

## IMPLEMENTASI EKSTRAKSI FITUR TEKSTUR DAN WARNA PADA KLASIFIKASI PENYAKIT DAUN JAGUNG MENGGUNAKAN RANDOM FOREST

Galan Ahmad Defanka<sup>1</sup>, Achmad Junaidi<sup>2</sup> dan Hendra Maulana<sup>3</sup>

<sup>1,2,3</sup> Jurusan Informatika, Fakultas Ilmu Komputer,

Universitas Pembangunan Nasional Veteran Jawa Timur

Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294.

\*Email: [20081010031@student.upnjatim.ac.id](mailto:20081010031@student.upnjatim.ac.id), [achmadjunaidi.if@upnjatim.ac.id](mailto:achmadjunaidi.if@upnjatim.ac.id), [Hendra.maulana.if@upnjatim.ac.id](mailto:Hendra.maulana.if@upnjatim.ac.id)

### Abstrak

Penelitian ini mengembangkan sistem untuk mengidentifikasi penyakit pada daun jagung menggunakan kombinasi ekstraksi fitur tekstur dan warna. Ekstraksi tekstur dilakukan dengan metode Local Binary Pattern (LBP), sedangkan karakteristik warna citra dianalisis menggunakan Fuzzy Color Histogram (FCH). Fitur yang dihasilkan dari kedua metode tersebut kemudian digunakan sebagai input pada algoritma Random Forest untuk proses klasifikasi. Dataset berupa citra daun jagung melalui beberapa tahapan pengolahan, yaitu preprocessing, segmentasi citra, ekstraksi fitur, serta pelatihan model. Hasil pengujian menunjukkan bahwa Random Forest memiliki performa terbaik dengan akurasi mencapai 93,75%. Penelitian ini membuktikan bahwa kombinasi fitur tekstur dan warna dengan algoritma Random Forest efektif dalam mendeteksi penyakit daun jagung.

**Kata kunci:** daun jagung, ekstraksi fitur, random forest

### 1. PENDAHULUAN

Produksi jagung memiliki peran besar dalam memenuhi kebutuhan pangan dan industri, karena jagung tidak hanya dikonsumsi sebagai bahan pangan, tetapi juga digunakan sebagai bahan baku utama pakan ternak, *bioetanol*, dan berbagai produk turunan lainnya. Tingginya permintaan membuat jagung menjadi komoditas strategis yang terus didorong peningkatan produktivitasnya untuk mendukung kemandirian pangan nasional (Eri Sutrisno, 2021).

Produksi jagung Indonesia menunjukkan peningkatan dari tahun ke tahun. Pada tahun 2023, luas tanam mencapai 4,8 juta hektar dengan total produksi sekitar 22 juta ton, dan meningkat menjadi 22,66 juta ton pada tahun 2024 seiring perluasan area tanam menjadi 4,9 juta hektar (Badan Pusat Statistik, 2024). Produktivitas jagung sangat dipengaruhi oleh kesehatan daun, karena daun menjadi organ utama proses *fotosintesis*. Serangan penyakit pada daun dapat menyebabkan berkurangnya area fotosintesis dan berujung pada penurunan hasil panen (Agnes Z. Yonatan, 2025).

Penyakit seperti bercak daun, hawar daun, dan karat daun masih sering muncul dan menyebabkan penurunan hasil panen. Identifikasi penyakit biasanya dilakukan secara manual, sehingga memerlukan waktu dan bergantung pada keahlian pengamat. Dengan berkembangnya *machine learning* dan pengolahan citra, deteksi penyakit dapat dilakukan secara otomatis. Penelitian ini menerapkan ekstraksi fitur tekstur dan warna pada citra daun jagung, kemudian melakukan klasifikasi penyakit menggunakan algoritma *Random Forest*.

### 2. METODOLOGI

#### 2.1. Dataset



Gambar 1. Sampel Foto Data Daun Jagung

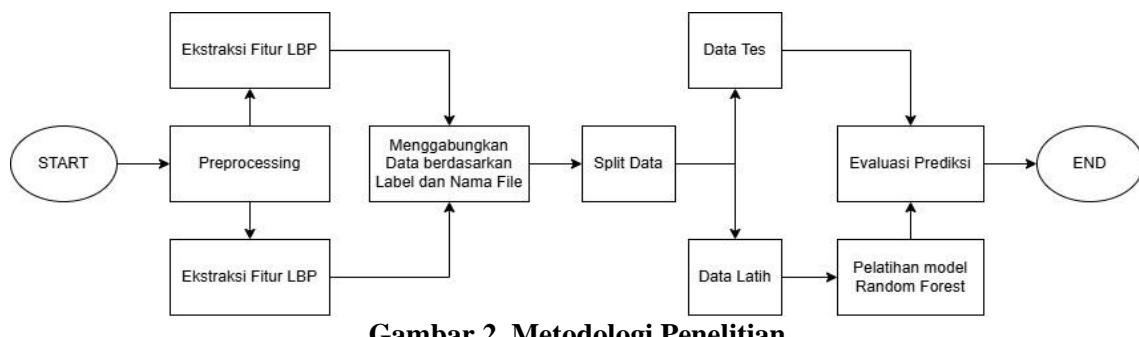
Dataset penelitian ini terdiri dari citra daun jagung yang dikelompokkan ke dalam beberapa kategori penyakit. Tiga kelas utama yang digunakan meliputi bercak daun, hawar daun, dan karat daun, sebagaimana ditunjukkan pada gambar di atas. Rincian jumlah citra pada masing-masing kelas disajikan pada tabel dibawah ini :

**Tabel 1. Dataset Daun Jagung**

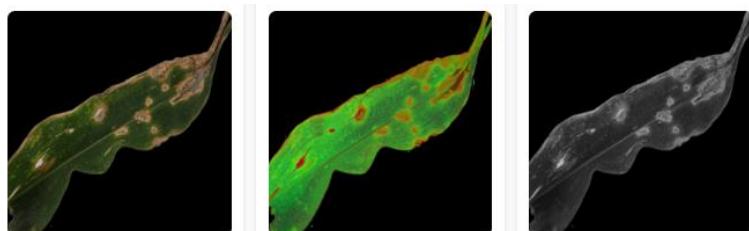
Kelas Daun Jagung	Jumlah
Bercak Daun	300
Hawar Daun	230
Karat Daun	320
Daun Sehat	150
<b>Jumlah</b>	<b>1000</b>

Tabel 1 menunjukkan distribusi jumlah citra pada masing-masing kelas daun jagung. Dataset mencakup empat kategori dengan total 1.000 citra, yang digunakan dalam proses pelatihan dan pengujian model.

## 2.2. Alur Penelitian

**Gambar 2. Metodologi Penelitian**

Tahap pertama adalah persiapan dataset, yang mencakup proses *preprocessing* terhadap citra daun jagung berupa *resize image* 512 x 512 lalu penghapusan *background* daun menggunakan *rembg* agar fokus kepada objek daun saja.

**Gambar 3. Persiapan Dataset Gambar**

## 2.3. Ekstraksi Fitur Tekstur

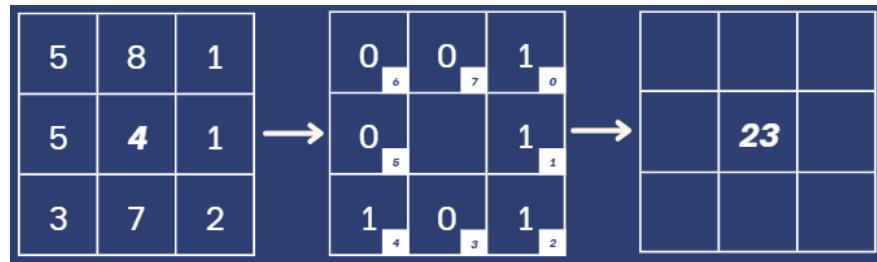
Pada tahapan ekstraksi fitur untuk mengenali tekstur, LBP bekerja dengan memberi label pada setiap piksel berdasarkan pola hubungan intensitas antara piksel pusat dan piksel-piksel tetangganya (Rachmad dkk., 2022). Proses LBP terdiri dari dua tahap, yaitu thresholding dan encoding. Pada tahap thresholding, nilai intensitas setiap piksel tetangga dibandingkan dengan nilai piksel pusat. Jika nilai piksel tetangga lebih besar atau sama dengan nilai piksel pusat, maka diberi nilai 1, dan jika lebih kecil diberi nilai 0. Kombinasi nilai biner hasil perbandingan tersebut kemudian diubah menjadi bilangan desimal pada tahap encoding. Nilai desimal inilah yang menjadi representasi pola tekstur.

$$s(x) = 1 \text{ jika } x \geq 0, \text{ dan } 0 \text{ jika } < 0 \quad (1)$$

Keterangan :

- $s(x)$  : nilai piksel pusat LBP.
- $x$  : nilai piksel tetangga.

Sebelum menerapkan LBP, citra harus dikonversi ke format *grayscale* agar variasi intensitas piksel dapat dianalisis tanpa dipengaruhi warna. Perhitungan LBP dilakukan menggunakan jendela berukuran  $3 \times 3$ , dengan satu piksel pusat dan delapan piksel tetangga. Nilai LBP yang dihasilkan dari seluruh piksel kemudian dihitung histogramnya, dan histogram tersebut menjadi fitur tekstur yang digunakan pada proses klasifikasi.



Gambar 4. Tahapan Ekstraksi Fitur LBP

Setelah nilai piksel pusat pada koordinat satu sudah diketahui, maka hal tersebut akan dilanjutkan ke piksel pusat selanjutnya.

#### 2.4. Ekstraksi Fitur Warna

Pada ekstraksi fitur warna FCH akan mengombinasikan konsep histogram warna dengan logika *fuzzy* untuk menghasilkan representasi warna yang lebih adaptif dan akurat. Pada histogram warna konvensional, setiap piksel hanya dimasukkan ke dalam satu kategori warna secara tegas (misalnya merah atau kuning).

$$\mu(h) = \exp\left(-\frac{(h-c)^2}{2\sigma^2}\right) \quad (2)$$

Keterangan :

- $\mu(h)$  : derajat keanggotaan piksel pada warna.
- $h$  : nilai *hue* piksel.
- $C$  : pusat center dari kategori warna.

Selanjutnya, dataset juga akan diubah ke ruang warna HSV (*Hue, Saturation, Value*). Penggunaan ruang warna HSV memungkinkan pemisahan informasi warna (*Hue*) dari tingkat pencahayaan (*Value*). Implementasi ekstraksi fitur *Fuzzy Color Histogram* (FCH) bertujuan untuk mendapatkan representasi fitur warna secara lebih adaptif. Dengan penerapan konsep logika fuzzy, setiap piksel tidak hanya dikategorikan ke satu warna tertentu, tetapi diberi derajat keanggotaan dan menggunakan *sigma* (jarak) pada nilai *Hue* pusat pada beberapa kategori warna yang bisa dilihat pada tabel dibawah ini :

Tabel 2. Kategori Warna

Warna	Hue Pusat	Sigma	Warna	Hue Pusat	Sigma
Hitam	-	-	Cyan	90	80 – 100
Coklat	0	0 – 10	Biru	120	110 – 130
Orange	20	10 – 30	Ungu	150	140 – 160
Kuning	30	20 – 40	Pink	170	160 – 180
Hijau	60	50 - 70	Merah	180	170 - 180

## 2.5. Penggabungan Hasil Ekstraksi Fitur

Setelah masing-masing ekstraksi fitur selesai, selanjutnya kedua data hasil ekstraksi fitur akan digabungkan berdasarkan *filename* dan *label*. Kemudian dilakukan pengecekan proporsi jumlah data agar menghindari adanya data yang hilang.

## 2.6. Klasifikasi Random Forest

*Random Forest* adalah algoritma *machine learning* berbasis *ensemble learning* yang dikembangkan oleh Leo Breiman pada tahun 2001 (Ashutosh Kumar Singh dkk., 2024). Algoritma ini bekerja dengan membangun banyak pohon keputusan (decision tree) melalui teknik *bootstrap sampling*, yaitu pengambilan sampel secara acak dari data *training* dengan pengembalian. Pada proses pembentukan *Decision Tree* dalam *Random Forest*, langkah pertama adalah menghitung tingkat ketidakpastian atau impurity dari suatu dataset. Ketidakpastian ini dihitung menggunakan rumus *entropy*, yang menunjukkan seberapa heterogen data dalam sebuah *node*. Semakin tinggi nilai *entropy*, semakin beragam kelas data tersebut dan semakin buruk pemisahannya. Sebaliknya, *entropy* bernilai 0 ketika data dalam *node* sudah homogen (semua masuk satu kelas). *Entropy* digunakan sebagai dasar untuk menentukan apakah suatu fitur layak digunakan untuk pemisahan (*splitting*).

$$\text{entropy}(S) = \sum_{i=1}^c p_i \log_2 p_i \quad (3)$$

Keterangan :

- S : dataset yang sedang dihitung *entropy*-nya.
- c : jumlah kelas pada dataset.
- $p_i$  : proporsi data yang termasuk dalam kelas ke-i.

Pada setiap pohon, hanya sebagian fitur yang dipilih secara acak untuk proses pemisahan *node* (*random feature selection*), sehingga setiap pohon bersifat unik dan tidak saling bergantung. Setelah nilai *entropy* awal diketahui, dataset kemudian dibagi berdasarkan salah satu fitur (misalnya warna daun, tekstur, dll.). Setiap hasil pembagian akan memiliki *entropy* masing-masing. Untuk menghitung seberapa baik fitur tersebut membagi dataset, digunakan *expected entropy*, yakni nilai rata-rata *entropy* dari setiap subset dengan memperhitungkan jumlah datanya (Rizky dan Andarsyah, 2023). Hasil ini digunakan untuk menentukan kualitas pemisahan *node*. Semakin rendah *expected entropy*, semakin baik fitur tersebut dijadikan titik pemisah dalam pohon keputusan.

$$\text{entropy}(A, B) = \sum p_{(c)} E_{(c)} \quad (4)$$

Keterangan :

- (A,B) : dataset A setelah dipisah oleh fitur B.
- $E_{(c)}$  : Nilai *entropy* pada subset kelas ke-c.
- $p_{(c)}$  : Probabilitas subset ke-c (jumlah data subset / total data).

Nilai fitur hasil ekstraksi LBP dan FCH yang telah digabungkan sebagai *input* ke model. *Random Forest* kemudian membangun sejumlah *decision tree* melalui teknik *bootstrap sampling* dan pemilihan fitur secara acak pada setiap node. Setiap pohon menghasilkan prediksi kelas, dan hasil akhir ditentukan melalui mekanisme voting, sehingga klasifikasi yang dihasilkan lebih stabil dan tidak mudah mengalami *overfitting*.

**Tabel 3. Testing Parameter**

Parameter	Value
bootstrap	True, False
n_estimators	100, 200
max_depth	10, 20
min_samples_split	2, 5, 10

Setelah semua pohon terbentuk, proses prediksi dilakukan melalui mekanisme *voting*, yaitu setiap pohon memberikan hasil klasifikasi, dan hasil yang paling banyak muncul menjadi hasil akhir.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Evaluasi Program berdasarkan Rasio Data

Selanjutnya dilakukan evaluasi performa model dengan menampilkan hasil akurasi dari proses klasifikasi, sehingga dapat diketahui sejauh mana model mampu mengidentifikasi jenis penyakit pada citra daun jagung.

Tabel 4. Akurasi Berdasarkan Rasio Data

Kelas Daun Jagung	Pembagian Rasio Data		
	60:40	70:30	80:20
Bercak Daun	88%	89%	92%
Hawar Daun	89%	90%	96%
Karat Daun	88%	88%	93%
Daun Sehat	91%	92%	94%
Rata-rata	89%	89,75%	93,75%

Berdasarkan hasil pengujian yang ditampilkan pada Tabel 4, didapatkan rasio 80:20 pembagian data latih dan data tes menghasilkan akurasi yang paling baik. Dapat disimpulkan semakin banyak data latih berbanding lurus dengan tingkat akurasi program.

#### 3.2. Evaluasi Perbandingan Metode Ekstraksi Fitur

Menggunakan rasio data 80:20 tadi kemudian dilakukan evaluasi performa model menggunakan perbedaan dataset, yakni hasil ekstraksi LBP saja, menggunakan hasil ekstraksi FCH saja dan menggunakan gabungan kedua metode ekstraksi yakni LBP dan FCH.

Tabel 5. Akurasi Berdasarkan Metode Ekstraksi Fitur

Ekstraksi Fitur	Akurasi
LBP	80,8%
FCH	90,5%
LBP + FCH	93,75%

Dari Tabel 5 dapat disimpulkan bahwa, penggunaan kedua metode ekstraksi LBP dan FCH dapat menambah fitur yang dibutuhkan untuk klasifikasi dan menghasilkan tingkat akurasi yang lebih baik daripada hanya menggunakan salah satunya saja.

#### 3.3. Evaluasi Perbandingan Metode Klasifikasi

Selanjutnya evaluasi perbandingan metode klasifikasi dilakukan dengan menggunakan dataset hasil ekstraksi fitur LBP dan FCH.

Tabel 6. Akurasi Berdasarkan Metode Klasifikasi

Evaluasi	KNN	SVM	Random Forest
Estimasi Waktu	18 detik	1 menit 15 detik	2 menit 5 detik
Akurasi	74,65%	68,25%	93,75%

Terlihat bahwa metode *Random Forest* juga menunjukkan hasil yang sangat baik daripada metode klasifikasi lain seperti KNN dan SVM (Ansori *et al.*, 2024). Akurasi tertinggi diperoleh pada metode *Random Forest* sebesar 93,75%. Hasil ini menunjukkan bahwa metode *Random Forest* menunjukkan akurasi yang lebih baik daripada KNN dan SVM walaupun estimasi *training* model sedikit lebih lama daripada KNN dan SVM.

#### 4. KESIMPULAN DAN SARAN

Penelitian ini berhasil menghasilkan sistem klasifikasi penyakit daun jagung berbasis pengolahan citra dan *machine learning*. Berdasarkan rangkaian pengujian yang dilakukan, dapat disimpulkan bahwa :

1. Metode LBP dan FCH mampu mengekstraksi karakteristik tekstur serta warna daun jagung secara akurat, sehingga keduanya efektif digunakan sebagai indikator pendekripsi penyakit.
2. *Random Forest* memberikan hasil klasifikasi paling optimal dalam memanfaatkan kedua jenis fitur tersebut, dengan akurasi terbaik sebesar 93,75% pada pembagian data 80:20.
3. Pendekatan yang menggabungkan fitur tekstur dan warna dengan algoritma *Random Forest* terbukti meningkatkan kemampuan identifikasi penyakit daun jagung, sekaligus menjadi alternatif lebih cepat dan objektif dibandingkan pemeriksaan manual.

Adapun beberapa saran agar penelitian yang dilakukan dapat dikembangkan jadi lebih baik lagi, berikut beberapa saran yang bisa dipertimbangkan :

1. Menambah jumlah dataset mungkin akan membuat data pelatihan jauh lebih bervariatif dan lebih dapat konsisten dalam menangani berbagai macam foto daun jagung.
2. Mengkaji metode ekstraksi fitur lain seperti GLCM atau HOG untuk melihat apakah kombinasi fitur dapat memberikan hasil akurasi yang lebih baik.
3. Mencoba algoritma klasifikasi lain seperti *Gradient Boosting* atau *XGBoost* untuk dibandingkan performanya terhadap *Random Forest*.

#### DAFTAR PUSTAKA

- Agnes Z. Yonatan (2025) *Produksi Jagung Indonesia Naik 3% pada 2024*. Available at: <https://data.goodstats.id/statistic/produksi-jagung-indonesia-naik-3-pada-2024-2QuuO> (Accessed: 31 March 2025).
- Ansori, N. et al. (2024) ‘CORN STALK DISEASE CLASSIFICATION USING RANDOM FOREST COMBINATION OF EXTRACTION FEATURES’, *Communications in Mathematical Biology and Neuroscience*, 2024. Available at: <https://doi.org/10.28919/cmbn/8404>.
- Ashutosh Kumar Singh et al. (2024) ‘Retraction: Hybrid Feature-Based Disease Detection in Plant Leaf Using Convolutional Neural Network, Bayesian Optimized SVM, and Random Forest Classifier (*Journal of Food Quality* (2022) 2022 (2845320) DOI: 10.1155/2022/2845320)’, *Journal of Food Quality*. Hindawi Limited. Available at: <https://doi.org/10.1155/2024/9890349>.
- Badan Pusat Statistik (2024) *Luas Panen, Produksi, dan Produktivitas Jagung Menurut Provinsi - Tabel Statistik - Badan Pusat Statistik Indonesia*. Available at: <https://www.bps.go.id/id/statistics-table/2/MjIwNCMy/luas-panen-produksi-dan-produktivitas-jagung-menurut-provinsi.html> (Accessed: 5 November 2024).
- Eri Sutrisno (2021) *Indonesia.go.id - Agar Jagung tak lagi Impor*. Available at: <https://indonesia.go.id/kategori/editorial/3494/agar-jagung-tak-lagi-impor#:~:text=Perlu%20diketahui%20awal%202021%20pemerintah%20melalui%20Kementerian%20Pertanian,pasokannya%20aman%20atau%20bahkan%20mencapai%20suarplus%20untuk%20ekspor>. (Accessed: 5 November 2024).
- Rachmad, A. et al. (2022) ‘Corn Leaf Disease Classification Using Local Binary Patterns (LBP) Feature Extraction’, in *Journal of Physics: Conference Series*. Institute of Physics. Available at: <https://doi.org/10.1088/1742-6596/2406/1/012020>.
- Rizky, M. and Andarsyah, R. (2023) ‘Klasifikasi MIT-BIH Arrhythmia Database Metode Random Forest dan CNN dengan Model ResNet-50: A Systematic Literature Review’, *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(3), pp. 190–196. Available at: <https://doi.org/10.47233/jteksis.v5i3.825>.