Analisis Performa *Gradient Boosting* dalam Klasifikasi Kepadatan Gulma Berbasis Fitur Citra

[1st] Kholifahdina
S1 Rekayasa Perangkat Lunak
Universitas Telkom
Purwokerto, Indonesia
kholifahdina@student.telkomuniversity.ac.id

[2nd] Gita Fadila Fitriana
S1 Rekayasa Perangkat Lunak
Universitas Telkom
Purwokerto, Indonesia
gitafadila@telkomuniversity.ac.id

[3nd] Paradise
S1 Teknik Informatika
Universitas Telkom
Purwokerto, Indonesia
paradise@telkomuniversity.ac.id

***Abstrak* — Klasifikasi kepadatan gulma merupakan aspek penting dalam mendukung pengelolaan pertanian presisi, namun variasi visual citra seperti pencahayaan, tekstur, dan tumpang tindih vegetasi sering menyebabkan ketidakstabilan hasil klasifikasi. Penelitian ini bertujuan untuk membandingkan kinerja beberapa algoritma *machine learning* dengan pendekatan *Gradient Boosting* pada klasifikasi kepadatan gulma berbasis citra. Metodologi penelitian meliputi *preprocessing* citra, ekstraksi fitur warna (HSV), tekstur (GLCM), dan bentuk (Hu Moments), serta seleksi fitur menggunakan Information Gain hingga diperoleh fitur optimal. Dataset yang digunakan berjumlah 2.097 citra yang terbagi dalam tiga kelas, yaitu renggang, sedang, dan padat. Model yang diuji meliputi *Logistic Regression*, *Support Vector Machine*, *Decision Tree*, dan *Random Forest*, baik sebagai model dasar maupun dalam kombinasi dengan *Gradient Boosting*. Hasil pengujian menunjukkan bahwa model dasar menghasilkan akurasi pada rentang 0,77–0,98, sedangkan integrasi *Gradient Boosting* mampu meningkatkan performa hingga mencapai akurasi tertinggi sebesar 0,9905. Peningkatan paling signifikan terjadi pada model dengan performa awal rendah, menunjukkan bahwa pendekatan *boosting* efektif dalam mengurangi bias dan meningkatkan kemampuan generalisasi model. Dengan demikian, *Gradient Boosting* terbukti mampu meningkatkan akurasi dan stabilitas klasifikasi kepadatan gulma pada data citra dengan kompleksitas tinggi.**

***Kata kunci*— citra gulma, *Gradient Boosting*, klasifikasi kepadatan gulma, *machine learning*, ekstraksi fitur**

I. PENDAHULUAN

Pertanian memiliki peran penting dalam menjaga ketahanan pangan, namun produktivitas tanaman sering mengalami penurunan akibat kompetisi gulma terhadap sumber daya seperti nutrisi, air, dan cahaya. Tingkat kepadatan gulma menjadi indikator penting dalam menentukan kebutuhan pengendalian, karena semakin tinggi kepadatan gulma, semakin besar pula potensi gangguan terhadap pertumbuhan tanaman. Identifikasi kepadatan gulma secara manual masih mengandalkan pengamatan visual, sehingga berpotensi menghasilkan penilaian yang subjektif, kurang konsisten, serta tidak efisien untuk diterapkan pada lahan berskala luas [1], [2], [3], [4], [5].

Pemanfaatan computer vision dan *machine learning* telah banyak digunakan dalam proses deteksi dan klasifikasi gulma berbasis citra. Namun, performa model sering mengalami penurunan ketika dihadapkan pada kondisi lapangan yang bervariasi, seperti perubahan pencahayaan, latar belakang tanah yang tidak homogen, serta adanya tumpang tindih antar vegetasi [6], [7], [8]. Pendekatan deep learning memang mampu memberikan akurasi yang tinggi, tetapi umumnya memerlukan jumlah dataset yang besar serta sumber daya komputasi yang tinggi [1], [3], [7]. Oleh karena itu, pada kondisi data terbatas, pendekatan *machine learning* berbasis ekstraksi fitur visual masih menjadi alternatif yang relevan karena lebih ringan dan mudah diimplementasikan [5], [9].

Beberapa algoritma *machine learning* klasik seperti *Logistic Regression*, *Support Vector Machine* (SVM), *Decision Tree*, dan *Random Forest* telah banyak digunakan sebagai model dasar dalam klasifikasi citra. Meskipun demikian, performa algoritma tersebut dapat bervariasi tergantung pada kompleksitas karakteristik citra yang digunakan [10], [11], [12], [13] Untuk mengatasi permasalahan tersebut, metode *Gradient Boosting* dipilih karena mampu membangun model secara iteratif dengan memperbaiki kesalahan prediksi pada tahap sebelumnya, sehingga lebih efektif dalam menangkap pola non-linear dan meningkatkan stabilitas klasifikasi pada berbagai kondisi data [14], [15] [16], [17].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan performa *Gradient Boosting* dengan beberapa algoritma *machine learning* klasik dalam klasifikasi kepadatan gulma berbasis citra jarak dekat. Selain itu, model yang dihasilkan diimplementasikan ke dalam prototipe berbasis web guna mendukung proses pelatihan, pengujian, serta visualisasi hasil klasifikasi secara interaktif.

II. KAJIAN TEORI

A. Gulma dan Citra Digital

Gulma merupakan tumbuhan yang tumbuh pada lokasi yang tidak diinginkan dan dapat mengganggu pertumbuhan tanaman utama karena bersaing dalam memperoleh air, cahaya, dan unsur hara. Dalam konteks pertanian, keberadaan gulma perlu dikendalikan karena dapat menurunkan produktivitas tanaman secara signifikan. Dalam citra digital, gulma dapat dikenali melalui karakteristik visual seperti distribusi yang tidak teratur, variasi warna hijau, serta tekstur yang kompleks akibat tumpang tindih daun. Tingkat kepadatan gulma pada citra umumnya diklasifikasikan menjadi tiga kategori, yaitu renggang, sedang, dan padat, berdasarkan luas tutupan vegetasi pada area pengamatan [1], [2], [3], [5], [8].

B. Ekstraksi Fitur Citra

Ekstraksi fitur merupakan proses penting dalam mengubah citra menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Fitur yang umum digunakan dalam klasifikasi citra meliputi fitur warna, tekstur, dan bentuk. Fitur warna digunakan untuk menangkap karakteristik distribusi warna objek, fitur tekstur untuk menggambarkan pola permukaan objek, sedangkan fitur bentuk digunakan untuk merepresentasikan karakteristik geometris objek. Kombinasi ketiga jenis fitur tersebut dapat meningkatkan kemampuan model dalam membedakan kelas pada citra [9].

C. *Machine learning*

Machine learning merupakan metode yang memungkinkan sistem untuk mempelajari pola dari data dan menghasilkan prediksi tanpa perlu diprogram secara eksplisit. Dalam tugas klasifikasi, pendekatan yang digunakan adalah supervised learning, di mana model dilatih menggunakan data yang telah memiliki label [9], [19].

D. *Gradient Boosting*

Gradient Boosting merupakan salah satu metode ensemble learning yang membangun model secara bertahap dengan memperbaiki kesalahan dari model sebelumnya. Pendekatan ini menggabungkan beberapa model sederhana untuk menghasilkan model yang lebih kuat dan akurat. Keunggulan utama *Gradient Boosting* terletak pada kemampuannya dalam menangkap pola non-linear dan meningkatkan stabilitas prediksi. Oleh karena itu, metode ini banyak digunakan dalam berbagai permasalahan klasifikasi, terutama pada data dengan karakteristik kompleks [17], [20].

E. Evaluasi Model Klasifikasi

Evaluasi model klasifikasi bertujuan untuk mengukur kinerja model dalam memprediksi kelas secara akurat. Beberapa metrik yang umum digunakan meliputi *accuracy*, *precision*, *recall*, dan F1-*score*. Penggunaan beberapa metrik evaluasi secara bersamaan memberikan gambaran yang lebih komprehensif terhadap performa model, terutama dalam kondisi data yang tidak seimbang [16] [21]. Selain itu, waktu komputasi juga dapat digunakan sebagai indikator efisiensi model dalam proses pelatihan dan pengujian [22].

III. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membangun dan mengevaluasi model klasifikasi kepadatan gulma berbasis citra. Tahapan penelitian meliputi *preprocessing* citra, ekstraksi dan seleksi fitur, pelatihan model, serta evaluasi kinerja model.

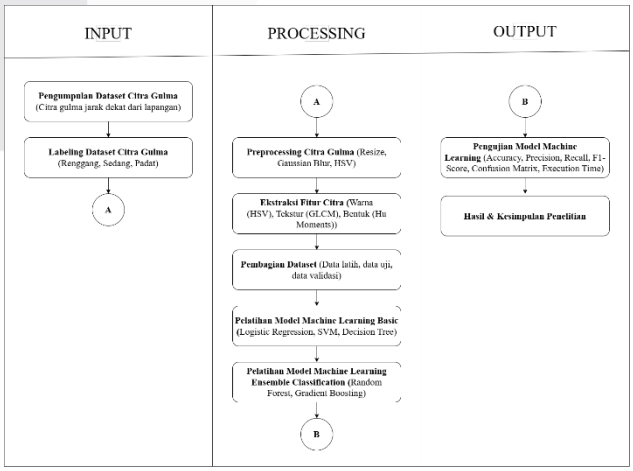
A. Data Penelitian

Data yang digunakan dalam penelitian ini berupa data primer citra gulma sebanyak 2.097 citra yang dikumpulkan pada periode 24 Juli hingga 25 Agustus 2024 di wilayah Purwokerto, Kabupaten Banyumas. Akuisisi citra dilakukan menggunakan kamera smartphone dengan jarak pengambilan sekitar ±15 cm secara tegak lurus (*nadir view*) terhadap objek, sehingga karakteristik visual seperti bentuk, tekstur, dan tingkat kepadatan gulma dapat terekam dengan jelas. Pengambilan citra dilakukan pada kondisi lingkungan alami dengan variasi pencahayaan.

TABEL 1
(Dataset Citra Gulma)

		
Renggang (610 citra)	**Sedang** (869 citra)	**Padat** (618 citra)

Dataset diklasifikasikan ke dalam tiga kelas kepadatan, yaitu renggang, sedang, dan padat berdasarkan tingkat tutupan vegetasi dan jarak antar gulma. Proses pelabelan dilakukan secara manual berdasarkan karakteristik visual pada setiap citra, kemudian diverifikasi oleh pakar di bidang agronomi untuk memastikan kesesuaian dengan kondisi riil di lapangan. Data tersebut yang akan digunakan untuk penelitian berikut:



Gambar 1
(Diagram Alir Penelitian)

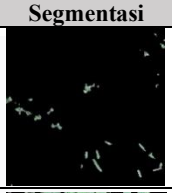
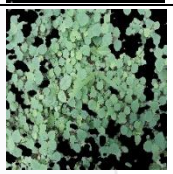
Diagram alur penelitian ditunjukkan pada Gambar 1, yang menggambarkan tahapan penelitian secara terstruktur mulai dari pengumpulan data, *preprocessing* citra, ekstraksi dan seleksi fitur, hingga pelatihan dan evaluasi model. Setiap tahapan dirancang untuk memastikan kualitas data dan optimalisasi performa model dalam melakukan klasifikasi kepadatan gulma berbasis citra.

B. *Preprocessing* Citra

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas citra dan menyeragamkan format data sebelum proses ekstraksi fitur. Proses ini meliputi penyesuaian ukuran citra (*resize*) menjadi 224×224 piksel untuk menjaga keseimbangan antara representasi fitur visual dan efisiensi komputasi. Selanjutnya, dilakukan reduksi *noise* menggunakan *Gaussian Blur* serta konversi ruang warna dari RGB ke HSV untuk mendukung proses segmentasi berbasis warna.

Segmentasi citra dilakukan dengan memanfaatkan rentang warna hijau pada ruang warna HSV untuk memisahkan objek gulma dari latar belakang seperti tanah dan elemen non-vegetasi lainnya. Proses ini menghasilkan citra yang menonjolkan area vegetasi sehingga karakteristik visual objek, seperti bentuk dan tekstur gulma, dapat dianalisis dengan lebih jelas pada tahap berikutnya.

TABEL 2
(Hasil *Preprocessing)*

Kelas Gulma	Citra Asli	Citra Hasil Segmentasi
Renggang		
Sedang		
Padat		

Berdasarkan hasil *preprocessing*, objek gulma berhasil dipisahkan dari latar belakang dengan cukup baik, terutama pada citra dengan tingkat kepadatan sedang dan padat, sedangkan pada citra dengan kepadatan renggang masih terdapat sebagian latar belakang yang terlihat namun tidak mengganggu proses ekstraksi fitur secara signifikan. Secara keseluruhan, tahapan *preprocessing* mampu meningkatkan kejelasan objek dan konsistensi data citra, serta menghasilkan citra dengan ukuran dan format yang seragam sehingga dapat digunakan secara optimal sebagai input pada proses ekstraksi fitur. Pada penelitian ini tidak dilakukan augmentasi data karena dataset yang digunakan bertujuan untuk merepresentasikan kondisi nyata di lapangan.

C. Ekstraksi Fitur

Ekstraksi fitur dilakukan untuk mengubah citra menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Pada penelitian ini digunakan kombinasi fitur warna (HSV), tekstur (GLCM), dan bentuk (Hu Moments) karena masing-masing mampu merepresentasikan karakteristik visual gulma secara saling melengkapi. Fitur warna digunakan untuk menangkap distribusi warna vegetasi, fitur tekstur merepresentasikan pola kerapatan dan hubungan antar piksel, sedangkan fitur bentuk digunakan untuk menggambarkan karakteristik geometris objek yang bersifat invarian terhadap rotasi dan skala. Proses ekstraksi menghasilkan total 39 fitur yang membentuk dataset numerik sebagai input model.

Namun, tidak seluruh fitur memiliki kontribusi yang sama dalam proses klasifikasi. Oleh karena itu, dilakukan seleksi fitur menggunakan metode *Information Gain* untuk mengukur tingkat relevansi setiap fitur terhadap kelas kepadatan gulma. Hasil perankingan menunjukkan bahwa fitur berbasis tekstur, khususnya *homogeneity* dan *energy* pada berbagai sudut, memiliki kontribusi paling dominan dalam membedakan kelas. Berdasarkan hasil tersebut, dipilih 14 fitur dengan nilai *Information Gain* tertinggi untuk digunakan pada tahap pelatihan model. Pemilihan ini bertujuan untuk mengurangi dimensi data serta meningkatkan efisiensi komputasi tanpa menghilangkan informasi penting yang dibutuhkan dalam proses klasifikasi.

D. Pembagian Data

Dataset yang digunakan dalam penelitian ini berjumlah 2.097 citra yang telah melalui proses validasi oleh pakar agronomi. Untuk kebutuhan pelatihan dan evaluasi model, data dibagi menjadi tiga bagian, yaitu data latih, data validasi, dan data uji dengan proporsi masing-masing sebesar 80%, 10%, dan 10%.

TABEL 3
(Pembagian Data)

Kelas Kepadatan	Total Data	Latih (80%)	Validasi (10%)	Uji (10%)
Renggang	610	488	61	61
Sedang	869	695	87	87
Padat	618	494	62	62
Total	**2.097**	**1.677**	**210**	**210**

Pembagian data dilakukan secara proporsional berdasarkan kelas kepadatan gulma untuk menjaga distribusi setiap kelas tetap seimbang pada setiap subset. Dengan pendekatan ini, model dapat dilatih menggunakan data yang representatif serta diuji secara objektif pada data yang belum pernah dilihat sebelumnya. Distribusi jumlah data pada masing-masing kelas dan subset ditunjukkan pada Tabel 3.

E. Pemodelan

Algoritma *machine learning* yang digunakan dalam penelitian ini, yaitu *Logistic Regression*, *Support Vector Machine* (SVM), *Decision Tree*, *Random Forest*, dan *Gradient Boosting*. Keempat algoritma pertama digunakan sebagai model pembanding karena mewakili pendekatan linear, *margin-based*, dan berbasis pohon, sedangkan *Gradient Boosting* digunakan sebagai model utama.

Logistic Regression merupakan model linear yang digunakan untuk memodelkan probabilitas suatu data termasuk ke dalam

kelas tertentu menggunakan fungsi *sigmoid*. Probabilitas klasifikasi dinyatakan sebagai:

$$P = (Y = 1|X) = \frac{1}{1+e^{-z}} \quad (1)$$

***Support Vector Machine* (SVM)** merupakan algoritma klasifikasi yang bekerja dengan mencari hyperplane optimal untuk memisahkan data antar kelas dengan margin maksimum. Persamaan hyperplane dinyatakan sebagai:

$$w \cdot x + b = 0 \quad (2)$$

Decision Tree membangun model klasifikasi dalam bentuk struktur pohon keputusan dengan membagi data secara rekursif. *Decision Tree* dalam penelitian ini menggunakan algoritma CART, proses pemisahan data dilakukan berdasarkan kriteria *impurity* Proses pemisahan menggunakan ukuran impurity yang dihitung dengan Gini Index:

$$Gini = 1 - \sum_{i=1}^{c} p_i^2 \quad (3)$$

Random Forest merupakan metode ensemble yang menggabungkan banyak *Decision Tree* untuk meningkatkan stabilitas dan akurasi model. Prediksi akhir diperoleh melalui mekanisme voting:

$$\hat{y} = mode(h_1(x), (h_2(x), \dots, (h_x(x)) \quad (4)$$

Gradient Boosting digunakan sebagai model utama karena membangun model secara bertahap dengan memperbaiki kesalahan dari model sebelumnya. Proses pembaruan model dinyatakan sebagai:

$$F_m(x) = F_{m-1}(x) +\cdot h_m(x) \quad (5)$$

Pelatihan model dilakukan dengan variasi parameter pada masing-masing algoritma untuk memperoleh konfigurasi optimal. Pada *Logistic Regression* dilakukan variasi parameter *max_iter*, pada SVM dilakukan variasi kernel (linear, RBF, dan polynomial), pada *Decision Tree* digunakan variasi *max_depth*, serta pada *Random Forest* digunakan variasi jumlah pohon (*n_estimators*). Selain itu, pada *Gradient Boosting* dilakukan variasi parameter *n_estimators* dan *learning_rate* untuk mengoptimalkan performa model.

F. Evaluasi Model

Evaluasi model dilakukan menggunakan data uji yang telah dipisahkan sebelumnya untuk mengukur kemampuan model dalam mengklasifikasikan tingkat kepadatan gulma. Hasil prediksi dari masing-masing model dibandingkan dengan label sebenarnya untuk menilai kinerja klasifikasi.

Metrik evaluasi yang digunakan meliputi ***accuracy*, *precision*, *recall*, dan F1-*score***, yang memberikan gambaran terhadap ketepatan dan kemampuan model dalam mengenali setiap kelas. Selain itu, digunakan ***confusion matrix*** untuk menganalisis distribusi kesalahan klasifikasi, serta ***execution time*** sebagai indikator efisiensi komputasi model.

IV. HASIL DAN PEMBAHASAN

Implementasi model klasifikasi untuk mendeteksi tingkat kepadatan gulma telah dilakukan sesuai dengan tahapan metodologi yang dirancang pada penelitian ini. Evaluasi dilakukan secara menyeluruh untuk mengukur performa model berdasarkan metrik akurasi serta menganalisis pengaruh kombinasi fitur dan parameter terhadap hasil klasifikasi. Selain itu, dilakukan analisis visual terhadap pola performa model untuk memahami karakteristik hasil yang diperoleh. Hasil eksperimen menunjukkan adanya variasi performa antar model serta pengaruh signifikan dari proses optimasi yang diterapkan.

A. Hasil *Preprocessing* dan Ekstraksi Fitur

Tahap *preprocessing* dan ekstraksi fitur telah dilakukan sebagaimana dijelaskan pada Bab III untuk menghasilkan data yang siap digunakan pada proses klasifikasi. Proses segmentasi berbasis HSV mampu memisahkan objek gulma dari latar belakang dengan baik, sehingga meningkatkan kejelasan karakteristik visual citra. Selanjutnya, ekstraksi fitur menghasilkan representasi numerik yang terdiri dari fitur warna, tekstur, dan bentuk. Berdasarkan hasil seleksi fitur, diperoleh bahwa penggunaan 14 fitur memberikan performa yang paling optimal dibandingkan jumlah fitur lainnya, sehingga digunakan pada tahap pelatihan model. $w \cdot x + b$

B. Hasil Pelatihan Model *Machine learning*

Hasil pelatihan model menunjukkan bahwa performa setiap algoritma dipengaruhi oleh jenis fitur yang digunakan. Secara umum, seluruh model menunjukkan performa yang lebih baik pada skenario tanpa penggunaan fitur tekstur (GLCM).

Pada ***Logistic Regression***, model menunjukkan stabilitas performa pada seluruh konfigurasi, dengan akurasi sebesar **0,8142 tanpa GLCM**, lebih tinggi dibandingkan **0,7809 pada skenario dengan GLCM**. Hal ini menunjukkan bahwa model linear kurang mampu memanfaatkan fitur tekstur yang kompleks dan cenderung lebih optimal menggunakan fitur yang lebih sederhana.

Pada ***Support Vector Machine* (SVM)**, performa model dipengaruhi oleh jenis kernel yang digunakan. Kernel linear menghasilkan akurasi terbaik pada skenario tanpa GLCM sebesar **0,8476**, sedangkan kernel RBF lebih optimal pada penggunaan fitur lengkap. Hal ini menunjukkan bahwa kompleksitas fitur memengaruhi kebutuhan pemodelan non-linear pada SVM.

Pada ***Decision Tree***, variasi parameter menghasilkan performa yang konsisten, dengan akurasi sebesar **0,8095 tanpa GLCM**. Hasil ini menunjukkan bahwa fitur warna dan bentuk sudah cukup representatif untuk memisahkan kelas dengan struktur model yang sederhana.

Sementara itu, ***Random Forest*** menghasilkan performa tertinggi di antara model dasar dengan akurasi mencapai **0,9857 tanpa GLCM**, yang menunjukkan keunggulan metode ensemble dalam menangkap pola kompleks pada data citra.

C. Hasil Optimasi Pelatihan Model *Machine learning*

Selanjutnya, dilakukan optimasi menggunakan metode *Gradient Boosting* pada seluruh model. Hasil menunjukkan bahwa penerapan *Gradient Boosting* secara konsisten meningkatkan performa model dibandingkan model dasar.

Pada ***Logistic Regression* berbasis *Gradient Boosting* (LR-GB)**, akurasi meningkat signifikan dari **0,8142 menjadi 0,9857** pada skenario tanpa GLCM. Namun, pada skenario dengan GLCM, akurasi maksimum hanya mencapai **0,9714**, yang menunjukkan bahwa penambahan fitur tekstur tidak memberikan peningkatan performa.

Pada **SVM berbasis *Gradient Boosting* (SVM-GB)**, akurasi meningkat dari **0,8476 menjadi 0,9904** pada skenario

tanpa GLCM. Sebaliknya, pada skenario dengan GLCM, akurasi menurun menjadi **0,9666**, yang mengindikasikan bahwa fitur tekstur tidak memberikan kontribusi signifikan terhadap peningkatan performa.

Pada ***Decision Tree* berbasis *Gradient Boosting* (DT-GB)**, akurasi meningkat dari **0,8095 menjadi 0,9904**, dengan waktu pelatihan yang relatif lebih efisien dibandingkan model lainnya. Hal ini menunjukkan bahwa kombinasi model sederhana dengan metode *boosting* mampu menghasilkan performa yang sangat optimal.

Pada ***Random Forest* berbasis *Gradient Boosting* (RF-GB)**, akurasi meningkat dari **0,9857 menjadi 0,9904** pada skenario tanpa GLCM. Namun, peningkatan ini relatif kecil dibandingkan model lainnya, yang menunjukkan bahwa *Random Forest* sudah memiliki performa tinggi sejak awal.

Secara keseluruhan, hasil ini menunjukkan bahwa penerapan *Gradient Boosting* mampu meningkatkan performa seluruh model secara signifikan, terutama pada model dengan performa awal yang rendah. Namun, peningkatan performa cenderung terbatas pada titik tertentu, yang mengindikasikan bahwa model telah mencapai batas optimal dalam memanfaatkan informasi dari fitur yang digunakan.

D. Perbandingan Performa Model

Untuk memberikan gambaran yang lebih komprehensif, perbandingan performa model dan parameter terbaik dirangkum pada Tabel 4.

TABEL 4
(Perbandingan Performa Model dan Parameter Terbaik)

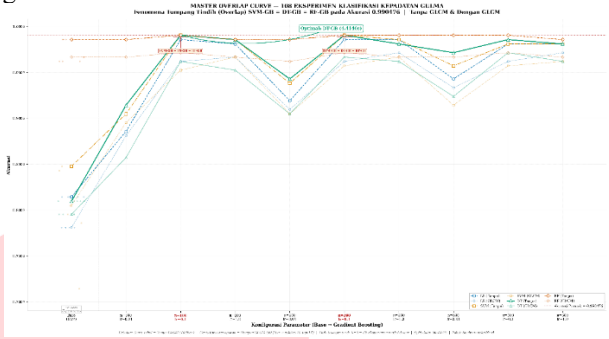
Model	Parameter Terbaik	*Accuracy* (tanpa GLCM)	*Accuracy* (GLCM)
Logistic Regression	max_iter = 100	0,8142	0,7809
SVM	kernel = linear, RBF (GLCM)	0,8476	0,8047
Decision Tree	max_depth = 3	0,8095	0,7952
Random Forest	n_estimators = 100	0,9857	0,9666
LR-GB	n=100, lr=0.1	0,9857	0,9714
SVM-GB	n=100, lr=0.1	0,9904	0,9666
DT-GB	n=100, lr=0.1	0,9904	0,9666
RF-GB	n=100, lr=0.1	0,9904	0,9666

Berdasarkan Tabel 4, terlihat bahwa seluruh model mengalami peningkatan performa setelah penerapan *Gradient Boosting*. Selain itu, skenario tanpa GLCM secara konsisten menghasilkan akurasi yang lebih tinggi dibandingkan dengan penggunaan fitur lengkap. Model **DT-GB tanpa GLCM** dipilih sebagai model terbaik karena mampu mencapai akurasi maksimum dengan waktu pelatihan yang lebih efisien dibandingkan model lainnya.

Berdasarkan Tabel 4, terlihat bahwa seluruh model mengalami peningkatan performa setelah penerapan *Gradient Boosting*. Selain itu, skenario tanpa GLCM secara konsisten menghasilkan akurasi yang lebih tinggi dibandingkan dengan penggunaan fitur lengkap. Hal ini menunjukkan bahwa fitur tekstur tidak memberikan kontribusi signifikan terhadap peningkatan performa model, bahkan cenderung menambah kompleksitas tanpa meningkatkan akurasi. Model DT-GB tanpa GLCM dipilih sebagai model terbaik karena mampu mencapai akurasi maksimum dengan waktu pelatihan yang lebih efisien dibandingkan model lainnya.

E. Visualisasi Performa Model

Visualisasi performa model menggunakan *Master Overlap Curve* menunjukkan pola peningkatan akurasi pada seluruh model setelah dilakukan optimasi. Selain itu, terdapat fenomena *overlap* antar model, di mana beberapa model menghasilkan akurasi yang sama pada konfigurasi tertentu. Hal ini menunjukkan bahwa model telah mencapai batas optimal dalam memanfaatkan informasi dari fitur yang digunakan.



GAMBAR 2
(Grafik Kurva Performa Model)

Visualisasi performa model menggunakan *Master Overlap Curve* menunjukkan adanya peningkatan akurasi yang signifikan pada seluruh model setelah dilakukan optimasi menggunakan *Gradient Boosting*. Selain itu, terdapat fenomena *overlap* antar kurva performa, di mana beberapa model menghasilkan nilai akurasi yang identik pada konfigurasi tertentu. Kondisi ini mengindikasikan bahwa model telah mencapai titik saturasi informasi, yaitu ketika fitur yang digunakan sudah cukup representatif sehingga perbedaan algoritma tidak lagi memberikan peningkatan performa yang signifikan.

Berdasarkan hasil penelitian, performa model dipengaruhi oleh kombinasi fitur dan metode optimasi yang digunakan. Fitur warna dan bentuk terbukti sudah cukup representatif, sedangkan penambahan fitur tekstur (GLCM) tidak memberikan peningkatan performa yang signifikan. Penerapan *Gradient Boosting* mampu meningkatkan akurasi seluruh model, meskipun peningkatan tersebut cenderung mencapai batas optimal. Model *Decision Tree* berbasis *Gradient Boosting* tanpa GLCM dipilih sebagai model terbaik karena memberikan akurasi tertinggi dengan efisiensi komputasi yang baik.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa kombinasi fitur warna (HSV) dan bentuk (Hu Moments) mampu merepresentasikan karakteristik citra gulma secara efektif dalam proses klasifikasi kepadatan. Penambahan fitur tekstur (GLCM) tidak memberikan peningkatan performa yang signifikan dan cenderung menambah kompleksitas model. Penerapan metode *Gradient Boosting* terbukti mampu meningkatkan akurasi seluruh model dibandingkan model dasar. Model terbaik diperoleh pada *Decision Tree* berbasis *Gradient Boosting* tanpa GLCM, yang mampu mencapai akurasi tertinggi sebesar 0,9904 dengan efisiensi waktu pelatihan yang baik. Hasil penelitian ini menunjukkan bahwa pemilihan fitur yang tepat serta penggunaan metode optimasi yang sesuai memiliki peran penting dalam meningkatkan performa model klasifikasi kepadatan gulma berbasis citra.

REFERENSI

[1] N. Y. Murad, T. Mahmood, A. R. M. Forkan, A. Morshed, P. P. Jayaraman, and M. S. Siddiqui, "Weed Detection Using Deep Learning: A Systematic Literature Review," Apr. 01, 2023, *MDPI*. doi: 10.3390/s23073670.

[2] F. D. Adhinata, Wahyono, and R. Sumiharto, "A Comprehensive Survey on Weed And Crop Classification Using *Machine learning* And Deep Learning," *Artificial Intelligence in Agriculture*, vol. 13, pp. 45–63, Sep. 2024, doi: 10.1016/j.aiia.2024.06.005.

[3] A. M. Mishra, M. P. Singh, P. Singh, M. Diwakar, I. Gupta, and A. Bijalwan, "Hybrid Deep Learning Model For Density And Growth Rate Estimation on Weed Image Dataset," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-86357-5.

[4] S. Dan *et al.*, *Pengembangan Pertanian Presisi*. 2024. [Online]. Available: https://epublikasi.pertanian.go.id/pertanianpress

[5] G. F. Fitriana, Paradise, D. P. Rakhmadani, P. Widyasunu, Kholifahdina, and F. D. Adhinata, "Visual Feature Combination Based on Feature Selection for Weed Density Classification," *International Journal of Computing and Digital Systems*, vol. 18, no. 1, 2025, doi: 10.12785/ijcds/1571111277.

[6] S. M. Jensen, M. J. Akhter, S. Azim, and J. Rasmussen, "The Predictive Power of Regression Models to Determine Grass Weed Infestations in Cereals Based on Drone Imagery—Statistical and Practical Aspects," *Agronomy*, vol. 11, no. 11, Nov. 2021, doi: 10.3390/agronomy11112277.

[7] S. G C, C. Koparan, M. R. Ahmed, Y. Zhang, K. Howatt, and X. Sun, "A Study on Deep Learning Algorithm Performance on Weed And Crop Species Identification Under Different Image Background," *Artificial Intelligence in Agriculture*, vol. 6, pp. 242–256, Jan. 2022, doi: 10.1016/j.aiia.2022.11.001.

[8] Z. Wu, Y. Chen, B. Zhao, X. Kang, and Y. Ding, "Review of Weed Detection Methods Based on Computer Vision," *Sensors*, vol. 21, no. 11, Jun. 2021, doi: 10.3390/s21113647.

[9] F. Marpaung, F. Aulia, N. Suryani SKom, and R. Cyra Nabila SKom, *Computer Vision dan Pengolahan Citra Digital*. Pustaka Askara, 2022. [Online]. Available: www.pustakaaksara.co.id

[10] P. Hu *et al.*, "Comparison of Conventional *Logistic Regression* and *Machine learning* Methods for Predicting Delayed Cerebral Ischemia After Aneurysmal Subarachnoid Hemorrhage: A Multicentric Observational Cohort Study," *Front. Aging Neurosci.*, vol. 14, Jun. 2022, doi: 10.3389/fnagi.2022.857521.

[11] Q. Li and J. Zhou, "A Comparative Analysis of Extreme *Gradient Boosting*, *Decision Tree*, *Support Vector Machine*s, and *Random Forest* Algorithm in Data Analysis of College Students' Psychological Health," *Informatica (Slovenia)*, vol. 49, no. 15, pp. 127–134, 2025, doi: 10.31449/inf.v49i15.7004.

[12] Z. Shao, M. N. Ahmad, and A. Javed, "Comparison of *Random Forest* and XGBoost Classifiers Using Integrated Optical and SAR Features for Mapping Urban Impervious Surface," *Remote Sens. (Basel).*, vol. 16, no. 4, Feb. 2024, doi: 10.3390/rs16040665.

[13] T. Huang *et al.*, "Comparing Algorithms for Estimation of Aboveground Biomass in Pinus yunnanensis," *Forests*, vol. 14, no. 9, Sep. 2023, doi: 10.3390/f14091742.

[14] Q. T. Bui *et al.*, "*Gradient Boosting* Machine and Object-based CNN for Land Cover Classification," *Remote Sens. (Basel).*, vol. 13, no. 14, Jul. 2021, doi: 10.3390/rs13142709.

[15] A. Lo Schiavo, W. Liu, A. Ramadan, K. El Khamlichi Drissi, C. Pasquier, and K. Tehrani, "An Analytical Investigation of Multi-Transmitter Dynamic Wireless Power Transfer for Electric Vehicles," *MDPI*, 2025, doi: 10.3390/app152413131.

[16] X. Y. Liew, N. Hameed, and J. Clos, "An Investigation of XGBoost-based Algorithm for Breast Cancer Classification," *Machine learning with Applications*, vol. 6, p. 100154, Dec. 2021, doi: 10.1016/j.mlwa.2021.100154.

[17] O. Alshboul, A. Shehadeh, G. Almasabha, and A. S. Almuflih, "Extreme *Gradient Boosting*-Based *Machine learning* Approach for Green Building Cost Prediction," *Sustainability (Switzerland)*, vol. 14, no. 11, Jun. 2022, doi: 10.3390/su14116651.

[18] R. Kursun and M. Koklu, *Optimized Classification of Cotton Leaf Diseases Using Machine learning and Feature Selection Techniques*. Çizgi Kitabevi Publishing, 2024.

[19] A. I. Adler and A. Painsky, "Feature Importance in *Gradient Boosting* Trees with Cross-Validation Feature Selection," *Entropy*, vol. 24, no. 5, May 2022, doi: 10.3390/e24050687.

[20] B. S. Waluyo Poetro, Eny Maria, H. Zein, E. Najwaini, and D. H. Zulfikar, "Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification," *Indonesian Journal of Data and Science*, vol. 5, no. 1, pp. 15–22, Mar. 2024, doi: 10.56705/ijodas.v5i1.123.

[21] A. Sanga and V. Zsófia, "Evaluation of *Machine learning* Algorithm's Selection Criteria for Satellite Image Classification, Leveraging *Accuracy* and *Execution time*s," *Egyptian Journal of Remote Sensing and Space Science*, vol. 28, no. 3, pp. 534–541, Sep. 2025, doi: 10.1016/j.ejrs.2025.08.001.