

Evaluasi Robustness Model YOLO Lintas Generasi untuk Klasifikasi Kode Resin Plastik pada Berbagai Tingkat Deformasi Fisik dan Ketidakseimbangan Kelas

Nyoman Sarasuartha Mahajaya^{1*}, I Gusti Made Ngurah Desnanjaya², I Nyoman Darma Kotama³

^{1,2}Rekayasa Sistem Komputer, Fakultas Teknologi Informatika, Institut Bisnis dan Teknologi Indonesia Denpasar, Bali, Indonesia

³Department of Information and Communication Systems, Okayama University, Okayama 700-8530, Japan

^{1*}nyoman_sarasuartha@instiki.ac.id, ²ngurah.desnanjaya@instiki.ac.id, ³p9363bg2@s.okayama-u.ac.id

INFO ARTIKEL

Article history:
Received Juni 2026
Published Juli 2026

ABSTRAK

Abstrak Klasifikasi kode resin plastik berbasis visi komputer merupakan komponen penting dalam otomatisasi proses pemilahan sampah untuk meningkatkan efisiensi daur ulang. Namun, sebagian besar penelitian sebelumnya hanya berfokus pada akurasi keseluruhan pada kondisi ideal dan belum mengevaluasi ketahanan model terhadap deformasi fisik objek serta ketidakseimbangan kelas yang umum ditemukan pada lingkungan daur ulang nyata. Penelitian ini menganalisis robustness tiga generasi model YOLO, yaitu YOLOv8n, YOLOv11n, dan YOLOv12n, menggunakan dataset WaDaBa yang memiliki metadata tingkat deformasi fisik. Evaluasi dilakukan secara multidimensi dengan mengukur akurasi Top-1, Macro-F1, recall per kelas, latensi inferensi, efisiensi Pareto, serta kinerja pada berbagai tingkat deformasi. Hasil eksperimen menunjukkan bahwa YOLOv8n memberikan performa terbaik dengan akurasi 85,42%, Macro-F1 0,8243, dan latensi inferensi 5,02 ms per citra (199,2 FPS), sehingga menjadi satu-satunya model yang berada pada frontier Pareto akurasi-latensi. Analisis deformasi menunjukkan bahwa seluruh model mengalami penurunan kinerja pada kondisi deformasi berat, namun YOLOv8n tetap mempertahankan akurasi tertinggi sebesar 76,61%. Selain itu, kelas PP menjadi kategori yang paling sulit diklasifikasikan oleh seluruh model. Temuan ini menunjukkan bahwa generasi YOLO yang lebih baru tidak selalu menghasilkan performa yang lebih baik serta menegaskan pentingnya evaluasi robustness di luar metrik akurasi konvensional.

Kata Kunci: YOLO, plastik, klasifikasi, deformasi, efisiensi, robustness

ABSTRAK

Abstracts Computer vision-based plastic resin code classification is a crucial component of automated waste sorting systems to improve recycling efficiency. However, most previous studies have focused primarily on overall accuracy under ideal conditions and have not evaluated model robustness against physical object deformation and class imbalance, both of which are commonly encountered in real-world recycling environments. This study analyzes the robustness of three generations of YOLO models, namely YOLOv8n, YOLOv11n, and YOLOv12n, using the WaDaBa dataset, which provides metadata on physical deformation levels. A multidimensional evaluation was conducted by measuring Top-1 accuracy, Macro-F1 score, per-class recall, inference latency, Pareto efficiency, and performance across different deformation levels. Experimental

results show that YOLOv8n achieved the best overall performance, with an accuracy of 85.42%, a Macro-F1 score of 0.8243, and an inference latency of 5.02 ms per image (199.2 FPS), making it the only model located on the accuracy–latency Pareto frontier. Deformation analysis revealed that all models experienced performance degradation under severe deformation conditions; however, YOLOv8n maintained the highest accuracy at 76.61%. In addition, the PP class was consistently the most difficult category for all models to classify. These findings indicate that newer YOLO generations do not necessarily deliver superior performance and highlight the importance of robustness-oriented evaluation beyond conventional accuracy metrics.

Keywords: YOLO, plastic, classification, deformation, efficiency, robustness.

©2026 Authors. Licensed Under [CC-BY-NC-SA 4.0](#)

1. Pendahuluan

Sampah plastik telah menjadi salah satu tantangan lingkungan terbesar di dunia akibat tingginya tingkat konsumsi plastik dan rendahnya efektivitas proses pemilahan sebelum daur ulang. Keberhasilan proses daur ulang sangat bergantung pada kemampuan mengidentifikasi jenis resin plastik secara akurat karena setiap jenis polimer memerlukan perlakuan pengolahan yang berbeda. Oleh karena itu, sistem klasifikasi otomatis berbasis visi komputer menjadi komponen penting dalam mendukung proses pemilahan sampah yang lebih cepat, konsisten, dan efisien. Berbagai penelitian menunjukkan bahwa penerapan kecerdasan buatan dan pembelajaran mendalam mampu meningkatkan akurasi serta efisiensi proses klasifikasi dan penyortiran sampah dibandingkan metode konvensional.[1]

Dalam beberapa tahun terakhir, model deep learning berbasis Convolutional Neural Network (CNN) dan keluarga You Only Look Once (YOLO) telah banyak digunakan untuk klasifikasi dan deteksi sampah plastik. Model-model tersebut menawarkan keseimbangan yang baik antara akurasi dan kecepatan inferensi sehingga berpotensi diterapkan pada sistem penyortiran real-time maupun perangkat tepi (edge devices). Beberapa penelitian melaporkan akurasi klasifikasi yang tinggi pada dataset sampah plastik menggunakan arsitektur CNN, transfer learning, maupun model YOLO generasi terbaru.[2]

Penelitian ini menggunakan WaDaBa (Waste Database) sebagai sumber data utama. WaDaBa merupakan dataset publik yang diperkenalkan oleh Bobulski dan Piątkowski untuk mendukung pengembangan metode otomatis dalam pengenalan dan pemilahan sampah plastik berbasis pengolahan citra. Dataset ini berisi sekitar 4.000 citra sampah plastik dengan variasi jenis resin, warna, kondisi pencahayaan, dan tingkat deformasi fisik yang terkode pada metadata berkas. Keberadaan informasi deformasi tersebut menjadikan WaDaBa unik dibandingkan dataset sampah plastik lainnya dan telah dimanfaatkan sebagai benchmark dalam berbagai penelitian klasifikasi plastik berbasis deep learning terkini. [3], [4]

Meskipun berbagai penelitian sebelumnya melaporkan tingkat akurasi yang tinggi, sebagian besar evaluasi dilakukan pada kondisi yang relatif ideal, yaitu objek masih utuh, latar belakang terkendali, dan distribusi kelas yang tidak dianalisis secara mendalam. Padahal pada lingkungan daur ulang nyata, sampah plastik sering mengalami deformasi fisik seperti penyok, terlipat, tertekan, atau rusak akibat proses penggunaan dan pengumpulan. Selain itu, distribusi data pada dataset resin plastik umumnya tidak seimbang, di mana beberapa kategori memiliki jumlah sampel jauh lebih besar dibanding kategori lainnya. Kondisi deformasi fisik dan ketidakseimbangan kelas tersebut dapat menyebabkan penurunan performa model yang signifikan meskipun akurasi keseluruhan terlihat tinggi. [5]

Penelitian terdahulu juga masih berfokus pada pelaporan metrik akurasi keseluruhan tanpa mengevaluasi ketahanan model terhadap variasi kondisi objek. Studi yang menggunakan dataset WaDaBa umumnya membandingkan model berdasarkan akurasi atau mean Average Precision (mAP), namun belum menganalisis secara khusus bagaimana performa model berubah ketika tingkat deformasi fisik meningkat. Selain itu, belum ditemukan penelitian yang melakukan evaluasi lintas generasi model YOLO pada tugas klasifikasi resin plastik menggunakan protokol pelatihan dan pengujian yang identik, sehingga sulit untuk menilai apakah peningkatan generasi arsitektur benar-benar menghasilkan peningkatan robustness dalam kondisi nyata. [4]

Penelitian mengenai klasifikasi sampah plastik berbasis deep learning telah berkembang pesat dalam beberapa tahun terakhir. Berbagai pendekatan telah diusulkan, mulai dari model CNN konvensional, transfer learning, hingga arsitektur YOLO yang dirancang untuk kebutuhan inferensi real-time. Sebagian besar penelitian melaporkan peningkatan akurasi yang signifikan pada kondisi pengujian yang terkontrol, namun evaluasi umumnya masih berfokus pada metrik akurasi keseluruhan. Aspek robustness terhadap deformasi fisik objek, ketidakseimbangan distribusi kelas, serta efisiensi komputasi untuk implementasi pada perangkat tepi masih relatif jarang dibahas secara bersamaan, padahal ketiga faktor tersebut sangat menentukan keberhasilan penerapan sistem pemilahan sampah otomatis di lingkungan operasional nyata. [1], [5]

Berdasarkan kesenjangan penelitian tersebut, studi ini mengusulkan kerangka evaluasi robustness untuk menganalisis performa tiga generasi model YOLO, yaitu YOLOv8n, YOLOv11n, dan YOLOv12n, pada tugas klasifikasi kode resin plastik menggunakan dataset WaDaBa. Berbeda dengan penelitian sebelumnya yang berfokus pada akurasi sebagai indikator utama, penelitian ini mengevaluasi model secara multidimensi melalui pengukuran akurasi Top-1, Macro-F1, recall per kelas, latensi inferensi, efisiensi Pareto, serta ketahanan terhadap berbagai tingkat deformasi fisik objek. Selain menyajikan benchmark lintas generasi YOLO dengan protokol eksperimen yang identik, penelitian ini juga memanfaatkan metadata deformasi unik pada WaDaBa untuk melakukan analisis robustness yang lebih mendekati kondisi operasional nyata. Hasil penelitian diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai hubungan antara akurasi, efisiensi komputasi, dan ketahanan model terhadap deformasi fisik, sekaligus menjadi dasar dalam pemilihan model yang optimal untuk sistem pemilahan sampah plastik berbasis perangkat tepi.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimental komparatif untuk mengevaluasi robustness tiga generasi model YOLO pada tugas klasifikasi kode resin plastik. Seluruh eksperimen dilakukan menggunakan platform Kaggle Notebook dengan akselerasi GPU NVIDIA yang disediakan oleh lingkungan komputasi berbasis cloud. Pendekatan ini memungkinkan evaluasi yang terkontrol terhadap performa model menggunakan protokol pelatihan dan pengujian yang identik sehingga perbandingan antar model dapat dilakukan secara objektif [1], [5].

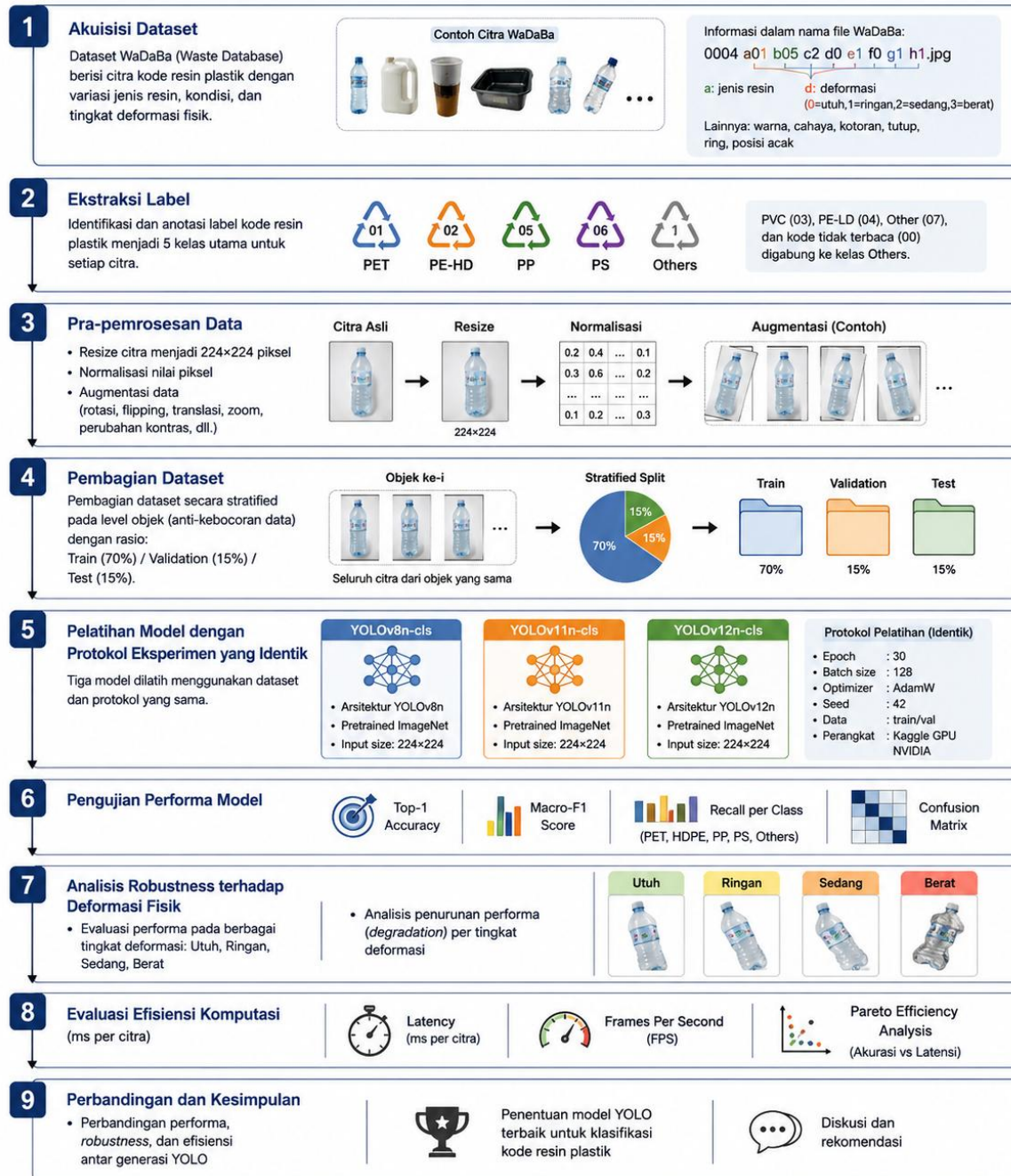
Secara umum, tahapan penelitian meliputi akuisisi dataset, ekstraksi metadata dan label, pra-pemrosesan data, pembagian dataset, pelatihan model, evaluasi performa, analisis robustness terhadap deformasi fisik, serta evaluasi efisiensi komputasi. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.

Dataset yang digunakan adalah WaDaBa (Waste Database), yaitu dataset publik yang dikembangkan untuk mendukung penelitian pengenalan dan pemilahan sampah plastik berbasis visi computer [3]. Dataset ini telah digunakan sebagai benchmark dalam berbagai penelitian klasifikasi dan deteksi sampah plastik berbasis deep learning karena menyediakan variasi jenis resin, kondisi pencahayaan, dan karakteristik fisik objek yang cukup beragam[4], [6]. Setelah proses parsing metadata, diperoleh 4.000 citra yang merepresentasikan 100 objek unik. Kode resin PET (01), PE-HD (02), PP (05), dan PS (06) dipetakan sebagai kelas utama, sedangkan PVC (03), PE-LD (04), Other (07), dan objek yang tidak dapat dibaca (00) digabungkan ke dalam kelas Others.

Tabel 1. Distribusi Dataset WaDaBa Setelah Pemetaan Kelas

Kelas	Jumlah Citra	Jumlah Objek
PET	2200	55
PE-HD	600	15
PP	640	16
PS	520	13
Others	40	1

Berdasarkan Tabel 1, dataset menunjukkan distribusi yang tidak seimbang, di mana kelas PET mendominasi sebanyak 2.200 citra sedangkan kelas Others hanya terdiri atas 40 citra. Kondisi ini umum ditemukan pada dataset klasifikasi sampah dan dapat menyebabkan bias model terhadap kelas mayoritas [5], [7].



Gambar 1. Alur Penelitian

Sumber: Penulis, 2026

Tahap pra-pemrosesan diawali dengan ekstraksi metadata dari nama berkas untuk memperoleh identitas objek, jenis resin, dan tingkat deformasi fisik. Seluruh citra kemudian diubah ukurannya menjadi 224×224 piksel dan dinormalisasi sebelum digunakan dalam proses pelatihan. Untuk meningkatkan kemampuan generalisasi model, dilakukan augmentasi data yang mencakup rotasi, translasi, zoom, perubahan kontras, dan pembalikan horizontal. Teknik augmentasi telah terbukti mampu meningkatkan robustness model klasifikasi sampah berbasis deep learning terhadap variasi visual objek dan kondisi lingkungan [5], [8], [9]

Dataset dibagi menjadi data pelatihan, validasi, dan pengujian menggunakan rasio 70:15:15 dengan pendekatan stratified split pada level objek. Berbeda dengan pembagian pada level citra, pendekatan ini memastikan bahwa citra dari objek yang sama tidak muncul pada lebih dari satu subset sehingga risiko data leakage dapat diminimalkan [4].

Tabel 2. Distribusi Dataset WaDaBa Setelah Pemetaan Kelas

Kelas	Jumlah Citra	Jumlah Objek
PET	2200	55
PE-HD	600	15
PP	640	16
PS	520	13
Others	40	1

Untuk mengurangi pengaruh ketidakseimbangan kelas, dilakukan oversampling pada subset pelatihan hingga jumlah sampel setiap kelas menjadi setara dengan kelas mayoritas. Setelah proses oversampling, masing-masing kelas pada data pelatihan memiliki 1.520 citra sehingga distribusi data menjadi lebih seimbang.

Penelitian ini membandingkan tiga generasi model YOLO yang menyediakan mode klasifikasi resmi, yaitu YOLOv8n-cls, YOLOv11n-cls, dan YOLOv12n-cls. Keluarga YOLO dipilih karena memiliki keseimbangan yang baik antara akurasi dan kecepatan inferensi sehingga banyak digunakan pada aplikasi klasifikasi dan penyortiran sampah berbasis visi komputer yang membutuhkan kemampuan real-time[1], [2], [7]

Tabel 3. Konfigurasi Model

Parameter	Nilai
Platform	Kaggle Notebook
GPU	Nvidia T4
Image Size	224x224
Batch Size	128
Epoch	30
Seed	42
Train Ratio	70%
Validation Ratio	15%
Test Ratio	15%

Seluruh model dilatih menggunakan konfigurasi yang identik sebagaimana ditunjukkan pada Tabel 3. Pendekatan ini bertujuan memastikan bahwa perbedaan performa yang diperoleh berasal dari karakteristik arsitektur model, bukan dari variasi parameter pelatihan.

Evaluasi dilakukan secara multidimensi menggunakan metrik akurasi Top-1, Precision, Recall, Macro-F1, latensi inferensi, throughput dalam satuan frame per second (FPS), jumlah parameter, dan ukuran model. Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh sampel pengujian dan dihitung menggunakan Persamaan (1).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Precision digunakan untuk mengukur proporsi prediksi positif yang benar sebagaimana ditunjukkan pada Persamaan (2).

$$Precision = TP / (TP + FP) \quad (2)$$

Recall digunakan untuk mengukur kemampuan model dalam mengenali seluruh sampel dari suatu kelas dan dihitung menggunakan Persamaan (3).

$$Recall = TP / (TP + FN) \quad (3)$$

Selanjutnya, F1-Score digunakan untuk mengevaluasi keseimbangan antara precision dan recall sebagaimana ditunjukkan pada Persamaan (4).

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

Selain evaluasi klasifikasi konvensional, penelitian ini melakukan analisis robustness dengan memanfaatkan metadata deformasi yang tersedia pada WaDaBa. Data uji dikelompokkan ke dalam empat tingkat deformasi, yaitu utuh, ringan, sedang, dan berat. Akurasi kemudian dihitung secara terpisah pada setiap kelompok deformasi untuk mengukur kemampuan model mempertahankan performa ketika

kondisi objek semakin terdegradasi. Evaluasi robustness menjadi penting karena objek sampah pada lingkungan operasional nyata sering mengalami deformasi fisik yang dapat memengaruhi kemampuan model dalam mengekstraksi fitur visual secara konsisten [2], [5].

Untuk mengevaluasi efisiensi komputasi, dilakukan pengukuran latensi inferensi pada citra tunggal dan throughput model dalam satuan FPS. Hasil evaluasi kemudian dianalisis menggunakan pendekatan Pareto frontier untuk mengidentifikasi model yang memberikan keseimbangan terbaik antara akurasi dan biaya komputasi. Selain itu, confusion matrix dan recall per kelas digunakan untuk mengidentifikasi pola kesalahan klasifikasi serta mengetahui kategori resin yang paling sulit dibedakan oleh masing-masing model. Pendekatan evaluasi ini memungkinkan analisis yang lebih komprehensif dibandingkan penggunaan metrik akurasi saja, sebagaimana direkomendasikan dalam penelitian klasifikasi sampah berbasis deep learning terbaru [1], [5], [7].

3. Hasil dan Pembahasan

Pada tahap ini dilakukan evaluasi terhadap performa YOLOv8, YOLOv11, dan YOLOv12 menggunakan dataset citra X-ray paru yang telah diproses pada tahap sebelumnya. Pengujian bertujuan untuk mengukur kemampuan setiap model dalam melakukan klasifikasi penyakit paru secara akurat dan konsisten. Analisis dilakukan berdasarkan metrik evaluasi global maupun performa per kelas untuk memperoleh gambaran yang lebih komprehensif mengenai keunggulan dan keterbatasan masing-masing model. Selain itu, hasil yang diperoleh juga dibandingkan untuk mengetahui perkembangan performa antar generasi model YOLO yang digunakan dalam penelitian ini.

Tabel 4. Perbandingan Performa Model

Model	Accuracy	Macro-F1	Latensi	FPS
YOLOv8n	85.42	0.8243	5.02	199.2
YOLOv11n	82.51	0.6884	6.48	154.3
YOLOv12n	69.68	0.5719	10.34	96.7

Tabel 4 menunjukkan perbandingan performa keseluruhan tiga generasi model YOLO pada tugas klasifikasi kode resin plastik. Hasil pengujian menunjukkan bahwa YOLOv8n memperoleh performa terbaik dengan akurasi Top-1 sebesar 85,42% dan Macro-F1 sebesar 0,8243. Selain menghasilkan akurasi tertinggi, YOLOv8n juga memiliki latensi inferensi paling rendah, yaitu 5,02 ms per citra atau setara dengan throughput sebesar 199,2 FPS. Hasil ini menunjukkan bahwa YOLOv8n mampu memberikan keseimbangan terbaik antara akurasi dan efisiensi komputasi.

YOLOv11n menempati posisi kedua dengan akurasi 82,51% dan Macro-F1 sebesar 0,6884. Meskipun mampu mempertahankan tingkat akurasi yang relatif tinggi, terjadi penurunan sebesar 2,91 poin persentase dibandingkan YOLOv8n. Selain itu, latensi inferensinya meningkat menjadi 6,48 ms per citra.

Sebaliknya, YOLOv12n menunjukkan penurunan performa yang cukup signifikan dengan akurasi hanya 69,68% dan Macro-F1 sebesar 0,5719. Model ini juga memiliki latensi inferensi tertinggi sebesar 10,34 ms per citra. Temuan ini menunjukkan bahwa peningkatan generasi arsitektur YOLO tidak selalu menghasilkan peningkatan performa pada tugas klasifikasi resin plastik. Pada dataset WaDaBa, YOLOv8n justru mampu memberikan kombinasi terbaik antara akurasi, kemampuan generalisasi, dan efisiensi komputasi.

Tabel 5. Recall Per Kelas

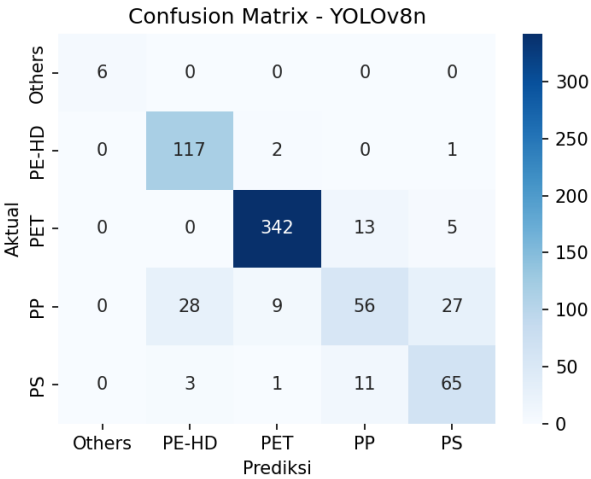
Model	Others	PE-HD	PET	PP	PS
YOLOv8n	100	97.5	95.0	46.67	81.25
YOLOv11n	100	100	93.89	44.17	61.25
YOLOv12n	100	39.17	95.28	19.17	73.75

Analisis recall per kelas dilakukan untuk mengevaluasi kemampuan model dalam mengenali setiap kategori resin plastik secara individual. Hasil pengujian menunjukkan bahwa seluruh model mampu mengidentifikasi kelas Others dengan recall sempurna sebesar 100%. Namun, performa antar model menunjukkan perbedaan yang cukup besar pada kelas PE-HD, PP, dan PS.

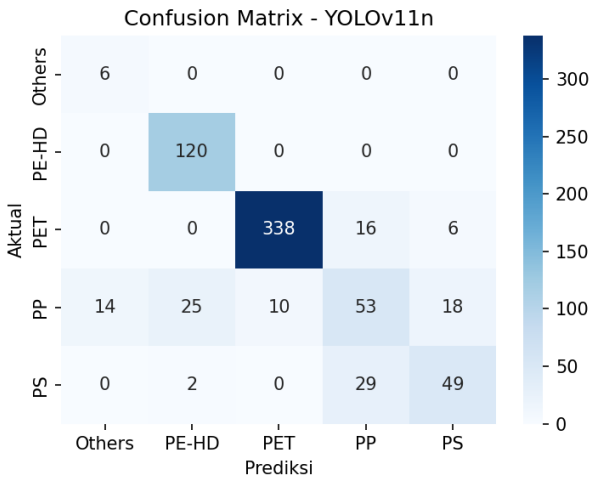
YOLOv11n menghasilkan recall tertinggi pada kelas PE-HD sebesar 100%, sedangkan YOLOv8n memperoleh recall sebesar 97,50%. Sebaliknya, YOLOv12n mengalami penurunan yang signifikan pada

kelas tersebut dengan recall hanya 39,17%. Pada kelas PET, seluruh model menunjukkan performa yang relatif tinggi dengan recall di atas 93%.

Kelas PP merupakan kategori yang paling sulit diklasifikasikan oleh seluruh model. Recall tertinggi hanya mencapai 46,67% pada YOLOv8n, sedangkan YOLOv11n dan YOLOv12n masing-masing hanya mencapai 44,17% dan 19,17%. Hasil ini mengindikasikan bahwa karakteristik visual PP memiliki kemiripan yang tinggi dengan beberapa kelas resin lainnya sehingga lebih sulit dibedakan oleh model. Temuan ini konsisten dengan nilai Macro-F1 yang menunjukkan bahwa ketidakseimbangan performa antar kelas masih menjadi tantangan utama dalam klasifikasi resin plastik.

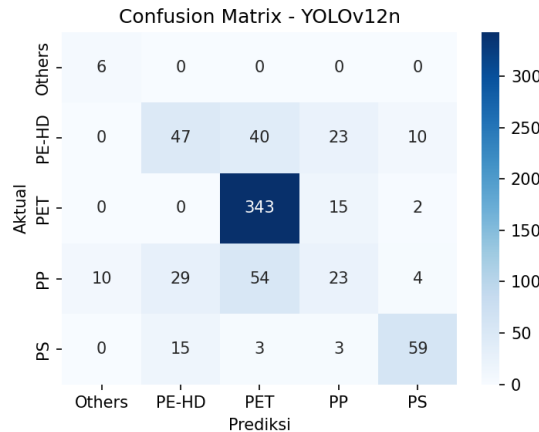


Gambar 2. Confusion Matrix YOLOv8n



Gambar 3. Confusion Matrix YOLOv11n

Sumber: Penulis, 2026



Gambar 4. Confusion Matrix YOLOv12n

Sumber: Penulis, 2026

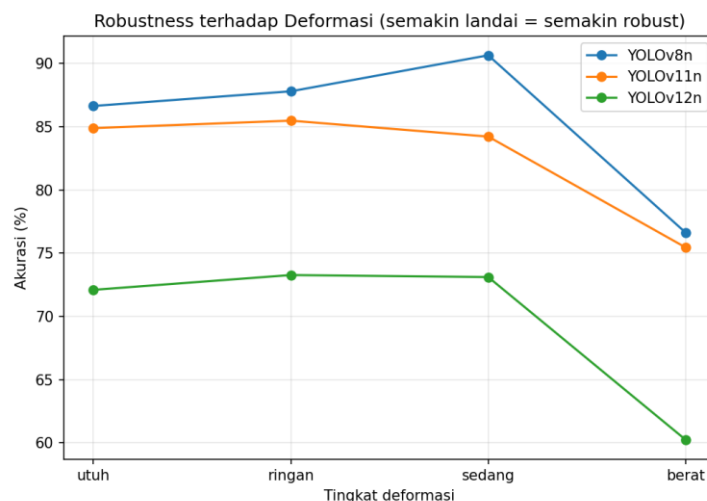
Analisis confusion matrix digunakan untuk mengidentifikasi pola kesalahan klasifikasi yang dilakukan oleh masing-masing model. Pada Gambar 2, YOLOv8n menunjukkan kemampuan klasifikasi yang baik pada kelas PET dan PE-HD dengan jumlah prediksi benar yang tinggi. Namun demikian, kesalahan klasifikasi masih sering terjadi pada kelas PP yang cenderung diprediksi sebagai PE-HD atau PS.

Pola yang serupa juga terlihat pada YOLOv11n sebagaimana ditunjukkan pada Gambar 3. Meskipun model ini mampu mencapai recall sempurna pada kelas PE-HD, sejumlah sampel PP masih salah diklasifikasikan ke dalam kelas PE-HD dan PS. Selain itu, sebagian sampel PS juga diprediksi sebagai PP sehingga menyebabkan penurunan recall pada kelas tersebut.

Gambar 4 menunjukkan bahwa YOLOv12n mengalami tingkat kebingungan antar kelas yang lebih tinggi dibandingkan dua model lainnya. Kesalahan terbesar terjadi pada kelas PE-HD dan PP. Sebanyak 40 sampel PE-HD diprediksi sebagai PET, sedangkan lebih dari separuh sampel PP salah diklasifikasikan menjadi PET atau PE-HD. Kondisi ini menjelaskan rendahnya nilai Macro-F1 dan recall yang diperoleh YOLOv12n.

Tabel 6. Stratifikasi Deformasi

Model	Utuh	Ringan	Sedang	Berat
YOLOv8n	86.63	87.79	90.64	76.61
YOLOv11n	84.88	85.47	84.21	75.44
YOLOv12n	72.09	73.26	73.1	60.23



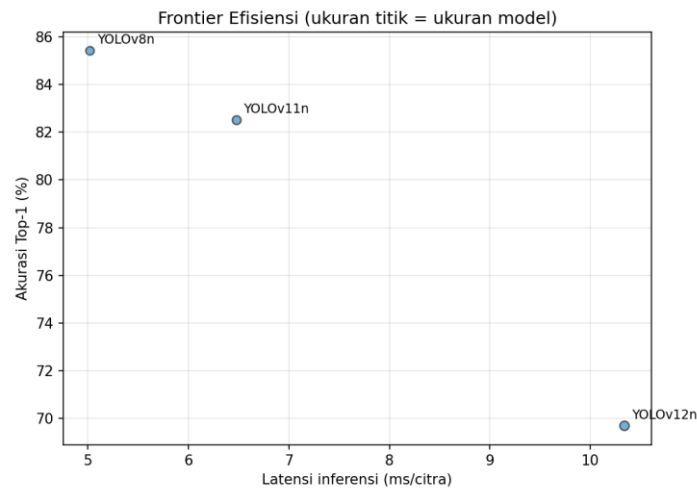
Gambar 5. Confusion Matrix YOLOv12n

Sumber: Penulis, 2026

Tabel 6 dan Gambar 5 menunjukkan performa model pada berbagai tingkat deformasi fisik objek. Secara umum, seluruh model mengalami penurunan akurasi ketika tingkat deformasi meningkat, terutama pada kategori deformasi berat.

YOLOv8n menunjukkan robustness terbaik dengan akurasi 86,63%, 87,79%, 90,64%, dan 76,61% masing-masing pada kategori utuh, ringan, sedang, dan berat. Menariknya, performa model justru meningkat pada kategori deformasi sedang sebelum mengalami penurunan pada deformasi berat. Hal ini menunjukkan bahwa model masih mampu mempertahankan kemampuan ekstraksi fitur pada kondisi objek yang mengalami perubahan bentuk moderat.

YOLOv11n menunjukkan pola yang relatif stabil hingga deformasi sedang, namun mengalami penurunan menjadi 75,44% pada deformasi berat. Sementara itu, YOLOv12n mengalami degradasi performa yang paling signifikan. Akurasinya turun dari 73,10% pada deformasi sedang menjadi hanya 60,23% pada deformasi berat. Hasil ini menunjukkan bahwa YOLOv12n memiliki sensitivitas yang lebih tinggi terhadap perubahan bentuk objek dibandingkan dua model lainnya.



Gambar 6. Pareto Frontier

Sumber: Penulis, 2026

Gambar 6 menunjukkan hubungan antara akurasi dan latensi inferensi untuk ketiga model yang diuji. Analisis Pareto digunakan untuk mengidentifikasi model yang memberikan keseimbangan terbaik antara performa klasifikasi dan biaya komputasi.

YOLOv8n berada pada posisi paling optimal dengan akurasi tertinggi sebesar 85,42% dan latensi terendah sebesar 5,02 ms per citra. Model ini menjadi satu-satunya solusi yang berada pada frontier Pareto karena tidak ada model lain yang mampu memberikan akurasi lebih tinggi dengan latensi yang lebih rendah secara bersamaan.

YOLOv11n berada di bawah frontier Pareto karena memiliki akurasi lebih rendah dan latensi lebih tinggi dibandingkan YOLOv8n. Kondisi yang sama juga terjadi pada YOLOv12n yang menunjukkan kombinasi performa dan efisiensi paling rendah. Dengan demikian, dari perspektif implementasi pada perangkat tepi dan sistem pemilahan sampah real-time, YOLOv8n merupakan pilihan yang paling layak untuk diterapkan.

4. Kesimpulan

Penelitian ini telah melakukan evaluasi komparatif terhadap tiga generasi model YOLO, yaitu YOLOv8n, YOLOv11n, dan YOLOv12n, untuk tugas klasifikasi kode resin plastik menggunakan dataset WaDaBa. Evaluasi dilakukan secara multidimensi dengan mempertimbangkan akurasi Top-1, Macro-F1, recall per kelas, latensi inferensi, efisiensi Pareto, serta robustness terhadap berbagai tingkat deformasi fisik objek. Hasil penelitian menunjukkan bahwa perbedaan generasi arsitektur YOLO tidak selalu menghasilkan peningkatan performa pada tugas klasifikasi resin plastik.

Berdasarkan hasil pengujian, YOLOv8n memberikan performa terbaik dengan akurasi Top-1 sebesar 85,42%, Macro-F1 sebesar 0,8243, dan latensi inferensi sebesar 5,02 ms per citra (199,2 FPS). Selain memiliki akurasi tertinggi, YOLOv8n juga menunjukkan ketahanan terbaik terhadap deformasi fisik dengan tetap mempertahankan akurasi sebesar 76,61% pada kategori deformasi berat. Analisis Pareto menunjukkan bahwa YOLOv8n merupakan satu-satunya model yang berada pada frontier Pareto karena mampu memberikan kombinasi optimal antara akurasi dan efisiensi komputasi.

Analisis confusion matrix dan recall per kelas menunjukkan bahwa kelas PP merupakan kategori yang paling sulit diklasifikasikan oleh seluruh model, sedangkan kelas PET dan PE-HD cenderung memiliki

tingkat pengenalan yang lebih baik. Selain itu, seluruh model mengalami penurunan performa ketika tingkat deformasi objek meningkat, khususnya pada kategori deformasi berat. Temuan ini menegaskan bahwa evaluasi robustness terhadap kondisi fisik objek perlu dipertimbangkan sebagai bagian penting dalam pengembangan sistem klasifikasi sampah plastik, karena metrik akurasi keseluruhan saja belum mampu menggambarkan performa model secara komprehensif.

Secara keseluruhan, hasil penelitian menunjukkan bahwa YOLOv8n merupakan model yang paling sesuai untuk implementasi sistem pemilahan sampah plastik berbasis perangkat tepi dan aplikasi real-time. Penelitian selanjutnya dapat difokuskan pada pengembangan teknik penanganan ketidakseimbangan kelas yang lebih efektif, perluasan jumlah objek pada kelas minoritas, serta eksplorasi metode augmentasi dan domain adaptation untuk meningkatkan robustness model terhadap kondisi operasional yang lebih beragam.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Institut Bisnis dan Teknologi Indonesia (INSTIKI) atas dukungan fasilitas dan sumber daya penelitian. Ucapan terima kasih juga disampaikan kepada penyelenggara skema INSTIKI Research and Development Program (IRDP) yang telah memberikan dukungan pendanaan, serta kepada seluruh kolaborator yang telah membantu kelancaran penelitian ini.

Daftar Pustaka

- [1] G. Ahmad *et al.*, “Intelligent waste sorting for urban sustainability using deep learning,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-08461-w.
- [2] I. Ranjbar, Y. Ventikos, and M. Arashpour, “Deep learning-based construction and demolition plastic waste classification by resin type using RGB images,” *Resour. Conserv. Recycl.*, vol. 212, Jan. 2025, doi: 10.1016/j.resconrec.2024.107937.
- [3] J. Bobulski and J. Piatkowski, “PET waste classification method and Plastic Waste DataBase WaDaBa.”
- [4] S. Kunwar, B. R. Owabumoye, and A. S. Alade, “Plastic Waste Classification Using Deep Learning: Insights from the WaDaBa Dataset,” 2024.
- [5] F. R. Sayem *et al.*, “Enhancing waste sorting and recycling efficiency: robust deep learning-based approach for classification and detection,” *Neural Comput. Appl.*, vol. 37, no. 6, pp. 4567–4583, Feb. 2025, doi: 10.1007/s00521-024-10855-2.
- [6] S. Kunwar, B. R. Owabumoye, and A. S. Alade, “Plastic Waste Detection Using Deep Learning: Insights from the WaDaBa Dataset,” *Industrial and Domestic Waste Management*, vol. 5, no. 1, pp. 1–11, Mar. 2025, doi: 10.53623/idwm.v5i1.580.
- [7] A. A. P. Chazhoor, E. S. L. Ho, B. Gao, and W. L. Woo, “Deep transfer learning benchmark for plastic waste classification,” *Intelligence and Robotics*, vol. 2, no. 1, pp. 1–19, 2022, doi: 10.20517/ir.2021.15.
- [8] S. M. M. A. N. D. H. Md. R. A. N. D. U. Md. P. A. N. D. M. Md. Al Islam Md. Minhazul AND Hasan, “Towards sustainable solutions: Effective waste classification framework via enhanced deep convolutional neural networks,” *PLoS One*, vol. 20, no. 6, pp. 1–49, Jun. 2025, doi: 10.1371/journal.pone.0324294.
- [9] L. Listyalina, M. Sarisky, D. Ellianto, U. F. Arifin, R. Utarianingrum, and H. I. Chandra, “Development of a real-time plastic waste detection system based on deep learning to support the automation of industrial waste sorting processes,” vol. 18, no. 2, 2025, doi: 10.31315/opsi.v18i2.