

KLASIFIKASI TINGKAT KEMATANGAN LADA MENGGUNAKAN ENSEMBLE LEARNING BERDASARKAN CITRA WARNA KULIT

**Jihan Izzathul Mujidah*¹⁾, Rizki Yusliana Bakti²⁾, Lukman³⁾, Muhammad Faisal⁴⁾,
Muhammad Syafaat⁵⁾, Muhyiddin AM Hayat⁶⁾, Andi Makbul Syamsuri⁷⁾**

1. Informatika, Universitas Muhammadiyah Makassar
Jihanizzathul1234@gmail.com
2. Informatika, Universitas Muhammadiyah Makassar
rizkiyusliana@unismuh.ac.id
3. Informatika, Universitas Muhammadiyah Makassar
lukman@unismuh.ac.id
4. Informatika, Universitas Muhammadiyah Makassar
muhfaisal@unismuh.ac.id
5. Teknik Pengairan, Universitas Muhammadiyah Makassar
syafaat_skuba@unismuh.ac.id
6. Informatika, Universitas Muhammadiyah Makassar
muhyiddin@unismuh.ac.id
7. Teknik Pengairan, Universitas Muhammadiyah Makassar
amakbulsyamsuri@unismuh.ac.id

Abstract

*Pepper fruit (*Piper nigrum L.*) is an agricultural commodity whose market value strongly depends on its ripeness level at harvest. Ripeness determination, which is still commonly performed through visual observation, tends to be inaccurate and subjective. This study aims to classify the ripeness level of pepper fruit based on skin color using an ensemble learning approach. The dataset consists of 1,996 pepper fruit images categorized into four ripeness levels unripe, semi ripe, ripe, and overripe. Color features were extracted from the HSV color model using color moment statistics including mean, standard deviation, and skewness. Random Forest and XGBoost models were combined using a soft voting method. The results show that the ensemble model achieved 98.25% accuracy, 98.30% precision, 98.27% recall, and 98.26% F1-score. The ensemble approach proved superior to single models by providing more accurate and stable classification of pepper fruit ripeness.*

Kata Kunci : *Klasifikasi Kematangan Buah Lada, HSV Color Moment, Random Forest, XGBoost, Ensemble Soft Voting*

A. PENDAHULUAN

Lada (*Piper nigrum L.*) merupakan salah satu komoditas rempah utama Indonesia yang memiliki nilai strategis di sektor pertanian dan perdagangan. Mutu

hasil panen lada sangat dipengaruhi oleh tingkat kematangan buah, yang berkaitan langsung dengan karakteristik warna, aroma, serta daya simpannya. Pemanenan pada tahap kematangan yang tidak tepat

dapat menurunkan kualitas produk dan nilai jual [1]. Penentuan tingkat kematangan buah lada secara konvensional masih dilakukan secara visual, yang sangat bergantung pada subjektivitas pengamat serta pencahayaan di lapangan [2].

Perubahan warna kulit buah selama proses pematangan menjadikan warna sebagai indikator visual yang penting dalam proses klasifikasi. Warna hijau mengindikasikan buah belum matang, sedangkan merah menunjukkan kematangan penuh. Tingkat kematangan buah lada dibagi menjadi empat kategori utama, yaitu belum matang, setengah matang, matang, dan lewat matang [3]. Dalam klasifikasi berbasis visual, pendekatan pengolahan citra digital dapat memberikan solusi yang lebih objektif dibandingkan pengamatan manual. klasifikasi merujuk pada proses pengelompokan objek ke dalam kategori tertentu berdasarkan ciri-ciri yang dimilikinya.

Model warna HSV (*Hue, Saturation, Value*) dipilih karena kemampuannya dalam merepresentasikan warna sesuai persepsi manusia dan kestabilannya terhadap perubahan pencahayaan [4]. Untuk mengekstraksi informasi warna ini secara statistik, digunakan *metode Color Moment* yang menghitung nilai rata-rata (*mean*), standard deviation, dan *skewness* dari setiap kanal warna HSV [5]. Pendekatan ini menghasilkan representasi fitur warna yang kompak namun informatif, sehingga dapat digunakan sebagai *input* yang efektif dalam proses klasifikasi [4].

Penelitian sebelumnya telah menunjukkan efektivitas berbagai algoritma machine learning, seperti *Random Forest* dan *XGBoost*, dalam melakukan klasifikasi berbasis fitur warna [6]. Masing-masing algoritma memiliki keunggulan tersendiri, penggunaan satu model saja sering kali belum mampu menangani kompleksitas data visual secara

optimal [7]. Untuk mengatasi hal tersebut, dikembangkan pendekatan *ensemble learning*, yaitu metode yang menggabungkan beberapa model prediksi untuk menghasilkan output yang lebih akurat dan stabil [8]. Salah satu bentuk teknik *ensemble* ini adalah *soft voting*, yang menggabungkan probabilitas prediksi dari beberapa model, lalu menetapkan kelas berdasarkan nilai probabilitas tertinggi [9].

Berbagai studi sebelumnya mendukung keunggulan pendekatan *ensemble* dalam klasifikasi citra. Hidayat mencatat akurasi 98,5% dalam pendeteksian kualitas buah menggunakan kombinasi *ensemble learning* dengan augmentasi [10]. Argadianata memperoleh akurasi 88,5% untuk klasifikasi kualitas buah apel menggunakan *Random Forest* [11]. Penelitian Elzandy dengan pendekatan *VGG16-XGBoost* menghasilkan akurasi 89,58% dalam klasifikasi varietas anggur [12]. Penelitian Mardianto juga membuktikan bahwa *Random Forest* memberikan akurasi tertinggi, yaitu 96%, dibanding CNN dan SVM dalam klasifikasi jenis mangga [13].

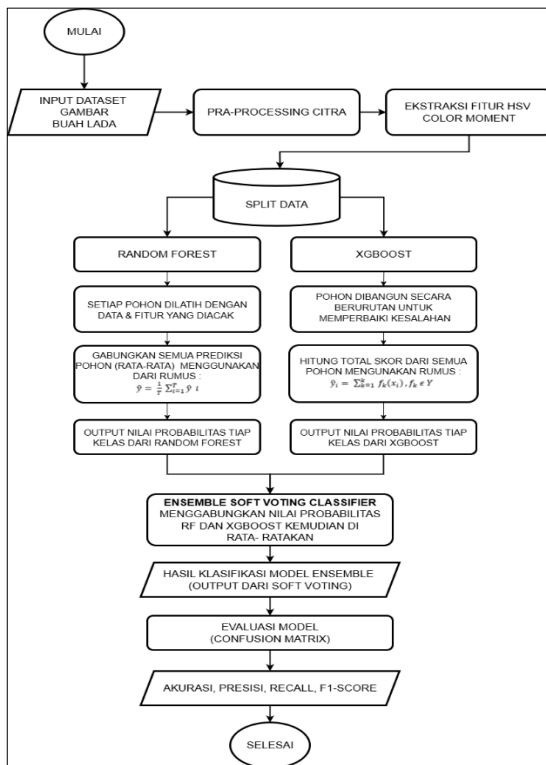
Penerapan metode *ensemble learning* yang menggabungkan *Random Forest* dan *XGBoost* untuk klasifikasi tingkat kematangan buah lada berbasis fitur warna HSV dan *Color Moment* masih belum banyak dibahas dalam literatur. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi tingkat kematangan lada menggunakan fitur warna dari ruang HSV yang diekstraksi melalui *Color Moment*, serta mengukur kinerja pendekatan *ensemble soft voting* dalam menghasilkan klasifikasi yang akurat. Penelitian ini diharapkan dapat memberikan kontribusi pada pemanfaatan *machine learning* dalam klasifikasi komoditas pertanian berbasis citra warna secara efektif dan terukur.

B. METODE PENELITIAN

Penelitian ini dilakukan menggunakan *Python* di *Visual Studio Code* untuk mengklasifikasikan tingkat kematangan buah lada berdasarkan fitur warna dari citra digital. Objek penelitian terdiri dari buah lada yang dikategorikan ke dalam empat kelas kematangan, yaitu belum matang, setengah matang, matang, dan lewat matang.

Penelitian ini menggunakan metode utama yaitu *ensemble soft voting*, yang menggabungkan keunggulan dua algoritma, yakni *Random Forest* dan *XGBoost* dengan cara menghitung rata-rata dari nilai probabilitas prediksi masing-masing model [14].

Berikut merupakan gambaran alur penelitian yang divisualisasikan pada Gambar 1 yang menyajikan tahapan proses secara terstruktur, mulai dari akuisisi citra, pra-pemrosesan, ekstraksi fitur, hingga proses klasifikasi dan evaluasi model menggunakan pendekatan *ensemble*.



Gambar 1. Flowchart Penelitian

Alur penelitian ini terdiri dari tahapan-tahapan sebagai berikut:

- a. **Input Dataset Citra**
Dataset citra buah lada diperoleh dari dokumentasi langsung di perkebunan.
- b. **Pra-pemrosesan Citra**
Citra diubah ukurannya dan dikonversi dari RGB ke HSV untuk keperluan ekstraksi fitur.
- c. **Ekstraksi Fitur**
Nilai *mean*, *standard deviation*, dan *skewness* dihitung dari masing-masing kanal HSV menggunakan metode Color Moment.
- d. **Pemisahan Data**
Dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20.
- e. **Pelatihan Model**
Model *Random Forest* dan *XGBoost* dilatih menggunakan data latih yang telah diekstraksi fiturnya.
- f. **Ensemble Soft Voting**
Prediksi akhir diperoleh dengan menggabungkan probabilitas dari kedua model menggunakan metode *soft voting*.
- g. **Evaluasi Model**
Model dievaluasi menggunakan *confusion matrix* dan metrik klasifikasi seperti akurasi, presisi, recall, dan F1-score.
- h. **Selesai**
Proses penelitian berakhir setelah memperoleh hasil evaluasi model dan interpretasi performa klasifikasi.

Penjelasan berikut mencakup tahapan inti dari proses klasifikasi:

1. Ekstraksi Warna HSV

Ekstraksi fitur HSV adalah model warna yang dirancang untuk mempermudah pemrosesan citra digital, terutama dalam aplikasi pengenalan warna [15]. Berdasarkan penjelasan mengenai

HSV di atas, nilai HSV dapat dihitung dari RGB dengan cara sederhana yang dirumuskan sebagai berikut [16]:

$$V = \max(R, G, B) \quad (1)$$

$$S = \begin{cases} 0 & \text{jika } v = 0 \\ \frac{v - \min(R, G, B)}{v} & \text{jika } v > 0 \end{cases} \quad (2)$$

$$H = \begin{cases} 0 & \text{jika } \Delta = 0 \\ 60^\circ \times \frac{G-B}{\Delta} & \text{jika } V = R \\ 60^\circ \times \left(2 + \frac{B-R}{\Delta}\right) & \text{jika } V = G \\ 60^\circ \times \left(4 + \frac{R-G}{\Delta}\right) & \text{jika } V = B \end{cases} \quad (3)$$

2. Color Moment

Color Moment merupakan metode ekstraksi ciri warna yang digunakan untuk mengukur kemiripan warna suatu citra [16]. Beberapa nilai dapat dihitung dengan [5]:

a. Mean

Mean adalah nilai rata-rata dari semua intensitas warna dalam suatu kanal gambar yang merepresentasikan warna dominan. Untuk itu, digunakan rumus berikut:

$$\mu = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N x_{ij} \quad (4)$$

b. Standar Deviasi

Standar deviasi adalah ukuran beragam warna didalam gambar. Rumus deviasi dapat dituliskan sebagai berikut:

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (x_{ij} - \mu)^2} \quad (5)$$

c. Skewness

Skewness adalah ukuran warna di gambar lebih cenderung ke warna gelap atau terang. *Skewness* secara matematis dirumuskan sebagai berikut:

$$\text{Skewness} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(\frac{x_{ij} - \mu}{\sigma} \right)^3 \quad (6)$$

3. Random Forest

Secara umum, *Random Forest* adalah algoritma klasifikasi berbasis *ensemble* yang membangun banyak pohon keputusan dari subset data dan fitur yang dipilih

secara acak. Hasil prediksi dari seluruh pohon kemudian digabungkan melalui voting mayoritas [17]. Persamaan rumus *Random Forest* sebagai berikut:

$$\hat{y} = \frac{1}{T} \sum_{i=1}^T \hat{y}_i \quad (7)$$

4. XGBoost

Adapun *XGBoost* adalah algoritma boosting yang membangun model secara bertahap, di mana setiap pohon baru dibuat untuk mengoreksi kesalahan prediksi dari pohon sebelumnya [12]. Persamaan rumus *XGBoost* sebagai berikut:

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i), f_k \in Y \quad (8)$$

5. Ensemble Soft Voting

Ensemble soft voting merupakan teknik penggabungan beberapa model klasifikasi untuk meningkatkan kinerja prediksi secara keseluruhan. Pada metode *soft voting*, prediksi akhir ditentukan berdasarkan rata-rata probabilitas dari setiap model, sehingga memperhitungkan tingkat keyakinan masing-masing model terhadap kelas yang dipilih. Kelas dengan rata-rata nilai probabilitas tertinggi dipilih sebagai hasil klasifikasi. Persamaan rumus *ensemble soft voting* sebagai berikut [18]:

$$S = \frac{1}{N} \sum_{i=1}^N P_i \quad (9)$$

6. Confusion Matrix

Dalam pemodelan klasifikasi, proses utamanya adalah memprediksi label kelas yang telah ditentukan sebelumnya. Salah satu metode evaluasi yang umum digunakan adalah *confusion matrix*, yang menyajikan hasil prediksi dalam bentuk tabel berdasarkan perbandingan antara kelas aktual dan kelas prediksi [19]. Metode ini tidak hanya menunjukkan jumlah prediksi yang benar, tetapi juga memperlihatkan jenis kesalahan yang dilakukan oleh model. Oleh karena itu,

confusion matrix memberikan gambaran yang lebih komprehensif dibandingkan hanya menggunakan akurasi [20]. Matriks ini terdiri dari empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* [21]. Gambar 2 adalah visualisasi *confusion matrix*.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Gambar 2. Confusion Matrix

Akurasi adalah menghitung seberapa banyak prediksi yang sesuai dengan label sebenarnya. Presisi menunjukkan seberapa tepat model dalam memprediksi kelas positif. *Recall* menggambarkan seberapa baik model menemukan semua data positif, kemudian *F1-score* menyeimbangkan keduanya. Perhitungan tersebut menghasilkan rumus sebagai berikut [22]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall+Precision} \quad (13)$$

C. HASIL DAN PEMBAHASAN

1. Sumber Data

Dataset penelitian ini diperoleh dari dokumentasi langsung di perkebunan lada Desa Tokalimbo, Kecamatan Towuti, Kabupaten Luwu Timur, Sulawesi Selatan. Terdapat 1.996 citra buah lada yang terbagi dalam empat kategori tingkat kematangan, yaitu belum matang, setengah matang, matang, dan lewat matang, masing-masing terdiri dari sekitar 400 citra. Adapun ciri-

ciri visual dari setiap kelas adalah sebagai berikut:

- Lada Belum Matang**
Memiliki warna kulit hijau cerah dengan permukaan mengkilap dan tekstur keras.
- Lada Setengah Matang**
Warna kulit mulai berubah ke kekuningan atau oranye pucat, meski sebagian besar masih hijau.
- Lada Matang Sempurna**
Didominasi warna merah cerah dengan tekstur kulit lebih lunak dan kilap mulai berkurang.
- Lada Lewat Matang**
Warna berubah menjadi merah tua atau kecoklatan dengan tekstur lembek dan kulit tampak keriput.



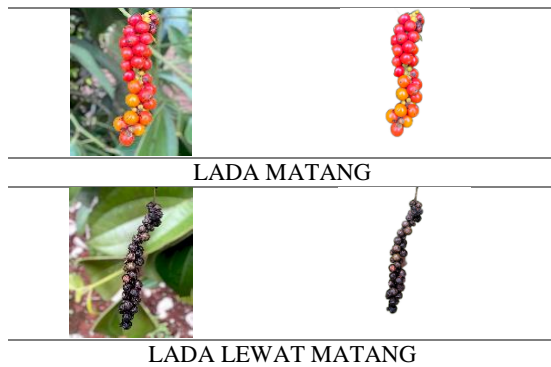
Gambar 3. Dataset Buah Lada

2. Processing Data

Sebelum klasifikasi, citra melalui proses segmentasi region of interest (ROI), penghapusan latar belakang, resize ke ukuran 224×224 piksel, dan normalisasi ke rentang 0–1.

Tabel 1. Hasil Processing Data

Gambah Asli	Hasil Segmentasi ROI
	
LADA BELUM MATANG	
	
LADA SETENGAH MATANG	



3. Ekstraksi Fitur

Metode ekstraksi fitur yang digunakan pada penelitian ini adalah *color moment* yang diterapkan pada ruang warna HSV. Pemilihan ruang warna HSV didasarkan pada kemampuannya merepresentasikan persepsi visual manusia terhadap warna serta memisahkan informasi warna murni, kejenuhan, dan tingkat kecerahan. *Color moment* menghitung tiga momen statistik utama, yaitu *mean*, *standard deviation*, dan *skewness* pada masing-masing kanal HSV. Dengan demikian, diperoleh total sembilan fitur warna dari setiap citra, yaitu Mean_H, Mean_S, Mean_V, Std_H, Std_S, Std_V, Skew_H, Skew_S, dan Skew_V, yang selanjutnya digunakan sebagai input untuk proses klasifikasi tingkat kematangan buah lada.

4. Python dan Library

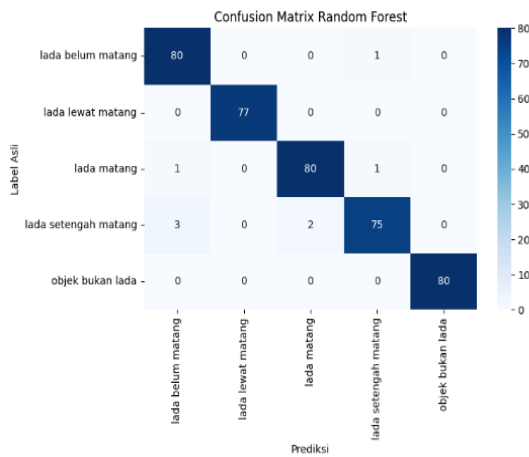
Python adalah bahasa pemrograman tingkat tinggi yang bersifat *open-source*, memiliki sintaks sederhana, dan banyak digunakan dalam bidang data science, *machine learning*, serta pengolahan citra digital. Dalam penggunaannya, *Python* didukung oleh berbagai *library*, yaitu kumpulan modul atau fungsi siap pakai yang dirancang untuk mempermudah proses pengembangan program, seperti dalam manipulasi data, ekstraksi fitur, pelatihan model, dan visualisasi hasil pada penelitian klasifikasi citra. Beberapa *library* penting yang digunakan dalam penelitian ini adalah :

- a. *NumPy* Digunakan untuk operasi numerik dan manipulasi array, terutama dalam pengolahan data fitur warna hasil ekstraksi HSV dan perhitungan statistik Color Moment.
- b. *OpenCV (cv2)* adalah *Library* utama untuk pengolahan citra digital, digunakan untuk membaca, menampilkan, mengubah ukuran citra, dan konversi ruang warna dari RGB ke HSV.
- c. *Pandas* Digunakan untuk mengelola dan memproses data dalam bentuk DataFrame, seperti menyimpan fitur-fitur warna, label kelas, serta ekspor hasil evaluasi ke format CSV.
- d. *Scikit-learn (sklearn)* Digunakan untuk preprocessing data (seperti *StandardScaler* dan *LabelEncoder*), membagi data menjadi data latih dan uji, serta implementasi model *Random Forest*, evaluasi performa model, dan metrik klasifikasi seperti akurasi, presisi, recall, dan F1-score.
- e. *XGBoost* Merupakan *library* khusus untuk algoritma *Extreme Gradient Boosting*, digunakan sebagai salah satu model utama dalam pendekatan ensemble klasifikasi.
- f. *Matplotlib* dan *Seaborn* Digunakan untuk visualisasi data dan hasil evaluasi model, seperti plotting *confusion matrix* untuk melihat performa pelatihan.

5. Hasil Pengujian Model

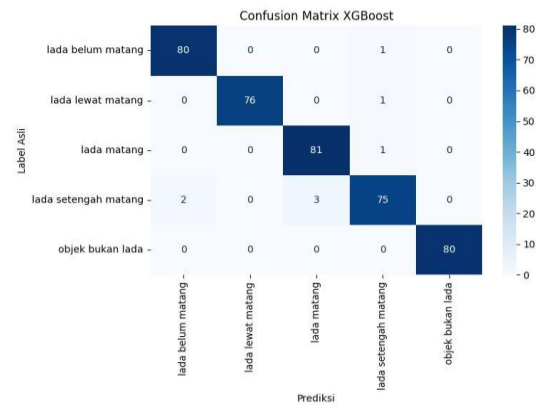
Penelitian ini menerapkan klasifikasi multi-kelas berdasarkan fitur warna HSV untuk membedakan empat tingkat kematangan buah lada. Fokus utama penelitian adalah penerapan metode *ensemble Soft Voting* yang menggabungkan dua algoritma, yaitu *Random Forest* dan *XGBoost*, guna meningkatkan akurasi, stabilitas, dan konsistensi prediksi. Kedua model dipilih karena mampu menangani data tabular secara efektif dan toleran terhadap korelasi antar fitur. Metode *Soft Voting* bekerja

dengan menggabungkan probabilitas keluaran dari masing-masing model, sehingga menghasilkan klasifikasi yang lebih andal dan mengurangi risiko kesalahan prediksi. Hasil pengujian menunjukkan bahwa metode *ensemble* ini memberikan akurasi tertinggi dengan distribusi prediksi yang lebih seimbang pada keempat kelas. Evaluasi dilakukan menggunakan 1.596 data latih dan 400 data uji. Kinerja model divisualisasikan melalui confusion matrix untuk memberikan gambaran yang jelas terhadap akurasi dan kesalahan klasifikasi pada setiap kelas.



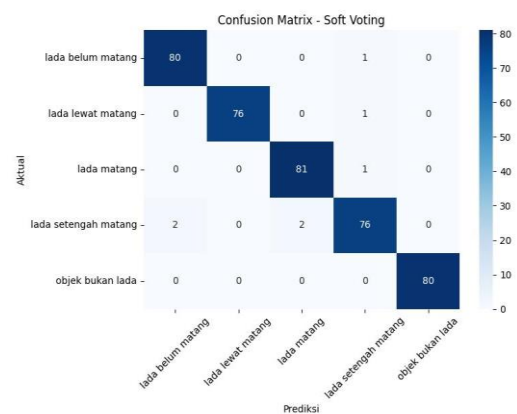
Gambar 4. Confusion Matrix Random Forest

Gambar 4 di atas menunjukkan confusion matrix dari model Random Forest yang digunakan untuk mengklasifikasikan tingkat kematangan lada. Hasil prediksi menunjukkan bahwa model memiliki kinerja yang sangat baik, dengan akurasi mencapai 98%. Nilai presisi sebesar 98,04%, *recall* 98,02%, dan *F1-score* 98,01% semakin menguatkan bahwa model memiliki performa klasifikasi yang optimal.



Gambar 5. Confusion Matrix XGBoost

Confusion matrix pada gambar 5 menunjukkan bahwa model XGBoost memiliki akurasi tinggi sebesar 98%. Mayoritas prediksi berada pada diagonal utama, menunjukkan akurasi klasifikasi yang tinggi. Nilai presisi 98,03%, *recall* 98,00%, dan *F1-score* 98,01% mencerminkan keseimbangan antara sensitivitas dan ketepatan model.



Gambar 6. Confusion matrix ensemble soft voting

Confusion matrix pada gambar 6 di atas menunjukkan kinerja model *ensemble* dengan pendekatan *soft voting*. Model ini mampu meningkatkan performa klasifikasi dengan akurasi mencapai 98%, presisi 98,04%, *recall* 98,02%, dan *F1-score* 98,01%. Dibandingkan model individu, ensemble menghasilkan prediksi yang

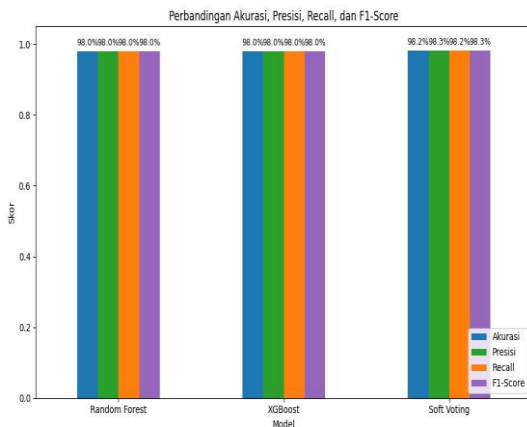
lebih stabil dan akurat, terutama dalam mengurangi kesalahan antar kelas.

Analisis akurasi per kelas dilakukan untuk mengevaluasi kemampuan model *Soft Voting* dalam mengenali setiap tingkat kematangan lada. Tabel 2 berikut menyajikan jumlah data uji, prediksi yang benar, dan persentase akurasi per kelas.

Tabel 2. Akurasi Per Kelas

Kelas	Data Uji	Prediksi Benar	Akurasi (%)
Lada Belum Matang	81	80	98.76
Lada Lewat Matang	77	76	98.70
Lada Matang	82	81	98.78
Lada Setengah Matang	80	76	95.00

Tabel 2 menunjukkan akurasi per kelas dari hasil prediksi model *Soft Voting* terhadap citra lada. Kelas lada setengah matang memiliki akurasi terendah sebesar 95,00% karena kemiripan warna dengan kelas lain, sehingga menyulitkan model dalam membedakannya.



Gambar 7. Perbandingan masing masing model

Tujuan dari diagram di atas adalah untuk memberikan visualisasi perbandingan kinerja model klasifikasi *Random Forest*, *XGBoost*, dan *Soft Voting* dalam mengidentifikasi tingkat kematangan lada. Melalui tampilan ini, pembaca dapat melihat secara langsung perbedaan hasil akurasi dari masing-

masing metode yang diuji. Selain itu, diagram ini memperjelas keunggulan pendekatan ensemble dibandingkan model tunggal, khususnya dalam hal stabilitas dan ketepatan prediksi. Visualisasi ini juga berfungsi sebagai alat bantu untuk menilai efektivitas setiap model secara menyeluruh.

Gambar 7 membandingkan kinerja *Random Forest*, *XGBoost*, dan *ensemble Soft Voting*. Kedua model tunggal mencatatkan akurasi 98,0%, sedangkan *Soft Voting* meningkatkannya menjadi 98,2%. Meskipun kenaikan akurasi tampak kecil, pendekatan ensemble menawarkan prediksi yang lebih stabil dan andal. Dengan menggabungkan keunggulan kedua model, *Soft Voting* mampu mereduksi bias dan varians, serta meningkatkan generalisasi terhadap data baru. Integrasi prediksi berbasis probabilitas memungkinkan sistem lebih tahan terhadap variasi dan noise. Oleh karena itu, pendekatan ini menjadi strategi yang efektif dalam meningkatkan kualitas model klasifikasi berbasis citra. Tujuan evaluasi ini adalah untuk menilai sejauh mana masing-masing model mampu mengklasifikasikan tingkat kematangan lada secara efektif dan konsisten.

Tabel 3. Perbandingan Akurasi

Model	Akurasi	Presisi	Recall	F1-Score
Random Forest	98%	98,04%	98,02%	98,01%
XGBoost	98%	98,03%	98%	98,01%
Ensemble Soft Voting	98,25%	98,27%	98,35%	98,26%

Dari tabel 4 di atas *Ensemble Soft Voting* mencatat skor tertinggi pada seluruh metrik evaluasi, termasuk akurasi, presisi, *recall*, dan *F1-score*. Hal ini menunjukkan bahwa pemilihan fitur yang tepat dan metode klasifikasi yang sesuai mampu menghasilkan model yang ringan

namun tetap andal untuk tugas pengenalan citra. Pemilihan fitur yang tepat dan metode klasifikasi yang sesuai mampu menghasilkan model yang ringan namun tetap andal untuk tugas pengenalan citra. Pendekatan ensemble, khususnya metode *Soft Voting*, memberikan keunggulan tambahan dengan menggabungkan kekuatan beberapa model dasar dalam hal ini *Random Forest* dan *XGBoost* untuk menghasilkan prediksi yang lebih stabil dan akurat. Hasil penelitian ini menunjukkan bahwa teknik *ensemble* bukan hanya adaptif, tetapi juga unggul dalam konteks klasifikasi citra berbasis warna seperti pada buah lada.

D. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan model klasifikasi tingkat kematangan buah lada berdasarkan warna menggunakan metode *Random Forest*, *XGBoost*, dan pendekatan *ensemble Soft Voting*. Hasil analisis menunjukkan bahwa fitur warna HSV, khususnya nilai mean, standar deviasi, dan skewness dari masing-masing kanal, efektif dalam membedakan tingkat kematangan lada. Ekstraksi sembilan fitur warna ini menjadi komponen penting dalam proses klasifikasi.

Model *Random Forest* dan *XGBoost* masing-masing mampu mencapai akurasi sebesar 98%, menunjukkan performa tinggi dalam mengenali pola visual dari citra. Penerapan *ensemble Soft Voting* yang menggabungkan kedua model tersebut menghasilkan akurasi tertinggi sebesar 98,25%, serta peningkatan presisi, recall, dan F1-score. Keunggulan metode *ensemble* terletak pada kemampuannya mengintegrasikan prediksi dari dua algoritma berbeda, menghasilkan model yang lebih stabil, akurat, dan andal dalam menggeneralisasi data baru. Evaluasi menggunakan *confusion matrix* dan akurasi per kelas memperkuat efektivitas sistem ini dalam tugas klasifikasi citra berbasis warna.

Untuk pengembangan lebih lanjut, disarankan menambahkan antarmuka web agar model lebih mudah diakses, memperkaya fitur dengan informasi tekstur, serta menerapkan metode *ensemble* stacking. Pendekatan stacking memungkinkan model meta mempelajari kombinasi prediksi *base learners*, sehingga meningkatkan akurasi dan generalisasi hasil klasifikasi.

E. REFERENSI

- [1] F. S. Alves *et al.*, “Evaluation of Antimicrobial Activity and Cytotoxicity Effects of Extracts of *Piper nigrum* L. and Piperine,” *Separations*, vol. 10, no. 1, pp. 1–18, 2023, doi: 10.3390/separations10010021.
- [2] X. Wu, X. Wu, H. Huang, F. Zhang, and Y. Wen, “Characterization of Pepper Ripeness in the Field Using Hyperspectral Imaging (HSI) with Back Propagation (BP) Neural Network and Kernel Based Extreme Learning Machine (KELM) Models,” *Anal. Lett.*, vol. 57, no. 3, pp. 409–424, 2024.
- [3] Pusat Data dan Informasi Pertanian, “OUTLOOK KOMODITAS LADA 2022 Pusat Data dan Sistem Informasi Pertanian i OUTLOOK LADA Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal-Kementerian Pertanian 2022,” 2022.
- [4] M. Ali Buchari, K. Haritami, and R. Passarella, “Implementasi Metode Hue Saturation Value Pada Identifikasi Citra Warna Kulit Studi Kasus: Pembuatan Surat Keterangan Catatan Kepolisian,” 2022.
- [5] B. Maulana Alfaruq, D. Erwanto, and I. Yanuartanti, “Klasifikasi Kematangan Buah Tomat Dengan

- Metode Support Vector Machine,” *Gener. J.*, vol. 7, no. 3, pp. 64–72, 2023, doi: 10.29407/gj.v7i3.21092.
- [6] R. Natras, B. Soja, and M. Schmidt, “Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting,” *Remote Sens.*, vol. 14, no. 15, pp. 1–34, 2022, doi: 10.3390/rs14153547.
- [7] I. D. Mienye and Y. Sun, “A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects,” *IEEE Access*, vol. 10, no. August, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
- [8] T. Kavzoglu and A. Teke, “Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost),” *Arab. J. Sci. Eng.*, vol. 47, no. 6, pp. 7367–7385, 2022, doi: 10.1007/s13369-022-06560-8.
- [9] S. Joses, D. Yulvida, and S. Rochimah, “Pendekatan Metode Ensemble Learning untuk Prakiraan Cuaca menggunakan Soft Voting Classifier,” *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 72–80, 2024, doi: 10.52158/jacost.v5i1.741.
- [10] S. Hidayat *et al.*, “Optimalisasi Model Ensemble Learning dengan Augmentasi dan SMOTE pada Sistem Pendeteksi Kualitas Buah,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 6, no. 1, pp. 27–36, Apr. 2024, doi: 10.35746/jtim.v6i1.406.
- [11] A. P. Argadianata *et al.*, “Klasifikasi Kualitas Buah Apel Menggunakan Metode Random Forest,” vol. 9, no. 2, pp. 2016–2022, 2025.
- [12] I. Elzandy *et al.*, “KLASIFIKASI VARIETAS ANGGUR BERDASARKAN CITRA DAUN MENGGUNAKAN VGG16-XGBOOST,” vol. 9, no. 3, pp. 5483–5491, 2025.
- [13] R. Mardianto, Stefanie Quinevera, and S. Rochimah, “Perbandingan Metode Random Forest, Convolutional Neural Network, dan Support Vector Machine Untuk Klasifikasi Jenis Mangga,” *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 63–71, 2024, doi: 10.52158/jacost.v5i1.742.
- [14] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, and J. Haider, “An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI,” *Sensors*, vol. 22, no. 19, 2022, doi: 10.3390/s22197268.
- [15] M. M. Maulana, “IMPLEMENTASI CITRA WARNA HSV UNTUK IDENTIFIKASI RHESUS GOLONGAN DARAH,” *Pharmacogn. Mag.*, vol. 75, no. 17, pp. 399–405, 2021.
- [16] A. I. Nurulrachman, R. C. Wihandika, and ..., “Ekstraksi Ciri pada Klasifikasi Citra Batik menggunakan Metode Gray Level Co-Occurrence Matrix, Local Binary Pattern, dan HSV Color Moment,” ... *Teknol. Inf. dan ...*, vol. 7, no. 1, pp. 374–383, 2023, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12197%0Ahttp://j-ptiik.ub.ac.id/index.php/j->

- ptiik/article/download/12197/5557
- [17] M. A. Amrozi, D. Figo SW, and R. Wahyusari, “Perbandingan Segmentasi Ruang Warna HSV dan YCbCr untuk Deteksi Objek,” *Infomatek*, vol. 26, no. 2, pp. 217–222, Dec. 2024, doi: 10.23969/infomatek.v26i2.19025.
- [18] C. C. Bey Lirna, T. Trimono, and A. T. Damaliana, “Employee Voluntary Attrition Prediction At Pt.Xyz: Ensemble Machine Learning Approach With Soft Voting Classifier,” *J. Tek. Inform.*, vol. 5, no. 5, pp. 1231–1239, 2024, doi: 10.52436/1.jutif.2024.5.5.2007.
- [19] E. Hernando, A. Chandra Saputra, and J. Parhusip, “KLASIFIKASI TINGKAT KEMATANGAN BUAH NAGA BERDASARKAN WARNA KULIT MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS,” Dec. 2024.
- [20] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.
- [21] S. E. Prasetyo, P. H. Prastyo, and S. Arti, “A Cardiotocographic Classification using Feature Selection: A comparative Study,” *JITCE (Journal Inf. Technol. Comput. Eng.*, vol. 5, no. 01, pp. 25–32, 2021, doi: 10.25077/jitce.5.01.25-32.2021.
- [22] E. Helmud, E. Helmud, F. Fitriyani, and P. Romadiana, “Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 1, pp. 92–97, 2024, doi: 10.32736/sisfokom.v13i1.1985.