

# Deteksi Cacat Biji Kopi Berdasarkan Spesifikasi Specialty Coffee Association dengan YOLOv8

1<sup>st</sup> Ibadurrahman Syahid  
Fakultas Informatika  
Universitas Telkom  
Bandung, Indonesia

imnotplying@students.telkomuniversity.ac.id

2<sup>nd</sup> Ema Rachmawati  
Fakultas Informatika  
Universitas Telkom  
Bandung, Indonesia

emarachmawati@telkomuniversity.ac.id

**Abstrak** — Produksi kopi diprediksi akan menambah sebanyak 5,8% pada tahun 2024. Kopi dapat dibagi menjadi berbagai kualitas berdasarkan kecacatan yang ditemukan. Pemeriksaan kualitas kopi biasanya dilakukan melalui inspeksi visual, yang memakan waktu dan subjektif. Penelitian lain yang telah dilakukan menerapkan metode yang hanya mendeteksi adanya cacat, atau menggunakan tekstur untuk penilaian kualitas kopi. Penelitian ini menggunakan pendekatan yang berbeda, dengan metode You Only Look Once versi 8 (YOLOv8) untuk mendeteksi cacat berdasarkan standar Specialty Coffee Association (SCA). Dataset yang disusun adalah kumpulan 204 citra yang menampilkan 300 gram biji kopi hijau arabika mandheling. Dengan menggunakan pendekatan di mana model akan mendeteksi dan mengklasifikasikan cacat berdasarkan standar SCA, model dapat memberikan hasil yang lebih akurat dalam mendeteksi cacat biji kopi hijau dan mempermudah inspeksi kualitas kopi. Kontribusi utama dari penelitian ini adalah model yang dapat mendeteksi biji kopi yang memiliki cacat berdasarkan standar SCA. Model yang dibuat memiliki mean average precision sebesar 0,14.

**Kata kunci**— biji kopi, deteksi objek, computer vision, kualitas kopi.

## I. PENDAHULUAN

Kopi adalah minuman populer dan komoditas global utama, dengan produksi melebihi 10 juta ton pada 2022/2023 dan diprediksi meningkat 5,8% pada 2024 [1]. Specialty Coffee Association (SCA) menetapkan standar kualitas [2] dan jenis cacat biji kopi [3] yang mempengaruhi cita rasa. Penentuan kualitas secara konvensional melalui inspeksi visual 300 gram sampel memakan waktu dan subjektif.

Penelitian ini bertujuan membangun model deteksi cacat biji kopi berdasarkan standar SCA [3]. Penelitian sebelumnya [4-11] telah mencoba mendeteksi cacat, misalnya Huang dkk. menggunakan Convolutional Neural Network (CNN) dengan akurasi 94,63% [4], dan Arboleda dkk. dengan akurasi 100% untuk cacat "Black" menggunakan image processing [9]. Namun, penelitian tersebut memiliki keterbatasan, seperti hanya mendeteksi cacat tertentu [4, 9, 10], menilai kualitas berdasarkan tekstur [8], atau mengklasifikasi biji kopi secara individual [5, 6, 7].

Penelitian ini berbeda karena akan mendeteksi seluruh jenis cacat biji kopi sesuai standar SCA [3] secara komprehensif, menggunakan metode You Only Look Once versi 8 (YOLOv8) [12]. YOLOv8, dengan model tanpa anchor dan head terpisah, efektif mendeteksi objek kecil dengan konteks gambar menyeluruh [13]. Kontribusi utama penelitian ini adalah kemajuan dalam deteksi cacat biji kopi, dengan model yang mampu mengidentifikasi berbagai jenis cacat sesuai standar SCA.

Berdasarkan latar belakang penelitian ini, permasalahan utama yang akan dibahas adalah bagaimana cara mempercepat dan meningkatkan akurasi proses inspeksi kualitas kopi dengan mendeteksi cacat biji kopi sesuai standar Specialty Coffee Association (SCA) menggunakan pendekatan model YOLOv8. Namun, penelitian ini memiliki beberapa batasan. Keterbatasan waktu berdampak pada jumlah dataset yang digunakan, yaitu hanya 204 citra. Dataset ini juga belum mencakup jenis cacat Foreign Matter. Selain itu, penelitian ini hanya akan menguji satu arsitektur model, yaitu YOLOv8.

Tujuan dalam penelitian ini adalah pengembangan model YOLOv8 untuk mempercepat dan meningkatkan akurasi proses inspeksi kualitas kopi dengan mendeteksi cacat biji kopi yang sesuai dengan arahan yang diberikan oleh SCA.

## II. KAJIAN TEORI

### A. Evaluasi Biji Kopi Melalui Visi Komputer

Penelitian evaluasi biji kopi menggunakan visi komputer telah dilakukan dengan beragam pendekatan [4, 8-11]. Huang dkk. (2020) menggunakan Convolutional Neural Network (CNN) dan YOLOv3 untuk klasifikasi real-time biji kopi baik/buruk pada konveyor dengan akurasi 94,63% [4], namun tidak dapat mengevaluasi biji yang tidak menghadap kamera.

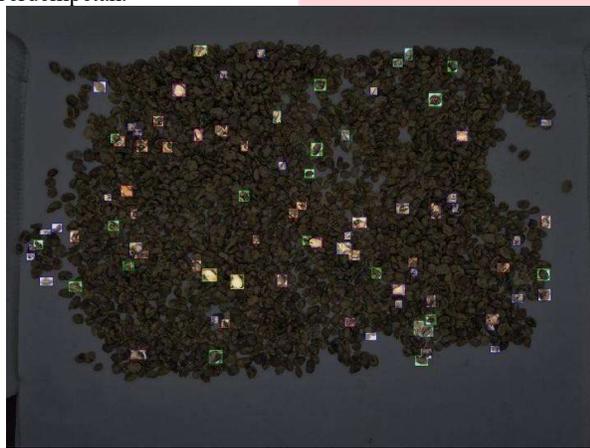
Chen dkk. (2020) menggunakan sensor Hyperspectral VIS-NIR, Constrained Energy Minimization (CEM), Support Vector Machine (SVM), dan CNN untuk mendeteksi kerusakan serangga dengan akurasi 96,48% [10]. Akbar dkk. (2020) mengevaluasi 300 gram biji kopi berdasarkan tekstur dan warna menggunakan K-Nearest Neighbor (KNN) dan Random Forest, dengan akurasi terbaik 86,56% (Random Forest) [8], namun datasetnya buram dan metodenya terbatas.

Pinto dkk. (2017) menggunakan CNN untuk klasifikasi enam tipe cacat dengan akurasi 72%-90% [11], namun evaluasinya terbatas pada dua citra (cacat dan normal), dengan akurasi tinggi hanya pada kelas "black" dan "sour" (perubahan warna signifikan). Arboleda dkk. (2018) menggunakan image processing untuk identifikasi biji hitam dengan akurasi 100% [9], tetapi hanya untuk cacat yang mempengaruhi warna.

Penelitian-penelitian tersebut [4, 9-11] umumnya hanya mengklasifikasi biji kopi cacat atau normal. Penelitian Pinto dkk. [11] mencoba klasifikasi berbagai jenis cacat, namun terbatas. Penelitian ini bertujuan mengatasi kekurangan tersebut dengan mendeteksi berbagai jenis cacat biji kopi sesuai spesifikasi Specialty Coffee Association (SCA).

### B. Deteksi Objek Berukuran Kecil

Model yang dibuat pada penelitian ini akan mendeteksi cacat biji kopi dari citra yang berisi biji kopi hijau mandheling sebanyak 300 gram. Gambar 1(A) menampilkan objek yang akan dideteksi oleh model sangat kecil dan berdempetan.



GAMBAR 1

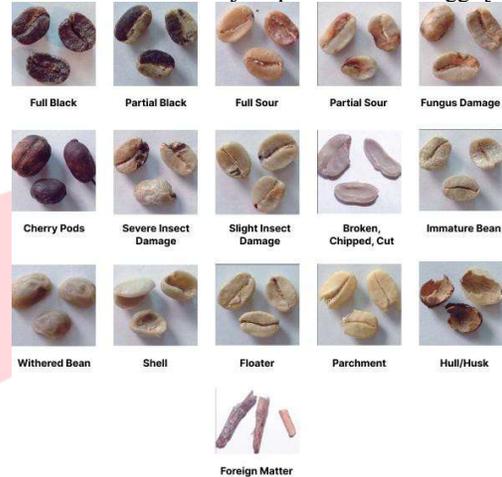
(A) Sampel citra yang sudah dianotasi.

Menurut Liu dkk [18], ada empat tantangan utama dalam deteksi objek berukuran kecil: pertama, keterbatasan informasi pada setiap feature layer di mana layer awal memiliki resolusi tinggi namun minim informasi semantik, sementara layer akhir kaya informasi semantik tetapi kehilangan resolusi spasial; kedua, minimnya informasi kontekstual pada objek kecil itu sendiri, padahal informasi ini penting untuk deteksi; ketiga, ketidakseimbangan kelas yang signifikan karena banyaknya anchor box yang dihasilkan region proposal network namun hanya sedikit yang cocok dengan objek kecil; dan keempat, tidak cukupnya contoh positif untuk melatih model secara efektif karena jumlah anchor box berskala kecil yang tidak memadai dan sulitnya mencocokkan anchor dengan ground truth. Kesimpulannya, kombinasi dari keterbatasan informasi fitur, minimnya konteks, ketidakseimbangan kelas, dan kelangkaan contoh positif berkontribusi pada sulitnya mendeteksi objek berukuran kecil menggunakan deep learning.

Pada penelitian yang dilakukan oleh Liu dkk [18], didapatkan performa model yang rendah untuk deteksi objek berukuran kecil, dengan performa model paling tinggi memiliki mAP sebesar 0,35 dan yang paling rendah dengan mAP 0,17.

### C. Efek Biji Kopi Cacat

Biji kopi yang cacat dapat memengaruhi kualitas kopi yang dihasilkan. Industri kopi bergantung pada mikrobiologi dan proses fermentasi untuk memungkinkan biji kopi mengembangkan karakteristik spesifik yang terkait dengan varietas kopi, proses yang digunakan, dan kondisi dalam pengeringan dan penyimpanan. Semua aspek ini harus ditangani untuk meningkatkan kualitas kopi dan mencoba memastikan ketersediaan biji kopi berkualitas tinggi [14].



GAMBAR 2

(B) Cacat biji kopi berdasarkan standar SCA [3].

Beberapa penelitian telah dilakukan untuk mengetahui efek dari cacat pada biji kopi terhadap kualitas kopi yang dihasilkan. Hale dkk pada tahun 2022 melakukan penelitian yang menunjukkan bahwa ada 4 macam fungi yang memengaruhi rasa kopi secara negatif [15]. Penelitian tersebut juga memberi hipotesis bahwa fungi mungkin tersebar melalui serangga antestia, yang sering kali ditemukan pada biji kopi yang cacat.

Diaz dkk melakukan penelitian pada tahun 2023 sekelompok kopi yang memiliki cacat memengaruhi komposisi kimia dari minuman yang dihasilkan [16]. Kelompok kopi dengan cacat "sour" memiliki kadar kahweol yang lebih tinggi, sementara konten vafestol dan trigoneline lebih tinggi pada kelompok kopi yang cacat "husk".

Biji kopi cacat menurunkan kualitas kopi [14]. Mikrobiologi dan fermentasi penting dalam mengembangkan karakteristik varietas kopi, proses, pengeringan, dan penyimpanan [14]. Penelitian menunjukkan bahwa fungi mempengaruhi rasa kopi secara negatif [15], dan cacat biji kopi mempengaruhi komposisi kimia minuman [16]. Misalnya, cacat "sour" meningkatkan kahweol, sedangkan "husk" meningkatkan vafestol dan trigoneline [16].

SCA mengklasifikasikan cacat biji kopi (Gambar 2(B)) dan ekuivalensinya terhadap cacat penuh (Tabel 1(A)) [3]. Contohnya, satu biji "Full Black" adalah satu cacat penuh, sedangkan tiga "Partial Black" juga dihitung satu cacat penuh [3].

Kualitas biji kopi dibagi menjadi lima kelas [17]:

- Specialty Grade (Grade 1): Maksimal 5 cacat penuh/300g, tanpa cacat kategori 1, toleransi ukuran layar 5%, minimal satu atribut khas, bebas kesalahan, tanpa quakers, kelembapan 9-13%.
- Premium Grade (Grade 2): Maksimal 8 cacat penuh/300g, cacat kategori 1 diperbolehkan, toleransi ukuran

layar 5%, minimal satu atribut khas, bebas kesalahan, maksimal tiga quakers, kelembapan 9-13%.

- Exchange Grade (Grade 3): 9-23 cacat penuh/300g, 50% di atas ukuran layar 15 (toleransi 5% di bawah 14), tanpa kesalahan cangkir, maksimal lima quakers, kelembapan 9-13%.

- Below Standard Grade (Grade 4): 24-86 cacat penuh/300g.

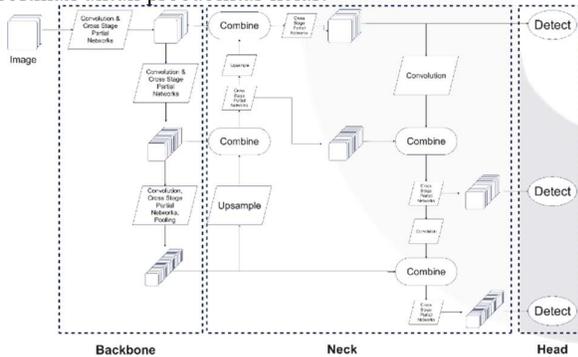
- Off Grade (Grade 5): Lebih dari 86 cacat penuh/300g.

TABEL 1  
(A) KESETARAAN CACAT BIJI KOPI [3].

Category 1 Defects		Category 2 Defects	
Defect	Full Defect Equivalents	Defect	Full Defect Equivalents
Full Black	1	Partial Black	3
Full Sour	1	Partial Sour	3
Dried Cherry/Pod	1	Parchment/Pergamino	5
Fungus Damaged	1	Floater	5
Foreign Matter	1	Immature/Unripe	5
Severe Insect Damage	5	Withered	5
		Shell	5
		Broken/Chipped/Cut	5
		Hull/Husk	5
		Slight Insect Damage	10

#### D. YOLOv8

You Only Look Once versi 8 (YOLOv8) [12], dirilis Ultralytics (pengembang YOLOv5) pada Januari 2023, menggunakan model tanpa anchor dengan head terpisah untuk tugas objectness, klasifikasi, dan regresi [13]. Desain ini meningkatkan akurasi. Fungsi aktivasi sigmoid digunakan untuk skor objectness (bounding box berisi objek), dan softmax untuk probabilitas kelas.

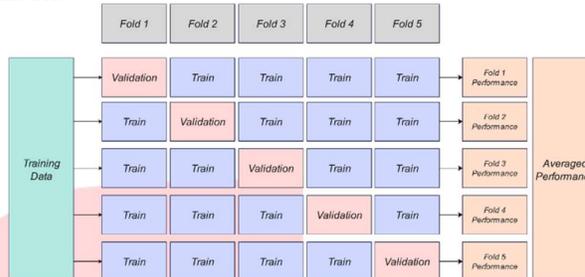


GAMBAR 3  
(C) Arsitektur YOLOv8.

Arsitektur YOLOv8 (Gambar 3 (C)) memiliki beberapa head, masing-masing menerima kombinasi fitur berbeda. Ini memungkinkan model mempertahankan detail objek kecil dan konteks objek besar. Head resolusi tinggi meningkatkan lokalisasi objek kecil, sedangkan head resolusi rendah menangkap informasi kontekstual untuk klasifikasi, meningkatkan efektivitas lokalisasi dan klasifikasi.

#### E. K-Fold Cross Validation

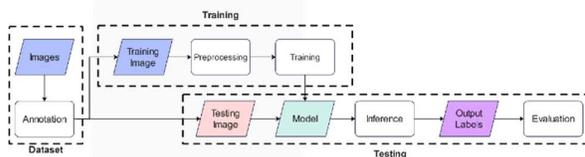
K-fold Cross Validation adalah metode pembagian dataset yang mana data dibagi menjadi k segmen atau fold yang ukurannya sama atau hampir sama. Pelatihan dan pengujian dilakukan dalam k iterasi, sehingga pada setiap iterasi, terdapat fold yang ditinggal untuk pengujian dan melatih model pada k-1 fold yang tersisa [19]. Performa dari semua fold akan dirata-ratakan untuk mendapatkan performa model.



GAMBAR 4  
(D) Proses K-fold Cross Validation.

Gambar 4(D) menunjukkan proses dari K-fold Cross Validation dengan jumlah 5 fold. Proses ini dapat meningkatkan robustness model yang dibuat dengan membuat model dengan beberapa kombinasi pelatihan dan pengujian, dan mengambil rata-rata dari model-model yang dibuat.

### III. METODE



GAMBAR 5  
(E) Desain sistem.

Penelitian ini membangun sistem deteksi cacat biji kopi yang terdiri dari tiga tahap utama: pembuatan dataset, pelatihan (training), dan pengujian (testing) (Gambar 5(E)).

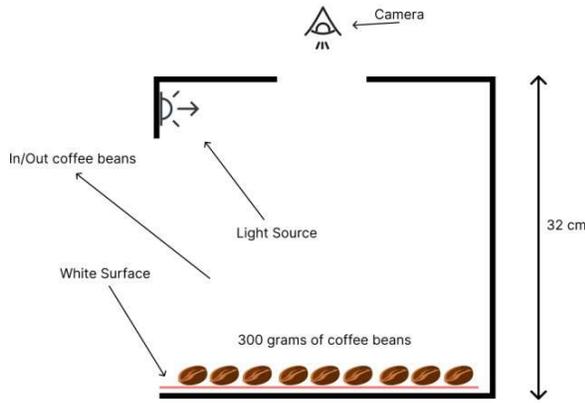
Pada tahap pembuatan dataset, gambar biji kopi dikumpulkan dan dianotasi secara manual untuk menentukan lokasi dan kelas cacat. Preprocessing dilakukan, meliputi perubahan ukuran dan augmentasi data.

Selanjutnya, model YOLOv8 dilatih secara otomatis menggunakan dataset yang telah diproses. Model ini belajar mengenali dan mendeteksi cacat pada biji kopi. Hasilnya adalah model terlatih yang mampu mendeteksi objek dalam gambar.

Tahap pengujian menggunakan set gambar berbeda yang juga dianotasi secara manual. Setelah preprocessing, model YOLOv8 melakukan inferensi (prediksi) pada gambar pengujian, menghasilkan kotak pembatas dan label kelas untuk objek yang terdeteksi. Hasil deteksi, berupa lokasi dan klasifikasi objek, kemudian ditampilkan.

Performa model dievaluasi dengan membandingkan hasil prediksi dengan anotasi manual. Proses ini memastikan akurasi dan keandalan model dalam mendeteksi cacat biji kopi.

## A. Citra Input



GAMBAR 6  
(F) Ilustrasi akuisisi citra dari 300 gram biji kopi hijau Arabika Mandheling.

Dataset penelitian ini terdiri dari 204 citra biji kopi hijau Mandheling (4000x3000 piksel), diambil dalam kotak dengan pencahayaan konsisten dan latar belakang putih (Gambar 6(F)). Setiap citra berisi 300 gram biji kopi yang diacak setelah setiap pengambilan untuk variasi. Metode 5-fold Cross Validation diterapkan pada 163 citra (130-131 training, 32-33 validasi per fold), dan 41 citra untuk pengujian terpisah.

Biji kopi dari "Narara Coffee" diklasifikasikan menjadi grade 1 (terbaik), 2, dan 3 berdasarkan jumlah dan jenis cacat [standar kualitas tidak disebutkan, tapi diasumsikan mengacu pada referensi sebelumnya]. Dataset terdiri dari 133 citra grade 1, 34 grade 2, dan 37 grade 3 (Gambar 7(G)). Variasi grade ini penting karena bentuk dan frekuensi cacat berbeda antar grade.

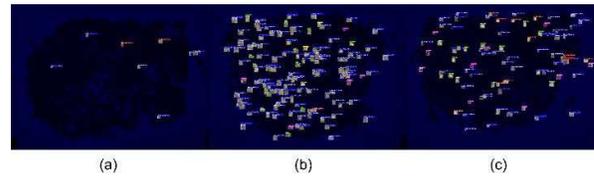


GAMBAR 7  
(G) Sampel citra biji kopi: (a) biji kopi dengan grade 1, (b) biji kopi dengan grade 2, (c) biji kopi dengan grade 3.

Karena keterbatasan jumlah citra, augmentasi data dilakukan pada citra pelatihan untuk meningkatkan performa model [20]. Augmentasi meliputi horizontal flipping, vertical flipping, dan rotasi 90 derajat. Hal ini bertujuan untuk memperkaya variasi data latih dan meningkatkan kemampuan generalisasi model.

## B. Anotasi

Anotasi manual dilakukan pada citra berdasarkan standar Specialty Coffee Association (SCA). Jenis cacat yang dianotasi meliputi: Full Black, Partial Black, Full Sour, Partial Sour, Fungus Damage, Cherry Pods, Severe Insect Damage, Slight Insect Damage, Broken/Chipped/Cut, Immature Bean, Withered Bean, Shell, Floater, Parchment, Hull/Husk, dan Foreign Matter (Gambar 2(B)). Namun, Foreign Matter tidak ditemukan dalam dataset ini. Total terdapat 8.152 label (Tabel 2(B)).



GAMBAR 8  
(H) Sampel citra yang dianotasi: (a) biji kopi dengan grade 1, (b) biji kopi dengan grade 2, (c) biji kopi dengan grade 3.

Format label yang digunakan adalah format YOLOv8, menggunakan kotak pembatas (bounding box) untuk menandai lokasi spasial objek. Setiap kotak label berisi informasi kelas objek dan empat koordinat:  $x_{min}$ ,  $y_{min}$  (sudut kiri atas), dan  $x_{max}$ ,  $y_{max}$  (sudut kanan bawah).

Gambar 8(H) menunjukkan sampel citra yang telah dianotasi. Biji kopi cacat ditandai dengan kotak berwarna, dengan kode warna sebagai berikut: Broken/Chipped/Cut (Biru), Full Black (Hijau), Full Sour (Merah), Fungus Damaged (Sian), Hull/Husk (Magenta), Immature/Uripe (Kuning), Partial Black (Biru Tua), Partial Sour (Hijau Tua), Severe Insect Damage (Merah Tua), Shell (Sian Tua), Slight Insect Damage (Magenta Tua). Kombinasi warna dan label ini memfasilitasi pelatihan model untuk mendeteksi dan mengklasifikasikan berbagai jenis cacat.

TABEL 2  
(A) DISTRIBUSI LABEL DATASET.

Type	Jumlah	Persentase
Full Black	902	11,1%
Partial Black	74	0,9%
Full Sour	137	1,7%
Partial Sour	111	1,4%
Fungus Damage	139	1,7%
Cherry Pods	11	0,1%
Severe Insect Damage	946	11,6%
Slight Insect Damage	416	5,1%
Broken/Chipped/Cut	4219	51,8%
Immature Bean	448	5,5%
Withered Bean	19	0,2%
Shell	477	5,9%
Floater	7	0,1%
Parchment	6	0,1%
Hull/Husk	240	2,9%

## C. Evaluasi Model

Performa model dievaluasi menggunakan metrik Precision, Recall, dan Mean Average Precision (mAP) [21].

$$\text{Precision} = \frac{\text{TruePositive}(TP)}{\text{TruePositive}(TP) + \text{FalsePositive}(FP)} \quad (1)$$

$$\text{Recall} = \frac{\text{TruePositive}(TP)}{\text{TruePositive}(TP) + \text{FalseNegative}(FN)} \quad (2)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (3)$$

Penelitian ini menggunakan mAP50, yang hanya mempertimbangkan prediksi dengan Intersection over Union (IoU) minimal 50%. IoU mengukur tumpang tindih antara bounding box prediksi dan anotasi (ground truth). Jika  $\text{IoU} \geq 50\%$ , prediksi dianggap True Positive (TP). Jika  $\text{IoU} < 50\%$ , maka dihitung sebagai False Positive, dan jika tidak ada prediksi pada ground truth object, maka dihitung sebagai False Negative. Singkatnya mAP50 mengukur seberapa baik model mendeteksi semua cacat biji dengan overlap minimal 50% antara area prediksi, dengan area pada data latih.

#### IV. HASIL DAN PEMBAHASAN

Hasil evaluasi dari kelima model yang dibuat dan rata-rata dari semua model dapat dilihat dalam Tabel 3(C). Rata-rata mean Average Precision pada 0,50 IoU (mAP50) hanya berjumlah 0,135, dengan model dari fold 1 yang memiliki mAP50 yang paling tinggi, berjumlah 0,140. Model yang dibuat memiliki waktu inferensi yang sangat cepat, dengan rata-rata waktu inferensi hanya 16 ms. Model fold 1 akan digunakan untuk analisis dalam performa dan keterbatasan model yang dibuat. Tabel 4(D) menampilkan metrik dari model fold 1 untuk semua jenis cacat yang ada pada dataset pengujian.

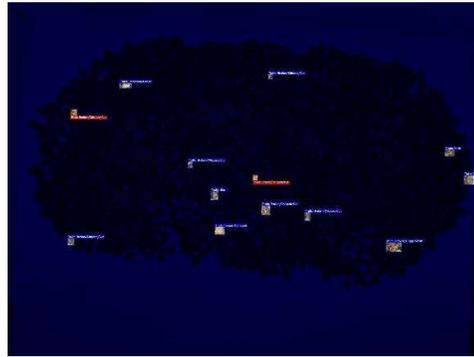
TABEL 3  
(B) HASIL PENGUJIAN MODEL PADA DATA TEST.

Fold	Precision	Recall	mAP <sub>50</sub>	Inference Time (ms)
1	<b>0,329</b>	<b>0,128</b>	<b>0,140</b>	<b>25,7</b>
2	0,155	0,137	0,139	12,8
3	0,333	0,109	0,132	11,4
4	0,133	0,140	0,128	15,2
5	0,419	0,128	0,136	14,3
<b>Rata-rata</b>	<b>0,274</b>	<b>0,128</b>	<b>0,135</b>	<b>16</b>

TABEL 4  
(C) HASIL PENGUJIAN MODEL 1 PADA DATA TEST.

Cacat	Precision	Recall	mAP <sub>50</sub>
all	0,329	0,128	0,140
Broken/Chipped/Cut	0,273	0,083	0,100
Dried Cherry/Pod	1,000	0,000	0,000
Full Black	0,512	0,591	0,609
Full Sour	0,220	0,104	0,147
Fungus Damaged	0,159	0,019	0,026
Hull/Husk	0,291	0,245	0,166
Immature/Unripe	0,510	0,327	0,385
Parchment/Pergamino	0,000	0,000	0,000
Partial Black	0,149	0,043	0,030
Partial Sour	0,688	0,178	0,225
Severe Insect Damage	0,258	0,085	0,105
Shell	0,327	0,047	0,069
Slight Insect Damage	0,224	0,063	0,091
Withered	0,000	0,000	0,000

Evaluasi model menggunakan dataset test yang berbeda dari dataset training. Gambar 9(I), Gambar 10(J), Gambar 11(K) menunjukkan hasil inferensi model fold 1 pada citra grade 1, 2, dan 3 (label biru: ground truth, merah: prediksi).



GAMBAR 9

(I) Hasil inferensi model fold 1 pada citra biji kopi grade 1, label biru adalah ground truth dan label merah adalah hasil inferensi model.



GAMBAR 10

(J) Hasil inferensi model fold 1 pada citra biji kopi grade 2, label biru adalah ground truth dan label merah adalah hasil inferensi model.

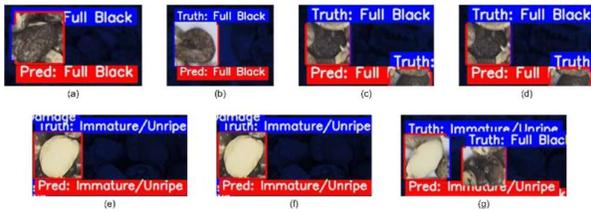


GAMBAR 11

(K) Hasil inferensi model fold 1 pada citra biji kopi grade 3, label biru adalah ground truth dan label merah adalah hasil inferensi model.

Tantangan utama adalah mendeteksi cacat yang kecil dan banyak dalam satu citra (Gambar 9(I), Gambar 10(J), Gambar 11(K)), yang membatasi konteks yang dapat dipelajari model [18]. Hal ini menyebabkan performa model rendah (Tabel 3(C), Tabel 4(D)), kemungkinan karena kurangnya variasi dataset training.

Model unggul dalam mendeteksi cacat dengan perbedaan warna signifikan, seperti full black (mAP50 0.609) dan immature/unripe (mAP50 0.385) (Tabel 4(D), Gambar 12(L)).



GAMBAR 12

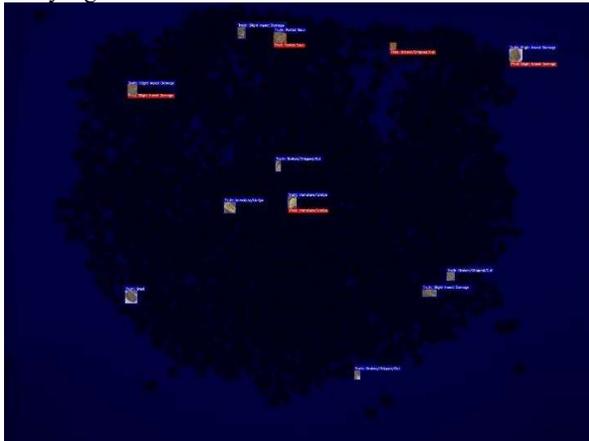
(L) Hasil inferensi model fold 1, label biru adalah ground truth dan label merah adalah hasil inferensi model: (a), (b), (c), (d) Inferensi akurat untuk cacat full black, (e), (f), (g) Inferensi akurat untuk cacat immature/unripe.



GAMBAR 13

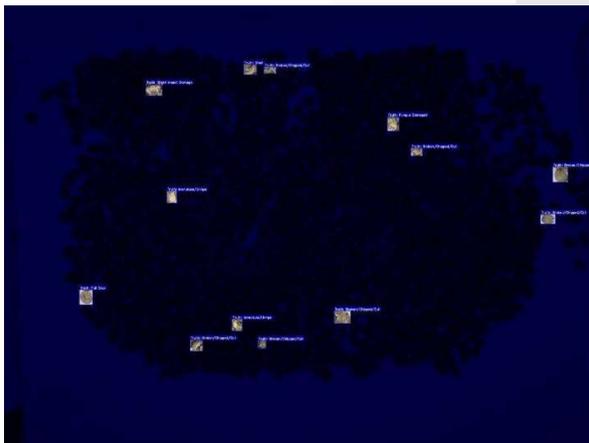
(M) Cacat yang tidak terdeteksi oleh model: (a) cacat slight insect damage, dengan tampilan titik hitam pada biji kopi, (b) cacat shell dengan tampilan biji kopi seperti kerang laut, (c) cacat broken/chipped/cut dengan tampilan setengah biji kopi penuh, (d) cacat broken/chipped/cut dengan tampilan potongan pada tengah biji kopi.

Model kesulitan mendeteksi cacat kecil seperti broken/chipped/cut, slight insect damage, dan shell (Gambar 13(M), Tabel 4(D)) karena kelainan bentuk atau titik hitam kecil yang sulit dikenali.



GAMBAR 14

(M) Hasil inferensi model fold 1 pada citra biji kopi grade 1, label biru adalah ground truth dan label merah adalah hasil inferensi model.

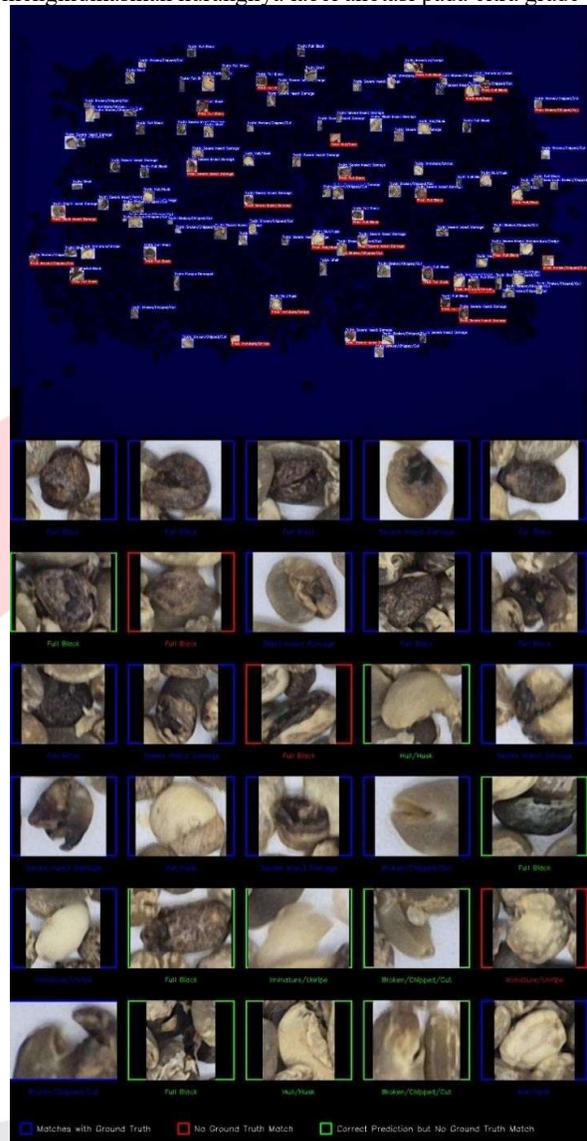


GAMBAR 15

(N) Sampel hasil inferensi yang tidak mengeluarkan prediksi.

Biji kopi grade 1 didominasi cacat broken/chipped/cut dan slight insect damage (Gambar 14(M)), menyebabkan model hanya mendeteksi sebagian cacat. Bahkan, beberapa

inferensi tidak menghasilkan prediksi (Gambar 15(O)), mengindikasikan kurangnya label anotasi pada citra grade 1.



GAMBAR 16

(P) Sampel analisis inferensi model fold 1 pada dataset pengujian grade 3, warna biru adalah deteksi yang cocok dengan ground truth, warna merah adalah deteksi yang tidak memiliki pasangan ground truth (False Positive) dan bukan deteksi akurat, warna hijau adalah deteksi yang tidak memiliki pasangan ground truth (False Positive) tetapi merupakan deteksi akurat.

Banyak deteksi False Positive (FP) pada grade 3 (Gambar 16(P), 17(Q)), grade 2 (Gambar 18(R), 19(S)) dan grade 1 sebenarnya akurat, tetapi tidak ada dalam anotasi ground truth. Ini menunjukkan kesulitan anotasi objek kecil dan banyak, serta potensi peningkatan performa model dengan anotasi yang lebih lengkap.



GAMBAR 17

(Q) Sampel analisis inferensi model fold 1 pada dataset pengujian grade 3, warna biru adalah deteksi yang cocok dengan ground truth, warna merah adalah deteksi yang tidak memiliki pasangan ground truth (False Positive) dan bukan deteksi akurat, warna hijau adalah deteksi yang tidak memiliki pasangan ground truth (False Positive) tetapi merupakan deteksi akurat.

Dari hasil yang ditampilkan pada Tabel 5(E), bisa dilihat bahwa setelah pengecekan mendalam pada hasil pengujian, 202 dari 290 FP sebenarnya adalah prediksi yang akurat yang dianggap salah yang diakibatkan oleh anotasi dataset yang kurang lengkap. Ini juga menunjukkan bahwa dari 542 prediksi model, didapatkan 454 deteksi yang akurat, yaitu 83,8% dari semua prediksi model merupakan prediksi yang benar, dan hanya 88 deteksi model yang salah.

TABEL 5

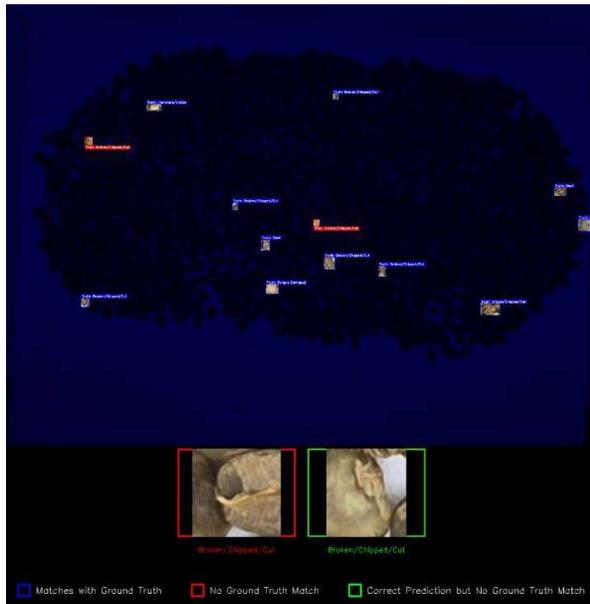
(D) TOTAL JUMLAH PREDIKSI PENGUJIAN, TRUE POSITIVE, FALSE POSITIVE, DAN FALSE POSITIVE (ACCURATE) YANG MENUNJUKKAN DETEKSI YANG TIDAK MEMILIKI PASANGAN GROUND TRUTH (FALSE POSITIVE) TETAPI MERUPAKAN DETEKSI AKURAT.

Grade	Jumlah Prediksi	True Positive	False Positive	False Positive (Accurate)	TP+FP(A)
Grade 1	43	20	23	11	31
Grade 2	182	70	112	78	148
Grade 3	317	162	155	113	275
Semua	542	252	290	202	454



GAMBAR 18

(R) Sampel analisis inferensi model fold 1 pada dataset pengujian grade 2, warna biru adalah deteksi yang cocok dengan ground truth, warna merah adalah deteksi yang tidak memiliki pasangan ground truth (False Positive) dan bukan deteksi akurat, warna hijau adalah deteksi yang tidak memiliki pasangan ground truth (False Positive) tetapi merupakan deteksi akurat.



GAMBAR 19

(S) Analisis inferensi model fold 1 pada dataset pengujian grade 1, warna biru adalah deteksi yang cocok dengan ground truth, warna merah adalah deteksi yang tidak memiliki pasangan ground truth (False Positive) dan bukan deteksi akurat, warna hijau adalah deteksi yang tidak memiliki pasangan ground truth (False Positive) tetapi merupakan deteksi akurat.

## V. KESIMPULAN

Penelitian ini mengusulkan implementasi YOLOv8 untuk mendeteksi cacat biji kopi arabika sesuai dengan standar SCA. Model yang dibuat relatif mahir dalam mendeteksi cacat yang memiliki perbedaan warna, tetapi memiliki kendala dalam mendeteksi jenis cacat yang hanya bersifat bentuk dari biji kopi dan tidak memiliki kelainan warna. Model yang dibuat pada penelitian ini menunjukkan performa yang kurang memuaskan. Beberapa alasan utama dari metrik yang rendah adalah objek deteksi yang berukuran kecil dan berjumlah banyak, dan kelengkapan dataset yang masih rendah. Namun, setelah pengecekan ulang secara manual hasil prediksi model pada dataset pengujian, 83,8% dari prediksi model adalah prediksi yang benar.

Untuk penelitian selanjutnya, diharapkan untuk melakukan pendalaman lebih lanjut dengan menggunakan metode pemrosesan yang berbeda, maupun menggunakan model dengan arsitektur yang berbeda, serta mengembangkan dataset, baik dengan menambah citra tambahan agar variasi citra lebih baik, augmentasi data dengan sintesis, dan mengecek kelengkapan anotasi.

## REFERENSI

- [1] International Coffee Organization. Coffee report and outlook. Technical report, International Coffee Organization, 2023.
- [2] SCAA green coffee beans classification, <http://www.coffeeresearch.org/coffee/scaaclass.htm>, 2006. Accessed: 2024-04-01.
- [3] James Kosalos, Rob Stephen, Steven Diaz, Paul Songer, Man'e Alves, Marty Curtis, and Steven Sung-

- Yong Kil. Arabica Green Coffee Defect Handbook. Specialty Coffee Association of America, 2013.
- [4] Nen-Fu Huang, Dong-Lin Chou, Chia-An Lee, Feng-Ping Wu, An-Chi Chuang, Yi-Hsien Chen, and Yin-Chun Tsai. Smart agriculture: real-time classification of green coffee beans by using a convolutional neural network. *IET Smart Cities*, 2, 10 2020.
- [5] Serawork Walleign, Mihai Polceanu, Towfik Jemal, and C'edric Buche. Coffee grading with convolutional neural networks using small datasets with high variance. 2019.
- [6] Edwin R Arboleda, Arnel C Fajardo, and Ruji P Medina. Classification of coffee bean species using image processing, artificial neural network and k nearest neighbors. In 2018 IEEE international conference on innovative research and development (ICIRD), pages 1–5. IEEE, 2018.
- [7] JR Arunkumar and Tagele berihun Mengist. Developing ethiopian yirga-cheffe coffee grading model using a deep learning classifier. *International Journal of Innovative Technology and Exploring Engineering*, 9(4):3303–3309, 2020.
- [8] Muhammad NS Akbar, Ema Rachmawati, and Febryanti Sthevanie. Visual feature and machine learning approach for arabica green coffee beans grade determination. In *Proceedings of the 6th International Conference on Communication and Information Processing*, pages 97–104, 2020.
- [9] Edwin R. Arboleda, Arnel C. Fajardo, and Ruji P. Medina. An image processing technique for coffee black beans identification. In 2018 IEEE International Conference on Innovative Research and Development (ICI-RD), pages 1–5, 2018.
- [10] Shih-Yu Chen, Chuan-Yu Chang, Cheng-Syue Ou, and Chou-Tien Lien. Detection of insect damage in green coffee beans using vis-nir hyperspectral imaging. *Remote Sensing*, 12(15), 2020.
- [11] Carlito Pinto, Junya Furukawa, Hidekazu Fukai, and Satoshi Tamura. Classification of green coffee bean images based on defect types using convolutional neural network (cnn). In 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), pages 1–5, 2017.
- [12] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO. <https://github.com/ultralytics/ultralytics>, 2023. Accessed: 2023-04-11.
- [13] Juan Terven, Diana-Margarita C'ordova-Esparza, and Julio-Alejandro Romero-Gonz'alez. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, 5(4):1680–1716, 2023.
- [14] Palmiro Poltronieri and Franca Rossi. Challenges in specialty coffee processing and quality assurance. *Challenges*, 7(2), 2016.
- [15] Amanda R Hale, Paul M Ruegger, Philippe Rolshausen, James Borneman, and Jiue-in Yang. Fungi associated with the potato taste defect in coffee beans from rwanda. *Botanical studies*, 63(1):17, 2022.
- [16] Rafael Carlos Eloy Dias, Sebastian Ed Wieland Opitz, and Chahan Ye-retzian. Bioactive compounds in blends

- of coffee defects originating from the harvesting. *Coffee Science* - ISSN 1984-3909, 17:e172027, Jan. 2023.
- [17] Coffee Research Institute. SCAA Coffee Beans Classification. <http://www.coffeeresearch.org/coffee/scaaclass.htm>. Accessed: 2023-04-11.
- [18] Liu, Y., Sun, P., Wergeles, N., & Shang, Y. (2021). A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications*, 172, 114602.
- [19] S. Yadav and S. Shukla, Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification, 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, 2016, pp. 78-83, doi: 10.1109/IACC.2016.25.
- [20] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1), 1-48.
- [21] J. Solawetz, What is Mean Average Precision (mAP) in Object Detection?, Roboflow, pp. 1–11, 2020.

