

DETEKSI PERUNDUNGAN FISIK DI LINGKUNGAN SEKOLAH MENGGUNAKAN DETEKSI OBJEK BERBASIS YOLOV11

Yasmin Maulidia Andini Putri¹, Rania Ayu Praesmawardhani², Hamim Thohari Mahfudhillah³, Annisa Ahmada Atusta⁴

^{1,2,3}MTsN 6 Malang, Jl. Raya Sukoraharjo No.37, RW.01, Kepanjen, Kabupaten Malang, Jawa Timur 65163, Indonesia

Keywords:

Computer vision
Deep learning
Deteksi perundungan fisik
Object detection
YOLOv11

Correspondent Email:

hamimtm@gmail.com

Abstrak. Perundungan fisik merupakan salah satu bentuk kekerasan di sekolah yang paling umum terjadi, yang berdampak serius secara fisik, psikologis, dan akademis terhadap siswa. Sistem pengawasan konvensional mengandalkan pemantauan manual terhadap rekaman CCTV, sehingga menyulitkan deteksi dini insiden perundungan. Kemajuan dalam bidang *computer vision* dan *deep learning* membuka peluang untuk mengembangkan sistem deteksi kekerasan otomatis yang dapat meningkatkan keamanan sekolah. Penelitian ini mengusulkan model deteksi objek berbasis YOLOv11 untuk mengidentifikasi tindakan perundungan fisik di lingkungan sekolah. *Dataset* khusus dibuat dengan merekam simulasi skenario perundungan yang melibatkan siswa, mencakup sembilan kategori perundungan fisik—*memukul, menendang, menampar, mencakar, mendorong, menjegal, berkelahi, melempar benda, dan menyikut*, serta satu kategori aktivitas normal. *Dataset* diberi anotasi manual di Label Studio. Model YOLOv11 Nano dilatih selama 100 epoch menggunakan gambar berukuran 640×640 dan berbagai teknik augmentasi data pada *workstation* yang dilengkapi GPU NVIDIA GeForce RTX 4060. Model ini mencapai presisi 93,48%, *Recall* 96,29%, *mAP@0,5* sebesar 96,68%, dan *mAP@0,5:0,95* sebesar 78,54%. Hasil ini menunjukkan bahwa pendekatan yang diusulkan mampu mendeteksi berbagai tindakan perundungan fisik secara akurat dan layak digunakan untuk pengawasan *real-time*, sehingga mendukung intervensi yang lebih cepat serta peningkatan keamanan siswa di sekolah.

Abstract. Physical bullying is one of the most common forms of school violence, leading to serious physical, psychological, and academic impacts on students. Traditional surveillance systems rely on manual monitoring of Closed-Circuit Television (CCTV) footage, making early detection of bullying incidents difficult. Advances in computer vision and deep learning provide opportunities to develop automated violence detection systems that can enhance school safety. This study proposes a YOLOv11-based object detection model to identify physical bullying actions in school environments. A custom dataset was created by recording simulated bullying scenarios involving students, covering nine categories of physical bullying—hitting, kicking, slapping, scratching, pushing, tripping, fighting, throwing objects, and elbowing—along with a normal activity class. Dataset manually annotated with bounding boxes in Label Studio. The YOLOv11 Nano model was trained for 100 epochs using 640×640 images and multiple data augmentation techniques on a workstation equipped with an NVIDIA GeForce RTX 4060 GPU. The model achieved 93.48% Precision, 96.29% Recall, 96.68% *mAP@0.5*, and 78.54% *mAP@0.5:0.95*. These results demonstrate that the proposed approach can accurately detect various physical bullying actions and is suitable for real-time surveillance, supporting faster intervention and improved student safety in schools.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

1. PENDAHULUAN

Sekolah merupakan tempat yang seharusnya memberikan lingkungan belajar yang aman, nyaman, dan mendukung perkembangan akademik, sosial, serta emosional siswa. Namun, perundungan (*bullying*) masih menjadi masalah yang banyak terjadi di berbagai negara. Menurut UNICEF, sekitar satu dari tiga siswa berusia 13–15 tahun pernah mengalami perundungan di sekolah [1]. Data ini menunjukkan bahwa kekerasan antar siswa masih menjadi persoalan yang perlu mendapat perhatian serius.

Salah satu bentuk perundungan yang paling sering terjadi adalah perundungan fisik, seperti *memukul, menendang, mendorong, menampar*, dan tindakan kekerasan lainnya. Perundungan fisik tidak hanya menyebabkan cedera pada korban, tetapi juga dapat menimbulkan dampak jangka panjang, seperti trauma, rendahnya rasa percaya diri, kecemasan, depresi, serta menurunnya prestasi belajar [2]. Oleh karena itu, deteksi dan penanganan sejak dini sangat penting untuk mengurangi dampak negatif yang ditimbulkan.

Kondisi di Indonesia juga menunjukkan permasalahan yang sama. Data dari Komisi Perlindungan Anak Indonesia (KPAI) menunjukkan bahwa kasus kekerasan antar siswa masih menjadi salah satu laporan yang paling sering diterima di lingkungan pendidikan [3]. Selain itu, berbagai survei pendidikan menunjukkan bahwa banyak siswa Indonesia pernah mengalami perundungan selama bersekolah, dengan perundungan fisik sebagai salah satu bentuk yang paling dominan. Berbagai upaya pencegahan telah dilakukan oleh sekolah, seperti layanan konseling, pendidikan karakter, dan peningkatan pengawasan guru. Namun, perundungan sering terjadi di area yang sulit diawasi secara langsung, seperti koridor, lapangan, area parkir, dan ruang terbuka lainnya. Akibatnya, banyak kasus baru diketahui setelah kejadian berlangsung sehingga penanganannya menjadi kurang efektif.

Untuk meningkatkan keamanan, banyak sekolah telah memasang kamera *Closed-Circuit Television* (CCTV). Namun, CCTV konvensional hanya berfungsi sebagai alat perekam sehingga tetap memerlukan

pemantauan terus-menerus oleh petugas. Mengawasi banyak layar CCTV dalam waktu yang lama bukanlah pekerjaan yang mudah. Operator dapat kehilangan fokus karena kelelahan, gangguan, atau banyaknya kamera yang harus dipantau. Kondisi ini menyebabkan beberapa kejadian penting tidak terdeteksi tepat waktu. Oleh karena itu, diperlukan sistem pengawasan yang lebih cerdas dan mampu mendeteksi tindakan mencurigakan secara otomatis serta memberikan peringatan ketika terjadi potensi kekerasan.

Perkembangan teknologi *artificial intelligence* (AI), khususnya pada bidang *computer vision* dan *deep learning*, telah membuka peluang untuk mengembangkan sistem pengawasan yang lebih cerdas. Model *deep learning* mampu mengenali pola visual dari gambar maupun video secara otomatis tanpa perlu membuat aturan secara manual [4]. Salah satu teknologi yang banyak digunakan adalah *object detection*, yaitu metode yang dapat mengenali jenis objek sekaligus menentukan posisinya pada gambar [5]. Kemampuan ini memungkinkan analisis video secara *real-time* dan telah diterapkan pada berbagai bidang, seperti pemantauan lalu lintas, keamanan publik, industri, kesehatan, hingga kota pintar.

Salah satu model deteksi objek yang paling banyak digunakan adalah *You Only Look Once* (YOLO). Model ini dikenal memiliki keseimbangan yang baik antara kecepatan dan akurasi. YOLO dapat mendeteksi dan mengklasifikasikan objek dalam satu proses sehingga lebih cepat digunakan untuk aplikasi *real-time* [6]. Sejak pertama kali diperkenalkan, YOLO terus mengalami pengembangan hingga versi terbaru, yaitu YOLOv11. Versi ini memiliki kemampuan deteksi yang lebih baik dengan kebutuhan komputasi yang tetap efisien. Oleh karena itu, YOLOv11 sangat sesuai digunakan pada sistem pengawasan sekolah yang membutuhkan deteksi secara cepat dan akurat.

Berbagai penelitian sebelumnya telah mengembangkan sistem deteksi kekerasan menggunakan *deep learning*. Namun, sebagian besar penelitian tersebut masih berfokus pada deteksi kekerasan secara umum, deteksi kerumunan, atau pengenalan aktivitas

menggunakan *dataset* dari film, pertandingan olahraga, maupun rekaman CCTV public [7]. Kondisi tersebut berbeda dengan situasi nyata yang terjadi di lingkungan sekolah. Selain itu, banyak penelitian menggunakan metode pengenalan aksi berbasis video yang memerlukan analisis beberapa *frame* sekaligus sehingga membutuhkan komputasi yang lebih besar dan proses deteksi menjadi lebih lambat. Penelitian yang secara khusus membahas deteksi perundungan fisik di sekolah menggunakan model deteksi objek yang ringan dan dapat bekerja secara *real-time* masih relatif terbatas. Oleh karena itu, diperlukan model yang dilatih menggunakan data yang benar-benar menggambarkan kondisi perundungan di lingkungan sekolah agar dapat diterapkan secara nyata.

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pengawasan sekolah berbasis AI, khususnya untuk mendeteksi perundungan fisik secara otomatis. Model yang dikembangkan diharapkan dapat membantu sekolah melakukan deteksi lebih cepat sehingga tindakan pencegahan dan penanganan dapat dilakukan sebelum dampak yang lebih besar terjadi.

Untuk mencapai tujuan tersebut, penelitian ini mengusulkan model deteksi objek berbasis YOLOv11 untuk mendeteksi perundungan fisik secara otomatis di lingkungan sekolah. *Dataset* dikumpulkan melalui perekaman simulasi perundungan yang diperagakan oleh siswa. *Dataset* ini mencakup sembilan jenis tindakan perundungan fisik, yaitu *memukul*, *menendang*, *menampar*, *mencakar*, *mendorong*, *menjegal*, *berkelahi*, *melempar benda*, dan *menyikut*, serta satu kategori aktivitas normal. Video kemudian diubah menjadi kumpulan gambar (*frame*) dan setiap objek diberi anotasi secara manual menggunakan *bounding box* untuk menghasilkan *dataset* deteksi objek. *Dataset* tersebut selanjutnya digunakan untuk melatih dan menguji model YOLOv11 Nano agar mampu mendeteksi tindakan perundungan fisik secara *real-time*.

2. TINJAUAN PUSTAKA

Artificial Intelligence (AI) atau kecerdasan buatan merupakan cabang ilmu komputer yang bertujuan mengembangkan sistem yang mampu meniru cara manusia dalam belajar, mengenali

pola, mengambil keputusan, dan menyelesaikan masalah [8]. Dengan memanfaatkan AI, komputer dapat mengolah data dalam jumlah besar, mengenali pola secara otomatis, serta menghasilkan prediksi atau klasifikasi dengan cepat. Saat ini, AI telah banyak diterapkan di berbagai bidang, seperti sistem keamanan, pengenalan wajah, kendaraan otonom, serta analisis gambar dan video [9].

Salah satu teknologi yang banyak digunakan dalam pengembangan AI adalah *deep learning*. *Deep learning* merupakan bagian dari *machine learning* yang menggunakan jaringan saraf tiruan (*neural network*) berlapis untuk mempelajari pola yang kompleks dari data [10]. Berbeda dengan metode konvensional, *deep learning* mampu mengekstraksi fitur secara otomatis dari data tanpa memerlukan proses perancangan fitur secara manual. Pada analisis gambar dan video, model yang paling banyak digunakan adalah *Convolutional Neural Network* (CNN) karena memiliki kemampuan yang baik dalam mengenali objek dan pola visual [11].

Perkembangan *deep learning* juga mendorong kemajuan pada bidang *computer vision*, yaitu teknologi yang memungkinkan komputer memahami informasi visual dari gambar maupun video. Melalui *computer vision*, sistem dapat mendeteksi objek, mengenali aktivitas manusia, serta menganalisis interaksi yang terjadi pada rekaman video secara *real-time* [12]. Kemampuan ini banyak dimanfaatkan dalam sistem pengawasan cerdas (*intelligent surveillance system*) untuk mendeteksi berbagai kejadian secara otomatis, termasuk perilaku agresif atau tindakan perundungan (*bullying*). Oleh karena itu, teknologi AI, *deep learning*, dan *computer vision* menjadi dasar utama dalam pengembangan sistem deteksi *bullying* berbasis video.

YOLOv11 (*You Only Look Once version 11*) merupakan model deteksi objek yang dikembangkan oleh *Ultralytics* untuk mendeteksi berbagai objek pada gambar maupun video secara *real-time*. Berbeda dengan beberapa metode deteksi objek lainnya, YOLO hanya memerlukan satu kali proses melalui jaringan saraf untuk menentukan jenis objek sekaligus lokasinya pada gambar. Pendekatan ini membuat YOLO memiliki kecepatan yang tinggi dengan tetap

mempertahankan akurasi yang baik, sehingga banyak digunakan pada berbagai aplikasi yang membutuhkan deteksi secara cepat, seperti sistem keamanan, kendaraan otonom, dan pengawasan berbasis video [13].

YOLOv11 merupakan pengembangan dari versi-versi YOLO sebelumnya dengan peningkatan pada akurasi deteksi, efisiensi komputasi, dan kecepatan inferensi [14]. Model ini mampu mengenali objek berukuran kecil maupun besar dengan lebih baik serta tetap memberikan performa yang stabil pada berbagai kondisi pencahayaan dan lingkungan. Selain itu, YOLOv11 mendukung pendekatan *anchor-free detection*, sehingga proses deteksi menjadi lebih sederhana dan efisien dibandingkan beberapa model sebelumnya [15]. Karakteristik tersebut menjadikan YOLOv11 sesuai untuk diterapkan pada sistem pengawasan cerdas (*intelligent surveillance system*) yang memerlukan respons cepat terhadap suatu kejadian. Dalam penelitian ini, YOLOv11 Nano digunakan untuk mendeteksi tindakan perundungan fisik pada rekaman CCTV sekolah. Model ini diharapkan mampu mengenali berbagai bentuk interaksi agresif antarsiswa secara otomatis sehingga dapat membantu proses pemantauan dan memberikan peringatan lebih dini terhadap potensi terjadinya *bullying*.

3. METODE PENELITIAN

3.1 Kerangka Penelitian

Penelitian ini menggunakan pendekatan penelitian rekayasa untuk mengembangkan dan mengevaluasi model kecerdasan buatan untuk deteksi otomatis perundungan fisik di lingkungan sekolah. Alur kerja penelitian secara keseluruhan terdiri dari lima tahap berurutan: (1) akuisisi dataset, (2) ekstraksi dan anotasi frame, (3) pengembangan model, (4) pelatihan model, dan (5) evaluasi kinerja. Gambar 1 mengilustrasikan kerangka kerja penelitian secara lengkap.

**Perekaman Video → Ekstraksi Bingkai
→ Anotasi → Pelatihan YOLOv11 →
Evaluasi**

Gambar 1 Alur kerja penelitian secara keseluruhan

Awalnya, skenario simulasi perundungan fisik direkam dalam lingkungan sekolah yang

terkontrol. Video yang direkam kemudian dikonversi menjadi *frame* gambar menggunakan program ekstraksi bingkai berbasis *Python*. Setiap gambar dianotasi secara manual menggunakan software *Label Studio* untuk menghasilkan label *bounding-box* yang kompatibel dengan format deteksi objek YOLO. *Dataset* yang telah dianotasi kemudian digunakan untuk melatih model deteksi objek YOLOv11 Nano. Terakhir, model yang telah dilatih dievaluasi menggunakan metrik deteksi objek standar, termasuk *Precision*, *Recall*, *mean Average Precision* (mAP@0.5), dan mAP@0.5:0.95.

3.2 Akuisisi Dataset

Karena kumpulan data yang tersedia untuk umum yang berisi aktivitas perundungan fisik di lingkungan sekolah sangat terbatas, kumpulan data khusus dikembangkan khusus untuk penelitian ini. Kumpulan data tersebut dibangun dengan merekam interaksi simulasi yang dilakukan secara sukarela oleh siswa dalam kondisi terkontrol yang memastikan keselamatan peserta sekaligus mempertahankan gerakan fisik dan interaksi interpersonal yang realistis. Sepuluh kelas objek didefinisikan, terdiri dari sembilan kategori tindakan perundungan fisik dan satu kelas aktivitas normal. Kategori perundungan fisik meliputi *memukul*, *menendang*, *menampar*, *mencakar*, *mendorong*, *menjegal*, *berkelahi*, *melempar benda*, dan *menyikut*, sedangkan kelas siswa mewakili aktivitas sekolah normal tanpa perilaku agresif.



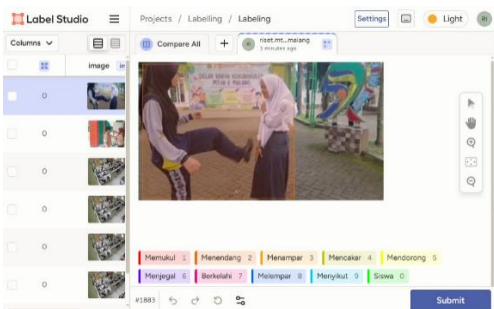
Gambar 2 Contoh *dataset* menunjukkan gambar siswa yang sedang *memukul* dan *menendang*

Untuk meningkatkan keragaman *dataset* dan generalisasi model, setiap aktivitas direkam dari berbagai sudut pandang kamera, jarak, dan orientasi tubuh. Variasi dalam postur siswa, pola interaksi, skala objek, dan kondisi pencahayaan sengaja disertakan untuk lebih mewakili skenario pengawasan nyata yang ditemui di lingkungan sekolah.

Video yang direkam kemudian diproses menjadi *frame*, menghasilkan *dataset* yang terdiri dari 2.195 gambar yang telah dianotasi. *Dataset* tersebut dibagi menjadi 1.756 gambar pelatihan (80%) dan 439 gambar validasi (20%) untuk pengembangan model terawasi. *Dataset* validasi berisi 537 *instance objek* yang telah dianotasi dan didistribusikan ke dalam sepuluh kelas yang telah ditentukan sebelumnya.

3.3 Ekstraksi Frame dan Anotasi Data

Video yang direkam dikonversi menjadi kumpulan data gambar melalui proses ekstraksi bingkai yang diimplementasikan dalam Python. Proses ini mengubah rangkaian video kontinu menjadi *frame* individual yang cocok untuk pelatihan deteksi objek terawasi sambil mempertahankan variasi gerakan dan interaksi manusia sepanjang rangkaian yang direkam. Semua *frame* yang diekstrak dianotasi secara manual menggunakan Label Studio.



Gambar 3 Contoh proses pelabelan *dataset* siswa yang sedang menendang menggunakan Label Studio

Setiap siswa yang muncul dalam gambar dikelilingi oleh *bounding box* dan diberi label aktivitas yang sesuai menurut definisi kelas yang telah ditentukan sebelumnya. Anotasi diekspor dalam format deteksi objek YOLO, di mana setiap objek direpresentasikan oleh pengidentifikasi kelas bersama dengan koordinat yang dinormalisasi yang menggambarkan posisi tengah, lebar, dan tinggi kotak pembatas.

Anotasi manual yang cermat dilakukan untuk memastikan lokalisasi objek dan pelabelan kelas yang akurat, karena kualitas anotasi secara langsung memengaruhi kemampuan pembelajaran dan akurasi deteksi model deteksi objek berbasis pembelajaran mendalam.

3.4 Model Deteksi Objek Berbasis YOLOv11

Studi ini menggunakan YOLOv11 Nano (YOLOv11n) sebagai arsitektur deteksi objek karena memberikan keseimbangan yang efektif antara akurasi deteksi dan efisiensi komputasi untuk aplikasi pengawasan waktu nyata. YOLOv11 merupakan generasi terbaru dari keluarga YOLO, yang melakukan lokalisasi dan klasifikasi objek secara simultan melalui satu kali proses maju (*forward pass*) pada jaringan saraf. Dibandingkan dengan detektor dua tahap konvensional, strategi deteksi satu tahap ini secara substansial mengurangi kompleksitas komputasi sambil mempertahankan akurasi deteksi yang tinggi, sehingga YOLOv11 cocok untuk sistem pengawasan cerdas yang membutuhkan inferensi cepat. Varian Nano yang ringan dipilih karena menawarkan kebutuhan komputasi yang lebih rendah sambil mempertahankan kinerja deteksi yang kompetitif, sehingga memungkinkan penerapan di masa mendatang pada perangkat keras komputasi terjangkau yang umum tersedia di lembaga pendidikan [16].

3.5 Pengaturan Eksperimental

Pelatihan model dilakukan secara *offline* menggunakan kerangka kerja Ultralytics YOLOv11. Eksperimen dilakukan pada *workstation* yang dilengkapi dengan prosesor Intel Core i7-8700 (Generasi ke-8, hingga 4,60 GHz), RAM DDR4 32 GB, SSD Samsung NVMe (512 GB), dan GPU NVIDIA GeForce RTX 4060 dengan VRAM 8 GB. Lingkungan perangkat lunak terdiri dari Python 3.12.0, kerangka kerja Ultralytics 8.4.78, dan CUDA 12.4 untuk pembelajaran mendalam yang dipercepat GPU.

Model pra-terlatih YOLOv11 Nano (yolo11n.pt) diadopsi sebagai bobot model awal. Pelatihan dilakukan selama 100 epoch menggunakan resolusi gambar masukan 640×640 piksel. *Optimizer* dipilih secara otomatis

oleh kerangka kerja Ultralytics, sedangkan laju pembelajaran awal ditetapkan menjadi 0,01.

Beberapa teknik augmentasi data digunakan selama pelatihan untuk meningkatkan kekokohan model dan mengurangi *overfitting*, termasuk *augmentasi Mosaic*, *augmentasi MixUp*, *augmentasi Copy-Paste*, *augmentasi warna HSV*, pembalikan horizontal, dan rotasi acak. Strategi augmentasi ini meningkatkan keragaman sampel pelatihan dengan memperkenalkan variasi dalam penampilan objek, orientasi, skala, dan pencahayaan, sehingga meningkatkan kemampuan model untuk melakukan generalisasi dalam kondisi pengawasan yang berbeda. Konfigurasi pelatihan lengkap dirangkum dalam Tabel 1.

Tabel 1 Konfigurasi pelatihan model YOLOv11

Parameter	Nilai
Model	YOLOv11 Nano (yolo11n.pt)
Kerangka	Ultralytics 8.4.78
Python	3.12.0
CUDA	12.4
Ukuran gambar masukan	640 × 640 piksel
Jumlah epoch	100
Ukuran batch	14
Tingkat pembelajaran awal	0,01
Pengoptimal	Otomatis (AdamW)
Penambahan data	Mosaik, MixUp, Salin-Tempel, HSV, Balik Horizontal, Rotasi
Prosesor	Intel Core i7-8700 (Generasi ke-8)
Ingatan	RAM DDR4 32 GB
GPU	NVIDIA GeForce RTX 4060 (8 GB VRAM)
Penyimpanan	SSD NVMe Samsung 512 GB

3.6 Evaluasi Kinerja

Performa model yang diusulkan dievaluasi menggunakan metrik deteksi objek standar yang dihitung pada *dataset* validasi. *Precision* mengukur proporsi objek yang terdeteksi dengan benar di antara semua deteksi yang diprediksi dan mencerminkan kemampuan model untuk meminimalkan prediksi positif palsu.

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

Di mana TP menunjukkan jumlah *True Positives* dan FP mewakili *False Positives*.

Recall mengukur proporsi objek yang terdeteksi dengan benar di antara semua objek sebenarnya dan menunjukkan kemampuan model untuk mengidentifikasi aktivitas perundungan yang relevan.

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

Di mana FN menunjukkan *False Negatives* (Kesalahan Negatif).

Performa deteksi objek secara keseluruhan dievaluasi lebih lanjut menggunakan *mean Average Precision* (mAP). Dua protokol evaluasi digunakan: mAP@0.5, yang menggunakan ambang batas *Intersection over Union* (IoU) sebesar 0,50, dan mAP@0.5:0.95, yang merata-ratakan *Average Precision* di seluruh ambang batas IoU yang berkisar dari 0,50 hingga 0,95 dengan interval 0,05. Sementara mAP@0.5 mengukur performa deteksi umum, mAP@0.5:0.95 memberikan penilaian yang lebih ketat terhadap akurasi lokalisasi. Selain metrik kuantitatif ini, proses pembelajaran dipantau menggunakan kurva *loss* pelatihan dan validasi, kurva *Precision-Recall*, dan kurva mAP sepanjang 100 epoch pelatihan untuk mengevaluasi perilaku konvergensi dan stabilitas model selama optimasi.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pelatihan Model

Model deteksi objek berbasis YOLOv11 Nano berhasil dilatih menggunakan *dataset* yang terdiri atas 2.195 citra hasil ekstraksi *frame* dan anotasi manual. Proses pelatihan dilakukan selama 100 epoch menggunakan ukuran citra masukan 640 × 640 piksel dengan memanfaatkan GPU NVIDIA GeForce RTX 4060. Selama proses pelatihan, model menunjukkan konvergensi yang stabil hingga diperoleh bobot terbaik (*best.pt*) yang digunakan pada tahap evaluasi.

Proses pelatihan berlangsung selama sekitar 24 menit 34 detik dengan memanfaatkan teknik

augmentasi data yang meliputi Mosaic, MixUp, Copy-Paste, transformasi HSV, pembalikan horizontal, dan rotasi. Penggunaan teknik augmentasi tersebut bertujuan untuk meningkatkan keragaman data pelatihan sehingga model memiliki kemampuan generalisasi yang lebih baik terhadap berbagai kondisi pengamatan.

4.2 Evaluasi Kinerja Model

Kinerja model dievaluasi menggunakan metrik standar pada deteksi objek, yaitu Precision, Recall, *mean Average Precision* (mAP@0.5), dan mAP@0.5:0.95. Hasil evaluasi ditunjukkan pada Tabel 2.

Tabel 2 Hasil evaluasi model YOLOv11 Nano

Metrik	Nilai
Precision	93,48%
Recall	96,29%
mAP@0.5	96,68%
mAP@0.5:0.95	78,54%

Berdasarkan Tabel 2, model memperoleh nilai Precision sebesar 93,48% dan Recall sebesar 96,29%. Nilai tersebut menunjukkan bahwa model mampu mendeteksi sebagian besar objek target dengan tingkat kesalahan deteksi (*false positive*) yang relatif rendah. Selain itu, model mencapai nilai mAP@0.5 sebesar 96,68% dan mAP@0.5:0.95 sebesar 78,54%, yang menunjukkan kemampuan lokalisasi objek yang baik pada berbagai tingkat *Intersection over Union* (IoU).

Selama proses pelatihan, nilai mAP@0.5 terbaik sebesar 97,72% dicapai pada epoch ke-54, sedangkan nilai mAP@0.5:0.95 terbaik sebesar 79,45% dicapai pada epoch ke-76. Bobot model terbaik (*best.pt*) dipilih secara otomatis berdasarkan performa validasi tersebut dan digunakan sebagai model akhir untuk proses pengujian.

4.3 Analisis Kinerja Berdasarkan Kelas

Selain evaluasi secara keseluruhan, performa model juga dianalisis pada setiap kelas aktivitas untuk mengetahui kemampuan deteksi terhadap masing-masing tindakan perundungan fisik. Hasil evaluasi ditampilkan pada Tabel 3.

Tabel 3 Kinerja deteksi untuk setiap kelas aktivitas

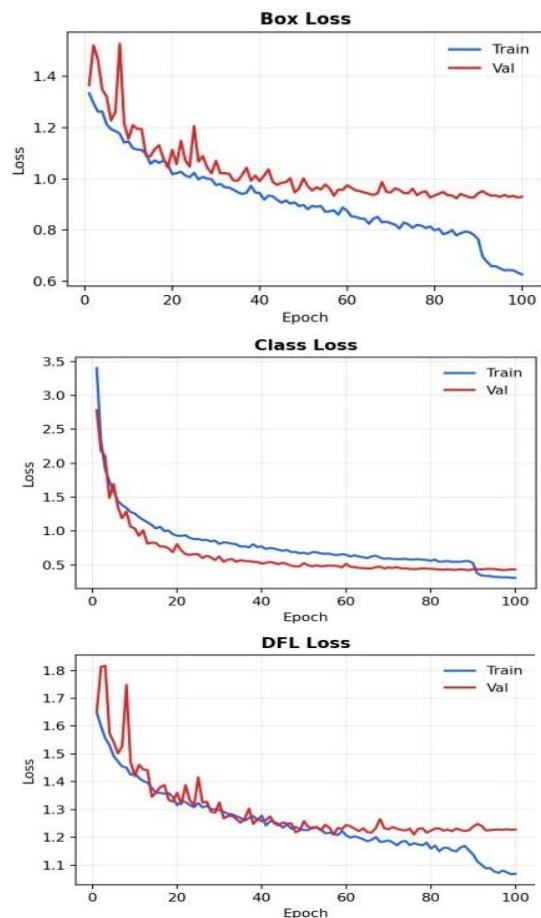
Kelas	Precision	Recall	mAP@0.5	mAP@0.5:0.95
<i>Memukul</i>	96.3	100.0	99,5	94.2
<i>Sepakan</i>	96,5	91.3	98,9	80.1
<i>Menampar</i>	96,6	100.0	99,5	85,7
<i>Goresan</i>	97.3	98.1	98,5	76,5
<i>Mendorong</i>	94,5	89,7	95.1	72.4
<i>Tersandung</i>	99,9	100.0	99,5	84.7
<i>Berkelahi</i>	99,3	100.0	99,5	88.1
<i>Melempar benda</i>	85,8	81.8	92,6	74,9
<i>Menyikut</i>	91,5	98.1	99,2	84,9
<i>Murid</i>	90.4	87,9	91.0	53.1

Berdasarkan Tabel 3, kelas *Memukul*, *Menampar*, *Menjegal*, dan *Berkelahi* memperoleh nilai mAP@0.5 mendekati 100%, menunjukkan bahwa model mampu mengenali aktivitas tersebut dengan tingkat akurasi yang sangat tinggi. Aktivitas-aktivitas tersebut memiliki pola gerakan dan interaksi antarsiswa yang relatif konsisten sehingga lebih mudah dipelajari oleh model.

Sebaliknya, kelas *melempar benda* dan *siswa* menunjukkan nilai mAP@0.5:0.95 yang relatif lebih rendah dibandingkan kelas lainnya. Kondisi ini menunjukkan bahwa lokalisasi objek pada kedua kelas tersebut masih lebih menantang akibat variasi posisi tubuh, arah gerakan, ukuran objek, maupun kondisi tumpang tindih (*occlusion*) antarobjek.

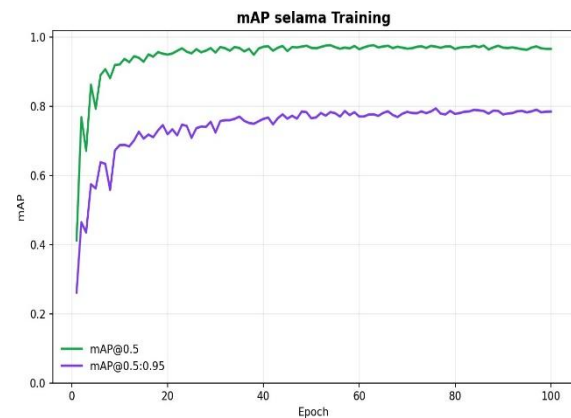
4.4 Analisis Proses Pelatihan

Perkembangan proses pelatihan diamati melalui kurva *loss*, mAP, Precision, Recall, dan *learning rate* yang dihasilkan selama 100 epoch pelatihan.



Gambar 4 Perkembangan nilai box loss, classification loss, dan distribution focal loss (DFL)

Gambar 4 menunjukkan perkembangan nilai *training loss* dan *validation loss*. Ketiga komponen *loss*, yaitu *box loss*, *classification loss*, dan *distribution focal loss* (DFL), mengalami penurunan secara bertahap selama proses pelatihan. Sementara itu, *validation loss* mulai menunjukkan kondisi stabil setelah sekitar epoch ke-50 hingga ke-60. Penurunan *training loss* yang diikuti oleh *validation loss* menunjukkan bahwa model mampu mempelajari pola pada data pelatihan tanpa mengalami peningkatan kesalahan pada data validasi, sehingga tidak terdapat indikasi *overfitting* yang berarti.



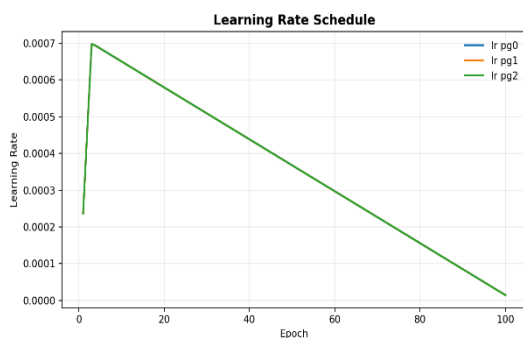
Gambar 5 Perkembangan nilai mAP

Perkembangan nilai mAP ditunjukkan pada Gambar 5. Nilai mAP@0.5 maupun mAP@0.5:0.95 mengalami peningkatan yang cukup cepat pada awal pelatihan, kemudian cenderung stabil setelah mendekati nilai optimum. Hal tersebut menunjukkan bahwa proses optimasi berhasil menemukan parameter model yang stabil sehingga kemampuan generalisasi meningkat pada data validasi.



Gambar 6 Perkembangan nilai Precision dan Recall.

Gambar 6 memperlihatkan perkembangan nilai Precision dan Recall. Kedua metrik tersebut meningkat secara bertahap dan mencapai nilai di atas 90% sejak pertengahan proses pelatihan. Kondisi tersebut menunjukkan keseimbangan yang baik antara kemampuan model dalam mendeteksi objek yang relevan dan meminimalkan kesalahan deteksi.

Gambar 7 Perubahan *Learning Rate* Selama Pelatihan

Perubahan *learning rate* selama pelatihan ditampilkan pada Gambar 7. Nilai *learning rate* mengikuti mekanisme *warm-up* pada tahap awal pelatihan kemudian menurun secara bertahap mengikuti strategi penyesuaian bawaan Ultralytics. Strategi ini membantu proses optimasi berlangsung lebih stabil sehingga model dapat mencapai performa terbaik pada akhir pelatihan.

4.5 Implementasi Model pada Citra Uji

Untuk mengevaluasi kemampuan model secara visual, dilakukan pengujian menggunakan beberapa citra yang tidak digunakan pada proses pelatihan. Hasil pengujian menunjukkan bahwa model mampu mendeteksi tindakan perundungan fisik sekaligus menentukan lokasi objek melalui *bounding box* beserta nilai *confidence* untuk setiap prediksi.

Gambar 8 Model berhasil mendeteksi aktivitas *memukul* dan *menendang*

Sebagai contoh, pada Gambar 8a model berhasil mendeteksi aktivitas *memukul* dengan tingkat *confidence* sebesar 94%, sedangkan pada Gambar 8b model berhasil mendeteksi aktivitas *menendang* dengan tingkat *confidence* sebesar 92%. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi dan lokalisasi objek secara bersamaan pada berbagai aktivitas perundungan fisik yang terdapat pada citra uji.

4.9 Pembahasan

Hasil eksperimen menunjukkan bahwa model deteksi objek berbasis YOLOv11 yang dilatih dapat secara efektif mengenali berbagai kategori aktivitas perundungan fisik di lingkungan sekolah. Model ini mencapai Presisi 93,48%, *Recall* 96,29%, mAP@0,5 sebesar 96,68%, dan mAP@ 0,5 : 0,95 sebesar 78,54%, yang menunjukkan kemampuan deteksi yang sangat baik di berbagai interaksi fisik. Temuan ini menunjukkan bahwa pendekatan yang diusulkan mampu membedakan perilaku agresif dari aktivitas siswa normal secara akurat sambil mempertahankan kinerja lokalisasi yang andal. Nilai *Recall* yang tinggi sangat penting untuk aplikasi pengawasan sekolah karena melewatkan insiden perundungan yang sebenarnya (negatif palsu) dapat menunda intervensi oleh guru atau administrator sekolah. Sebaliknya, presisi yang tinggi menunjukkan bahwa sistem menghasilkan relatif sedikit alarm palsu [17], sehingga mengurangi pemberitahuan yang tidak perlu dan meningkatkan kepraktisan penerapan di lingkungan pendidikan dunia nyata.

Performa deteksi yang tinggi dapat dikaitkan dengan beberapa faktor. Pertama, *dataset* khusus dikembangkan secara spesifik untuk skenario perundungan fisik di lingkungan sekolah, alih-alih mengandalkan *dataset* yang tersedia secara publik yang dikumpulkan dari domain yang tidak terkait. *Dataset* tersebut berisi sepuluh kelas aktivitas dan menggabungkan variasi sudut pandang kamera, orientasi tubuh, pola interaksi, dan kondisi pencahayaan, sehingga memungkinkan model untuk mempelajari fitur visual representatif untuk berbagai perilaku perundungan. Kedua, penerapan teknik augmentasi Mosaic, MixUp, dan Copy-Paste meningkatkan keragaman data pelatihan [18] sehingga model lebih mampu mengenali variasi pose dan interaksi antarsiswa,

yang berkontribusi terhadap tingginya nilai Recall. Terakhir, arsitektur YOLOv11 Nano memberikan keseimbangan yang efisien antara efisiensi komputasi dan akurasi deteksi, sehingga sangat cocok untuk aplikasi pengawasan waktu nyata.

Meskipun kinerja keseluruhan sangat baik, variasi diamati di antara kelas aktivitas individu. Kelas *memukul*, *menampar*, *menjegal*, dan *berkelahi* mencapai kinerja deteksi tertinggi, dengan nilai mAP@0.5 mendekati 100%. Aktivitas-aktivitas ini melibatkan gerakan tubuh yang khas dan interaksi fisik yang dekat yang menghasilkan pola spasial yang konsisten, memungkinkan model untuk mempelajari fitur visual yang sangat diskriminatif. Sebaliknya, kelas *melempar benda* dan *siswa* memperoleh kinerja lokalisasi yang relatif lebih rendah, terutama di bawah kriteria evaluasi mAP@0.5 :0.95 yang lebih ketat. Kelas *melempar benda* berisi gerakan tubuh yang cepat disertai dengan objek yang bergerak di sepanjang lintasan yang berbeda, meningkatkan variabilitas intra-kelas dan membuat lokalisasi yang tepat menjadi lebih menantang. Sementara itu, kelas *siswa* mewakili perilaku kelas normal dengan keragaman yang substansial dalam postur, skala, orientasi, dan oklusi. Karakteristik ini meningkatkan kompleksitas visual dan secara alami mengurangi akurasi lokalisasi meskipun mempertahankan kinerja deteksi keseluruhan yang memuaskan.

Hasil yang diperoleh dalam penelitian ini konsisten dengan penelitian sebelumnya yang menunjukkan efektivitas arsitektur berbasis YOLO untuk pengawasan kelas dan analisis aktivitas manusia. Chen dkk. (2023) mengusulkan model YOLOv8 yang ditingkatkan untuk deteksi perilaku di kelas dengan mengintegrasikan modul Res2Net, *multi-head self-attention* (MHSA), dan *efficient multi-scale attention* (EMA). Pendekatan mereka meningkatkan deteksi perilaku di kelas dalam kondisi yang melibatkan populasi siswa yang padat dan oklusi objek, menyoroti kesesuaian detektor objek berbasis YOLO untuk lingkungan pendidikan [19].

Berbeda dengan penelitian Chen dkk., yang berfokus terutama pada perilaku umum di kelas seperti mengangkat tangan dan membaca, penelitian ini secara khusus membahas deteksi perundungan fisik, yang melibatkan interaksi

interpersonal yang kompleks seperti *memukul*, *menendang*, *mendorong*, *berkelahi*, dan *menyikut*. Tindakan-tindakan ini jauh lebih sulit dikenali karena melibatkan banyak individu yang berinteraksi secara bersamaan dengan tumpang tindih tubuh yang substansial dan gerakan dinamis. Meskipun demikian, model YOLOv11 yang diusulkan mencapai akurasi deteksi yang sangat baik tanpa memerlukan modifikasi arsitektur di luar implementasi standar YOLOv11 Nano. Temuan ini menunjukkan bahwa *dataset* yang disusun dengan cermat dan anotasi berkualitas tinggi dapat secara substansial meningkatkan kinerja deteksi bahkan ketika menggunakan arsitektur deteksi objek yang ringan [20].

Temuan saat ini juga sebanding dengan yang dilaporkan oleh kerangka kerja Violence-YOLO yang diusulkan oleh Xu et al. yang meningkatkan arsitektur GELAN untuk meningkatkan deteksi kekerasan dalam skenario pengawasan publik [21]. Studi mereka menunjukkan bahwa peningkatan arsitektur dapat meningkatkan akurasi deteksi di lingkungan yang kompleks. Namun, tidak seperti Violence-YOLO, yang menargetkan kekerasan umum di tempat umum seperti bandara dan fasilitas transportasi, model yang diusulkan dilatih secara khusus pada skenario perundungan fisik berbasis sekolah. Akibatnya, fitur visual yang dipelajari lebih representatif terhadap interaksi yang umum diamati di antara siswa di lingkungan pendidikan, sehingga meningkatkan relevansi praktis dari pendekatan yang diusulkan untuk pengawasan sekolah yang cerdas.

Studi relevan lainnya dilakukan oleh Sepulau dkk. yang mengembangkan sistem pemantauan kelas berbasis YOLOv8 untuk mendeteksi perilaku siswa seperti makan dan minum selama perkuliahan. Hasil mereka menunjukkan bahwa deteksi objek berbasis YOLO dapat beroperasi secara efisien pada platform komputasi edge sambil mempertahankan akurasi yang memuaskan [22]. Meskipun perilaku target mereka berbeda secara substansial dari perundungan fisik, kedua studi tersebut menunjukkan bahwa arsitektur YOLO yang ringan mampu mengenali aktivitas manusia di lingkungan kelas yang ramai dengan kecepatan yang cukup untuk penerapan secara *real-time*.

Dibandingkan dengan penelitian sebelumnya, kontribusi utama penelitian ini terletak tidak hanya pada akurasi deteksi yang dicapai, tetapi juga pada pengembangan *dataset* perundungan fisik khusus yang dirancang khusus untuk lingkungan sekolah. *Dataset* publik yang ada untuk deteksi kekerasan biasanya berisi video yang dikumpulkan dari film, acara olahraga, atau rekaman pengawasan publik, yang sangat berbeda dari interaksi di kelas. Dengan membangun *dataset* yang secara akurat mewakili perilaku siswa yang sebenarnya dalam lingkungan pendidikan, model yang diusulkan mempelajari pola visual yang secara langsung relevan dengan aplikasi pengawasan sekolah. *Dataset* khusus domain ini merupakan kontribusi penting karena ketersediaan data pelatihan yang representatif tetap menjadi salah satu tantangan utama dalam mengembangkan sistem deteksi perundungan berbasis AI yang praktis.

Terlepas dari hasil yang menggembirakan, beberapa keterbatasan perlu diakui. Pertama, *dataset* dikumpulkan melalui simulasi terkontrol, bukan dari insiden perundungan yang terjadi secara alami. Meskipun simulasi memungkinkan pengumpulan data yang aman dan dapat diterima secara etis, perilaku perundungan yang sebenarnya mungkin menunjukkan variabilitas yang lebih besar dalam pola gerakan, respons peserta, dan kondisi lingkungan. Kedua, model saat ini melakukan deteksi objek pada setiap *frame* gambar dan oleh karena itu tidak secara eksplisit memodelkan hubungan temporal antar *frame* yang berurutan. Menggabungkan informasi temporal melalui model pengenalan aksi berbasis video atau teknik pelacakan multi-objek dapat lebih meningkatkan pengenalan interaksi perundungan yang kompleks. Terakhir, ukuran *dataset* masih relatif kecil dibandingkan dengan tolok ukur deteksi objek skala besar, menunjukkan bahwa studi di masa mendatang dapat memperoleh manfaat dari perluasan *dataset* dengan sekolah tambahan, posisi kamera, kondisi pencahayaan, dan demografi peserta.

Oleh karena itu, penelitian selanjutnya harus fokus pada pengumpulan *dataset* multi-sekolah yang lebih besar, mengintegrasikan pengenalan aksi temporal dengan deteksi objek, dan mengevaluasi model yang diusulkan menggunakan rekaman CCTV nyata yang

diperoleh dari sistem pengawasan sekolah operasional. Menggabungkan deteksi objek YOLOv11 dengan pelacakan multi-objek dan analisis peristiwa cerdas dapat lebih meningkatkan kemampuan untuk secara otomatis mengidentifikasi interaksi agresif yang berkelanjutan sekaligus mengurangi alarm palsu. Perkembangan tersebut akan berkontribusi pada terciptanya sistem pengawasan cerdas yang mampu mendukung intervensi dini dan meningkatkan keselamatan siswa di lingkungan pendidikan.

5. KESIMPULAN

Penelitian ini berhasil mengembangkan model deteksi objek berbasis YOLOv11 Nano untuk mendeteksi perundungan fisik secara otomatis di lingkungan sekolah. Untuk mendukung pengembangan model, penelitian ini juga menghasilkan *dataset* khusus yang dibangun melalui simulasi interaksi siswa dan mencakup sembilan kategori tindakan perundungan fisik serta satu kategori aktivitas normal. *Dataset* tersebut diproses melalui tahap ekstraksi *frame* dan anotasi manual menggunakan Label Studio sebelum digunakan pada proses pelatihan model.

Hasil evaluasi menunjukkan bahwa model yang diusulkan mencapai nilai Precision sebesar 93,48%, Recall sebesar 96,29%, mAP@0.5 sebesar 96,68%, dan mAP@0.5:0.95 sebesar 78,54%. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali berbagai tindakan perundungan fisik sekaligus melakukan lokalisasi objek secara akurat. Dengan memanfaatkan arsitektur YOLOv11 Nano yang ringan, model juga memiliki potensi untuk diimplementasikan pada sistem pengawasan video cerdas yang memerlukan deteksi objek secara *real-time*.

Kontribusi utama penelitian ini terletak pada pengembangan *dataset* perundungan fisik yang spesifik untuk lingkungan sekolah serta penerapan model deteksi berbasis YOLOv11 Nano yang mampu mencapai kinerja tinggi tanpa memerlukan modifikasi arsitektur yang kompleks. Temuan ini menunjukkan bahwa kombinasi antara *dataset* yang representatif dan model deteksi objek yang efisien dapat menjadi solusi yang menjanjikan untuk mendukung deteksi dini perundungan fisik dan membantu

pengambilan tindakan yang lebih cepat di lingkungan pendidikan.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. *Dataset* yang digunakan diperoleh melalui skenario simulasi sehingga belum sepenuhnya merepresentasikan variasi kondisi nyata di lingkungan sekolah. Selain itu, model yang dikembangkan masih melakukan deteksi berbasis *frame* tanpa mempertimbangkan hubungan temporal antarframe. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas *dataset* menggunakan rekaman dari berbagai sekolah dengan kondisi yang lebih beragam serta mengintegrasikan metode analisis temporal, seperti action recognition dan multi-object tracking, agar sistem mampu mengenali rangkaian perilaku perundungan secara lebih akurat. Evaluasi menggunakan rekaman CCTV nyata juga diperlukan untuk menguji ketahanan model sebelum diimplementasikan pada lingkungan operasional.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan selama pelaksanaan penelitian ini. Apresiasi yang sebesar-besarnya disampaikan kepada rekan-rekan peneliti yang telah berkontribusi dalam proses pembuatan dan anotasi *dataset* video. Penulis juga menyampaikan terima kasih kepada seluruh sivitas akademika, Kepala Madrasah, dan Komite MTsN 6 Malang atas dukungan, fasilitas, serta pendanaan yang diberikan sehingga penelitian ini dapat terlaksana dengan baik.

DAFTAR PUSTAKA

- [1] UNICEF, "Half of world's teens experience peer violence in and around school – UNICEF." Accessed: Aug. 21, 2025. [Online]. Available: https://www.unicef.org/eca/press-releases/half-worlds-teens-experience-peer-violence-and-around-school-unicef?utm_source=chatgpt.com
- [2] M. D. Pratama, V. W. Wati, F. Fadhilah, S. Syarifuddin, and S. Safitri, "The effects of bullying on early childhood emotions and its impact on psychological development," *J. Ilm. Konseling Pendidik.*, vol. 4, no. 2, 2025.
- [3] F. M. Rodzi, "Penanganan Kasus Kekerasan Seksual Pada Anak Oleh Pemerintah Kabupaten Bima," PhD Thesis, Sekolah Tinggi Pembangunan Masyarakat Desa" APMD", 2025.
- [4] S. Yahya and others, "Kajian pemanfaatan deep learning dalam pembelajaran," *transformasi*, vol. 7, no. 1, pp. 25–41, 2025.
- [5] D. H. Saputra, B. Imran, and others, "Object Detection Untuk Mendeteksi Citra Buah-Buahan Menggunakan Metode Yolo," *J. Kecerdasan Buatan Dan Teknol. Inf.*, vol. 2, no. 2, pp. 70–80, 2023.
- [6] M. F. Golfantara, "Penggunaan Algoritma Yolo V8 Untuk Identifikasi Rempah-Rempah," *J. Inform. Dan Tek. Elektro Terap.*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5221.
- [7] M. Y. N. Hakim, "Pemodelan Machine Learning Untuk Deteksi Kekerasan Fisik Berbasis Convolutional Neural Network," 2025.
- [8] A. Nurlita and M. Munawaroh, "Pengembangan Chatbot Dengan Metode Natural Language Processing Untuk Layanan Pelanggan (Studi Kasus PT Masterlink Internet Solution)," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 3S1, Oct. 2025, doi: 10.23960/jitet.v13i3S1.8176.
- [9] I. Zaenudin and A. B. Riyan, "Perkembangan kecerdasan buatan (AI) dan dampaknya pada dunia teknologi," *J. Inform. Utama*, vol. 2, no. 2, pp. 128–153, 2024.
- [10] H. A. Pratiwi, M. Cahyanti, and M. Lamsani, "Implementasi deep learning flower scanner menggunakan metode convolutional neural network," *J. Inform.*, vol. 2, no. 1, pp. 41–59, 2021.
- [11] S. S. Rambe, A. Asriyanik, and P. Prajoko, "Penerapan Model Convolutional Neural Network (CNN) Berbasis Mobilenetv2 untuk Klasifikasi Tingkat Kesegaran Ikan Nila," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 3, Jul. 2025, doi: 10.23960/jitet.v13i3.6744.
- [12] N. Khairunisa, . C., and A. Jamaludin, "Analisis Perbandingan Algoritma CNN dan Yolo Dalam Mengidentifikasi Kerusakan Jalan," *J. Inform. Dan Tek. Elektro Terap.*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4434.
- [13] G. Agustiguna, "Deteksi Tindak Kekerasan Pada Anak Di Sekolah Menggunakan Yolo V8 Untuk Meningkatkan Keamanan Lingkungan Pendidikan Berbasis Web," PhD Thesis, Universitas Muhamamdiyah Ponorogo, 2025.
- [14] D. E. P. Sahtio, M. A. P. Putra, and I. B. K. Sudiarmika, "Performance Evaluation and Comparison of YOLO11 Model Variants for Indonesian Sign Language (BISINDO)," *J. Inform. Dan Tek. Elektro Terap.*, vol. 14, no. 2, Apr. 2026, doi: 10.23960/jitet.v14i2.9348.

- [15] I. A. Zulkarnain, "Yolov11 Optimization Through Hyperparameter Tuning and Data Augmentation to Improve Vehicle Detection Accuracy at Night," vol. 5, 2025.
- [16] V. B. Djusmin, E. Sukma, and F. E. Susilawati, "Analisis Kinerja Model Yolov12 Nano Terhadap Ketidakseimbangan Data pada Deteksi Tandan Kelapa Sawit di Lingkungan Perkebunan," vol. 15, 2025.
- [17] M. S. Başarslan and H. Canlı, "UAV-based forest fire early warning and intervention simulation system with high-accuracy hybrid AI model," *Appl. Sci.*, vol. 16, no. 3, p. 1201, 2026.
- [18] A. Zamani and Z. Abedini, "Optimizing Data Augmentation for Real-Time Small UAV Detection: A Lightweight Context-Aware Approach," *ArXiv Prepr. ArXiv260419999*, 2026.
- [19] H. Chen, G. Zhou, and H. Jiang, "Student Behavior Detection in the Classroom Based on Improved YOLOv8," *Sensors*, vol. 23, no. 20, p. 8385, Oct. 2023, doi: 10.3390/s23208385.
- [20] N. F. Y. P. Putra, A. Kurniawardhani, and M. Kom, "Analisis Pengaruh Kelengkapan Anotasi terhadap Kinerja Model YOLOv11 dalam Deteksi Multi Objek Buah," *AUTOMATA*, vol. 6, no. 2, 2025.
- [21] W. Xu, D. Zhu, R. Deng, K. Yung, and A. W. H. Ip, "Violence-YOLO: Enhanced GELAN Algorithm for Violence Detection," *Appl. Sci.*, vol. 14, no. 15, p. 6712, Jan. 2024, doi: 10.3390/app14156712.
- [22] I. I. Sepulau, K. Kusumanto, and N. L. Husni, "Student Behavior Monitoring System in Classroom Environment Using YOLOv8," *Indones. J. Artif. Intell. Data Min.*, vol. 8, no. 2, p. 437, Aug. 2025, doi: 10.24014/ijaidm.v8i2.37086.