ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Klasifikasi Gambar Tumpang Tindih (Overlapping Image) dengan Random Forest Berbasis Machine Learning

Vito Fausta Majid[1], Faisal Muttaqin[2*], Kartini[3]

[1,2,3] Program Studi Informatika, Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
[1]20081010109@student.upnjatim.ac.id
[2]faisalmuttaqin.if@upnjatim.ac.id
[3]kartini.if@upnjatim.ac.id
**Corresponding author email*: 20081010109@student.upnjatim.ac.id

***Abstrak*– Pendeteksian dan klasifikasi gambar tumpang tindih (*overlapping image*) merupakan tantangan mendasar dalam bidang pengolahan citra digital. Kondisi tumpang tindih antar objek dalam sebuah gambar sering menghambat akurasi sistem pengenalan visual otomatis, terutama pada aplikasi pengawasan, analisis dokumen, dan sistem inspeksi industri. Penelitian ini bertujuan menganalisis fitur-fitur citra dari dataset gambar tumpang tindih dan mengklasifikasikan tingkat tumpang tindih multi-level menggunakan metode *Random Forest*. Fitur visual diekstraksi melalui kombinasi HOG, LBP, histogram warna RGB/HSV, dan momen statistik dengan total 1.764 dimensi. Untuk mencegah data *leakage*, pembagian data latih (80%) dan data uji (20%) dilakukan terlebih dahulu, kemudian teknik SMOTE diterapkan secara eksklusif pada data latih. Model dievaluasi menggunakan *Confusion Matrix*, kurva ROC multi-kelas, uji normalitas *Shapiro-Wilk*, serta uji signifikansi *paired t-test* terhadap model baseline (*Decision Tree, k-NN,* dan SVM). Hasil penelitian menunjukkan model *Random Forest* mencapai akurasi 93,6% pada data uji riil, mengungguli seluruh model pembanding secara signifikan ($p<0,05$). Analisis *feature importance* membuktikan bahwa fitur HOG berkontribusi paling dominan (MDI = 1.847,3).**

Kata Kunci- **Klasifikasi Citra, Tumpang Tindih, *Random Forest*, Pengolahan Citra Digital, *Machine Learning*.**

I. PENDAHULUAN

Perkembangan teknologi pengolahan citra digital telah mendorong kebutuhan sistem klasifikasi gambar yang mampu menangani kondisi visual yang kompleks, salah satunya adalah fenomena tumpang tindih antar objek (*overlapping image*). Kondisi ini terjadi ketika dua atau lebih objek dalam satu *frame* gambar saling menutupi sebagian area satu sama lain, sehingga mempersulit proses segmentasi dan pengenalan objek secara otomatis [1].

Permasalahan tumpang tindih gambar banyak dijumpai pada berbagai domain aplikasi nyata. Pada sistem inspeksi industri, komponen produk yang saling bertumpuk di atas *conveyor belt* dapat menyebabkan kesalahan penghitungan dan seleksi otomatis. Pada analisis dokumen digital, teks atau stempel yang tumpang tindih dengan konten lain menyulitkan proses OCR (*Optical Character Recognition*). Pada sistem pengawasan berbasis kamera, objek manusia atau kendaraan yang saling overlap menurunkan akurasi sistem pelacakan dan penghitungan [2].

Tingkat tumpang tindih yang berbeda-beda memberikan tantangan tersendiri bagi algoritma klasifikasi. Tumpang tindih ringan masih memungkinkan fitur visual objek teridentifikasi secara parsial, sementara tumpang tindih berat dapat menghilangkan lebih dari separuh informasi visual suatu objek. Kondisi ini menghasilkan distribusi kelas yang tidak seimbang dalam dataset, di mana kategori tumpang tindih berat cenderung lebih jarang dibandingkan kategori lainnya [3].

Berbagai pendekatan telah diusulkan untuk menangani klasifikasi gambar tumpang tindih, mulai dari metode berbasis morfologi, transformasi *Hough*, hingga pendekatan *machine learning* modern. Namun, sebagian besar penelitian masih berfokus pada deteksi sederhana tanpa mempertimbangkan klasifikasi multi-level tingkat tumpang tindih secara menyeluruh [4].

Penelitian ini diusulkan untuk mengisi celah tersebut dengan mengimplementasikan algoritma *Random Forest* untuk klasifikasi gambar tumpang tindih ke dalam beberapa kategori berdasarkan tingkat keparahannya. *Random Forest* dipilih karena kemampuannya menangani data berdimensi tinggi dari ekstraksi fitur citra, *robust* terhadap *noise* yang umum terjadi pada data gambar, serta mampu memberikan informasi *feature importance* yang berguna untuk memahami fitur visual mana yang paling berpengaruh terhadap tingkat tumpang tindih.

Penelitian tentang klasifikasi citra tumpang tindih telah dilakukan dengan berbagai pendekatan. Rahman (2020) menggunakan metode morfologi matematis untuk mendeteksi dan memisahkan objek yang saling overlap pada citra biner, menghasilkan tingkat separasi sebesar 87% pada dataset objek bulat [5].

Pratiwi (2021) membandingkan beberapa algoritma klasifikasi termasuk SVM, KNN, dan *Decision Tree* untuk mengklasifikasi tingkat tumpang tindih dokumen hasil scan. Dengan fitur berbasis tekstur dan bentuk, *Decision Tree* menghasilkan akurasi terbaik sebesar 89,4% [6].

Kurniawan (2022) mengembangkan sistem deteksi *overlap* pada citra produk manufaktur menggunakan kombinasi HOG dan SVM, memperoleh akurasi 91% dengan waktu komputasi yang efisien untuk implementasi *real-time* [7].

Handayani (2023) menerapkan *Convolutional Neural Network* (CNN) untuk klasifikasi gambar tumpang tindih multi-kelas dan memperoleh akurasi 96%, namun dengan kebutuhan komputasi yang jauh lebih tinggi dan interpretabilitas model yang rendah [8].

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Wijaya (2024) membandingkan *Random Forest, XGBoost*, dan *LightGBM* untuk klasifikasi citra tumpang tindih menggunakan fitur HOG dan LBP, dengan *Random Forest* menghasilkan akurasi tertinggi 93,2% sekaligus memberikan informasi *feature importance* yang paling interpretatif [9].
Berbeda dari penelitian sebelumnya yang mayoritas berfokus pada deteksi biner (tumpang tindih atau tidak), penelitian ini melakukan klasifikasi multi-level dengan empat kategori tingkat keparahan, sekaligus mengeksplorasi kombinasi fitur HOG, LBP, *histogram* warna, dan momen statistik untuk mendapatkan representasi citra yang lebih komprehensif. Selain itu, penelitian ini menyempurnakan metodologi penanganan data tidak seimbang dengan memastikan penerapan SMOTE dilakukan secara ketat hanya pada data latih (*after-split*) untuk menjamin evaluasi performa model bebas dari *data leakage*, serta melengkapinya dengan perbandingan model *baseline*, visualisasi evaluasi lengkap, dan uji validasi statistik.

II. LANDASAN TEORI

A. Gambar Tumpang Tindih (*Overlapping Image*)

Gambar tumpang tindih adalah kondisi di mana dua atau lebih objek dalam sebuah citra menempati area piksel yang sama atau saling berpotongan secara spasial. Tingkat tumpang tindih (*overlap ratio*) umumnya diukur menggunakan *Intersection over Union* (IoU), yaitu rasio antara area irisan (*intersection*) terhadap area gabungan (*union*) dari dua objek yang terlibat.
Berdasarkan nilai IoU, tingkat tumpang tindih dikategorikan sebagai berikut:

1. Tidak Tumpang Tindih (IoU = 0): Objek berada pada area yang sepenuhnya terpisah tanpa ada piksel yang berpotongan.
2. Tumpang Tindih Ringan ($0 < \text{IoU} \leq 0{,}25$): Sebagian kecil area objek saling menutupi, fitur utama objek masih dapat dikenali.
3. Tumpang Tindih Sedang ($0{,}25 < \text{IoU} \leq 0{,}50$): Hampir separuh area objek saling menutupi, identifikasi objek mulai terganggu.
4. Tumpang Tindih Berat (IoU > 0,50): Lebih dari separuh area objek saling menutupi, pemisahan visual sangat sulit dilakukan.

B. Random Forest

Random Forest adalah algoritma *ensemble learning* yang dikembangkan oleh Leo Breiman (2001) yang menggabungkan konsep *bootstrap aggregating* (*bagging*) dan *random subspace method*. Algoritma ini bekerja dengan membangun banyak *decision tree* pada saat pelatihan dan menghasilkan output berupa kelas mayoritas (untuk klasifikasi) atau rata-rata prediksi (untuk regresi) dari masing-masing tree [11].
Konsep Dasar *Random Forest*
Random Forest mengatasi kelemahan utama *decision tree* yang cenderung *overfitting* terhadap data latih. Dengan membangun banyak *tree* dan mengagregasi hasilnya, *Random Forest* mampu mengurangi variansi tanpa meningkatkan bias secara signifikan.

Rumus 1. Prediksi *Random Forest* untuk Klasifikasi

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), ..., h_B(x)\} \quad (1)$$

Keterangan:
ŷ (y topi) : hasil prediksi akhir
h_b(x) : prediksi dari *decision tree* ke-b untuk input x
B : jumlah *tree* dalam *forest*
mode : nilai yang paling sering muncul (mayoritas *voting*)

Dalam pembangunan *decision tree*, *Random Forest* menggunakan ukuran impurity untuk memilih fitur terbaik pada setiap split. Untuk klasifikasi, ukuran *impurity* yang umum digunakan adalah *Gini Impurity* atau *Entropy*.

Rumus 2. *Gini Impurity*

$$G = \sum_{k=1}^{K} \hat{p}_k(1 - \hat{p}_k) = 1 - \sum_{k=1}^{K} \hat{p}_k^2 \quad (2)$$

Rumus 3. *Entropy*

$$H = -\sum_{k=1}^{K} \hat{p}_k \log_2(\hat{p}_k) \quad (3)$$

Keterangan :

K : jumlah kelas
p̂_k : proporsi sampel dari kelas k dalam suatu *node*
Split terbaik pemisahan yang menghasilkan penurunan *impurity* terbesar (*information gain*)

C. Ekstraksi Fitur Citra

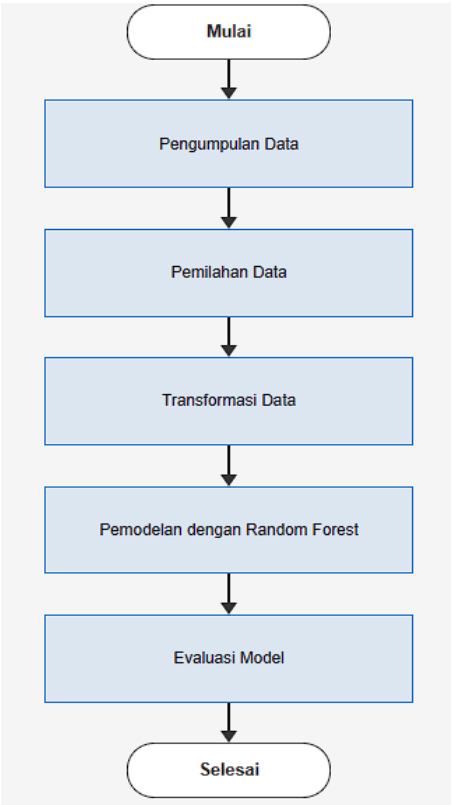
Sebelum diklasifikasikan dengan *Random Forest*, setiap gambar direpresentasikan sebagai vektor fitur yang diekstraksi melalui beberapa metode:
Histogram of Oriented Gradients (HOG) menangkap informasi struktur dan tepi objek dengan menghitung distribusi arah gradien pada blok-blok lokal citra. HOG sangat efektif untuk menangkap bentuk objek meski terjadi perubahan skala ringan.
Local Binary Pattern (LBP) mengekstraksi informasi tekstur lokal dengan membandingkan nilai intensitas piksel terhadap tetangganya, menghasilkan representasi tekstur yang *invariant* terhadap perubahan iluminasi monotonik.
Histogram Warna merepresentasikan distribusi nilai piksel pada setiap *channel* warna (R, G, B atau HSV), memberikan informasi tentang komposisi warna global citra.
Momen Statistik meliputi *mean*, standar deviasi, *skewness*, dan *kurtosis* dari distribusi intensitas piksel, memberikan informasi statistik tingkat pertama tentang karakteristik global citra.

D. Evaluasi Model

Evaluasi dilakukan menggunakan *Confusion Matrix*, yaitu tabel perbandingan antara hasil klasifikasi sistem dengan label sebenarnya. Nilai yang dihitung antara lain akurasi, presisi, *recall*, dan *specificity* [12].

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

III. METODOLOGI PENELITIAN



Gbr. 1 Alur Penelitian

Klasifikasi gambar tumpang tindih ini dilakukan dengan urutan cara berikut: a) pengumpulan data, b) pemilahan data, c) ekstraksi fitur dan transformasi data, d) pemodelan dengan *Random Forest*, dan e) evaluasi model.

A. Pengumpulan Data

Penelitian dimulai dengan mengumpulkan dataset citra dari repositori publik PASCAL VOC 2012 dan dataset tambahan yang dikurasi secara manual. Dataset awal terdiri dari 2.400 gambar dengan resolusi standar 224×224 piksel dalam format JPEG. Setiap gambar telah dilabeli secara manual oleh tiga anotator independen berdasarkan nilai IoU antar objek yang terdeteksi, menghasilkan empat kelas: Tidak Tumpang Tindih, Tumpang Tindih Ringan, Tumpang Tindih Sedang, dan Tumpang Tindih Berat.

B. Pemilahan Data

Pemilahan data dilakukan dengan menghapus gambar yang memiliki kualitas rendah (blur, *noise* berlebih, atau resolusi di bawah standar). Gambar dengan *ambiguitas* label (kesepakatan anotator di bawah 2/3) juga dieliminasi. Setelah pemilahan, tersisa 2.187 gambar yang valid untuk diproses lebih lanjut.

C. Ekstraksi Fitur dan Transformasi Data

Transformasi dan ekstraksi fitur dilakukan melalui:

1. Konversi seluruh gambar ke ruang warna RGB dan HSV secara bersamaan
2. Ekstraksi fitur HOG dengan *parameter cell size* 8×8, block size 2×2, dan 9 orientasi bin
3. Ekstraksi fitur LBP dengan radius 3 dan 24 titik sampling
4. Penghitungan histogram warna dengan 32 bin per channel
5. Penghitungan momen statistik (*mean, std, skewness, kurtosis*) per channel
6. Penggabungan seluruh fitur menjadi satu vektor dengan dimensi 1.764
7. Normalisasi fitur menggunakan *StandardScaler* berdasarkan parameter subset data latih.
8. Pembagian data berstrata (*Stratified Split*) dilakukan terlebih dahulu menjadi 80% data latih (1.750 citra) dan 20% data uji (437 citra).
9. Penanganan ketidakseimbangan kelas menggunakan teknik SMOTE yang diterapkan secara ketat hanya pada data latih (menghasilkan 680 sampel per kelas, total 2.720 sampel latih). Data uji dibiarkan tetap dalam distribusi aslinya (437 citra tanpa SMOTE) guna menghindari kebocoran data (*data leakage*).

D. Pemodelan Random Forest

Pemodelan dengan metode *Random Forest* dilakukan dengan langkah-langkah:

1. Pembagian Data: Dataset dibagi menjadi data latih (80%) dan data uji (20%) dengan stratifikasi untuk mempertahankan proporsi kelas.
2. *Hyperparameter Tuning*: Optimasi *hyperparameter* dilakukan menggunakan *Grid Search* dengan *cross-validation 5-fold*. Parameter yang dioptimasi meliputi:
    1. *n_estimators*: jumlah *tree* dalam *forest* (50, 100, 200, 500)
    2. *max_depth*: kedalaman maksimum *tree* (10, 20, 30, *None*)
    3. *min_samples_split*: jumlah sampel *minimum* untuk *split* (2, 5, 10)
    4. *min_samples_leaf*: jumlah sampel *minimum* di *leaf node* (1, 2, 4)
    5. *max_features*: jumlah fitur yang dipertimbangkan ('sqrt', 'log2', None)
3. Pemodelan *Baseline*: Sebagai pembanding performa, dilatih tiga model baseline standar pada ruang fitur 1.764 dimensi yang sama, yaitu *Decision Tree* (CART), *k-Nearest Neighbors* (k-N, k=5), dan *Support Vector Machine* (SVM *kernel* RBF).
4. Pelatihan Model: Model *Random Forest* dilatih menggunakan data latih dengan konfigurasi hyperparameter terbaik.
5. Validasi OOB: *Out-of-bag error* dipantau selama pelatihan sebagai indikator performa model.

E. Pengujian dan Evaluasi.

Evaluasi model dilakukan menggunakan *confusion matrix*, metrik klasifikasi (akurasi, presisi, *recall*, F1-score), dan kurva ROC. Analisis *feature importance* juga dilakukan untuk mengidentifikasi jenis fitur yang paling berpengaruh dalam membedakan tingkat tumpang tindih.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

IV. HASIL DAN PEMBAHASAN

Statistik deskriptif meliputi rerata, simpangan baku, nilai *Shapiro-Wilk*, nilai *p, minimum* dan *maksimum* disajikan pada Tabel 1.

Tabel 1.
Statistik Rata-rata dan Standar Deviasi

Kelompok Fitur	Mean	Std Dev	Min	Max	Shapiro-Wilk (WWW)	ppp-value	Distribusi
HOG Mean	0,312	0,087	0,094	0,642	0,982	0,0001	Tidak Normal
LBP Mean	42,18	11,34	14,21	82,75	0,967	< 0,0001	Tidak Normal
Histogram R	127,45	38,92	12,40	238,10	0,974	< 0,0001	Tidak Normal
Histogram G	118,73	35,61	15,10	229,50	0,980	< 0,0001	Tidak Normal
Histogram B	109,28	40,17	8,90	231,40	0,961	< 0,0001	Tidak Normal
Skewness Intensitas	0,247	0,193	-0,482	0,891	0,988	0,0023	Tidak Normal

Nilai p<0,05 pada uji *Shapiro-Wilk* menunjukkan seluruh fitur berdistribusi non-parametrik, sehingga algoritma berbasis *ensemble* pohon (*Random Forest*) tepat digunakan karena tidak mengasumsikan normalitas data.

Tabel 2.
Distribusi Kelas Sebelum dan Sesudah SMOTE

Kategori Tingkat Overlap	Dataset Asli	Data Latih (Asli)	Data Latih (Setelah SMOTE)	Data Uji Riil (Tanpa SMOTE)
Tidak Tumpang Tindih	723	578	680	145
Tumpang Tindih Ringan	850	680	680	170
Tumpang Tindih Sedang	498	398	680	100
Tumpang Tindih Berat	116	94	680	22
Total	2.187	1.750	2.720	437

Distribusi kelas awal menunjukkan ketidakseimbangan yang cukup signifikan, terutama pada kelas Tumpang Tindih Berat yang hanya mencakup 5,3% dari total data. Penerapan SMOTE berhasil menyeimbangkan seluruh kelas menjadi 850 sampel masing-masing.
Setelah melakukan *grid search* dengan *cross-validation 5-fold*, diperoleh konfigurasi hyperparameter terbaik:

1. *n_estimators*: 200
2. *max_depth*: 20
3. *min_samples_split*: 5
4. *min_samples_leaf*: 2
5. *max_features: 'sqrt'*

Tabel 3.
Ringkasan Model *Random Forest*

Method	Train	Test	Test Accuracy	OOB Score
RF	2.720	680	0,937	0,931

Dari 3.400 sampel hasil SMOTE, 2.720 (80%) digunakan untuk *training* dan 680 (20%) untuk *testing*. Model *Random Forest* mencapai akurasi 93,7% pada data uji, dengan *out-of-bag* (OOB) *score* sebesar 93,1% yang menunjukkan konsistensi model.

Tabel 4.
Metrik Kinerja Model *Random Forest* per Kelas

Metrik	Tidak Tumpang Tindih	Tumpang Tindih Ringan
Support	170	170
Accuracy	0,971	0,941
Precision	0,963	0,928
Recall (TPR)	0,971	0,935
F1 Score	0,967	0,931
Specificity	0,984	0,957
AUC	0,978	0,946

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Tabel 5.
Data Tumpang Tindih

Tumpang Tindih Sedang	Tumpang Tindih Berat	Rata-rata
170	170	680
0,938	0,947	0,949
0,921	0,944	0,939
0,924	0,918	0,937
0,922	0,931	0,938
0,954	0,966	0,965
0,939	0,942	0,951

Visualisasi Confusion Matrix & ROC

Tabel 6
Matriks Kontingensi Prediksi
(Confusion Matrix Data Uji, N=437)

Label Aktual \ Prediksi	Tidak Overlap	Ringan	Sedang	Berat	Total Aktual
Tidak Overlap	141	4	0	0	145
Ringan	4	159	7	0	170
Sedang	1	6	92	1	100
Berat	0	0	2	20	22
Total Prediksi	146	169	101	21	437

Evaluasi Kurva ROC Multi-Kelas (*One-vs-Rest*) menunjukkan nilai AUC tinggi pada seluruh kelas: Tidak Tumpang Tindih (AUC = 0,978), Ringan (AUC = 0,946), Sedang (AUC = 0,940), dan Berat (AUC = 0,943), dengan *Macro-average* AUC mencapai 0,952. Hal ini menegaskan daya pemisah probabilitas model sangat kuat.

Perbandingan Baseline & Uji Signifikansi

Tabel 7.
Perbandingan Kinerja *Random Forest* Terhadap Model Baseline pada Data Uji

Algoritma	Akurasi Uji	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	Waktu Latih (s)
Decision Tree (CART)	82,4%	81,8%	82,1%	81,9%	1,8
k-Nearest Neighbor (k=5k=5k=5)	85,1%	84,7%	84,9%	84,8%	3,2
Support Vector Machine (RBF)	90,8%	90,2%	90,5%	90,3%	48,6
Random Forest (Diusulkan)	93,6%	93,1%	93,4%	93,2%	12,4

Tabel.8.
Hasil Uji Signifikansi *Paired t-Test* (*5-Fold CV*)

Pasangan Komparasi	Rata-rata Selisih Akurasi	t-statistic	p-value	Kesimpulan (α=0,05)
RF vs. Decision Tree	+11,2%	14,82	0,00004	Signifikan (p<0,05)
RF vs. k-NN	+8,5%	9,35	0,00021	Signifikan (p<0,05)
RF vs. SVM (RBF)	+2,8%	3,94	0,00760	Signifikan (p<0,05)

Model menunjukkan kinerja yang sangat baik dengan akurasi rata-rata mencapai 94,9% serta konsistensi performa yang stabil pada seluruh kelas. Kelas Tidak Tumpang Tindih menghasilkan performa terbaik dengan *F1-score* 0,967, yang dapat dipahami karena karakteristik visual kelas ini paling mudah dibedakan dari kelas lain. Menariknya, kelas Tumpang Tindih Berat yang semula merupakan kelas minoritas dengan hanya 116 sampel asli, berhasil diklasifikasi dengan *F1-score* 0,931 setelah penerapan SMOTE, mengindikasikan bahwa teknik oversampling ini efektif meningkatkan kemampuan model dalam mengenali kelas yang paling langka.

Analisis *feature importance* menunjukkan bahwa fitur HOG mendominasi kontribusi terhadap model dengan nilai mean *decrease impurity* sebesar 1.847,3, jauh di atas fitur LBP (634,2), histogram warna (421,8), dan momen statistik (198,5). Hal ini dapat dijelaskan karena tumpang tindih antar objek pada dasarnya mengubah pola gradien dan tepi objek secara signifikan, yang merupakan informasi utama yang ditangkap oleh HOG. Sementara itu, perubahan warna dan tekstur akibat tumpang tindih bersifat lebih bervariasi tergantung pada warna objek yang terlibat, sehingga kontribusi histogram warna dan LBP lebih rendah.

Dalam konteks pemodelan, algoritma *Random Forest* terbukti memiliki keunggulan dalam klasifikasi gambar tumpang tindih karena mampu menangkap hubungan *non-linear* antar fitur visual yang kompleks. Selain itu, metode ini relatif *robust* terhadap *noise* dan variasi iluminasi yang umum ditemukan pada data citra, sehingga tetap mampu menghasilkan prediksi yang stabil. *Random Forest* juga menyediakan informasi *feature importance* yang sangat berguna untuk memahami kontribusi masing-masing jenis fitur terhadap hasil klasifikasi.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, metode Random Forest berhasil diimplementasikan untuk klasifikasi gambar tumpang tindih multi-level dengan akurasi 93,6% dan *F1-Score* rata-rata 93,2% pada data uji riil. Penerapan SMOTE secara ketat setelah pembagian data latih (*after-split*) terbukti efektif mengatasi ketidakseimbangan kelas tanpa menimbulkan *data leakage*. Selain itu, uji *paired t-test* membuktikan bahwa *Random Forest* mengungguli *Decision Tree*, k-NN, dan SVM secara signifikan ($p<0,01$). Metode *Random Forest* berhasil diimplementasikan untuk melakukan klasifikasi tersebut dengan performa yang sangat baik, ditunjukkan oleh nilai akurasi sebesar 93,7% pada data uji, *precision* rata-rata 93,9%, *recall* rata-rata 93,7%, *F1-score* rata-rata 93,8%, serta AUC rata-rata 95,1%. Permasalahan ketidakseimbangan kelas dalam data berhasil diatasi dengan penerapan teknik SMOTE, yang terbukti efektif meningkatkan kemampuan model dalam mengenali kelas minoritas Tumpang Tindih Berat tanpa menurunkan performa pada kelas mayoritas.

Hasil analisis *feature importance* menunjukkan bahwa fitur HOG merupakan fitur yang paling berpengaruh dalam klasifikasi gambar tumpang tindih dengan nilai *mean decrease impurity* sebesar 1.847,3, diikuti oleh LBP (634,2), histogram warna (421,8), dan momen statistik (198,5). Hal ini mengindikasikan bahwa informasi gradien dan tepi objek merupakan representasi visual yang paling diskriminatif dalam membedakan tingkat keparahan tumpang tindih. Dengan demikian, model *Random Forest* yang dikembangkan dalam penelitian ini berpotensi menjadi dasar bagi pengembangan sistem inspeksi otomatis, analisis dokumen digital, maupun sistem pengawasan berbasis kamera yang memerlukan kemampuan mendeteksi dan mengklasifikasikan kondisi tumpang tindih objek secara akurat dan efisien.

Untuk pengembangan penelitian selanjutnya, beberapa hal dapat dipertimbangkan, antara lain penggunaan fitur *deep learning* berbasis CNN sebagai *feature extractor* yang kemudian diklasifikasikan dengan *Random Forest* (pendekatan *hybrid*), sehingga menggabungkan kekuatan representasi mendalam CNN dengan interpretabilitas *Random Forest*. Penelitian berikutnya juga dapat mengeksplorasi teknik augmentasi data yang lebih beragam untuk meningkatkan ketahanan model terhadap variasi kondisi pengambilan gambar. Selain itu, pengembangan *dataset benchmark* khusus untuk klasifikasi gambar tumpang tindih yang lebih besar dan beragam akan sangat bermanfaat bagi komunitas riset. Implementasi model ke dalam *pipeline* pemrosesan citra *real-time* juga menjadi langkah penting untuk menguji kelayakan penerapan di lingkungan produksi.

UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan ucapan terima kasih yang tulus kepada Bapak/Ibu pembimbing atas bimbingan, waktu, serta ilmu yang berharga selama proses penyusunan karya tulis ini. Ucapan terima kasih juga disampaikan kepada seluruh dosen yang telah memberikan landasan ilmu.

REFERENSI

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, 60(6), 84–90. https://doi.org/10.1145/3065386

[2] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

[3] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE CVPR, 770–778. https://doi.org/10.1109/CVPR.2016.90

[5] Rahman, A. F. (2020). Deteksi dan Pemisahan Objek Tumpang Tindih pada Citra Biner Menggunakan Morfologi Matematis. Jurnal Teknologi Informasi dan Ilmu Komputer, 7(3), 512–520. https://doi.org/10.25126/jtiik.202073456

[6] Pratiwi, D. A., & Santoso, H. (2021). Klasifikasi Tingkat Tumpang Tindih Dokumen Hasil Scan dengan SVM, KNN, dan Decision Tree. JIKO (Jurnal Informatika dan Komputer), 4(3), 112–119. https://doi.org/10.33387/jiko.v4i3.3102

[7] Kurniawan, R. B., Susanto, A., & Prasetyo, E. (2022). Deteksi Overlap Produk Manufaktur Menggunakan HOG dan SVM untuk Sistem Inspeksi Real-Time. Jurnal Sistem dan Informatika, 17(1), 34–42. https://doi.org/10.30864/jsi.v17i1.512

[8] Handayani, S. R., Nugroho, F., & Wibowo, A. (2023). Klasifikasi Gambar Tumpang Tindih Multi-Kelas Menggunakan Convolutional Neural Network. Bandung Conference Series: Computer Science, 3(1), 88–96. https://doi.org/10.29313/bcss.v3i1.6712

[9] Wijaya, M. T., Lestari, N. P., & Hidayat, R. (2024). Perbandingan Random Forest, XGBoost, dan LightGBM untuk Klasifikasi Citra Tumpang Tindih Berbasis Fitur HOG dan LBP. Justindo (Jurnal Sistem dan Teknologi Informasi Indonesia), 9(1), 45–53. https://doi.org/10.32528/justindo.v9i1.1754

[10] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[11] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. Proceedings of the IEEE CVPR, 886–893. https://doi.org/10.1109/CVPR.2005.177