JURTI, Vol.10 No.3, September 2026, ISSN: 2579-8790

Deteksi dan Klasifikasi Penyakit Daun Jeruk Menggunakan Model *You Only Look Once*

Ummul Hairah*[1], Andi Tejawati[2], Anindita Septiarini[3], Masna Wati[4], Novianti Puspitasari[5], Diky Dwicandra[6], Faza Alameka[7]
[1,2,3,4,5,6] Program Studi Informatika, Universitas Mulawarman, Samarinda
[7]Program Studi Sistem Informasi, Universitas Mulia, Samarinda
e-mail: [1]*ummulhairah@unmul.ac.id, [2]anditejawati@unmul.ac.id, [3]anindita@unmul.ac.id, [4]masnawati@unmul.ac.id, [5]novipuspitasari@unmul.ac.id, [6]dikydwicandra9c@gmail.com, [7]faza.alameka@gmail.com

Abstrak

Identifikasi penyakit daun jeruk secara cepat diperlukan untuk mendukung pengendalian penyakit sebelum gejala berkembang lebih luas. Penelitian ini mengevaluasi YOLOv8n, YOLOv9t, YOLOv10n, dan YOLO11n pada dataset Citrus Common Leaf 2020 (CCL'20) yang terdiri atas 2.684 citra dan 10.332 anotasi. Tiga kelas penyakit, yaitu Anthracnose, Bacterial Brown Spot, dan Melanose, dievaluasi pada level objek, sedangkan 427 citra Healthy diperlakukan sebagai citra latar dan dianalisis bersama ketiga kelas penyakit pada level citra. Pengujian menggunakan tiga rasio pembagian data, yaitu 80:10:10, 70:15:15, dan 60:20:20. Pada skenario 80:10:10, YOLOv10n memperoleh mAP50-95 sebesar 0,9373 dan YOLOv9t sebesar 0,9368. Kedua nilai tersebut tergolong tinggi dan kompetitif dalam konteks benchmark CCL'20 karena melampaui YOLOv7 (0,9260), walaupun masih di bawah YOLOv8s teroptimasi (0,9610) pada studi lain; perbandingan ini bersifat kontekstual karena konfigurasi eksperimen berbeda. YOLOv9t menunjukkan rentang performa paling kecil pada tiga skenario, sedangkan YOLOv10n mencapai 151,7 FPS dengan waktu inferensi 6,59 ms per citra. Evaluasi level citra menghasilkan akurasi 85,13%, dengan seluruh citra Healthy teridentifikasi secara tepat. Hasil penelitian menunjukkan bahwa peningkatan generasi YOLO tidak selalu menghasilkan peningkatan akurasi yang konsisten, sehingga pemilihan model perlu mempertimbangkan akurasi, kestabilan, dan efisiensi komputasi secara bersamaan.

Kata kunci*— CCL'20, Daun Jeruk, Deteksi Penyakit, Deep Learning, YOLO*

1. PENDAHULUAN

Tanaman jeruk merupakan komoditas hortikultura penting yang produktivitasnya dapat menurun akibat penyakit daun [1]. Gejala seperti perubahan warna, bercak, dan lesi sulit dibedakan karena variasi ukuran, keparahan, pencahayaan, latar, dan kemiripan visual antarpenyakit [2]. Visi komputer dan deep learning menawarkan analisis citra yang lebih objektif dengan mempelajari fitur warna, tekstur, bentuk, dan pola spasial langsung dari data [3], [4]. Pada penyakit tanaman, object detection memberikan kelas sekaligus lokasi gejala melalui bounding box [5]-[8], dan telah diterapkan pada sayuran, apel, anggur, dan teh untuk menangani objek kecil, latar kompleks, serta keterbatasan komputasi [9]-[11]. Riset terkini menekankan keseimbangan akurasi dan efisiensi, melalui model ringan, mekanisme perhatian, dan fusi multi-skala, sehingga kecepatan inferensi dan ukuran model perlu dievaluasi bersama akurasi [12]-[16]. Arsitektur YOLO terus berkembang: YOLOv7 memperkenalkan trainable bag-of-freebies [17], YOLOv8 menggunakan ekosistem anchor-free [18], YOLOv9 mengusulkan Programmable Gradient Information dan GELAN [19], YOLOv10 menerapkan consistent dual assignments untuk inferensi end-to-end tanpa NMS [20], dan YOLO11 melanjutkan optimasi efisiensi [21].

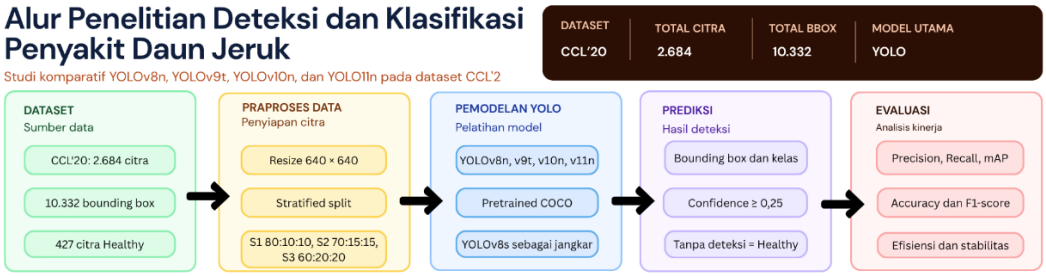
JURTI, Vol.10 No.3, September 2026, ISSN: 2579-8790

Peningkatan pada benchmark umum belum tentu berulang pada penyakit tanaman, sehingga perlu diuji pada domain spesifik. CCL'20 menyediakan anotasi penyakit daun jeruk tingkat objek dan menjadi benchmark bagi beberapa detektor [1], dengan mAP50-95 tinggi pada keluarga YOLO setelah augmentasi dan penyetelan hiperparameter [22]. Namun, perbandingan seragam YOLOv9, YOLOv10, dan YOLO11 dalam konfigurasi ringan belum tersedia [1], [22]. Citra Healthy tidak memiliki bounding box [1] sehingga dapat menjadi contoh negatif saat pelatihan, tetapi evaluasi level citra perlu dipisahkan dari evaluasi bounding box, terutama pada citra dengan banyak lesi atau beberapa label [23].

Penelitian ini membandingkan YOLOv8n, YOLOv9t, YOLOv10n, dan YOLO11n pada CCL'20 dengan tiga rasio pembagian data, serta mengevaluasi level objek, level citra, dan efisiensi komputasi. Satu run YOLOv8s digunakan sebagai pembanding terhadap benchmark terdahulu [22]. Tujuannya menilai konsistensi keuntungan antargenerasi serta kompromi antara ketepatan, kestabilan, ukuran bobot, dan kecepatan inferensi. Kestabilan didefinisikan sebagai konsistensi mAP50-95 antarskenario, dinilai secara deskriptif melalui rentang (selisih maksimum dan minimum pada S1, S2, S3); rentang yang lebih kecil menunjukkan perubahan performa yang lebih kecil saat proporsi data latih berubah.

2. METODE PENELITIAN

Penelitian terdiri atas lima tahap: penyiapan dataset, praproses, pemodelan, pascaproses prediksi, dan evaluasi. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian

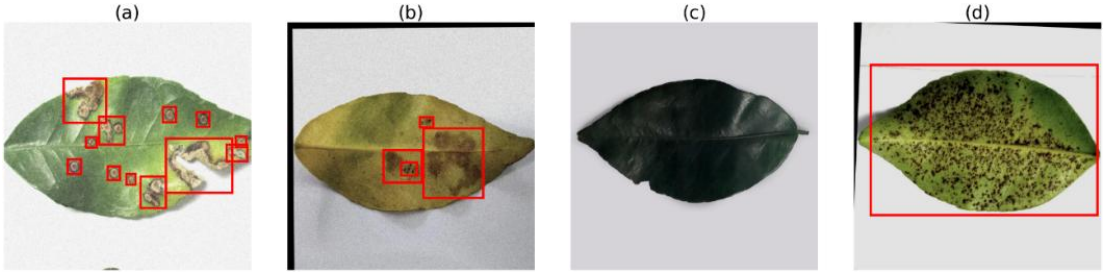
2.1. Dataset

CCL'20 memuat 2.684 citra (2.257 berpenyakit dan 427 Healthy) dengan 10.332 bounding box pada kelas Anthracnose, Bacterial Brown Spot, dan Melanose [1]. Distribusinya adalah 1.443 citra Anthracnose, 481 Bacterial Brown Spot, 333 Melanose, dan 427 Healthy; 518 citra berpenyakit memuat lebih dari satu jenis penyakit [1]. Label citra multi-penyakit untuk stratifikasi dan evaluasi level citra ditetapkan dari kelas bounding box dominan, sedangkan seluruh bounding box tetap dipakai pada evaluasi level objek. Dua sumber anotasi resmi dinormalisasi sebelum parsing karena sintaks JSON tidak sepenuhnya sesuai, dan pemeriksaan akhir tidak menemukan bounding box invalid atau keluar batas citra (Tabel 1, Gambar 2).

Tabel 1. Distribusi dataset CCL'20

Kelas	Jumlah Citra	Persentase	Total Bounding Box
Anthracnose	1.443	53,76%	7.927
Bacterial Brown Spot	481	17,92%	1.998
Healthy	427	15,91%	0
Melanose	333	12,41%	407
Total	2.684	100%	10.332

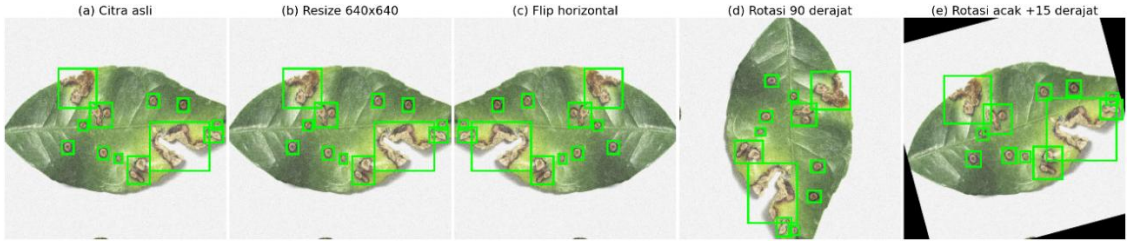
JURTI, Vol.10 No.3, September 2026, ISSN: 2579-8790



Gambar 2. Contoh citra CCL'20 pada kelas (a) Anthracnose, (b) Bacterial Brown Spot, (c) Healthy, dan (d) Melanose

2.2. Praproses Data

Citra diseragamkan menjadi 640 × 640 piksel dan anotasi COCO dikonversi ke format YOLO ternormalisasi, sedangkan citra Healthy dipertahankan sebagai latar berlabel kosong untuk contoh negatif. Augmentasi memakai konfigurasi bawaan framework dan hanya diterapkan pada data latih. Data dibagi secara stratified (seed 0) menjadi S1 80:10:10, S2 70:15:15, dan S3 60:20:20 untuk menjaga proporsi kelas dan mengamati kestabilan saat data latih berkurang (Gambar 3, Tabel 2).



Gambar 3. Contoh hasil praproses citra dan transformasi bounding box pada tahap inspeksi data

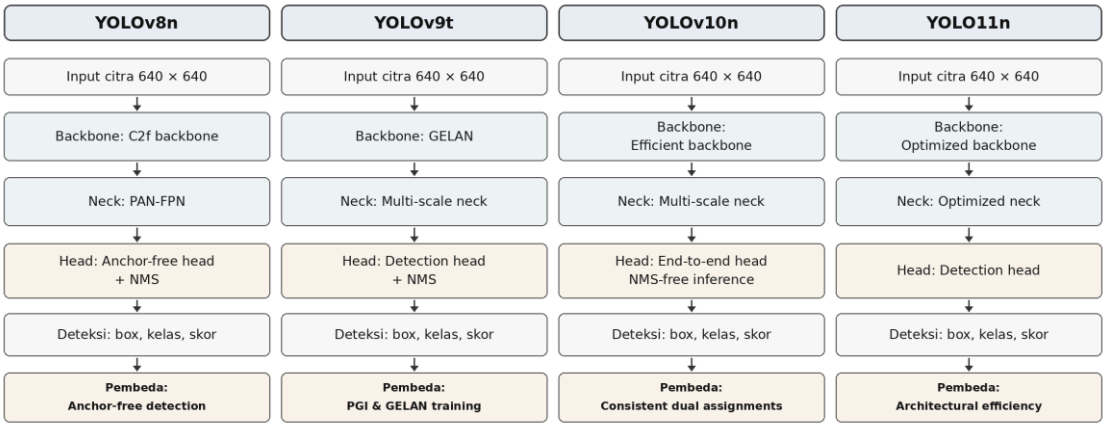
Tabel 2. Pembagian dataset pada tiga skenario eksperimen

Skenario	Rasio	Training	Validation	Testing	Total
S1	80:10:10	2.147	268	269	2.684
S2	70:15:15	1.878	403	403	2.684
S3	60:20:20	1.610	537	537	2.684

2.3. Pemodelan

Empat varian ringan yang diuji adalah YOLOv8n, YOLOv9t, YOLOv10n, dan YOLO11n [18]-[21], seluruhnya dengan bobot pra-latih COCO, masukan 640 × 640 piksel, batch size 16, maksimum 100 epoch (patience 20), seed 0, dan optimizer otomatis pada GPU NVIDIA T4. Tiga skenario per model menghasilkan 12 run utama, ditambah satu run YOLOv8s pada S1 sebagai pembanding skala terhadap benchmark terdahulu [22]. Arsitektur asli dipertahankan agar perbandingan berfokus pada perilaku model dalam konfigurasi seragam. Perbedaan utamanya (Gambar 4) YOLOv8n memakai kepala deteksi anchor-free, YOLOv9t berbasis GELAN dan PGI, YOLOv10n memakai consistent dual assignments untuk inferensi end-to-end tanpa NMS, dan YOLO11n mengoptimalkan efisiensi arsitektur [18]-[21].

JURTI, Vol.10 No.3, September 2026, ISSN: 2579-8790



Gambar 4. Diagram pemodelan komparatif YOLOv8n, YOLOv9t, YOLOv10n, dan YOLO11n

2.4. Pascaproses

Pascaproses menghasilkan koordinat bounding box, kelas, dan confidence score. Prediksi model konvensional disaring menggunakan NMS bawaan framework, sedangkan YOLOv10 mempertahankan inferensi end-to-end dengan consistent dual assignments [20]. Pada evaluasi level citra, citra tanpa deteksi penyakit pada confidence threshold 0,25 dikategorikan Healthy; jika terdapat deteksi, kelas ditentukan dari confidence tertinggi. Aturan ini merupakan evaluasi turunan detektor, bukan pengganti klasifikasi citra khusus [23].

2.5. Evaluasi

Evaluasi dilakukan pada level objek dan citra. Level objek menggunakan Intersection over Union (IoU), precision, recall, mAP50, dan mAP50-95. IoU mengukur tumpang tindih bounding box prediksi dan ground truth, sedangkan precision dan recall mengukur ketepatan prediksi positif dan kemampuan menemukan objek. IoU, precision, recall, dan AP berturut-turut ditunjukkan pada Persamaan (1), (2), (3), dan (4).

$$IoU = \frac{B_p \cap B_{gt}}{B_p \cup B_{gt}} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$AP = \int_0^1 P\,(R) dR \tag{4}$$

Average Precision merupakan luas kurva precision-recall tiap kelas. mAP merata-ratakan AP antarkelas, sedangkan mAP50-95 menghitung rata-rata mAP pada IoU 0,50-0,95 dengan interval 0,05. Formulasinya ditunjukkan pada Persamaan (5).

$$mAP = \frac{1}{N} \sum_{i=1}^{N} A\,P_i \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

JURTI, Vol.10 No.3, September 2026, ISSN: 2579-8790

Evaluasi level citra menggunakan accuracy, precision, recall, F1-score, dan confusion matrix. Accuracy dihitung dari proporsi prediksi diagonal confusion matrix sebagaimana Persamaan (6), sedangkan F1-score dari precision dan recall sebagaimana Persamaan (7).

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Efisiensi dinilai melalui ukuran bobot, waktu inferensi, FPS, dan waktu pelatihan. Karena mAP50 hampir seluruhnya mencapai 0,995, mAP50-95 digunakan sebagai indikator pembeda utama. Setiap konfigurasi dijalankan sekali dengan seed tetap sehingga hasil ditafsirkan secara deskriptif, bukan sebagai estimasi statistik antar-run

3. HASIL DAN PEMBAHASAN

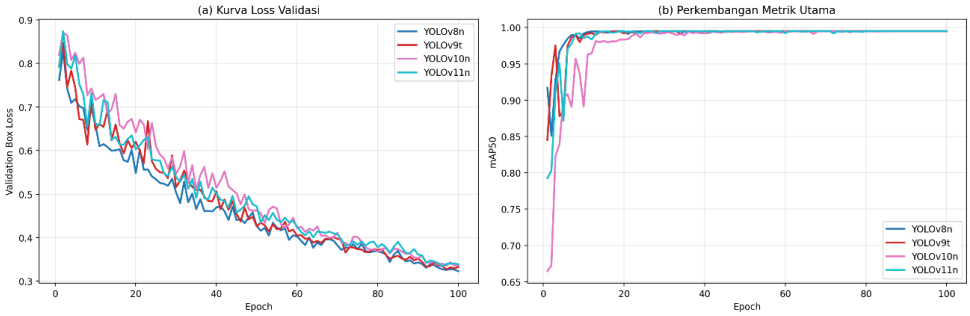
3.1. Perbandingan Kinerja Deteksi

Pada S1, seluruh varian ringan mencapai mAP50 sekitar 0,995 sehingga metrik ini kurang sensitif membedakan model. mAP50-95 memberikan pemisahan lebih jelas: YOLOv10n mencapai 0,9373, YOLOv9t 0,9368, YOLO11n 0,9337, dan YOLOv8n 0,9335. Selisih YOLOv10n dan YOLOv9t hanya 0,0005 atau 0,05 poin persentase, sehingga kinerjanya lebih tepat dinilai sebanding daripada berbeda substantif. Rangkuman S1 disajikan pada Tabel 3.

Tabel 3. Perbandingan kinerja model pada skenario S1 80:10:10

Model	Precision	Recall	mAP50	mAP50-95
YOLOv8n	0,9991	0,9977	0,9950	0,9335
YOLOv9t	0,9994	0,9986	0,9950	0,9368
YOLOv10n	0,9941	0,9898	0,9950	0,9373
YOLO11n	0,9990	0,9981	0,9950	0,9337

Kurva pelatihan menunjukkan validation box loss menurun bertahap, sedangkan mAP50 meningkat cepat lalu bertahan sekitar 0,995. Pola konvergensi serupa menunjukkan seluruh model mempelajari objek penyakit tanpa indikasi kegagalan pelatihan. Namun, mAP tinggi tetap perlu ditafsirkan bersama struktur, distribusi kelas, dan kemiripan visual data karena faktor tersebut memengaruhi generalisasi [3], [4]. Kurva ditunjukkan pada Gambar 5.



Gambar 5. Kurva validation box loss dan perkembangan mAP50 pada skenario S1

3.2. Kinerja per Kelas dan Evaluasi Level Citra

Perbedaan AP antarkelas lebih tampak pada AP50-95 daripada AP50. Anthracnose memiliki AP50-95 terendah pada seluruh model meskipun jumlah citra dan bounding box-nya terbesar, sedangkan Melanose yang paling sedikit datanya justru tertinggi pada beberapa model. Hal ini menunjukkan kesulitan deteksi tidak hanya ditentukan ukuran sampel, tetapi juga bentuk lesi, skala, latar, dan karakter visual [12], [13]. Nilai AP dirangkum pada Tabel 4.

JURTI, Vol.10 No.3, September 2026, ISSN: 2579-8790

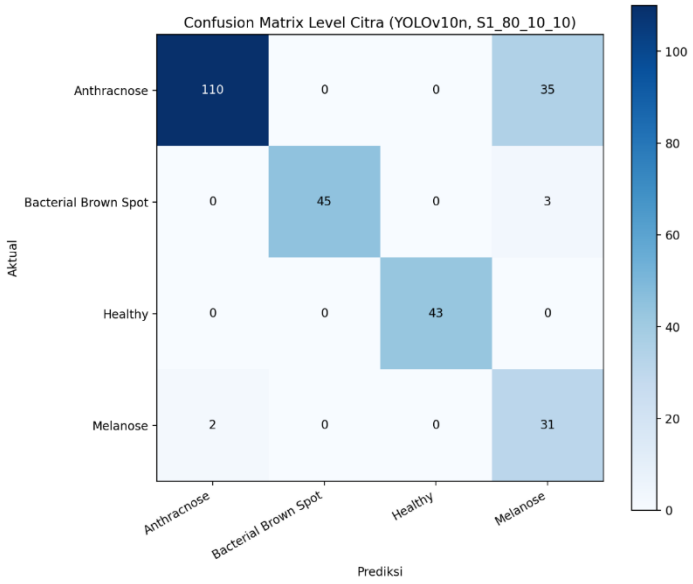
Tabel 4. Average Precision setiap kelas penyakit pada skenario S1

Model	AP50			AP50-95		
	Anthracnose	BBS	Melanose	Anthracnose	BBS	Melanose
YOLOv8n	0,9950	0,9950	0,9950	0,9105	0,9368	0,9532
YOLOv9t	0,9950	0,9950	0,9950	0,9101	0,9525	0,9477
YOLOv10n	0,9949	0,9949	0,9950	0,9070	0,9505	0,9545
YOLO11n	0,9950	0,9950	0,9950	0,9092	0,9366	0,9552

Evaluasi level citra YOLOv10n S1 menghasilkan accuracy 0,8513 pada 269 citra. Healthy memperoleh precision, recall, dan F1-score 1,0000; Bacterial Brown Spot memiliki F1-score 0,9677, Anthracnose 0,8560, dan Melanose 0,6078. F1-score Melanose terutama dibatasi precision 0,4493 meskipun recall 0,9394. Hasil ini menunjukkan bahwa deteksi objek tinggi tidak selalu menghasilkan keputusan level citra yang setara ketika kelas ditentukan dari confidence tertinggi [8], [23]. Confusion matrix ditunjukkan pada Gambar 6 dan metriknya pada Tabel 5.

Tabel 5. Hasil evaluasi level citra pada YOLOv10n skenario S1

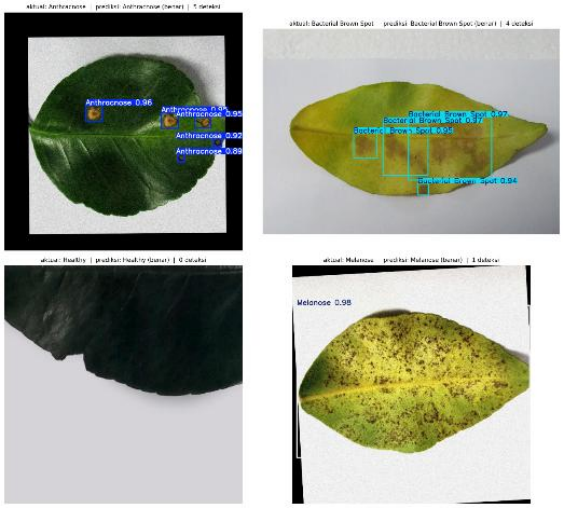
Kelas	Precision	Recall	F1-score	Support
Anthracnose	0,9821	0,7586	0,8560	145
Bacterial Brown Spot	1,0000	0,9375	0,9677	48
Healthy	1,0000	1,0000	1,0000	43
Melanose	0,4493	0,9394	0,6078	33
Accuracy			0,8513	269
Macro average	0,8579	0,9089	0,8579	269
Weighted average	0,9228	0,8513	0,8685	269



Gambar 6. Confusion matrix evaluasi level citra YOLOv10n pada skenario S1

Confusion matrix menunjukkan 31 dari 33 citra Melanose dikenali benar, tetapi 35 Anthracnose dan tiga Bacterial Brown Spot diprediksi sebagai Melanose; dua Melanose diprediksi sebagai Anthracnose. Dengan demikian, kelemahan utama Melanose adalah false positive, bukan kegagalan deteksi. Seluruh 43 citra Healthy tidak menghasilkan deteksi penyakit pada threshold yang digunakan. Contoh deteksi ditunjukkan pada Gambar 7.

JURTI, Vol.10 No.3, September 2026, ISSN: 2579-8790



Gambar 7. Contoh hasil deteksi pada kelas Anthracnose, Bacterial Brown Spot, Healthy, dan Melanose

3.3. Efisiensi Model

Efisiensi model menunjukkan karakter berbeda. YOLOv10n memiliki inferensi tercepat, 6,59 ms per citra atau 151,7 FPS; YOLO11n mencapai 124,5 FPS, YOLOv8n 113,9 FPS, dan YOLOv9t 104,6 FPS. YOLOv9t memiliki bobot terkecil 4,65 MB, sedangkan YOLOv10n 5,76 MB. Ukuran bobot tidak selalu sejalan dengan latency karena operasi internal dan strategi inferensi juga memengaruhi efisiensi [16], [20]. Ringkasannya disajikan pada Tabel 6.

Tabel 6. Perbandingan efisiensi model pada skenario S1

Model	Ukuran Bobot (MB)	Inferensi (ms)	FPS	Waktu Latih (menit)
YOLOv10n	5,76	6,59	151,7	51,7
YOLO11n	5,48	8,03	124,5	42,6
YOLOv8n	6,26	8,78	113,9	40,4
YOLOv9t	4,65	9,56	104,6	59,3
YOLOv8s	22,53	13,81	72,4	64,0

YOLOv8s menghasilkan mAP50-95 0,9433, sekitar 0,60 poin persentase di atas YOLOv10n, tetapi bobotnya 22,53 MB dan inferensinya 13,81 ms. Dibandingkan YOLOv10n, kebutuhan penyimpanan hampir empat kali dan inferensi sekitar dua kali lebih lambat. Karena itu, peningkatan akurasi kecil pada implementasi real-time perlu dipertimbangkan bersama biaya komputasi dan keterbatasan perangkat [6], [16].

3.4. Kestabilan pada Variasi Pembagian Data

mAP50-95 menunjukkan perubahan peringkat ketika proporsi data latih berkurang. YOLOv10n terbaik pada S1, sedangkan YOLOv9t terbaik pada S2 dan S3. Rentang YOLOv9t paling kecil, yaitu 0,0092, diikuti YOLOv8n 0,0117, YOLO11n 0,0139, dan YOLOv10n 0,0149. Secara deskriptif, YOLOv9t paling stabil terhadap perubahan rasio data, tetapi hasil ini bukan uji signifikansi karena setiap konfigurasi hanya menggunakan satu seed. Nilai lintas skenario disajikan pada Tabel 7.

Tabel 7. Kestabilan mAP50-95 pada tiga skenario pembagian data

Model	S1 80:10:10	S2 70:15:15	S3 60:20:20	Rentang
YOLOv8n	0,9335	0,9250	0,9218	0,0117
YOLOv9t	0,9368	0,9315	0,9276	0,0092
YOLOv10n	0,9373	0,9233	0,9224	0,0149
YOLO11n	0,9337	0,9258	0,9197	0,0139

JURTI, Vol.10 No.3, September 2026, ISSN: 2579-8790

Pengurangan data latih dari 2.147 citra pada S1 menjadi 1.610 pada S3 menurunkan mAP50-95 seluruh model dengan besar yang berbeda. Perubahan peringkat menunjukkan bahwa satu skenario belum cukup untuk menyatakan superioritas generasi tertentu. Pada data tidak seimbang dan bervariasi secara visual, interpretasi performa perlu mempertimbangkan struktur data selain metrik agregat [4].

3.5. Perbandingan dengan Penelitian Terdahulu

YOLOv10n penelitian ini mencapai mAP50-95 0,9373, lebih tinggi daripada Scaled YOLOv4 P7 0,8930, CenterNet2 0,9140, dan YOLOv7 0,9260 pada benchmark CCL'20 [1]. YOLOv8s penelitian ini mencapai 0,9433, sedangkan studi terdahulu dengan augmentasi dan penyetelan hiperparameter melaporkan 0,9610 [22]. Perbandingan ini bersifat kontekstual karena pembagian data, augmentasi, dan konfigurasi pelatihan tidak sepenuhnya identik [1], [22]. Rangkuman benchmark disajikan pada Tabel 8.

Tabel 8. Perbandingan kinerja dengan benchmark terdahulu pada CCL'20

Pendekatan	Varian	mAP50-95
Scaled YOLOv4 P7	-	0,8930
CenterNet2	-	0,9140
YOLOv7	-	0,9260
YOLOv5s	s	0,9420
YOLOv8s	s	0,9610
YOLOv8s, penelitian ini	s	0,9433
YOLOv10n, penelitian ini	n	0,9373

Pemeriksaan kemiripan menghasilkan 2.269 klaster dari 2.684 citra pada ambang perceptual hash 5, dengan rasio inflasi 1,183. Temuan ini menunjukkan redundansi visual. Karena pembagian masih berbasis stratifikasi kelas, citra sangat mirip berpotensi muncul pada subset berbeda. Potensi tersebut perlu dipertimbangkan saat menafsirkan mAP tinggi dan memperkuat kebutuhan pengujian pada citra lapangan independen [3], [4]. Secara keseluruhan, generasi YOLO yang lebih baru tidak selalu memberikan akurasi tertinggi. YOLOv10n unggul pada mAP50-95 S1 dan kecepatan inferensi, sedangkan YOLOv9t lebih stabil terhadap variasi pembagian data; YOLO11n tidak melampaui seluruh pendahulunya. Karena itu, pemilihan model sebaiknya berbasis kebutuhan domain dan implementasi, bukan urutan generasi.

4. KESIMPULAN

Keempat model, YOLOv8n, YOLOv9t, YOLOv10n, dan YOLO11n, mencapai mAP50 sekitar 0,995 dengan mAP50-95 pada skenario S1 (80:10:10; total 2.684 citra) masing-masing 0,9335, 0,9368, 0,9373, dan 0,9337. YOLOv10n mencatat mAP50-95 tertinggi (0,9373) dan inferensi tercepat (151,7 FPS), tetapi hanya unggul 0,05 poin persentase atas YOLOv9t. YOLOv9t terbaik pada S2 (70:15:15) dan S3 (60:20:20) dengan mAP50-95 0,9315 dan 0,9276, serta memiliki rentang antarskenario terkecil (0,0092), yang bersifat deskriptif dan belum dibuktikan secara statistik antar-seed. Evaluasi level citra mencapai akurasi 85,13%, dengan seluruh citra Healthy dikenali tepat dan kesalahan terutama terjadi antara Anthracnose dan Melanose. Dengan demikian, peningkatan generasi YOLO tidak menunjukkan tren performa linier, sehingga pemilihan model perlu menimbang akurasi, kestabilan, dan efisiensi komputasi.

5. SARAN

Penelitian lanjutan perlu menggunakan beberapa seed dan pengulangan untuk menganalisis variasi kinerja secara statistik, serta pembagian berbasis kelompok citra serupa guna mengurangi kebocoran visual. Evaluasi level citra dapat dikembangkan melalui confidence threshold berbasis validasi dan pendekatan multilabel. Generalisasi perlu diuji pada citra lapangan independen dengan variasi pencahayaan, latar, perangkat, dan tingkat keparahan, sedangkan

model ringan perlu diuji pada perangkat edge atau mobile untuk menilai latensi, memori, dan kestabilan inferensi.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Fakultas Teknik Universitas Mulawarman atas dukungan pendanaan dan fasilitas yang diberikan dalam pelaksanaan penelitian berjudul "Pengembangan Model Deteksi Penyakit Daun Jeruk Berbasis Deep Learning Menggunakan You Only Look Once (YOLO)" berdasarkan Kontrak Penelitian Nomor 12748/UN17.9/PT.00.03/2026. Penulis juga mengucapkan terima kasih kepada pengembang dataset Citrus Common Leaf 2020 (CCL'20) atas penyediaan citra dan anotasi yang digunakan dalam penelitian ini.

DAFTAR PUSTAKA

[1] S. Dananjayan, Y. Tang, J. Zhuang, C. Hou, and S. Luo, "Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images," Comput. Electron. Agric., vol. 193, art. no. 106658, 2022. https://doi.org/10.1016/j.compag.2021.106658.

[2] X. Qiu et al., "Detection of citrus diseases in complex backgrounds based on image-text multimodal fusion and knowledge assistance," Front. Plant Sci., vol. 14, art. no. 1280365, 2023. https://doi.org/10.3389/fpls.2023.1280365.

[3] M. Shoaib et al., "An advanced deep learning models-based plant disease detection: A review of recent research," Front. Plant Sci., vol. 14, art. no. 1158933, 2023. https://doi.org/10.3389/fpls.2023.1158933.

[4] I. Pacal et al., "A systematic review of deep learning techniques for plant diseases," Artif. Intell. Rev., vol. 57, art. no. 304, 2024. https://doi.org/10.1007/s10462-024-10944-7.

[5] J. Li, Y. Qiao, S. Liu, J. Zhang, Z. Yang, and M. Wang, "An improved YOLOv5-based vegetable disease detection method," Comput. Electron. Agric., vol. 202, art. no. 107345, 2022. https://doi.org/10.1016/j.compag.2022.107345.

[6] Y. Wang, Y. Wang, and J. Zhao, "MGA-YOLO: A lightweight one-stage network for apple leaf disease detection," Front. Plant Sci., vol. 13, art. no. 927424, 2022. https://doi.org/10.3389/fpls.2022.927424.

[7] Z. Zhang, Y. Qiao, Y. Guo, and D. He, "Deep learning based automatic grape downy mildew detection," Front. Plant Sci., vol. 13, art. no. 872107, 2022. https://doi.org/10.3389/fpls.2022.872107.

[8] Y. Peng and Y. Wang, "Leaf disease image retrieval with object detection and deep metric learning," Front. Plant Sci., vol. 13, art. no. 963302, 2022. https://doi.org/10.3389/fpls.2022.963302.

[9] M. J. A. Soeb et al., "Tea leaf disease detection and identification based on YOLOv7 (YOLO-T)," Sci. Rep., vol. 13, art. no. 6078, 2023. https://doi.org/10.1038/s41598-023-33270-4.

[10] S. Zhu, W. Ma, J. Wang, M. Yang, Y. Wang, and C. Wang, "EADD-YOLO: An efficient and accurate disease detector for apple leaf using improved lightweight YOLOv5," Front. Plant Sci., vol. 14, art. no. 1120724, 2023. https://doi.org/10.3389/fpls.2023.1120724.

[11] W. Xu and R. Wang, "ALAD-YOLO: An lightweight and accurate detector for apple leaf diseases," Front. Plant Sci., vol. 14, art. no. 1204569, 2023. https://doi.org/10.3389/fpls.2023.1204569.

[12] X. Wang and J. Liu, "Vegetable disease detection using an improved YOLOv8 algorithm in the greenhouse plant environment," Sci. Rep., vol. 14, art. no. 4261, 2024. https://doi.org/10.1038/s41598-024-54540-9.

[13] Y. Wang, P. Zhang, and S. Tian, "Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion," Front. Plant Sci., vol. 15, art. no. 1382802, 2024. https://doi.org/10.3389/fpls.2024.1382802.

[14] T. Li, L. Zhang, and J. Lin, "Precision agriculture with YOLO-Leaf: Advanced methods for detecting apple leaf diseases," Front. Plant Sci., vol. 15, art. no. 1452502, 2024. https://doi.org/10.3389/fpls.2024.1452502.

[15] J. Li et al., "Potato late blight leaf detection in complex environments," Sci. Rep., vol. 14, art. no. 31046, 2024. https://doi.org/10.1038/s41598-024-82272-3.

[16] W. Feng, J. Liu, Z. Li, and S. Lyu, "YOLO-Citrus: A lightweight and efficient model for citrus leaf disease detection in complex agricultural environments," Front. Plant Sci., vol. 16, art. no. 1668036, 2025. https://doi.org/10.3389/fpls.2025.1668036.

[17] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," arXiv:2207.02696, 2022. https://arxiv.org/abs/2207.02696.

[18] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," software, ver. 8.0.0, 2023. https://docs.ultralytics.com/models/yolov8/.

[19] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," arXiv:2402.13616, 2024. https://arxiv.org/abs/2402.13616.

[20] A. Wang et al., "YOLOv10: Real-time end-to-end object detection," arXiv:2405.14458, 2024. https://arxiv.org/abs/2405.14458.

[21] G. Jocher and J. Qiu, "Ultralytics YOLO11," software, ver. 11.0.0, 2024. https://docs.ultralytics.com/models/yolo11/.

[22] U. Ali, M. A. Ismail, R. A. A. Habeeb, and S. R. A. Shah, "Performance evaluation of YOLO models in plant disease detection," J. Informatics Web Eng., vol. 3, no. 2, pp. 199-211, 2024. https://doi.org/10.33093/jiwe.2024.3.2.15.

[23] B. Yang et al., "A novel plant type, leaf disease and severity identification framework using CNN and transformer with multi-label method," Sci. Rep., vol. 14, art. no. 11664, 2024. https://doi.org/10.1038/s41598-024-62452-x.