

# **Personal Protective Equipment Completeness Monitoring System Using YOLO-Based Computer Vision**

**Baasith Khoiruddin Akmal<sup>1\*</sup>, Wiji Lestari<sup>2\*</sup>, Afu Ichsan Pradana<sup>3\*</sup>**

\* Sistem Informasi, Universitas Duta Bangsa Surakarta

[210101007@mhs.udb.ac.id](mailto:210101007@mhs.udb.ac.id)<sup>1</sup>, [wiji\\_lestari@udb.ac.id](mailto:wiji_lestari@udb.ac.id)<sup>2</sup>, [afu\\_ichsan@udb.ac.id](mailto:afu_ichsan@udb.ac.id)<sup>3</sup>

---

## **Article Info**

### **Article history:**

Received 2025-07-10

Revised 2025-07-27

Accepted 2025-07-30

---

### **Keyword:**

*Computer Vision,  
Object Detection,  
Personal Protective Equipment  
(PPE),  
Workplace Safety,  
YOLOv8s,  
YOLOv11s.*

---

## **ABSTRACT**

Workplace safety in the construction sector remains a critical concern, primarily due to low compliance with Personal Protective Equipment (PPE) standards. To address this, this study develops and evaluates a real-time PPE monitoring system, conducting a comparative analysis of two state-of-the-art object detection models: YOLOv8s and YOLOv11s. The system is designed to detect three essential PPE items: helmets, masks, and vests, and both models were trained on a custom dataset of 9,202 augmented images over 200 epochs. The final evaluation on an unseen test set revealed highly competitive performance. While YOLOv8s achieved a marginally higher mAP@0.5 (90.8%), YOLOv11s demonstrated superior precision (92.0%) and better performance on the stricter mAP@0.5:0.95 metric (54.4%). Based on this nuanced trade-off and its significantly higher computational efficiency (15% fewer parameters), YOLOv11s was selected as the optimal model. The chosen model achieved a real-time inference speed of approximately 112 FPS. A functional web-based prototype was developed using Flask to demonstrate the system's practical application. These findings confirm that YOLOv11s offers a more balanced and efficient solution for automating PPE compliance monitoring and highlight that a holistic evaluation beyond a single metric is crucial for deploying robust computer vision systems in real-world safety applications.



This is an open access article under the [CC-BY-SA](#) license.

---

## **I. PENDAHULUAN**

Sektor konstruksi secara global dikenal sebagai salah satu lingkungan kerja dengan tingkat risiko tertinggi. Menurut *International Labour Organization (ILO)*, setiap tahunnya lebih dari 2,93 juta pekerja meninggal dunia akibat kecelakaan kerja atau penyakit terkait pekerjaan, serta lebih dari 395 juta pekerja mengalami cedera tidak fatal sepanjang tahun 2019 [1]. Di tingkat nasional, data dari Kementerian Ketenagakerjaan menunjukkan bahwa hanya dalam bulan Desember 2024, jumlah kecelakaan kerja di Indonesia tercatat lebih dari 1.300 kasus di Provinsi Aceh, serta masing-masing melebihi 3.000 kasus di Jawa Barat dan DKI Jakarta, dua wilayah dengan aktivitas konstruksi yang sangat padat. Jika dilihat secara keseluruhan, 462.241 kasus kecelakaan kerja tercatat sepanjang Januari hingga Desember 2024, dengan 91,65% di antaranya berasal dari peserta penerima upah, 7,43% dari peserta bukan penerima upah, dengan sektor jasa konstruksi menyumbang 0,92% dari total kasus [2]. Salah

satu penyebab utama tingginya angka kecelakaan kerja di sektor konstruksi adalah rendahnya kepatuhan terhadap penggunaan APD. Banyak pekerja enggan menggunakan APD karena alasan kenyamanan, kurangnya pengawasan, serta rendahnya kesadaran terhadap risiko di lingkungan kerja [3].

Implementasi regulasi Keselamatan dan Kesehatan Kerja (K3) di Indonesia masih menghadapi tantangan, terutama pada tingkat kepatuhan penggunaan Alat Pelindung Diri (APD) yang rendah di berbagai proyek. Studi oleh Bramistra pada proyek pembangunan Rumah Susun di Jawa Tengah menemukan tingkat penerapan APD secara keseluruhan hanya 39,445%, yang disebabkan oleh minimnya pengawasan, tidak adanya tenaga ahli K3 dari kontraktor, dan tidak dilaksanakannya *Toolbox Meeting* (TBM) [4]. Penelitian lain oleh Sartina dan Purnamawati di Jatinangor, Sumedang, menguatkan temuan ini dengan mencatat bahwa hanya 54% pekerja yang menggunakan APD secara lengkap

akibat faktor ketidaknyamanan dan kurangnya pengawasan [5].

Dampaknya, ketidakpatuhan ini berkorelasi langsung dengan meningkatnya angka kecelakaan kerja. Studi oleh Goma et al. di Kabupaten Bolaang Mongondow Utara membuktikan secara statistik bahwa 78% pekerja yang tidak menggunakan APD secara lengkap pernah mengalami kecelakaan kerja, sangat kontras dengan hanya 13,3% pada kelompok pekerja yang patuh. Penelitian ini menegaskan bahwa lemahnya pengawasan dan pengabaian prosedur menjadi pemicu utama insiden di lapangan [6].

Fakta bahwa mayoritas insiden berakar pada perilaku manusia dan diperparah oleh keterbatasan pengawasan manual inilah yang mendorong inovasi teknologi di bidang keselamatan kerja. Dalam konteks keselamatan kerja, *computer vision* berperan penting karena lebih dari 90% kecelakaan kerja di sektor konstruksi disebabkan oleh perilaku tidak aman dan kondisi kerja yang berbahaya, seperti tidak menggunakan alat pelindung diri (APD) yang sesuai [7].

Masalah kepatuhan terhadap penggunaan APD semakin kompleks karena diperparah oleh terbatasnya kapasitas pengawasan keselamatan kerja, baik dari segi jumlah personel maupun efektivitas sistem pelaporan. Data global dari indikator *Sustainable Development Goals* (SDG) 8.8.1 mencatat bahwa rata-rata insiden cedera kerja non-fatal mencapai 1.137 kasus per 100.000 pekerja, sementara cedera kerja fatal sebesar 5,78 kasus. Namun demikian, rasio petugas pengawas keselamatan kerja secara global hanya 0,81 orang per 10.000 pekerja, angka yang dinilai belum memadai untuk menjangkau seluruh lokasi kerja secara efektif [8].

Pengawasan penggunaan APD secara manual yang selama ini dilakukan oleh petugas lapangan memiliki keterbatasan dari segi jangkauan, akurasi, dan konsistensi. Oleh karena itu, muncul kebutuhan akan sistem otomatis yang mampu mengatasi keterbatasan ini, di mana *computer vision* menawarkan pendekatan solutif dengan kemampuannya menganalisis data visual untuk mendeteksi pelanggaran K3 secara *real-time* dan objektif [9]. *Computer vision* merupakan cabang dari kecerdasan buatan (*AI*) yang memungkinkan sistem untuk mengekstrak dan menganalisis informasi visual dari gambar atau video. Teknologi ini telah diterapkan secara luas dalam bidang keamanan, pengawasan, dan keselamatan kerja, termasuk dalam deteksi atribut APD seperti helm, rompi, dan masker [7].

Dalam beberapa tahun terakhir, pemanfaatan teknologi berbasis kecerdasan buatan (*Artificial Intelligence/AI*) terus meningkat di berbagai sektor, termasuk bidang keselamatan kerja. Integrasi *AI* dengan *Internet of Things (IoT)*, *edge computing*, dan *computer vision* telah mendorong terciptanya sistem pengawasan otomatis yang lebih cerdas, cepat, dan hemat sumber daya [9] [7]. Salah satu tren yang berkembang pesat adalah penggunaan model deteksi objek ringan yang mampu dijalankan secara *real-time* pada perangkat komputasi terbatas, seperti kamera CCTV atau *embedded systems* [10]. Tren ini membuka peluang baru untuk pengembangan sistem monitoring APD yang tidak hanya akurat, tetapi juga efisien

dan mudah diimplementasikan dalam skala luas, termasuk di lingkungan kerja konstruksi yang dinamis [11].

Beberapa studi sebelumnya telah mengembangkan sistem deteksi APD berbasis algoritma YOLO (*You Only Look Once*) karena keunggulannya dalam hal kecepatan dan akurasi deteksi objek secara *real-time*. Taufiqurrochman dan Februariyanti [12] merancang aplikasi deteksi APD untuk pekerja proyek menggunakan YOLOv5, dengan hasil akurasi deteksi mencapai 93,4% untuk objek helm, masker, rompi, dan sepatu. Sementara itu, penelitian oleh Nisa et al. [13] mengintegrasikan YOLOv8 dengan Django untuk membangun sistem monitoring APD di *Paiton Steam Power Plant*, menghasilkan mAP sebesar 78,8% dan precision sebesar 86,1%. Kedua studi tersebut menunjukkan efektivitas YOLO sebagai algoritma deteksi objek untuk kebutuhan keselamatan kerja.

Meskipun studi-studi tersebut menunjukkan efektivitas YOLO, perkembangan pesat dalam arsitektur *deep learning* terus menghadirkan model-model baru yang lebih efisien. Sebagian besar penelitian yang ada belum melakukan perbandingan langsung pada arsitektur termutakhir untuk menentukan keseimbangan optimal antara akurasi dan kecepatan inferensi untuk aplikasi di dunia nyata. Kemunculan model seperti YOLOv8 yang populer dan YOLOv11 yang diklaim lebih efisien [14] menciptakan sebuah celah penelitian (*research gap*) untuk mengevaluasi secara empiris mana arsitektur yang paling unggul untuk tugas pemantauan APD secara *real-time*.

Kesenjangan inilah yang mendasari tujuan penelitian ini, yaitu untuk merancang, melatih, dan mengevaluasi secara komparatif kinerja dua arsitektur deteksi objek mutakhir, yaitu YOLOv8s dan YOLOv11s, untuk sistem monitoring kelengkapan APD. Fokus sistem ini adalah mendeteksi tiga jenis APD krusial: helm, masker, dan rompi. Kinerja kedua model akan dievaluasi secara kuantitatif menggunakan metrik standar seperti *precision*, *recall*, dan *mean Average Precision* (mAP). Selanjutnya, model dengan performa terbaik akan diimplementasikan dalam sebuah prototipe aplikasi web fungsional untuk mendemonstrasikan kelayakannya dalam skenario penggunaan praktis. Studi komparatif ini diharapkan dapat memberikan rekomendasi berbasis bukti mengenai arsitektur paling efektif untuk otomatisasi pengawasan K3, sehingga dapat meningkatkan kepatuhan penggunaan APD di sektor konstruksi.

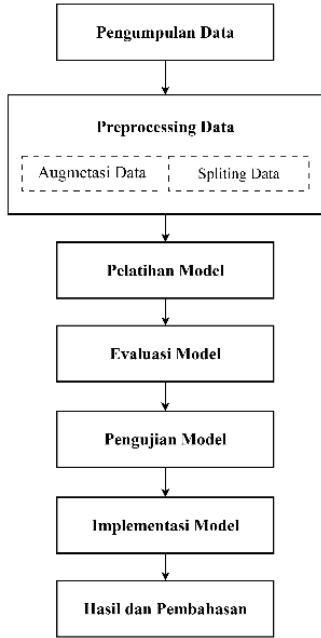
## II. METODE

Penelitian ini dilaksanakan melalui delapan tahapan sistematis guna memastikan proses berjalan terstruktur dan hasil yang didapat valid. Alur penelitian diilustrasikan pada Gambar 1.

### A. Pengumpulan Data

*Dataset* yang digunakan dalam penelitian ini merupakan kombinasi dari dua sumber utama untuk memastikan volume dan keanekaragaman data yang memadai. Sumber pertama adalah hasil pengambilan gambar secara langsung di salah

satu lokasi proyek konstruksi, sedangkan sumber kedua merupakan dataset publik bertema '*Safety Helmet, Mask, and Reflective Jacket*' yang tersedia di repositori seperti Kaggle dan GitHub.



Gambar 1. Alur Penelitian

Penggabungan data ini bertujuan untuk menciptakan dataset yang lebih representatif terhadap kondisi nyata di lapangan. Kumpulan data yang terkumpul terdiri dari 4.601 gambar.



Gambar 2. Statistik Dataset Awal Sebelum Augmentasi

Selanjutnya, proses anotasi dilakukan secara manual menggunakan platform *Roboflow* untuk memberi label dan *bounding box* pada tiga kelas objek yang menjadi fokus penelitian, yaitu *helmet* (helm), *mask* (masker), dan *vest* (rompi).

#### B. Preprocessing Data

Setelah proses anotasi, langkah berikutnya adalah *preprocessing* data. Tahapan ini diawali dengan augmentasi data yang bertujuan untuk meningkatkan variasi visual, mencegah *overfitting*, dan memperluas kemampuan generalisasi model. Teknik augmentasi yang diterapkan

secara otomatis melalui Roboflow meliputi *rotate*, *crop*, *shear*, *grayscale*, *saturation*, dan *brightness*.



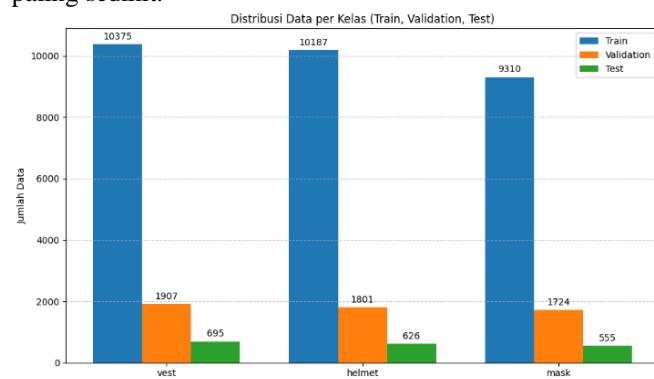
Gambar 3. Gambar yang Telah Melalui Proses Augmentasi

Setelah augmentasi, jumlah dataset meningkat signifikan dari 4.601 menjadi 9.202 gambar.

TABEL I  
DISTRIBUSI LABEL SEBELUM DAN SESUDAH AUGMENTASI

Kelas	Distribusi Label Sebelum Augmentasi	Distribusi Label Setelah Augmentasi
<i>helmet</i>	6.370	12.614
<i>mask</i>	5.839	11.589
<i>vest</i>	6.520	12.977
Total	18.729	37.180

Berdasarkan data distribusi label pada Tabel I, dapat dilihat bahwa jumlah instans objek antar kelas relatif seimbang. Kelas '*Vest*' memiliki jumlah label terbanyak (12.977 setelah augmentasi), hanya sekitar 1.12 kali lebih banyak dibandingkan kelas '*Mask*' (11.589 label) yang memiliki data paling sedikit.



Gambar 4. Distribusi Label Pada Dataset

Distribusi ini menunjukkan tingkat ketidakseimbangan data yang sangat ringan dan dapat diabaikan (*negligible imbalance*), yang merupakan kondisi ideal untuk melatih model deteksi objek.

Langkah selanjutnya adalah pembagian dataset ke dalam tiga subset, yaitu 80% untuk *train data* (7.362 gambar), 15% untuk *val data* (1.380 gambar), dan 5% untuk *test data* (460 gambar).

Dataset uji tidak digunakan selama proses pelatihan dan validasi guna menjamin objektivitas pada tahap evaluasi akhir model.

### C. Pelatihan Model

Penelitian ini menggunakan dua arsitektur model dari keluarga YOLO, yaitu YOLOv8s dan YOLOv11s, untuk dilatih dan dibandingkan kinerjanya. Pemilihan kedua model ini didasarkan pada popularitas YOLOv8 sebagai *baseline* yang kuat dan klaim efisiensi yang lebih tinggi pada YOLOv11 [14]. Keduanya merupakan arsitektur ringan (versi ‘s’ atau small) yang telah dilatih sebelumnya (pre-trained).

Untuk memastikan perbandingan yang objektif (fair comparison), kedua model dilatih menggunakan konfigurasi hyperparameter dan lingkungan komputasi yang identik. Rincian konfigurasi pelatihan dirangkum pada Tabel II.

TABEL II  
KONFIGURASI HYPERPARAMETER PELATIHAN

Parameter	Nilai / Spesifikasi
Epoch	200
Ukuran Gambar Input	640 x 640 piksel
Batch Size	64
Optimizer	AdamW
Learning Rate Awal (lr0)	0.0005

Seluruh proses pelatihan dijalankan pada platform Google Colaboratory (Colab) dengan memanfaatkan GPU NVIDIA A100-SXM4-40GB dan lingkungan CUDA versi 12.4. Selama proses pelatihan, bobot model dengan performa terbaik pada validation set untuk setiap arsitektur disimpan secara otomatis untuk digunakan pada tahap evaluasi akhir.

Pemilihan arsitektur ini didasarkan pada efisiensinya yang tinggi dalam hal jumlah parameter dan performa komputasi. YOLOv11 mampu mencapai nilai *mean Average Precision (mAP)* lebih tinggi dengan 22% parameter lebih sedikit dibandingkan arsitektur populer seperti YOLOv8 [15]. Efisiensi ini diperoleh melalui peningkatan arsitektur pada bagian *backbone* dan *neck*, yang memperkuat kemampuan ekstraksi fitur dan akurasi deteksi [16].

### D. Evaluasi Model

Evaluasi model dilakukan secara kuantitatif menggunakan metrik-metrik baku dalam tugas deteksi objek. Penilaian dimulai dari perhitungan *Intersection over Union (IoU)*, yang mengukur sejauh mana prediksi *bounding box* model sesuai dengan *ground truth*. Berdasarkan ambang batas *IoU*, deteksi diklasifikasikan sebagai *True Positive*, *False Positive*, atau *False Negative* [17].

Dari klasifikasi tersebut dihitunglah metrik *precision*, *recall*, dan *mean Average Precision (mAP)*. *Precision* menunjukkan proporsi prediksi benar dari seluruh deteksi yang dilakukan, sementara *recall* mengukur sejauh mana model dapat menemukan seluruh objek yang ada. *mAP* dihitung sebagai rata-rata area di bawah kurva *precision-recall* untuk seluruh kelas objek. Penelitian ini menggunakan dua varian *mAP*, yakni *mAP@0.5* untuk evaluasi umum, dan *mAP@0.5:0.95* yang lebih ketat sesuai standar COCO [18]. Selain metrik performa tersebut, penelitian ini juga membandingkan metrik efisiensi komputasi yang meliputi jumlah parameter, GFLOPs, ukuran model, dan waktu inferensi.

### E. Pengujian Model

Setelah model dari setiap arsitektur selesai dilatih dan bobot terbaik dipilih berdasarkan performa pada *validation set*, langkah selanjutnya adalah pengujian akhir. Tahap ini bertujuan untuk mengukur kinerja generalisasi model pada data yang sepenuhnya baru dan belum pernah dilihat sebelumnya (*unseen data*).

Untuk tujuan ini, digunakan *test set* yang telah disisihkan di awal, yang terdiri dari 460 gambar (5% dari total dataset). Metrik evaluasi yang dijelaskan pada sub-bab sebelumnya (Precision, Recall, dan mAP) dihitung berdasarkan hasil prediksi model pada *test set* ini. Hasil dari pengujian inilah yang menjadi acuan utama untuk analisis kuantitatif dan perbandingan kinerja final antara model YOLOv8s dan YOLOv11s yang disajikan pada Bab III.

### F. Implementasi Model

Untuk mendemonstrasikan penerapan praktis dari model yang dikembangkan, model akan diimplementasikan ke dalam bentuk prototipe aplikasi web berbasis *Flask*. Aplikasi ini dirancang untuk menerima input video dari webcam, memproses setiap frame menggunakan model YOLOv11s yang telah dilatih, dan menampilkan hasil deteksi secara *real-time*. Implementasi ini berfungsi sebagai *proof-of-concept* dari sistem monitoring APD yang kelak dapat diintegrasikan dengan sistem CCTV di lokasi kerja sebenarnya.

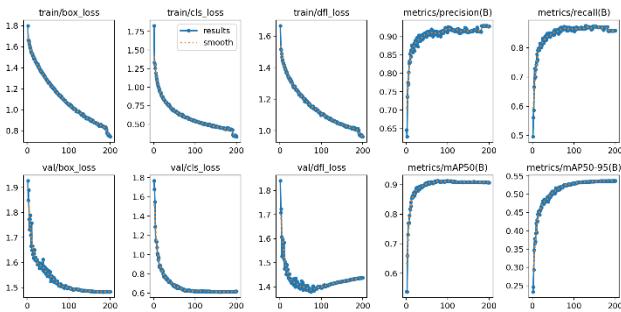
## III. HASIL DAN PEMBAHASAN

Setelah melalui tahapan pengumpulan, anotasi, augmentasi, dan pembagian dataset, tahap selanjutnya adalah menyajikan hasil dan pembahasan dari penelitian yang telah dilakukan. Penelitian ini menggunakan total 9.202 gambar yang telah diaugmentasi, yang terdiri dari 80% untuk *train data* (7.362 gambar), 15% untuk *val data* (1.380 gambar), dan 5% untuk *test data* (460 gambar untuk mendeteksi tiga kelas APD: *helmet* (helm), *mask* (masker), dan *vest* (rompi)).

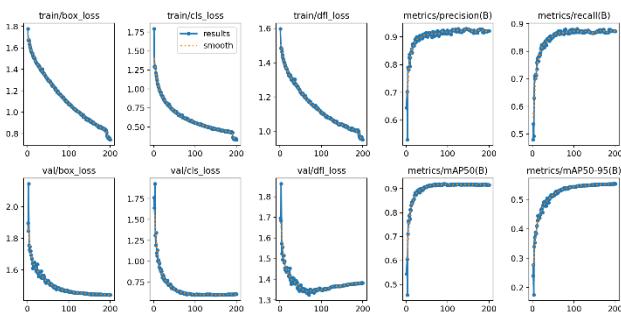
### A. Hasil Proses Pelatihan dan Validasi

Proses pelatihan untuk model YOLOv8s dan YOLOv11s dijalankan selama 200 *epoch*, dengan total waktu komputasi

masing-masing sekitar 1.98 jam untuk YOLOv8s dan 2.10 jam untuk YOLOv11s. Perkembangan proses belajar dan konvergensi kedua model dipantau melalui grafik kurva loss dan metrik pada *validation set*, yang disajikan pada Gambar 5 dan Gambar 6.



Gambar 5. Grafik Hasil Pelatihan Model YOLOv8s



Gambar 6. Grafik Hasil Pelatihan Model YOLOv11s

Analisis kurva *loss* pada Gambar 5 dan 6 menunjukkan bahwa kedua model berhasil belajar secara efektif, yang dibuktikan oleh tren penurunan konsisten pada kurva *loss* pelatihan (*train\_loss*). Namun, pada tahap akhir pelatihan (setelah *epoch* ke-150), terdapat indikasi *overfitting* ringan. Gejala ini teridentifikasi dari kurva *loss* validasi (*val/box\_loss* dan *val/dfl\_loss*) yang mulai menunjukkan tren kenaikan tipis. Hal ini menandakan bahwa model mulai terlalu 'menghafal' data latih, sehingga kemampuan generalisasinya untuk melokalisasi objek pada data baru sedikit menurun. Untuk melengkapi analisis visual dari grafik, kinerja puncak kedua model pada *validation set* dirangkum secara kuantitatif pada Tabel III.

TABEL III

PERBANDINGAN KINERJA DAN EFISIENSI MODEL PADA VALIDATION SET

Metrik	YOLOv8s	YOLOv11s
Presisi (P)	0.927	0.920
Recall (R)	0.859	0.873
mAP@0.5	0.907	0.916
mAP@0.5:0.95	0.537	0.554
Ukuran Model	22.5 MB	19.2 MB
Waktu Inferensi	0.7 ms/gambar	0.8 ms/gambar
Parameter	11.1 Juta	9.4 Juta
GFLOPs	28.4	21.3

Tabel ini membandingkan metrik performa dan efisiensi model berdasarkan bobot terbaik (*best.pt*) yang didapat

selama pelatihan. Berdasarkan Tabel III, YOLOv11s menunjukkan keunggulan pada sebagian besar metrik kunci, terutama pada akurasi deteksi (mAP) dan efisiensi model (ukuran, parameter, GFLOPs). Meskipun YOLOv8s sedikit lebih unggul dalam kecepatan inferensi, keunggulan akurasi pada YOLOv11s menjadikannya kandidat yang lebih andal. Kinerja final dari kedua model ini akan ditentukan melalui pengujian akhir menggunakan *test set* pada bagian selanjutnya.

#### B. Evaluasi Komparatif pada Data Uji

Tahap pengujian pada data yang belum pernah dilihat sebelumnya (*unseen test set*) adalah validasi paling krusial dalam menentukan kelayakan sebuah model. Pengujian ini dilakukan pada 460 gambar independen untuk mengukur kemampuan generalisasi arsitektur YOLOv8s dan YOLOv11s dalam skenario yang mendekati kondisi nyata. Hasil dari evaluasi komparatif ini, yang disajikan pada Tabel IV, memberikan sebuah temuan yang menarik.

TABEL IV  
PERBANDINGAN KINERJA FINAL MODEL PADA TEST SET

Metrik	YOLOv8s	YOLOv11s
Presisi (P)	0.91	0.92
Recall (R)	0.866	0.86
mAP@0.5	0.908	0.906
mAP@0.5:0.95	0.542	0.544
Waktu Inferensi	8.7 ms	8.9 ms

Data pada Tabel IV mengindikasikan bahwa kinerja kedua model sangat kompetitif, dengan perbedaan yang relatif kecil berbeda dari asumsi umum dalam literatur bahwa arsitektur yang lebih baru selalu menunjukkan dominasi performa. Perbedaan pada metrik utama seperti mAP@0.5 dan kecepatan inferensi berada pada margin yang sangat kecil, menegaskan bahwa YOLOv8s sebagai *baseline* yang matang tetap menjadi standar yang sulit untuk dilampaui.

Namun, analisis yang lebih granular mengidentifikasi adanya *trade-off* subtil yang krusial bagi penentuan model final. YOLOv11s menunjukkan keunggulan pada Presisi (0.920) dan metrik mAP@0.5:0.95 (0.544) yang lebih ketat. Ini mengindikasikan bahwa YOLOv11s lebih andal dalam meminimalkan deteksi keliru (*false positive*) dan mampu menghasilkan *bounding box* yang lebih presisi secara lokasi. Sebaliknya, YOLOv8s sedikit lebih unggul dalam Recall (0.866), yang berarti memiliki kemampuan lebih tinggi untuk mendeteksi setiap kemungkinan keberadaan objek APD.

Dalam sistem monitoring keselamatan, tingginya presisi menjadi prioritas untuk mengurangi alarm palsu dan menjaga kepercayaan pengguna. Oleh karena itu, keunggulan YOLOv11s dalam hal presisi, meskipun tipis, menjadi faktor penting untuk dipilih sebagai model final untuk analisis kualitatif dan implementasi prototipe lebih lanjut.

Dalam konteks sistem monitoring keselamatan, mengurangi alarm palsu adalah prioritas untuk menjaga

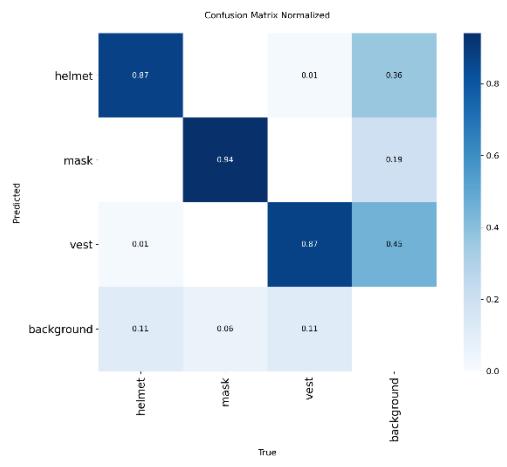
kepercayaan pengguna terhadap sistem. Oleh karena itu, keunggulan YOLOv11s dalam hal presisi, meskipun tipis, menjadi faktor penentu. Berdasarkan pertimbangan ini, YOLOv11s dipilih sebagai model final untuk analisis kualitatif dan implementasi prototipe lebih lanjut.

Temuan bahwa arsitektur yang lebih baru tidak serta-merta mendominasi pendahulunya pada semua metrik sejalan dengan tren yang diamati dalam literatur yang lebih luas. Sebagai contoh, sebuah studi komparatif eksternal oleh Sharma et al. (2024) juga mengungkap adanya *trade-off* krusial antara akurasi dan kecepatan pada model-model YOLO termutakhir. Dalam studi mereka, meskipun YOLOv9 mampu mencapai akurasi mAP@0.5 tertinggi (0.935), keunggulan tersebut harus dibayar dengan waktu inferensi yang jauh lebih lambat (54.2 ms), sehingga kurang ideal untuk aplikasi *real-time*.

Studi yang sama menyoroti bahwa YOLOv11 menawarkan keseimbangan yang lebih baik, dengan kecepatan inferensi superior (13.5 ms) dan tingkat presisi yang lebih tinggi (0.894) dibandingkan YOLOv8 (0.860) [19]. Temuan ini menegaskan bahwa seleksi arsitektur deteksi objek harus mempertimbangkan kebutuhan spesifik sistem secara strategis (seperti *trade-off* antara presisi dan *recall*), bukan semata-mata mengikuti versi model yang lebih baru.

### C. Analisis Kinerja Mendalam Model YOLOv11s

Setelah YOLOv11s dipilih sebagai model dengan keseimbangan performa optimal, analisis mendalam dilakukan untuk mengidentifikasi karakteristik kinerjanya secara spesifik. Tahap ini bertujuan untuk mengevaluasi kekuatan dan kelemahan model secara kuantitatif menggunakan alat diagnostik seperti *Confusion Matrix* dan kurva *Precision-Recall*. Untuk membedah jenis kesalahan yang dibuat oleh model, digunakan *Confusion Matrix* ternormalisasi yang disajikan pada Gambar 7.

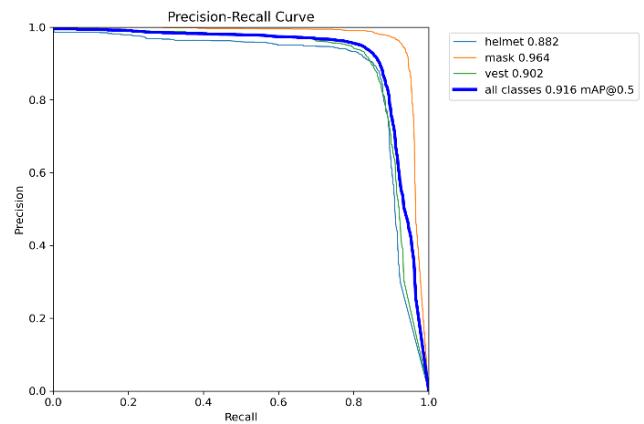


Gambar 7. Confusion Matrix Ternormalisasi untuk Model YOLOv11s pada Test Set

Analisis *Confusion Matrix* pada Gambar 7 memberikan beberapa wawasan penting. Diagonal utama, yang

merepresentasikan nilai *Recall*, menunjukkan bahwa kelas mask mencapai efektivitas deteksi tertinggi (0.94), sementara kelas helmet dan vest menunjukkan tingkat *Recall* yang sebanding (0.87). Sumber kesalahan utama yang teridentifikasi adalah *False Negative*, di mana 11% dari objek helmet dan vest gagal terdeteksi dan keliru diklasifikasikan sebagai background. Kegagalan deteksi ini kemungkinan besar disebabkan oleh tantangan visual di lapangan seperti oklusi parsial atau variasi warna. Keunggulan signifikan dari model ini adalah tingkat kebingungan antar-kelas yang sangat rendah, yang membuktikan kemampuannya untuk membedakan jenis-jenis APD secara akurat.

Selanjutnya, kinerja model dievaluasi lebih lanjut melalui kurva *Precision-Recall* (P-R) pada Gambar 8. Kurva ini memvisualisasikan *trade-off* antara presisi dan *recall* untuk setiap kelas.



Gambar 8. Kurva Precision-Recall (P-R) untuk Model YOLOv11s pada Test Set

Hasil pada Gambar 8 secara jelas mengonfirmasi superioritas deteksi pada kelas mask, yang mencapai nilai *Average Precision* (AP) tertinggi sebesar 0.964. Posisi kurvanya yang konsisten berada di level atas menunjukkan bahwa model mampu menjaga presisi deteksi yang sangat tinggi bahkan saat berusaha menemukan keseluruhan objek masker yang ada. Di sisi lain, kelas vest dan helmet menunjukkan kinerja yang solid namun lebih menantang, dengan masing-masing mencatatkan AP sebesar 0.902 dan 0.882. Bentuk kurva pada kedua kelas ini, terutama helmet, yang cenderung lebih cepat melandai mengindikasikan bahwa untuk mencapai *recall* yang lebih tinggi, model harus lebih banyak mengorbankan tingkat presisinya. Temuan ini konsisten dengan analisis pada *Confusion Matrix* sebelumnya, yang juga mengidentifikasi kedua kelas ini memiliki tantangan deteksi yang lebih besar.

Secara agregat, kinerja rata-rata model untuk semua kelas menghasilkan mAP@0.5 sebesar 91.6%. Nilai ini merepresentasikan tingkat akurasi keseluruhan yang sangat tinggi dan andal, sekaligus menegaskan bahwa model telah mencapai tingkat kematangan yang sangat baik pada akhir proses pelatihan dan validasi.

#### D. Analisis Kualitatif dan Studi Kasus

Analisis kualitatif dilakukan untuk mengevaluasi bagaimana model YOLOv11s berperilaku dalam berbagai skenario visual yang representatif. Beberapa studi kasus yang disajikan pada Gambar 9 digunakan untuk mendemonstrasikan kekuatan dan mengidentifikasi potensi kelemahan model secara praktis.

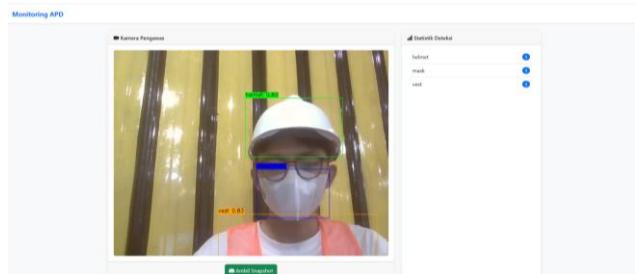


Gambar 9. Hasil Deteksi Model YOLOv11s pada Berbagai Skenario

Hasil studi kasus visual pada Gambar 9 mengonfirmasi bahwa model yang dikembangkan memiliki tingkat robustitas yang tinggi. Model menunjukkan kemampuan generalisasi yang kuat dengan berhasil mendeteksi APD secara akurat pada gambar tanpa informasi warna (*grayscale*), membuktikan bahwa pengenalan objek lebih didasarkan pada bentuk dan tekstur. Selain itu, model juga terbukti invariant terhadap perubahan orientasi objek, mampu menangani adegan yang padat dengan banyak pekerja (multi-objek), serta dapat mengenali APD dari jarak jauh (skala kecil). Bahkan pada kondisi yang lebih menantang seperti oklusi parsial, model tetap dapat mengidentifikasi helm dan rompi yang sebagian areanya terhalang oleh objek lain. Namun demikian, analisis ini juga mengungkapkan keterbatasan spesifik dari model. Kelemahan utama yang terungkap adalah potensi terjadinya *false positive*, di mana model keliru mengidentifikasi topi biasa sebagai helmet. Kasus ini menyoroti area perbaikan yang jelas, yaitu perlunya pengayaan dataset dengan lebih banyak contoh negatif (objek non-APD) untuk meningkatkan kemampuan diskriminatif model di masa depan. Secara keseluruhan, analisis kualitatif ini memvalidasi bahwa model YOLOv11s efektif dan praktis untuk implementasi di lapangan.

#### E. Implementasi dan Uji Coba Sistem

Sebagai bukti konsep (*proof-of-concept*), model YOLOv11s yang terpilih diimplementasikan ke dalam sebuah prototipe aplikasi web fungsional menggunakan *Framework* Flask. Sistem ini dirancang untuk memproses aliran video dari webcam secara langsung dan menampilkan hasil deteksi APD secara *real-time*, seperti yang ditunjukkan pada antarmuka dalam Gambar 10.



Gambar 10. Antarmuka Prototipe Aplikasi Web Saat Melakukan Deteksi

Tampilan pada Gambar 10 menunjukkan sistem yang sedang beroperasi, di mana ia mampu secara akurat menggambar kotak pembatas (*bounding box*) beserta label kelas (*helmet, mask, vest*) dan skor kepercayaan pada objek APD yang terdeteksi. Untuk menguji kinerja sistem secara praktis, serangkaian uji coba fungsional dilakukan menggunakan input video langsung dari kamera untuk menyimulasikan berbagai skenario di dunia nyata. Hasil dari uji coba kualitatif ini, yang menunjukkan respons sistem pada kondisi-kondisi spesifik, dirangkum pada Tabel V.

TABEL V  
HASIL UJI COBA KUALITATIF SISTEM PADA BERBAGAI SKENARIO

No	Skenario Pengujian	Gambar	Hasil Deteksi & Analisis Kinerja
1.	Kondisi Ideal (APD Lengkap)		Model berhasil mendeteksi semua APD (helmet, mask, vest) dengan skor kepercayaan tinggi. Ini menunjukkan kinerja dasar yang sangat baik pada kondisi optimal.
2.	Deteksi Multi-Objek		Model mampu mendeteksi APD lengkap pada dua pekerja secara simultan dalam satu frame. Ini membuktikan kemampuan model untuk menangani skenario ramai di lapangan.

3.	Oklusi Parsial		Model tetap mendeteksi semua APD meskipun sebagian terhalang. Namun, skor kepercayaan pada rompi menurun menjadi 0.55, menunjukkan performa bisa terpengaruh oleh tingkat oklusi.
4.	Pembedaan dengan Pakaian Biasa ( <i>True Negative</i> )		Model secara tepat tidak mengklasifikasi sikan rompi berwarna krem maupun rompi jeans sebagai rompi keselamatan. Ini menunjukkan kemampuan klasifikasi yang sangat baik dan risiko <i>false positive</i> yang rendah.

Kelayakan sistem untuk penerapan *real-time* tidak hanya dibuktikan melalui demonstrasi visual ini, tetapi juga didukung kuat oleh metrik efisiensi komputasi model YOLOv11s yang ringkas dan cepat, seperti yang disajikan pada Tabel VI.

TABEL VI  
METRIK EFISIENSI DAN KECEPATAN MODEL

Metrik Efisiensi	Nilai
Jumlah Parameter	9.4 Juta
GFLOPs	21.3
Waktu Inferensi	8.9 ms / gambar
Kecepatan (FPS)	~112 FPS

Efisiensi arsitektur YOLOv11s terlihat jelas dari jumlah parameter yang relatif kecil (9.4 Juta) dan beban komputasi yang rendah (21.3 GFLOPs). Hal ini memungkinkan model untuk mencapai kecepatan inferensi rata-rata 8.9 milidetik per gambar, atau setara dengan ~112 *Frames Per Second* (FPS). Angka ini jauh melampaui ambang batas minimal untuk pemrosesan video *real-time* (umumnya 24-30 FPS), yang mengonfirmasi bahwa model yang dikembangkan sangat cocok untuk diintegrasikan pada sistem pemantauan berbasis CCTV di dunia nyata.

#### F. Implikasi Etis dan Privasi

Penerapan sistem monitoring berbasis *computer vision* di lingkungan kerja menimbulkan pertanyaan penting mengenai etika dan privasi pekerja. Pengawasan konstan melalui kamera, meskipun bertujuan untuk keselamatan, berpotensi menciptakan perasaan tidak nyaman dan melanggar privasi individu. Sebagaimana ditekankan oleh Vukicevic et al., keberhasilan implementasi teknologi semacam ini sangat bergantung pada penerimaan oleh para pekerja, yang hanya dapat dicapai melalui transparansi dan kebijakan yang jelas [20].

Oleh karena itu, jika sistem ini akan diimplementasikan secara komersial, beberapa mitigasi perlu dilakukan. Pertama, kebijakan penggunaan sistem harus dikomunikasikan secara transparan kepada semua pekerja. Kedua, sistem idealnya hanya berfokus pada deteksi objek APD dan menghindari pengenalan wajah (*facial recognition*) untuk menjaga anonimitas. Data video yang direkam harus dikelola dengan kebijakan retensi yang ketat dan hanya dapat diakses oleh personel K3 yang berwenang untuk investigasi insiden, bukan untuk pengawasan kinerja. Dengan demikian, manfaat keselamatan dari sistem ini dapat dicapai tanpa mengorbankan hak-hak privasi pekerja.

#### IV. KESIMPULAN

Untuk menilai kepraktisan dan keandalan sistem yang diusulkan, penting untuk mengevaluasi secara kritis apakah hasil yang dicapai memadai untuk penerapan di dunia nyata. Evaluasi komparatif mengungkap bahwa arsitektur YOLOv8s dan YOLOv11s memberikan kinerja yang sangat kompetitif pada *test set*. Menariknya, berlawanan dengan ekspektasi bahwa arsitektur yang lebih baru akan unggul secara signifikan, perbedaan performa keduanya sangat tipis. YOLOv8s mencapai mAP@0.5 sedikit lebih tinggi (0.908), sementara YOLOv11s menunjukkan presisi yang lebih baik (0.920) dan kinerja yang lebih kuat pada metrik yang lebih ketat, mAP@0.5:0.95 (0.544).

Berdasarkan *trade-off* ini, YOLOv11s dipilih sebagai model final. Keputusan ini diambil berdasarkan evaluasi holistik, di mana presisi yang lebih tinggi menjadi faktor penting dalam meminimalkan *false positive*, yang dapat membangun kepercayaan pengguna terhadap sistem pemantauan otomatis. Hal ini, dikombinasikan dengan efisiensi komputasinya yang jauh lebih unggul (parameter 15% lebih sedikit dan ukuran model 14% lebih kecil), menjadikan sistem ini lebih andal dan layak untuk implementasi dunia nyata yang berskala besar. Kecepatan pemrosesan model yang tinggi (~112 FPS) lebih lanjut mengonfirmasi kemampuannya untuk analisis video secara *real-time*.

Meskipun sistem deteksi Alat Pelindung Diri (APD) yang dikembangkan dalam penelitian ini menunjukkan kinerja yang kuat, terdapat beberapa keterbatasan yang perlu dipertimbangkan, terutama terkait kompleksitas visual dan cakupan dataset. Salah satu tantangan utama, yang

teridentifikasi selama analisis kualitatif, adalah kerentanan model untuk menghasilkan *false positive* pada penutup kepala non-standar. Deteksi dapat terhambat jika seorang pekerja mengenakan penutup kepala dengan tampilan yang tidak umum (misalnya, kupluk atau topi biasa) yang tidak cukup terwakili sebagai contoh negatif dalam dataset pelatihan.

Selain itu, analisis kuantitatif mengungkap adanya tingkat *False Negative* sebesar 11% untuk helm dan rompi. Hal ini mengindikasikan bahwa sistem mungkin gagal mendeteksi sebagian kecil pekerja yang tidak patuh, terutama dalam skenario dengan oklusi objek yang tinggi, pencahayaan buruk, atau kondisi visual menantang lainnya. Oleh karena itu, kombinasi penampilan objek dan faktor lingkungan memainkan peran penting dalam efektivitas sistem.

Untuk pengembangan di masa depan, penting untuk mengeksplorasi skalabilitas sistem. Cakupan pengembangan mencakup integrasi backend yang komprehensif serta sistem notifikasi waktu nyata (*real-time*) untuk memperingatkan petugas keselamatan, serta *dashboard* analitik bagi manajemen untuk melacak tren kepatuhan. Pekerjaan di masa depan juga harus fokus pada pengayaan dataset dengan memasukkan lebih banyak contoh yang beragam, terutama dari kondisi menantang dan sampel negatif.

Terakhir, pengujian lapangan yang komprehensif di lingkungan yang lebih realistik sangat penting untuk mengevaluasi ketahanan jangka panjang sistem. Ini termasuk penerapan pada infrastruktur CCTV yang ada di lokasi konstruksi aktif untuk mendapatkan wawasan lebih dalam mengenai kinerja sistem di bawah berbagai kondisi lingkungan, penerimaan pengguna, dan potensi integrasi dengan sistem manajemen keselamatan yang sudah ada.

#### DAFTAR PUSTAKA

- [1] International Labour Organization, “Call for safer and healthier working environments,” Nov. 2023. [Online]. Available: <https://www.ilo.org/publications/call-safer-and-healthier-working-environments>
- [2] Kementerian Ketenagakerjaan Republik Indonesia, “Kasus Kecelakaan Kerja Tahun 2024,” Dec. 2024. Accessed: Jul. 19, 2025. [Online]. Available: <https://satudata.kemnaker.go.id/data/kumpulan-data/2447>
- [3] I. Mopio, R. Maulana, O. Heicqal Ardian, and S. Novita Sari, “Implementasi Keselamatan dan Kesehatan Kerja (K3) Pada Proyek Pembangunan Mall Pelayanan Publik (MPP) Dengan Target Zero Accident,” *Jurnal Ilmiah Teknik Unida*, vol. 4, no. 1, 2023.
- [4] R. O. Bramistra, “Analisis Penerapan Keselamatan dan Kesehatan Kerja (K3) Pada Proyek Gedung Bertingkat (Studi Kasus Penggunaan Alat Pelindung Diri Pada Paket Pembangunan Rumah Susun Pondok Pesantren Provinsi Jawa Tengah II TA. 2022),” Universitas Islam Indonesia, 2024.
- [5] I. Sartina and D. Purnamawati, “Evaluasi Penggunaan APD dalam Konteks Kesehatan dan Keselamatan Kerja di Industri Kontruksi,” in *Prosiding Seminar Nasional Kesehatan Masyarakat Universitas Muhammadiyah Surakarta*, 2024, pp. 131–144.
- [6] M. A. D. Goma, S. Sarman, H. Akbar, and F. Rumaf, “Hubungan Penerapan Standar Operasional Prosedur dan Penggunaan Alat Pelindung Diri Dengan Kecelakaan Kerja Pada Karyawan PT. X Kabupaten Bolang Mongondow Utara,” *Environmental Occupational Health And Safety Journal*, vol. 4, no. 2, pp. 35–42, 2024.
- [7] W. Fang *et al.*, “Computer vision applications in construction safety assurance,” *Autom Constr*, vol. 110, Feb. 2020, doi: 10.1016/j.autcon.2019.103013.
- [8] ILOSTAT, “Occupational Safety and Health Statistics (OSH) Database,” 2025. [Online]. Available: <https://ilostat.ilo.org/topics/safety-and-health-at-work/>
- [9] J. Li *et al.*, “A Review of Computer Vision-Based Monitoring Approaches for Construction Workers’ Work-Related Behaviors,” *IEEE Access*, vol. 12, pp. 7134–7155, 2024, doi: 10.1109/ACCESS.2024.3350773.
- [10] A. M. Vukicevic, M. Petrovic, P. Milosevic, A. Peulic, K. Jovanovic, and A. Novakovic, “A systematic review of computer vision-based personal protective equipment compliance in industry practice: advancements, challenges and future directions,” *Artif Intell Rev*, vol. 57, no. 12, Dec. 2024, doi: 10.1007/s10462-024-10978-x.
- [11] S. Arshad, O. Akinade, S. Bello, and M. Bilal, “Computer vision and IoT research landscape for health and safety management on construction sites,” *Journal of Building Engineering*, vol. 76, Oct. 2023, doi: 10.1016/j.jobe.2023.107049.
- [12] M. Alfin Taufiqurrochman, H. Februariyanti, and J. Homepage, “Rancang Bangun Aplikasi Deteksi Alat Pelindung Diri (APD) untuk Pekerja Proyek dengan Menggunakan Algoritma Yolov5 Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi),” *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 2, p. 2024, 2024, doi: 10.35870/jtik.
- [13] K. Nisa, F. Nur Fajri, and Z. Arifin, “Implementation of Personal Protective Equipment Detection Using Django and Yolo Web at Paiton Steam Power Plant (PLTU),” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 2, pp. 333–347, 2023, doi: 10.26555/jiteki.v9i2.26131.
- [14] Ultralytics, “Ultralytics YOLOv11 Documentation,” ultralytics. Accessed: Nov. 21, 2024. [Online]. Available: <https://docs.ultralytics.com/>
- [15] Ultralytics, “YOLO Model Comparison: YOLOv11 vs Previous.” [Online]. Available: <https://www.ultralytics.com/blog/comparing-ultralytics-yolov11-vs-previous-yolo-models>
- [16] Roboflow, “How to Train a YOLOv11 Object Detection Model on a Custom Dataset.” [Online]. Available: <https://blog.roboflow.com/yolov11-how-to-train-custom-data>
- [17] R. Padilla, S. L. Netto, and E. A. B. da Silva, “A Survey on Performance Metrics for Object-Detection Algorithms,” in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242. doi: 10.1109/IWSSIP48289.2020.9145130.
- [18] I.-H. and M. L. H.-Y. Wang Chien-Yao and Yeh, “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information,” in *Computer Vision – ECCV 2024*, E. and R. S. and R. O. and S. T. and V. G. Leonardis Aleš and Ricci, Ed., Cham: Springer Nature Switzerland, 2025, pp. 1–21.
- [19] A. Sharma, V. Kumar, and L. Longchamps, “Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species,” *Smart Agricultural Technology*, vol. 9, Dec. 2024, doi: 10.1016/j.atech.2024.100648.
- [20] A. M. Vukicevic, M. Petrovic, P. Milosevic, A. Peulic, K. Jovanovic, and A. Novakovic, “A systematic review of computer vision-based personal protective equipment compliance in industry practice: advancements, challenges and future directions,” *Artif Intell Rev*, vol. 57, no. 12, Dec. 2024, doi: 10.1007/s10462-024-10978-x.