuk meningkatkan jumlah dan variasi data, dilakukan teknik rotasi ( $\pm 15^\circ$ ), *horizontal flipping*, *zoom* ( $\pm 10\%$ ), dan translasi kecil ( $\pm 5\%$ )

## 2.4 Ekstraksi Fitur

Ekstraksi fitur bertujuan mengubah citra menjadi representasi numerik yang merepresentasikan karakteristik visual daun:

- Fitur Warna: *Histogram* komponen HSV (*Hue*, *Saturation*, *Value*) dan rata-rata nilai RGB
- Fitur Tekstur: Diukur menggunakan metode *Co-occurrence Matrix Gray-Level (GLCM)* yang memiliki empat metrik utama: kontras, homogenitas, energi, dan korelasi. Fitur vektor berdimensi tetap dibuat dan disimpan dalam file .csv untuk pelatihan model

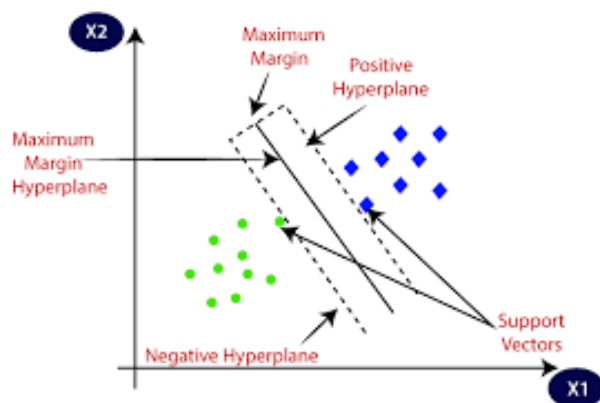
## 2.5 Pelatihan dan Pengujian Model

- Pelatihan Model (*Training*): Data latih yang telah diekstraksi fiturnya kemudian diproses oleh algoritma *SVM* dan *Random Forest*. Model dilatih untuk mengenali pola yang membedakan tiap kategori penyakit.
- Pengujian Model (*Testing*): Pada tahap ini, data uji diproses melalui model yang telah dilatih untuk mengukur kinerja klasifikasi.

## 2.6 Algoritma Klasifikasi

- Support Vector Machine (SVM)

*Support Vector Machines (SVM)* adalah salah satu metode pengajaran paling efektif yang digunakan untuk klasifikasi karena proses penggunaan model klasifikasi SVM melibatkan analisis setiap label atau kelas dan membuat margin sedekat mungkin dengan nol. Oleh karena itu, SVM biasanya digunakan untuk klasifikasi Regresi Vektor Pendukung atau Klasifikasi Vektor Pendukung. *Hyperplane* pada SVM adalah salah satu fitur yang dapat digunakan untuk menilai hubungan antara satu kelas dengan kelas lainnya; metode kerja SVM memungkinkan identifikasi *hyperplane* terbaik dengan identifikasi siswa yang paling mudah diidentifikasi. Untuk klasifikasinya, dua dimensi disebut garis, dan tiga dimensi disebut bidang. Fungsi ini biasanya digunakan untuk mengkategorikan dimensi yang lebih dari tiga.[6] Ilustrasi prinsip kerja algoritma *Support Vector Machine (SVM)* dalam memisahkan dua kelas data dapat dilihat pada Gambar 2. SVM bekerja dengan mencari sebuah *hyperplane* yang tidak hanya mampu membedakan dua kelas data, tetapi juga memaksimalkan jarak antar kelas (*margin*) untuk meningkatkan akurasi klasifikasi.



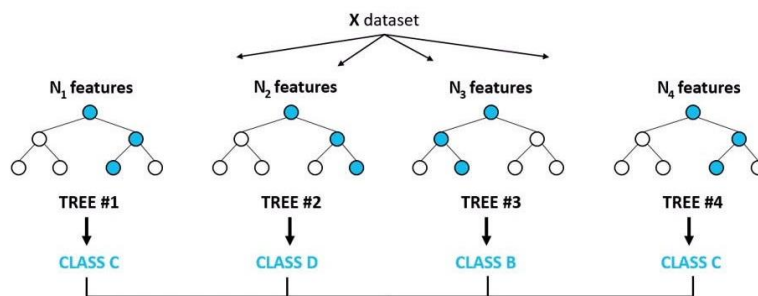
Gambar 2. Ilustrasi Algoritma *Support Vector Machine (SVM)*



Gambar 2 menunjukkan prinsip kerja algoritma *Support Vector Machine* (SVM) dalam memisahkan dua kelas data. Pada ilustrasi tersebut, terdapat dua kelompok data yang direpresentasikan oleh simbol berbeda (misalnya lingkaran biru dan segitiga oranye). SVM membentuk sebuah *hyperplane* sebagai garis pemisah utama yang membagi dua kelas tersebut. Dua garis paralel di kiri dan kanan *hyperplane* disebut sebagai *positive hyperplane* dan *negative hyperplane*, yang bersama-sama membentuk *maximum margin* yaitu jarak terlebar antara dua kelas yang tidak menyentuh titik data manapun kecuali *support vector*. Titik-titik data yang berada tepat di margin disebut *support vector*, karena peran krusialnya dalam menentukan posisi *hyperplane*. Tujuan utama SVM adalah memaksimalkan margin ini agar model memiliki kemampuan generalisasi yang lebih baik saat diuji pada data baru. Semakin lebar margin yang berhasil dibentuk tanpa melanggar batas kelas, maka semakin optimal performa klasifikasi yang dihasilkan.

b. *Random Forest*

Leo Breiman dan Adele Cutler mengembangkan algoritma pembelajaran mesin terbimbing yang dikenal sebagai *Random Forest*. menggunakan keputusan beberapa pohon keputusan untuk menghasilkan prediksi atau klasifikasi. Model berbasis pohon membagi fungsi dataset yang ditentukan menjadi dua kelompok sesuai dengan kriteria tertentu hingga kondisi yang ditentukan memburuk dengan metode untuk menggabungkan keluaran setiap pohon keputusan. Simpul daun atau simpul daun terletak di bawah pohon keputusan. Algoritma *random forest* menghasilkan hasil yang lebih konsisten dan andal. Pekerjaan algoritma ini melibatkan analisis dataset secara tidak bias terlebih dahulu, kemudian membuat pohon keputusan untuk setiap sampel sehingga dapat menghasilkan output dari analisis tersebut. Selanjutnya, algoritma akan membuat suara berdasarkan hasil prediksi untuk masalah klasifikasi dengan mengidentifikasi modus, dan untuk masalah regresi, algoritma akan menggunakan nilai rata-rata. Langkah terakhir adalah menentukan hasil berdasarkan hasil suara, yang dianggap sebagai prediksi output. Grafik berikut adalah contoh klasifikasi data menggunakan algoritma *random forest*. [7]



Gambar 3. Ilustrasi algoritma *Random Forest* untuk klasifikasi citra daun

Gambar 3 menunjukkan bahwa setiap pohon keputusan (*decision tree*) menerima subset data yang berbeda, baik dari sisi fitur maupun jumlah data latih. Masing-masing pohon membuat keputusan klasifikasinya sendiri berdasarkan struktur pohon yang dibentuk dari subset tersebut. Setelah semua pohon selesai melakukan prediksi, sistem akan melakukan *majority voting* untuk menentukan prediksi akhir. Metode ini membuat *Random Forest* sangat andal dalam menghadapi data berdimensi tinggi dan meminimalkan *overfitting*, karena variasi antar pohon memperkuat generalisasi model. Oleh karena itu, *Random Forest* sering menghasilkan performa klasifikasi yang lebih stabil dan akurat dibandingkan model tunggal seperti *decision tree* biasa.

2.7 Evaluasi Model

Untuk menilai kinerja model klasifikasi yang dibuat, evaluasi dilakukan. Dalam penelitian ini, *Confusion Matrix* digunakan untuk evaluasi, yang terdiri dari metrik berikut:

a. Akurasi: Mengukur persentase prediksi yang benar dari total prediksi.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

b. Presisi: Mengukur persentase prediksi positif yang benar.

$$\text{Presisi} = \frac{TP}{TP+FP} \tag{2}$$

c. Recall: Mengukur persentase kasus positif yang berhasil diidentifikasi.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

d. F1-Score: Kombinasi dari presisi dan recall yang memberikan gambaran seimbang terhadap performa model

$$\text{F1 - Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Akurasi, presisi, *recall*, dan F1-score merupakan metrik evaluasi yang digunakan untuk menilai kinerja model klasifikasi. Akurasi mengukur persentase prediksi yang benar dari total prediksi, dihitung dari jumlah *True Positive*

(TP) dan *True Negative* (TN) dibandingkan total keseluruhan kasus (TP+TN+FP+FN). Presisi adalah rasio prediksi positif yang benar (TP) terhadap seluruh prediksi positif (TP+FP), menunjukkan seberapa tepat model dalam mengklasifikasikan kelas positif. Recall mengukur seberapa banyak kasus positif yang berhasil dikenali oleh model, yaitu rasio TP terhadap jumlah kasus sebenarnya positif (TP+FN). Sementara itu, *F1-score* merupakan rata-rata harmonik dari presisi dan recall yang memberikan gambaran seimbang tentang performa model, terutama saat data tidak seimbang, dengan rumus  $F1 = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$ . Keempat metrik ini saling melengkapi untuk memberikan evaluasi menyeluruh terhadap hasil klasifikasi.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset *Kaggle* berikut: <https://www.kaggle.com/datasets/rahmi21/rice-datasets>. Dataset ini terdiri dari lima ratus gambar daun padi yang dikelompokkan ke dalam tiga jenis penyakit: *blast*, *blight*, dan *tungro*. Setiap gambar penyakit menunjukkan pola, warna, dan bentuk gejala yang berbeda pada permukaan daun. Untuk memudahkan pelabelan dan analisis lebih lanjut, setiap kelas penyakit mengungkapkannya dalam folder terpisah. Keseluruhan datanya lima ratus, dengan kelas *blast* 200, kelas *blight* 220, dan kelas *tungro* 80. Gambar 4 dibawah menampilkan contoh citra daun padi dari tiga kelas penyakit yang digunakan dalam penelitian ini, yaitu *Blast*, *Blight*, dan *Tungro*. Setiap kelas menunjukkan karakteristik visual yang berbeda-beda yang menjadi dasar dalam proses klasifikasi oleh algoritma pembelajaran mesin.



**Gambar 4.** Contoh citra daun padi untuk masing-masing kelas penyakit (*Blast*, *Blight*, dan *Tungro*)

Gambar 4 menunjukkan lima contoh citra daun dari masing-masing kelas penyakit.

- Blast* ditandai dengan bercak oval keabu-abuan atau coklat kehitaman pada permukaan daun.
- Blight* menampilkan bercak memanjang kekuningan atau nekrosis di sekitar urat daun.
- Tungro* cenderung menyebabkan daun menguning, mengecil, dan mengerut tidak beraturan.

Variasi warna, pola bercak, dan tekstur dari setiap kelas inilah yang akan diekstraksi menjadi fitur numerik dalam proses *machine learning*. Visualisasi ini membantu menjelaskan bagaimana sistem mengenali perbedaan antar kelas berdasarkan ciri visual yang tampak pada daun.

#### 3.2 Preprocessing Data

Pada tahap preprocessing dilakukan dalam beberapa tahap utama. Pertama, seluruh citra daun padi dari masing-masing kelas penyakit (*Blast*, *Blight*, dan *Tungro*) diubah ukurannya menjadi  $128 \times 128$  piksel. Proses *resize* ini bertujuan untuk menyeragamkan dimensi citra sehingga memudahkan proses ekstraksi fitur dan pelatihan model. Selanjutnya, dilakukan normalisasi nilai piksel dengan membagi nilai RGB setiap piksel dengan 255 agar berada pada rentang 0 hingga 1. Normalisasi ini penting untuk mempercepat proses konvergensi model dan mengurangi risiko vanishing gradient saat pelatihan. Salah satu tantangan utama pada dataset yang digunakan adalah ketidakseimbangan jumlah citra pada masing-masing kelas penyakit, di mana kelas *Tungro* memiliki jumlah citra yang jauh lebih sedikit dibandingkan kelas *Blast* dan *Blight*. Ketidakseimbangan ini dapat menyebabkan model cenderung bias dan kurang akurat dalam mengenali kelas minoritas. Untuk mengatasi hal tersebut, dilakukan augmentasi data secara selektif pada kelas dengan jumlah citra lebih sedikit. Teknik augmentasi yang digunakan meliputi rotasi ( $\pm 15^\circ$ ), *flipping* horizontal, *zoom* ( $\pm 10\%$ ), dan translasi kecil ( $\pm 5\%$ ). Dengan teknik ini, jumlah citra pada setiap kelas dapat disamakan, sehingga distribusi data menjadi lebih seimbang. Hasil akhir tahapan preprocessing menunjukkan bahwa seluruh kelas penyakit kini memiliki jumlah citra yang sama, yaitu 220 gambar per kelas. Gambar 5 dibawah memperlihatkan citra daun padi dari berbagai kelas penyakit sebelum dilakukan proses preprocessing. Tahapan ini penting untuk menyeragamkan kualitas visual citra agar dapat digunakan secara optimal dalam proses pelatihan model klasifikasi.



**Gambar 5.** Contoh citra daun padi sebelum dan sesudah *preprocessing* dan augmentasi (*Blast, Blight, Tungro*)

Gambar 5 menampilkan citra asli daun padi dari tiga kelas penyakit, yaitu *blast*, *blight*, dan *tungro*. Terlihat adanya perbedaan ukuran, orientasi, dan pencahayaan pada citra-citra tersebut. Oleh karena itu, pada tahap *preprocessing* dilakukan beberapa langkah, seperti:

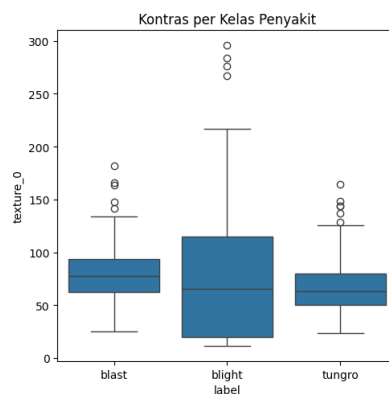
- Resize*: untuk menyamakan dimensi gambar ( $128 \times 128$  piksel).
- Normalisasi: mengubah skala nilai piksel ke rentang 0–1 agar mempercepat proses pelatihan model.
- Augmentasi: seperti rotasi, *flipping*, *zooming*, dan translasi kecil untuk memperbanyak variasi data dan menyeimbangkan jumlah antar kelas.

*Preprocessing* sangat penting untuk memastikan bahwa citra yang digunakan dalam pelatihan model tidak bias terhadap bentuk atau orientasi tertentu, serta menjaga generalisasi model saat mengklasifikasikan data uji.

### 3.3 Ekstraksi Fitur

Tujuan tahapan ekstraksi fitur adalah untuk mengubah gambar daun padi hasil *preprocessing* menjadi representasi numerik yang menunjukkan gejala penyakit secara visual. Dua kategori fitur utama yang digunakan dalam penelitian ini adalah fitur warna dan tekstur. Metode *Co-occurrence Matrix* Kelas Hitam (GLCM) dengan empat arah sudut ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , dan  $135^\circ$ ) dan jarak antar piksel 1 digunakan untuk mengekstraksi.

Fitur tekstur. Nilai RGB (*Red*, *Green*, dan *Blue*) rata-rata menggambarkan tingkat intensitas warna global dari gambar, dan histogram HSV (*Hue*, *Saturation*, *Value*) sebanyak 10 bins per kanal. Enam properti statistik utama diambil dari GLCM: kontras, homogenitas, energi, korelasi, ASM, dan dissimilarity. Fitur-fitur tersebut kemudian dikombinasikan untuk membentuk vektor fitur yang merepresentasikan setiap citra daun secara kuantitatif. Vektor ini menjadi masukan bagi algoritma klasifikasi seperti Support Vector Machine (SVM) dan Random Forest dalam membedakan jenis penyakit berdasarkan pola warna dan tekstur yang unik pada setiap kelas. Pemilihan kombinasi fitur warna dan tekstur bertujuan untuk meningkatkan akurasi klasifikasi, karena gejala penyakit pada daun padi sering kali muncul sebagai perubahan warna dan pola permukaan daun. Dengan demikian, tahap ekstraksi fitur memainkan peran krusial dalam memastikan bahwa informasi visual yang relevan berhasil ditangkap dan dimanfaatkan secara optimal dalam proses pembelajaran mesin. Gambar 6 berikut memperlihatkan hasil visualisasi distribusi nilai fitur tekstur kontras (GLCM) dari citra daun padi berdasarkan masing-masing kelas penyakit. Visualisasi ini dilakukan menggunakan boxplot untuk melihat persebaran dan perbedaan nilai antar kelas.



**Gambar 6.** Distribusi nilai fitur kontras (GLCM) untuk masing-masing kelas penyakit daun padi



Gambar 6 menunjukkan bahwa setiap kelas penyakit daun padi memiliki distribusi nilai kontras yang berbeda.

- Kelas *blight* cenderung memiliki variasi kontras yang lebih lebar dan nilai median lebih tinggi dibandingkan dua kelas lainnya.
- Kelas *blast* dan *tungro* memiliki sebaran yang lebih rapat, dengan beberapa *outlier* di nilai kontras tinggi.

Perbedaan ini menunjukkan bahwa fitur tekstur kontras cukup representatif dalam membedakan karakter visual antar penyakit. Nilai-nilai fitur ini kemudian digunakan sebagai input untuk algoritma klasifikasi seperti SVM dan Random Forest.

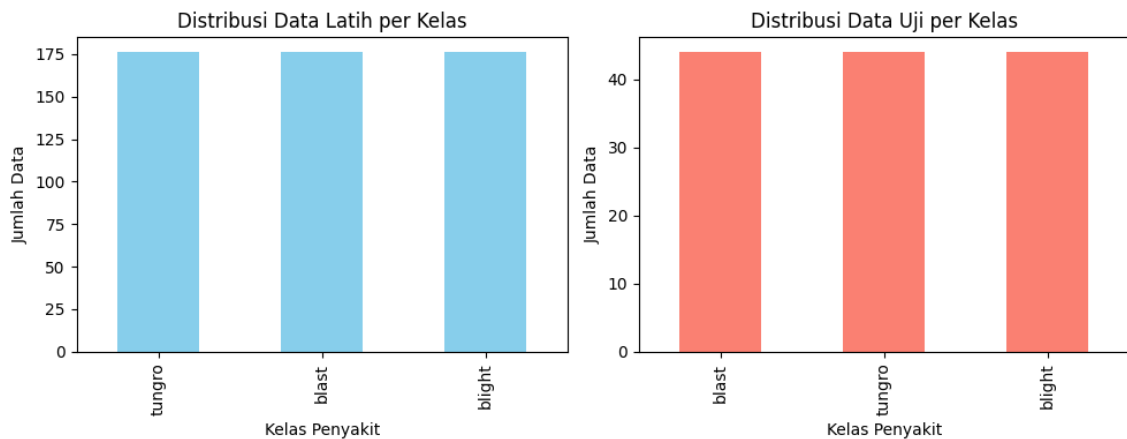
Visualisasi *boxplot* dari fitur tekstur menunjukkan distribusi yang berbeda antar kelas penyakit:

- Kelas *Blight* cenderung memiliki nilai kontras (*texture\_0*) tertinggi, mengindikasikan bahwa permukaan daun lebih tidak rata atau kompleks secara visual.
- Kelas *Tungro* memiliki nilai homogenitas (*texture\_2*) dan energi (*texture\_3*) tertinggi, yang menggambarkan tekstur permukaan daun lebih merata dan polos.
- Fitur korelasi (*texture\_4*) juga menunjukkan perbedaan antar kelas, di mana nilai korelasi *Blight* sedikit lebih tinggi daripada dua kelas lainnya.

Hasil visualisasi fitur memperkuat dugaan bahwa gabungan fitur warna dan tekstur yang digunakan memiliki daya diskriminasi yang baik. Perbedaan distribusi nilai antar kelas menunjukkan bahwa algoritma pengajaran mesin seperti Random Forest dan SVM dapat efektif menggunakan fitur-fitur ini untuk membedakan kategori penyakit. Vektor fitur hasil ekstraksi disimpan dalam format .csv, yang kemudian digunakan pada proses pelatihan dan pengujian model klasifikasi pada tahapan selanjutnya.

### 3.4 Pelatihan SVM dan Random Forest

Dua klasifikasi algoritma *Random Forest* dan *Support Vector Machine (SVM)* digunakan pada tahap model pelatihan. Model kedua dibor dengan dataset yang menghasilkan ekstraksi fitur warna dan tekstur dari gambar daun padi. Untuk menguji kemampuan model generalisasi, dataset tersebut dibagi menjadi 80% data latih dan 20% data uji. Dataset hasil ekstraksi fitur dibagi menjadi dua bagian pada tahap pembagian data (*split data*). Data latih berjumlah 528 data (80 persen) dan data uji berjumlah 132 data (20 persen). Pembagian ini dilakukan secara bertingkat, sehingga proporsi kelas penyakit *Blast*, *Blight*, dan *Tungro* tetap seimbang pada data uji dan latih. Gambar 7 dibawah menunjukkan distribusi data latih dan data uji berdasarkan kelas penyakit setelah dilakukan pembagian dataset dengan rasio 80:20. Tujuannya adalah untuk memastikan bahwa setiap kelas penyakit tetap terdistribusi secara merata dalam kedua subset data.



**Gambar 7.** Distribusi jumlah data latih dan data uji (rasio 80:20) untuk masing-masing kelas penyakit daun padi

Dari Gambar 7, terlihat bahwa jumlah citra daun padi untuk masing-masing kelas penyakit (*blast*, *blight*, dan *tungro*) dibagi secara proporsional ke dalam dua bagian:

- Data latih (training data) berjumlah 80% dari total dataset.
- Data uji (testing data) sebanyak 20% dari total dataset.

Distribusi yang seimbang ini penting agar model pembelajaran mesin dapat melakukan pelatihan secara adil tanpa bias terhadap kelas mayoritas. Proses ini juga memastikan bahwa performa model pada data uji dapat mencerminkan kemampuan generalisasi yang sebenarnya.

Algoritma *Support Vector Machine (SVM)* berbasis margin untuk klasifikasi. Tujuannya adalah menemukan garis (atau *hyperplane*) terbaik yang dapat memisahkan data dari berbagai kelas. Dalam kasus citra, SVM memetakan atribut hasil ekstraksi ke dalam ruang berukuran besar dan kemudian menemukan pemisah antar kelas terbaik. Penelitian ini menangani data yang tidak dapat dipisahkan secara linier dengan kernel RBF (*Function Basis Radial*). Model dibor dengan 80% data hasil ekstraksi fitur, dan diuji dengan 20% sisanya. Hasilnya menunjukkan akurasi 68,18%.

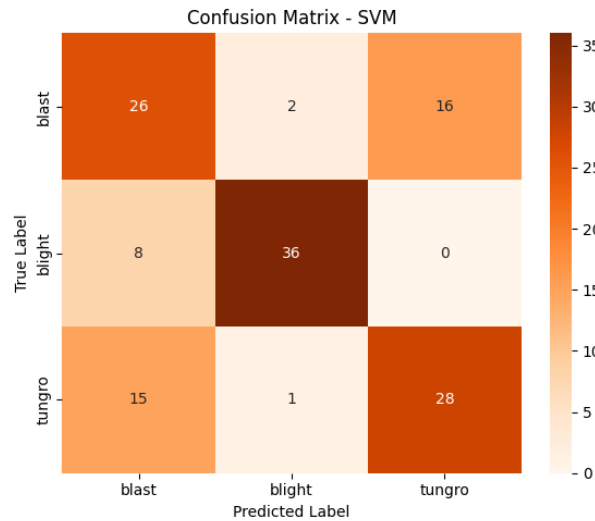
Algoritma pembelajaran kelompok *Random Forest* membangun banyak pohon keputusan (pohon keputusan) dan melakukan voting mayoritas untuk menentukan hasil klasifikasi akhir. Masing-masing pohon dibuat menggunakan subset



data acak. Setiap pohon memiliki fitur yang dipilih secara acak untuk meningkatkan kekeringan dan mengurangi *overfitting* Hasil akurasi mencapai 97,73%.

### 3.5 Evaluasi Model

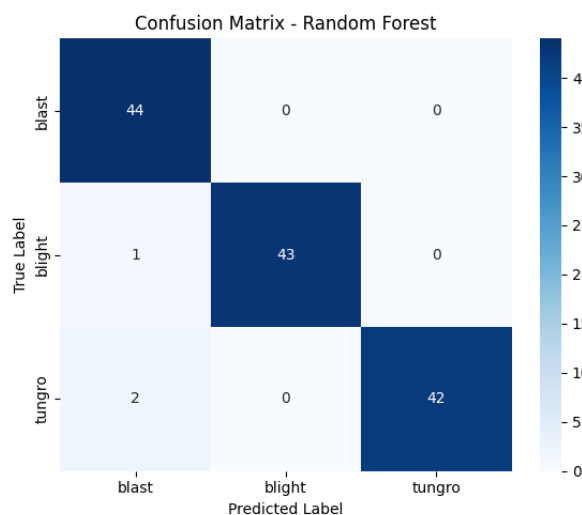
Dilakukan evaluasi model untuk menyalakan klasifikasi penyakit daun padi Blast, Blight, dan Tungro. Dengan menggunakan data uji sebesar 20% dari seluruh dataset hasil ekstraksi fitur, dua algoritma yang diuji adalah Random Forest dan Support Vector Machine (SVM). Untuk mengetahui performa model klasifikasi *Support Vector Machine* (SVM) dalam mengidentifikasi jenis penyakit padi, berikut disajikan gambar 8 *confusion matrix* hasil pengujian model terhadap data uji.



**Gambar 8.** *Confusion matrix* hasil klasifikasi menggunakan algoritma *Support Vector Machine* (SVM)

Gambar 8 menunjukkan *confusion matrix* hasil klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) terhadap tiga kelas penyakit daun padi: *Blast*, *Blight*, dan *Tungro*. Berdasarkan matrix tersebut, kelas *Blight* memiliki tingkat klasifikasi yang paling akurat dengan 36 citra diklasifikasikan secara benar dari total 44 data. Sementara itu, kelas *Blast* dan *Tungro* menunjukkan tingkat kesalahan klasifikasi yang lebih tinggi, dengan sebagian data *Blast* salah diklasifikasikan sebagai *Tungro* (16 kasus), dan sebagian data *Tungro* diklasifikasikan sebagai *Blast* (15 kasus). Hal ini menunjukkan bahwa model SVM cenderung bias terhadap kelas *Blight* dan mengalami kesulitan dalam membedakan karakteristik antara *Blast* dan *Tungro*.

Dari tiga kelas yang diuji, SVM paling baik dalam mengenali kelas *blight* dengan nilai precision sebesar 0,92 dan *f1-score* sebesar 0,87. Namun, performa SVM terhadap kelas *blast* dan *tungro* masih tergolong rendah, dengan *f1-score* masing-masing hanya mencapai 0,56 dan 0,64. Hal ini menunjukkan bahwa SVM kurang optimal dalam membedakan fitur visual antara kedua kelas tersebut, yang kemungkinan besar memiliki karakteristik warna dan tekstur yang mirip. Gambar 9 berikut menyajikan *confusion matrix* dari model *Random Forest* dalam mengklasifikasikan tiga jenis penyakit padi.



**Gambar 9.** *Confusion Matrix* Model *Random Forest* terhadap Tiga Jenis Penyakit Padi



Gambar 9 menampilkan hasil prediksi model *Random Forest* terhadap label penyakit *blast*, *blight*, dan *tungro*. Model menunjukkan performa klasifikasi yang sangat baik, khususnya dalam mengenali semua kelas dengan presisi dan akurasi tinggi. Kelas *blast* dan *tungro* berhasil dikenali secara tepat dengan jumlah prediksi benar masing-masing sebanyak 44 dan 42. Sementara itu, kelas *blight* juga diklasifikasikan secara akurat dengan jumlah prediksi benar sebanyak 43 dari total 45 data. Hanya terdapat sedikit kesalahan klasifikasi, yaitu 2 kasus *tungro* yang diprediksi sebagai *blast*. Hasil ini menunjukkan bahwa model *Random Forest* lebih optimal dibandingkan SVM dalam membedakan fitur visual antar ketiga kelas penyakit padi, bahkan untuk kelas yang cenderung memiliki karakteristik visual yang serupa.

Berbeda dengan SVM, model *Random Forest* menunjukkan performa yang sangat tinggi dengan akurasi mencapai 97,73%. Model ini mampu mengklasifikasikan semua kelas dengan sangat baik, ditunjukkan oleh nilai *precision* dan *recall* yang tinggi dan stabil di atas 0,94 untuk seluruh kelas. *F1-score* untuk masing-masing kelas, yakni *blast*, *blight*, dan *tungro*, berturut-turut adalah 0,97, 0,99, dan 0,98. Keunggulan ini menunjukkan bahwa *Random Forest* sangat efektif dalam menangani data citra berdimensi tinggi dan kompleks, berkat kemampuannya dalam menggabungkan banyak pohon keputusan dan melakukan voting mayoritas, sehingga dapat meminimalkan *overfitting* dan meningkatkan generalisasi. Berdasarkan hasil evaluasi ini, dapat disimpulkan bahwa *Random Forest* merupakan model yang paling andal dan layak digunakan sebagai sistem klasifikasi penyakit daun padi berbasis citra.

Hasil dari penelitian ini menunjukkan bahwa model *Random Forest* memberikan performa klasifikasi yang sangat baik dengan akurasi mencapai 97,73%. Jika dibandingkan dengan penelitian sebelumnya yang menggunakan metode *Naive Bayes* dalam klasifikasi penyakit padi melalui citra daun. Penelitian ini menghasilkan akurasi 76%, yang lebih rendah dari rata-rata model *deep learning* (CNN) yang bisa mencapai >80% pada tugas serupa. Hal ini menunjukkan bahwa kombinasi antara fitur warna (HSV, RGB) dan tekstur (GLCM) seperti yang digunakan dalam penelitian ini mampu meningkatkan akurasi model.

#### 4. KESIMPULAN

Hasil penelitian menunjukkan bahwa metode ekstraksi ciri warna dan tekstur pada gambar daun padi dapat memberikan informasi yang representatif untuk membedakan penyakit *Blast*, *Blight*, dan *Tungro*. *Preprocessing* dan augmentasi telah terbukti penting untuk menyeimbangkan jumlah data antar kelas. Ini menghasilkan dataset yang lebih siap untuk digunakan dalam proses pelatihan model. Dua mesin pembelajaran algoritma, *Random Forest* dan *Support Vector Machine* (SVM), telah diuji untuk melakukan klasifikasi. Hasil evaluasi menunjukkan bahwa model *Random Forest* jauh lebih baik daripada SVM. Dengan akurasi sebesar 97,73% dan nilai *f1-score* yang tinggi dan stabil pada kelas ketiga penyakit, SVM hanya memiliki akurasi sebesar 68,18% dan gagal membedakan antara kelas *blast* dan *tungro*. Oleh karena itu, dapat disimpulkan bahwa algoritma *Random Forest* adalah pilihan yang paling cocok untuk digunakan dalam sistem klasifikasi penyakit daun padi yang berbasis gambar digital.

#### REFERENCES

- [1] S. Agustiani, Y. Tajul Arifin, A. Junaidi, S. Khotimatul Wildah, and A. Mustopa, "Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram," *J. Komputasi*, vol. 10, no. 1, 2022, doi: 10.23960/komputasi.v10i1.2961.
- [2] P. Novantara, R. L. F., and M. Arismawati, "Deteksi Hama Penyakit Daun Padi Dengan Menggunakan Teknik Optimasi Deep Learning Convolutional Neural Network," vol. 7, no. 3, 2025, doi: 10.32877/bt.v7i3.2284.
- [3] T. Ayu, V. Dwi, and A. E. Minarno, "Pendiagnosa Daun Mangga Dengan Model Convolutional Neural Network," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 6, no. 2, p. 230, 2021, doi: 10.24114/cess.v6i2.22857.
- [4] M. Muhibbul, "Segmentasi Citra Penyakit Daun Bawang Merah Menggunakan K-Means Dan Otsu," *JAMI J. Ahli Muda Indones.*, vol. 4, no. 1, pp. 13–17, 2023, doi: 10.46510/jami.v4i1.141.
- [5] Ulfah Nur Oktaviana, Ricky Hendrawan, Alfian Dwi Khoirul Annas, and Galih Wasis Wicaksono, "Klasifikasi Penyakit Padi berdasarkan Citra Daun Menggunakan Model Terlatih Resnet101," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1216–1222, 2021, doi: 10.29207/resti.v5i6.3607.
- [6] K. Saputra and Z. Zuriati, "Identification of Rice Disease Types Based on Digital Images Leaves Using Algorithm Support Vector Machine (SVM)," *International Conference On Agriculture and Applied Science (ICoAAS)*, vol. 2418 no. November, pp. 9–16, 2022, doi: <https://doi.org/10.25181/icoaas.v3i3.2861>.
- [7] S. K. Wildah, A. Latif, A. Mustopa, S. Suharyanto, M. S. Maulana, and A. Sasongko, "Klasifikasi Penyakit Daun Kopi Menggunakan Kombinasi Haralick, Color Histogram dan Random Forest," *J. Sist. dan Teknol. Inf.*, vol. 11, no. 1, p. 35, 2023, doi: 10.26418/justin.v11i1.60985.
- [8] Ahmad Tohirin, Agung akurniawan, Mahfuad Al Hayat, Muhammad Fahmi, & Muhammad Kevin Naufal Fadillah. (2024). Literatur Review : Klasifikasi Penyakit Tanaman Padi Berdasarkan Citra Udara dengan Algoritma SVM. *Buletin Ilmiah Ilmu Komputer Dan Multimedia (BIKMA)*, 2(3), 596–602. Retrieved from <https://jurnalmahasiswa.com/index.php/biikma/article/view/1717>
- [9] M. Faturrachman, P. Studi, T. Informatika, F. Komputer, and T. Dan, "Deteksi Penyakit Pada Daun Singkong Mennggunakan Metode dan SVM ", *e-Proceeding of Engineering*, vol. 11, no. 6, pp. 5805–5813, 2022. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/24728>
- [10] N. Istiqomah and M. Murinto, "Klasifikasi Penyakit Tanaman Padi Berbasis Citra Daun Menggunakan Convolutional Neural Network (CNN)," *JSTIE (Jurnal Sarj. Tek. Inform.)*, vol. 12, no. 1, p. 18, 2024, doi: 10.12928/jstie.v12i1.27314.
- [11] Afis Julianto, Andi Sunyoto, and Ferry Wahyu Wibowo, "Optimasi Hyperparameter Convolutional Neural Network Untuk Klasifikasi Penyakit Tanaman Padi," *Tek. Teknol. Inf. dan Multimed.*, vol. 3, no. 2, pp. 98–105, 2022, doi: 10.46764/teknimedia.v3i2.77.



- [12] B. W. Kurniadi, H. Prasetyo, G. L. Ahmad, B. Aditya Wibisono, and D. Sandya Prasvita, "Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 1–11, 2021. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/1564/1336>
- [13] A. Purnamawati, W. Nugroho, D. Putri, and W. F. Hidayat, "Deteksi Penyakit Daun pada Tanaman Padi Menggunakan Algoritma Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN," *J. Nas. Inform. dan Teknol. Jar.*, vol. 5, no. 1, pp. 212–215, 2020, [Online]. Available: <https://doi.org/10.30743/infotekjar.v5i1.2934>
- [14] S. Keputusan *et al.*, "Terakreditasi SINTA Peringkat 3 Klasifikasi Citra Daun Anggur Menggunakan SVM Kernel Linear," *JOINTECS (Journal of Information Technology and Computer Science)* vol. 7, no. 1, pp. 19–26, 2026. Available: <https://www.researchgate.net/publication/377699006>
- [15] R. Adenia, A. E. Minarno, and Y. Azhar, "Implementasi Convolutional Neural Network Untuk Ekstraksi Fitur Citra Daun Dalam Kasus Deteksi Penyakit Pada Tanaman Mangga Menggunakan Random Forest," *J. Repos.*, vol. 4, no. 4, pp. 473–482, 2024, doi: 10.22219/repositor.v4i4.32287.
- [16] A. Faizin, A. Tri Arsanto, Moch. Lutfi, and A. Rochim Musa, "Deep Pre-Trained Model Menggunakan Arsitektur Densenet Untuk Identifikasi Penyakit Daun Padi," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 615–621, 2022, doi: 10.36040/jati.v6i2.5475.
- [17] S. Khotimatul Wildah and A. Latif, "Deteksi Infeksi pada Daun Kapas menggunakan Kombinasi Metode Ekstraksi Fitur Warna dan Tekstur," *Indones. J. Softw. Eng.*, vol. 9, no. 1, 2023, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijse72>
- [18] P. Rosyani, S. Saprudin, and R. Amalia, "Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 132, 2021, doi: 10.26418/justin.v9i2.44120.
- [19] U. Khultsum and A. Subekti, "Penerapan Algoritma Random Forest dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 186, 2021, doi: 10.30865/mib.v5i1.2624.
- [20] R. Suhendra, and I. Juliwardi, "Identifikasi dan Klasifikasi Penyakit Daun Jagung Menggunakan Support Vector Machine," *J. Teknol. Inf.*, vol. 1, no. 1, pp. 29–35, 2022, [Online]. Available: <http://jurnal.utu.ac.id/JTI>