

Deteksi Ekspresi Afektif Mahasiswa Pada Sistem Pembelajaran Pemrograman Menggunakan Yolov13n

Verell Hermawan^{1*}, Sawali Wahyu²

^{1,2}Fakultas Ilmu Komputer, Teknik Informatika, Universitas Esa Unggul, Jakarta, Indonesia

E-mail: ^{1*}varellhmwn@student.esaunggul.ac.id, ²sawaliwahyu@esaunggul.ac.id

(* : corresponding author)

Abstrak

Pembelajaran pemrograman menuntut proses pemecahan masalah berulang yang dapat memengaruhi kondisi afektif mahasiswa, tetapi perubahan tersebut sulit diamati secara kontinu melalui observasi manual atau laporan diri. Penelitian ini mengembangkan sistem Facial Emotion Recognition berbasis YOLOv13n untuk mendeteksi empat kondisi afektif, yaitu engagement, confusion, frustration, dan boredom, secara real-time serta mengintegrasikan hasil deteksi dengan aktivitas pembelajaran pada aplikasi web. Deteksi dilakukan saat mahasiswa membaca modul dan mengerjakan evaluasi berupa kuis pilihan ganda atau latihan kode interaktif. Model di-fine-tuning menggunakan dataset gabungan sebanyak 1.660 citra, terdiri atas 953 citra dari Roboflow Universe dan 707 hard samples, dengan pembagian data 80:10:10 secara terstratifikasi. Evaluasi terhadap 173 citra pengujian menghasilkan precision 0,994, recall 0,982, F1-score 0,988, mAP@0.5 0,994, dan mAP@0.5:0.95 0,982, dengan waktu inferensi 6,2 ms per citra. Penerapan confidence filtering, sliding window, dan majority voting menurunkan perubahan label sebesar 77,20%, sehingga meningkatkan kestabilan temporal. Pengujian black-box pada 27 skenario menunjukkan seluruh fungsi aplikasi berjalan sesuai rancangan. Sistem menyediakan informasi deskriptif berupa skor, durasi sesi, distribusi ekspresi, ekspresi dominan, dan timeline, bukan diagnosis psikologis. Hasil menunjukkan potensi sistem sebagai pendukung pemantauan pola ekspresi afektif dalam pembelajaran pemrograman.

Kata kunci: Ekspresi Afektif, Facial Emotion Recognition, YOLOv13n, Learning Analytics, Pembelajaran Pemrograman

Abstract

Programming learning requires repeated problem-solving processes that may influence students' affective states. However, these changes are difficult to continuously observe through manual observation or self-reporting. This study develops a YOLOv13n-based Facial Emotion Recognition system to detect four affective states—engagement, confusion, frustration, and boredom—in real time and integrate the detected results with learning activities in a web-based application. Detection is performed while students read learning modules and complete assessments through multiple-choice quizzes or interactive coding exercises. The model was fine-tuned using a combined dataset of 1,660 images, consisting of 953 images from Roboflow Universe and 707 hard samples, with stratified training, validation, and testing splits of 80:10:10. Evaluation on 173 test images achieved a precision of 0.994, recall of 0.982, F1-score of 0.988, mAP@0.5 of 0.994, and mAP@0.5:0.95 of 0.982, with an inference time of 6.2 ms per image. The application of confidence filtering, sliding window, and majority voting reduced label changes by 77.20%, improving temporal stability. Black-box testing across 27 scenarios confirmed that all application functions operated as designed. The system provides descriptive indicators rather than psychological diagnosis, demonstrating potential for monitoring affective expression patterns during programming learning.

Keywords: Affective Expressions, Facial Emotion Recognition, YOLOv13n, Learning Analytics, Programming Learning

1. PENDAHULUAN

Pembelajaran pemrograman tidak hanya menuntut penguasaan sintaks, tetapi juga melibatkan penalaran logis, pemahaman konsep, pengujian solusi, *debugging*, dan perbaikan solusi secara berulang. Selama proses tersebut, mahasiswa dapat mengalami perubahan kondisi afektif yang menyertai aktivitas belajar. Keterlibatan dapat muncul ketika mahasiswa memahami permasalahan dan mampu melanjutkan proses penyelesaian tugas, sedangkan *confusion* dapat muncul ketika mahasiswa menghadapi konsep atau kesalahan yang belum dipahami. Apabila hambatan berlangsung secara berulang, kondisi tersebut dapat berkembang menjadi *frustration*, sedangkan penurunan keterlibatan terhadap aktivitas dapat berkaitan dengan *boredom*. Dinamika kondisi afektif selama mahasiswa mengerjakan permasalahan pemrograman telah dilaporkan

dalam penelitian sebelumnya [1]. Dalam *Control-Value Theory*, respons afektif dalam aktivitas akademik juga berkaitan dengan penilaian individu terhadap tingkat kendali atas aktivitas dan nilai yang diberikan terhadap aktivitas tersebut [2].

Dalam penelitian ini, istilah pembelajaran pemrograman merujuk pada sesi pembelajaran berbasis web yang terdiri atas pembacaan materi, kuis pilihan ganda, dan latihan kode terstruktur. Aplikasi menyediakan tiga modul, yaitu Pengenalan Machine Learning, Deep Learning & Neural Networks, serta Computer Vision & Object Detection. Setelah membaca materi, mahasiswa dapat memilih evaluasi berupa sepuluh soal pilihan ganda yang mencakup soal konsep dan *problem solving*, atau lima latihan kode berbasis Python dalam bentuk *fill-in-the-blank*. Pada latihan kode, mahasiswa diminta melengkapi bagian sintaks atau nilai yang dikosongkan pada potongan program berdasarkan materi yang telah dipelajari. Dengan demikian, aktivitas yang diamati mencakup pembelajaran konseptual dan latihan pemrograman terstruktur, sehingga tidak seluruh durasi sesi merupakan aktivitas menulis kode secara bebas. Deteksi ekspresi berlangsung selama rangkaian aktivitas pembelajaran tersebut.

Pengukuran kondisi afektif dalam pembelajaran umumnya dilakukan melalui kuesioner, laporan diri, atau observasi pengajar. Pendekatan tersebut tetap penting, tetapi memiliki keterbatasan untuk menangkap perubahan yang berlangsung dalam waktu singkat ketika mahasiswa berinteraksi dengan komputer. *Facial Emotion Recognition* (FER) menawarkan pendekatan berbasis *computer vision* untuk mengolah informasi visual pada wajah dari citra atau video. Studi pada pembelajaran sains menunjukkan bahwa ekspresi wajah dapat digunakan sebagai salah satu indikator untuk menelusuri *emotional engagement* [3], sedangkan penelitian pada kelas daring menunjukkan bahwa perubahan ekspresi wajah dapat menangkap sebagian dinamika *engagement* mahasiswa [4]. Analisis multimodal pada aktivitas pemrograman juga menunjukkan bahwa keadaan kognitif dan afektif dapat berubah sepanjang proses pemecahan masalah [5], sedangkan pendekatan multimodal lain menggabungkan beberapa sumber informasi untuk menilai *engagement* [6]. Namun, pendekatan multimodal memerlukan lebih banyak sumber data dan alur pemrosesan. Oleh karena itu, FER dalam penelitian ini digunakan untuk memperoleh indikator ekspresi visual yang diasosiasikan dengan kondisi afektif, bukan untuk mengukur atau mendiagnosis keadaan psikologis mahasiswa secara langsung.

Empat kategori yang digunakan dalam penelitian ini, yaitu *engagement*, *confusion*, *frustration*, dan *boredom*, dipilih berdasarkan relevansinya terhadap dinamika aktivitas pembelajaran dan pemecahan masalah. Keempat kondisi tersebut juga digunakan sebagai kategori afektif pada DAiSEE, yaitu dataset *in-the-wild* untuk analisis kondisi afektif pengguna pada lingkungan digital yang menggunakan anotasi *boredom*, *confusion*, *engagement*, dan *frustration* [7]. Penggunaan kategori berbasis kondisi afektif dalam konteks pembelajaran juga didukung oleh dataset dan penelitian sebelumnya. DAiSEE menggunakan empat dimensi *boredom*, *confusion*, *engagement*, dan *frustration* pada video pengguna dalam lingkungan digital [7], sedangkan Xu et al. [8] menggunakan ekspresi wajah dan gestur pada aktivitas pembelajaran daring untuk mengenali kondisi *learning-centered* seperti *engaged*, *confused*, dan *frustrated*. Dalam penelitian ini, kategori tersebut digunakan sebagai label operasional pola ekspresi visual dan tidak diperlakukan sebagai pengukuran langsung terhadap kondisi psikologis mahasiswa. *Engagement* berkaitan dengan keterlibatan mahasiswa terhadap aktivitas pembelajaran dan telah dikaji menggunakan informasi ekspresi wajah dalam konteks pendidikan [3], [4]. Dalam aktivitas pemecahan masalah, *confusion* dapat muncul ketika mahasiswa menghadapi konsep, informasi, atau kesalahan yang belum dipahami [1], [5]. *Confusion* dapat menjadi bagian dari proses pencarian pemahaman, tetapi hambatan yang berlangsung dan tidak terselesaikan dapat berkembang menjadi *frustration* [5], [9]. Sementara itu, *boredom* berkaitan dengan menurunnya keterlibatan atau nilai yang dirasakan terhadap aktivitas belajar dan termasuk dalam pembahasan emosi akademik pada *Control-Value Theory* [2]. Dalam penelitian ini, keempat kategori tersebut digunakan sebagai label operasional ekspresi afektif berbasis visual yang dipelajari model dari data beranotasi, bukan sebagai penentuan langsung terhadap keadaan psikologis mahasiswa.

Keluarga You Only Look Once (YOLO) menarik untuk kebutuhan real-time karena mendeteksi lokasi objek dan kelas dalam satu proses. YOLOv5 telah diterapkan untuk mendeteksi

tiga kategori ekspresi mahasiswa dengan mAP terbaik 89,60% [10]. Perbandingan varian YOLOv8 pada ekspresi wajah mahasiswa menunjukkan bahwa model ringan dapat mencapai waktu inferensi rendah, dengan YOLOv8s menghasilkan mAP 0,840 [11]. Model YOLO ringan lain juga dilaporkan kompetitif untuk deteksi ekspresi [12], sedangkan YOLOv13 memperkenalkan Hypergraph-based Adaptive Correlation Enhancement dan Full-Pipeline Aggregation-and-Distribution untuk meningkatkan representasi fitur sekaligus mempertahankan efisiensi [13]. Penerapan YOLO secara real-time juga telah digunakan pada konteks object detection lain, dengan evaluasi menggunakan mAP, precision, recall, F1-score, dan FPS [14]. Di luar keluarga YOLO, penelitian terkait mencakup FER pada pembelajaran sains dan kelas daring [3], [4], analisis multimodal pada aktivitas pemrograman dan engagement [5], [6], serta sistem visualisasi keadaan afektif secara real-time [15], [16]. Ringkasan posisi penelitian terhadap studi terkait disajikan pada Tabel 1.

Tabel 1. Perbandingan penelitian terkait dan posisi penelitian

Referensi	Konteks/Model	Hasil utama	Kesenjangan terhadap penelitian ini
Tang et al. [3]	FER pada pembelajaran sains	Ekspresi wajah digunakan untuk menelusuri emotional engagement	Belum mengintegrasikan latihan pemrograman dan object detection real-time
Hu dan Gao [4]	FER pada kelas daring	FER menangkap perubahan engagement secara dinamis	Fokus pada korelasi, bukan aplikasi pembelajaran terpadu
Mangaroska et al. [5]	Multimodal pada aktivitas pemrograman	Mengamati perubahan keadaan kognitif dan afektif saat problem solving	Memerlukan beberapa sumber data dan alur analisis yang lebih kompleks
Yan et al. [6]	Multimodal deep learning	Penilaian engagement menggunakan beberapa sumber data	Kebutuhan komputasi dan sensor lebih kompleks untuk real-time
Ulandari et al. [10]	YOLOv5, tiga kelas	mAP terbaik 89,60%	Generasi YOLO dan kategori ekspresi masih terbatas
Tresnawati et al. [11]	Perbandingan YOLOv8	YOLOv8s: mAP 0,840 dan inferensi 2,4 ms	Belum menguji YOLOv13n dan integrasi dashboard pembelajaran
Aulia et al. [12]	Varian YOLO ringan untuk FER	Menunjukkan kelayakan model ringan untuk deteksi ekspresi	Belum berfokus pada empat emosi belajar dan aktivitas pemrograman
Lei et al. [13]	YOLOv13 pada deteksi objek umum	HyperACE dan FullPAD meningkatkan representasi fitur	Belum dievaluasi untuk ekspresi afektif mahasiswa
Hasnine et al. [15]	Dashboard afektif real-time	Mendeteksi dan menampilkan keadaan afektif pelajar daring	Tidak mengintegrasikan YOLOv13n dengan kuis pemrograman
Harb et al. [16]	Pengenalan emosi siswa jarak jauh	Deteksi emosi dan visualisasi untuk pemantauan pembelajaran	Belum menggabungkan latihan pemrograman, pasca-pemrosesan temporal, dan skor

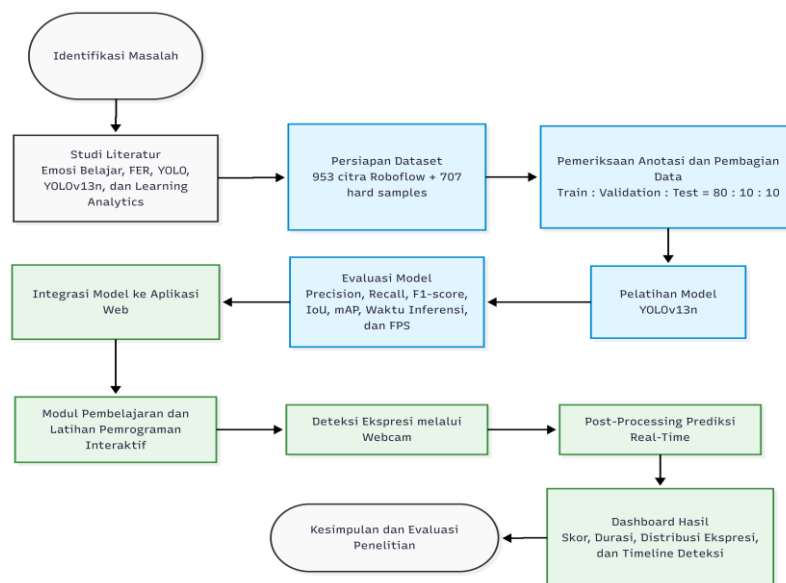
Interpretasi keluaran FER perlu dilakukan secara hati-hati. Kinerja detektor berbasis *computer vision* dapat dipengaruhi oleh representasi fitur dan variasi data [17]. Oklusi pada wajah dapat menurunkan reliabilitas pengenalan ekspresi [18], sedangkan ketidakseimbangan demografis dan perbedaan kelompok regional berpotensi menyebabkan perbedaan performa model [19], [20]. Selain itu, pose kepala, arah pandangan, pencahayaan, dan gestur di sekitar wajah dapat menghasilkan pola visual yang menyerupai kategori tertentu. Oleh karena itu, label *engaged*, *confused*, *bored*, dan *frustrated* pada penelitian ini diposisikan sebagai hasil klasifikasi pola ekspresi visual berdasarkan data yang dipelajari model. Keluaran tersebut tidak dimaknai sebagai diagnosis psikologis maupun sebagai bukti bahwa kondisi internal mahasiswa selalu sesuai dengan label yang dihasilkan sistem.

Selain performa model, penyajian hasil perlu mendukung keterbacaan informasi yang dihasilkan. Dashboard *real-time* dapat membantu menampilkan keadaan afektif pembelajar [15], tetapi desain dashboard perlu memperhatikan dukungannya terhadap proses belajar [21] dan tidak berhenti pada penyajian analitik semata [22]. Pengembangan dashboard yang berorientasi pada mahasiswa juga menekankan pentingnya informasi yang mudah dipahami dan relevan terhadap

aktivitas pengguna [23]. Berdasarkan kesenjangan tersebut, penelitian ini mengembangkan dan mengevaluasi YOLOv13n untuk mengklasifikasikan pola ekspresi visual ke dalam empat kategori afektif, yaitu *engagement*, *confusion*, *frustration*, dan *boredom*. Model diintegrasikan dengan sesi pembelajaran berbasis web yang mencakup pembacaan materi, kuis pilihan ganda, dan latihan kode Python terstruktur. Prediksi *real-time* kemudian melalui pasca-pemrosesan temporal untuk mengurangi perubahan label berdurasi singkat sebelum hasilnya disajikan bersama skor, durasi sesi, distribusi ekspresi, ekspresi dominan, dan *timeline*. Penelitian ini berfokus pada performa model, kestabilan temporal hasil deteksi, dan integrasi fungsi aplikasi, serta tidak digunakan untuk membuktikan dampak sistem terhadap hasil belajar maupun menentukan keadaan psikologis mahasiswa secara langsung.

2. METODE PENELITIAN

Alur penelitian secara keseluruhan tergambar dalam kerangka penelitian pada Gambar 1 berikut:



Gambar 1. Kerangka Penelitian

Gambar 1 menunjukkan alur metodologi penelitian yang dimulai dari identifikasi masalah dan studi literatur, kemudian dilanjutkan dengan persiapan dataset, pemeriksaan anotasi, pembagian data, serta pelatihan model YOLOv13n. Model yang telah dilatih dievaluasi menggunakan metrik seperti Precision, Recall, F1-score, IoU, mAP, waktu inferensi, dan FPS. Selanjutnya, model diintegrasikan ke dalam aplikasi web untuk mendukung pembelajaran pemrograman interaktif melalui deteksi ekspresi wajah secara real-time menggunakan webcam. Hasil deteksi diproses pada tahap post-processing dan ditampilkan dalam dashboard yang menyajikan skor, durasi, distribusi ekspresi, serta timeline deteksi sebagai bagian dari Learning Analytics, sebelum dilakukan evaluasi dan penarikan kesimpulan penelitian.

2.1. Rancangan Penelitian dan Dataset

Penelitian menggunakan pendekatan pengembangan sistem yang mencakup identifikasi masalah, studi literatur, persiapan dataset, *fine-tuning* YOLOv13n, evaluasi model, integrasi aplikasi web, penerapan pasca-pemrosesan temporal, dan pengujian fungsional. Sistem dikembangkan sebagai prototipe pembelajaran berbasis web dengan tiga modul, yaitu Pengenalan Machine Learning, Deep Learning & Neural Networks, serta Computer Vision & Object Detection. Aktivitas pembelajaran dibedakan menjadi pembacaan materi dan evaluasi. Evaluasi dapat dilakukan melalui kuis pilihan ganda sebanyak 10 soal yang mencakup soal konsep dan *problem solving*, atau lima latihan kode Python berbentuk *fill-in-the-blank*, yaitu mahasiswa melengkapi sintaks atau nilai yang dikosongkan pada potongan program. Proses deteksi ekspresi

berlangsung setelah webcam diaktifkan dan tetap berjalan selama mahasiswa membaca materi serta mengerjakan mode evaluasi yang dipilih. Pada pembacaan materi, sistem secara normal mengaktifkan akses menuju evaluasi setelah 60 detik, sedangkan kuis dan latihan kode masing-masing memiliki batas waktu lima menit. Dengan demikian, aktivitas yang diamati mencakup pembelajaran konseptual dan latihan pemrograman terstruktur, bukan seluruhnya aktivitas menulis kode secara bebas.

Tabel 2. Skenario Aktivitas Pembelajaran

Tahap	Aktivitas	Bentuk	Durasi	Deteksi ekspresi
Materi	Membaca materi modul	Teks dan contoh konsep	≥60 detik secara normal; tersedia fungsi lewati	Aktif
Kuis	Evaluasi konsep dan problem solving	10 soal pilihan ganda	Maks. 5 menit	Aktif
Latihan kode	Melengkapi potongan kode Python	5 soal fill-in-the-blank	Maks. 5 menit	Aktif
Hasil	Melihat hasil sesi	Skor, distribusi, timeline	Setelah evaluasi	Kamera dihentikan

Soal disusun berdasarkan materi pada masing-masing modul dan tidak dikategorikan secara formal ke dalam tingkat kesulitan mudah, sedang, atau sulit. Oleh karena itu, penelitian ini tidak menganalisis pengaruh tingkat kesulitan soal terhadap hasil deteksi ekspresi.

Dataset penelitian berjumlah 1.660 citra dan berasal dari total 21 subjek. Dataset publik dari Roboflow Universe terdiri atas 953 citra dari 11 subjek, sedangkan 707 *hard samples* berasal dari 10 subjek tambahan. *Hard samples* digunakan untuk menambah variasi kondisi penggunaan webcam, seperti perubahan posisi wajah, pencahayaan, kemiripan visual antarkelas, dan ekspresi yang tidak terlalu kuat. Setiap citra memiliki *bounding box* pada area wajah dan salah satu dari empat label teknis, yaitu *engaged*, *confused*, *bored*, dan *frustrated*.

Pada dataset publik, label dan anotasi *bounding box* mengikuti anotasi yang tersedia pada dataset sumber dan kemudian diperiksa serta dipetakan ulang agar konsisten dengan susunan kelas penelitian. Pada *hard samples*, anotasi dilakukan secara manual menggunakan MakeSense.ai dengan memberikan *bounding box* pada area wajah dan menetapkan salah satu kelas target. Citra dengan ekspresi ambigu, kualitas rendah, atau area wajah yang tidak terlihat secara memadai tidak digunakan.

Informasi mengenai jumlah anotator, kesepakatan antarpenilai, dan mekanisme persetujuan subjek pada proses pengumpulan dataset publik mengikuti dokumentasi dataset sumber; informasi yang tidak tersedia pada sumber tidak diasumsikan dalam penelitian ini.

Setiap citra memiliki bounding box dan label berformat YOLO. Penamaan dan ID kelas dari sumber data diperiksa, dipetakan ulang, dan digabungkan ke Master Combined Dataset. Pembagian data dilakukan dengan stratified random sampling agar proporsi setiap kelas tetap terjaga, penggunaan pembagian terstratifikasi penting karena rasio train/test dapat memengaruhi performa model pada dataset khusus [24]. Komposisi kelas dan pembagian subset ditunjukkan pada Tabel 3.

Tabel 3. Komposisi Kelas Dan Pembagian Dataset Penelitian

Kelas/Sumber	Data Publik (Roboflow)	Data Private	Jumlah	Persentase/Fungsi
<i>Engaged</i>	280	268	548	33,01%
<i>Confused</i>	225	140	365	21,99%
<i>Bored</i>	235	129	364	21,93%
<i>Frustrated</i>	213	170	383	23,07%
Total	953	707	1.660	100,00%
Pembagian Dataset				
Subset	Target	Aktual	Persentase aktual	Fungsi
Pelatihan	80%	1.319	79,46%	Pembaruan bobot
Validasi	10%	168	10,12%	Pemantauan dan pemilihan model
Pengujian	10%	173	10,42%	Evaluasi akhir independen

Tabel 3 menjelaskan pembagian dataset pada eksperimen ini dilakukan pada tingkat citra dengan mempertahankan proporsi setiap kelas dan belum menggunakan identitas subjek sebagai kelompok pembagian. Oleh karena itu, citra yang berasal dari individu yang sama berpotensi berada pada subset pelatihan, validasi, dan pengujian yang berbeda. Kondisi tersebut menimbulkan risiko *subject leakage*, sehingga hasil evaluasi penelitian ini diperlakukan sebagai evaluasi berbasis *image-level split* dan belum merepresentasikan evaluasi *subject-independent*. Penelitian selanjutnya perlu menerapkan pembagian berbasis identitas subjek agar seluruh citra dari satu individu hanya berada pada satu subset.

2.2. Fine-Tuning YOLOv13n

Model diinisialisasi dari bobot pralatih yolov13n.pt dan di-fine-tuning menggunakan PyTorch pada GPU RTX 4060. Transfer learning dipilih agar representasi dari pelatihan awal dapat disesuaikan dengan dataset ekspresi yang lebih kecil; pemilihan detektor pralatih yang sesuai berpengaruh terhadap efisiensi transfer [25]. Citra diproses pada ukuran 640×640 piksel. Augmentasi *on-the-fly* meliputi variasi hue, saturasi, kecerahan, rotasi, translasi, skala, horizontal flip, mosaic, dan mixup; teknik campuran citra telah digunakan untuk memperluas variasi pada FER [26]. Konfigurasi utama pelatihan disajikan pada Tabel 4 berikut:

Tabel 4. Konfigurasi Utama Pelatihan YOLOv13n

Parameter	Nilai	Keterangan
Bobot awal	yolov13n.pt	Inisialisasi dari model pralatih
Maksimum epoch	150	Batas siklus pelatihan
Ukuran citra	640×640	Ukuran masukan model
Batch size	16	Citra per iterasi
Patience	25	Early stopping tanpa peningkatan
Optimizer	AdamW	Pembaruan parameter
Initial learning rate	0,001	Tingkat pembelajaran awal
Final LR factor	0,01	Faktor tingkat pembelajaran akhir
Weight decay	0,0005	Regularisasi bobot
Augmentasi	HSV, rotasi, translasi, skala, flip, mosaic, mixup	Variasi data pelatihan

2.3. Integrasi Sistem dan Pasca-Pemrosesan

Aplikasi web menyediakan halaman identitas, pemilihan tiga modul pembelajaran, aktivasi webcam, materi, pilihan evaluasi berupa 10 soal kuis atau lima soal latihan kode, halaman hasil, dan riwayat sesi. Frame webcam dipra-proses, kemudian YOLOv13n menghasilkan bounding box, label, dan confidence. Prediksi mentah diteruskan ke pasca-pemrosesan sebelum disimpan bersama skor dan durasi sesi.

Pasca-pemrosesan dirancang untuk meningkatkan keterbacaan rangkaian waktu, bukan untuk mengubah akurasi dasar model. Sistem memilih wajah dengan bounding box terbesar dan mengabaikan wajah yang terlalu kecil. Prediksi stabil ditentukan dari kelas dominan pada sliding window, rasio voting, dan rata-rata confidence. Jika syarat minimum belum terpenuhi, aplikasi menampilkan status operasional neutral; status ini bukan kelas kelima yang dipelajari model. Konfigurasi disajikan pada Tabel 5 berikut:

Tabel 5. Konfigurasi Pasca-Pemrosesan Prediksi Real-Time

Komponen	Nilai	Fungsi
Ambang deteksi awal	0,20	Menyaring kandidat awal
Luas wajah minimum	2% luas frame	Mengabaikan wajah terlalu kecil
Pemilihan wajah	Bounding box terbesar	Menentukan wajah utama
Sliding window	30 frame	Menyimpan prediksi terbaru
Prediksi minimum	8 frame	Batas awal penentuan kelas dominan
Rasio voting minimum	0,40	Dukungan minimum kelas dominan
Rata-rata confidence	0,40	Keyakinan minimum kelas dominan
Status tidak stabil	neutral	Mencegah penetapan kelas sebelum stabil

Parameter pasca-pemrosesan ditetapkan sebagai konfigurasi operasional secara empiris selama pengembangan sistem dan tidak diperoleh melalui optimasi hiperparameter secara

sistematis. Ukuran *sliding window* 30 frame digunakan untuk mempertahankan riwayat prediksi sekitar satu detik pada target aliran 30 FPS. Minimum delapan prediksi memberikan periode observasi awal sekitar 0,27 detik sebelum kelas stabil dapat ditentukan. Rasio voting minimum 0,40 dan rata-rata *confidence* minimum 0,40 digunakan sebagai batas operasional agar kelas dominan tidak langsung ditetapkan ketika dukungan prediksi atau tingkat keyakinannya masih rendah. Seluruh nilai tersebut dipertahankan tetap selama evaluasi pasca-pemrosesan. Oleh karena itu, penelitian menilai efek konfigurasi tersebut terhadap kestabilan temporal dan tidak menyatakan bahwa nilai parameter yang digunakan merupakan kombinasi yang optimal.

2.4. Evaluasi Model dan Sistem

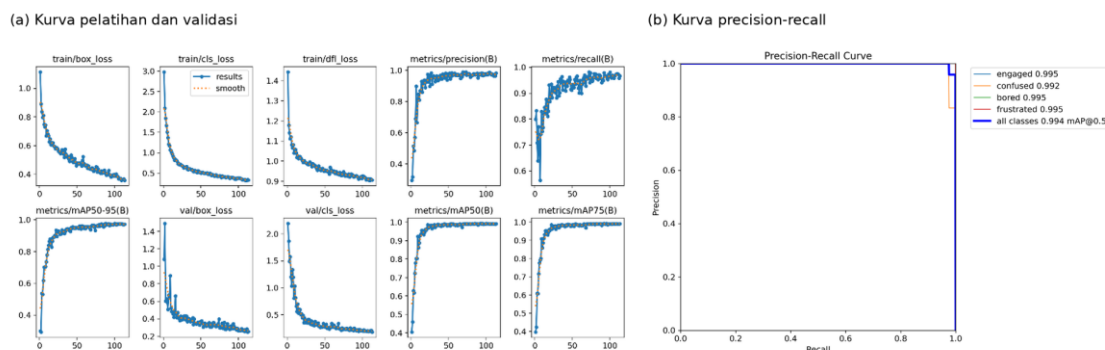
Bobot terbaik dievaluasi pada 173 citra pengujian yang tidak digunakan untuk optimasi maupun pemilihan model. Metrik yang digunakan meliputi precision, recall, F1-score, Intersection over Union, mAP@0.5, mAP@0.5:0.95, waktu inferensi, dan frames per second. Pasca-pemrosesan dievaluasi pada rekaman yang sama dengan membandingkan jumlah perubahan label, segmen singkat, panjang segmen, dan frame neutral sebelum serta sesudah stabilisasi. Pengujian aplikasi menggunakan black-box testing untuk memeriksa keluaran yang terlihat oleh pengguna tanpa menilai struktur internal kode; pendekatan serupa digunakan untuk memastikan respons dan alur layanan web sesuai spesifikasi [27]. Sebanyak 27 skenario menguji identitas, modul, kamera, deteksi, evaluasi, penyimpanan, riwayat, penanganan kesalahan, dan logout.

Data pengujian tidak digunakan dalam pembaruan bobot maupun pemilihan model. Namun, karena pembagian dataset dilakukan pada tingkat citra, pemisahan tersebut tidak menjamin bahwa identitas individu berbeda antar-subset. Evaluasi pasca-pemrosesan dilakukan setelah parameter operasional ditetapkan dengan membandingkan prediksi mentah dan hasil stabilisasi pada rangkaian frame yang sama. Evaluasi digunakan untuk mengukur perubahan jumlah pergantian label dan segmen berdurasi singkat, bukan untuk menentukan kombinasi parameter yang optimal atau menunjukkan peningkatan akurasi dasar YOLOv13n.

3. HASIL DAN PEMBAHASAN

3.1. Pelatihan dan Pemilihan Model

Pelatihan berlangsung sampai epoch ke-113 dan dihentikan melalui *early stopping* setelah tidak terjadi peningkatan selama 25 epoch. Bobot best.pt diperoleh pada epoch ke-88 dengan precision validasi 0,98217, recall 0,95700, mAP@0.5 0,99133, dan mAP@0.5:0.95 0,97674. Nilai train/box_loss, train/cls_loss, dan train/df_l_loss menurun, sedangkan precision, recall, dan mAP meningkat kemudian stabil. Pola pada Gambar 2(a) menunjukkan proses optimasi mencapai konvergensi tanpa menggunakan bobot epoch terakhir semata. Kurva precision-recall pada Gambar 2(b) berada di dekat kanan atas untuk seluruh kelas, dengan mAP@0.5 keseluruhan 0,994.



Gambar 2. (a) Kurva Pelatihan dan Validasi; (b) Kurva Precision-Recall Pada Data Pengujian

3.2. Evaluasi Model pada Data Pengujian

Hasil evaluasi per kelas disajikan pada Tabel 5. Secara keseluruhan model memperoleh precision 0,994, recall 0,982, F1-score 0,988, mAP@0.5 0,994, dan mAP@0.5:0.95 0,982. Kelas confused memiliki recall terendah, yaitu 0,952, sedangkan bored menghasilkan F1-score tertinggi sebesar 0,998. Perbedaan antarkelas relatif kecil pada data pengujian, tetapi hasil tersebut harus dibaca sesuai ruang lingkup dataset dan perangkat penelitian. Dijelaskan pada Tabel 5 berikut ini:

Tabel 5. Hasil Evaluasi YOLOv13n

Kelas	Citra	Precision	Recall	F1	mAP@0.5	mAP@0.75	mAP@0.5:0.95
Engaged	53	0,980	1,000	0,990	0,995	0,995	0,980
Confused	40	1,000	0,952	0,975	0,992	0,992	0,985
Bored	38	0,997	1,000	0,998	0,995	0,995	0,979
Frustrated	42	1,000	0,978	0,989	0,995	0,995	0,983
Keseluruhan	173	0,994	0,982	0,988	0,994	0,994	0,982

Tabel 5 menjelaskan evaluasi yang dilakukan pada 173 citra pengujian yang berasal dari dataset gabungan dengan total 21 subjek, terdiri atas 11 subjek dari dataset publik dan 10 subjek dari *hard samples*. Meskipun data pengujian tidak digunakan dalam pembaruan bobot maupun pemilihan model, pembagian dataset pada penelitian ini dilakukan pada tingkat citra menggunakan *stratified random sampling* dan belum memisahkan data berdasarkan identitas subjek. Dengan demikian, citra dari individu yang sama berpotensi terdapat pada lebih dari satu subset. Kondisi tersebut menimbulkan risiko *subject leakage* dan dapat menghasilkan estimasi performa yang lebih optimistis karena model mungkin telah mempelajari karakteristik visual individu yang juga muncul pada data pengujian. Oleh karena itu, nilai precision, recall, F1-score, dan mAP pada penelitian ini perlu dipahami sebagai hasil evaluasi berbasis *image-level split* dan belum menunjukkan kemampuan generalisasi terhadap subjek yang sepenuhnya tidak terlihat selama pelatihan.

Perbandingan anotasi ground truth dan hasil prediksi pada Gambar 3 menunjukkan bahwa model mampu menempatkan bounding box pada area wajah dan menghasilkan label dengan confidence tinggi pada sampel pengujian. Namun, ketepatan pada citra pengujian tidak menjamin klasifikasi selalu benar dalam penggunaan langsung. Variasi pose kepala, arah pandangan, oklusi, dan gestur tangan di sekitar wajah dapat membentuk pola visual yang mirip dengan kelas tertentu. Temuan ini konsisten dengan kebutuhan untuk memperhatikan oklusi dan potensi bias representasi data pada FER [18]-[20].



Anotasi ground truth

Hasil prediksi YOLOv13n

Gambar 3. Perbandingan Anotasi Ground Truth dan Prediksi YOLOv13n

3.3. Kecepatan Pemrosesan dan Efek Pasca-Pemrosesan

Pada evaluasi data citra, proses pra-pemrosesan membutuhkan waktu rata-rata sebesar 0,6 milidetik, inferensi model sebesar 6,2 milidetik, dan pasca-pemrosesan dasar detektor sebesar 0,9

milidetik. Dengan demikian, keseluruhan pipeline evaluasi membutuhkan waktu sekitar 7,7 milidetik per citra atau setara dengan throughput teoretis sekitar 129,9 citra per detik. Pada pengujian secara real-time selama 161,9 detik, sistem memproses sebanyak 4.857 frame dengan kecepatan rata-rata 51,77 FPS setelah tahap pemanasan awal dan median kecepatan sebesar 51,82 FPS. Hasil tersebut menunjukkan bahwa proses inferensi dan pasca-pemrosesan dapat dijalankan di atas target aliran aplikasi yang ditetapkan sekitar 30 FPS.

Evaluasi pasca-pemrosesan dilakukan dengan membandingkan urutan prediksi mentah YOLOv13n dengan hasil setelah penerapan confidence filtering, sliding-window smoothing, majority voting, dan penyaringan durasi. Perbandingan dilakukan terhadap 4.857 frame dari sesi pengujian yang sama agar kedua keluaran memiliki kondisi evaluasi yang konsisten. Hasil perbandingan disajikan pada Tabel 6.

Tabel 6. Ringkasan Kecepatan dan Perubahan Setelah Pasca-Pemrosesan

Indikator	Prediksi mentah	Setelah pasca-pemrosesan	Perubahan
Jumlah <i>frame</i>	4.857	4.857	Tetap
Perubahan label	522	119	Turun 77,20%
Jumlah segmen label	523	120	Turun 77,06%
Segmen kurang dari 8 <i>frame</i>	371	38	Turun 89,76%
Persentase segmen singkat	70,94%	31,67%	Turun 39,27 poin
Median panjang segmen	3 <i>frame</i>	23 <i>frame</i>	Meningkat
Median durasi segmen	0,10 detik	0,77 detik	Meningkat
<i>Frame neutral</i>	0	105	2,16%

Tabel 6 menjelaskan jumlah perubahan label berkurang dari 522 menjadi 119 kali atau turun sebesar 77,20%. Jumlah segmen label juga berkurang dari 523 menjadi 120 segmen atau sebesar 77,06%. Penurunan kedua indikator tersebut menunjukkan bahwa pasca-pemrosesan mampu mengurangi frekuensi pergantian label selama pemrosesan video secara real-time. Hubungan antara kedua indikator juga konsisten karena jumlah segmen pada suatu rangkaian prediksi pada umumnya merupakan jumlah perubahan label ditambah satu segmen.

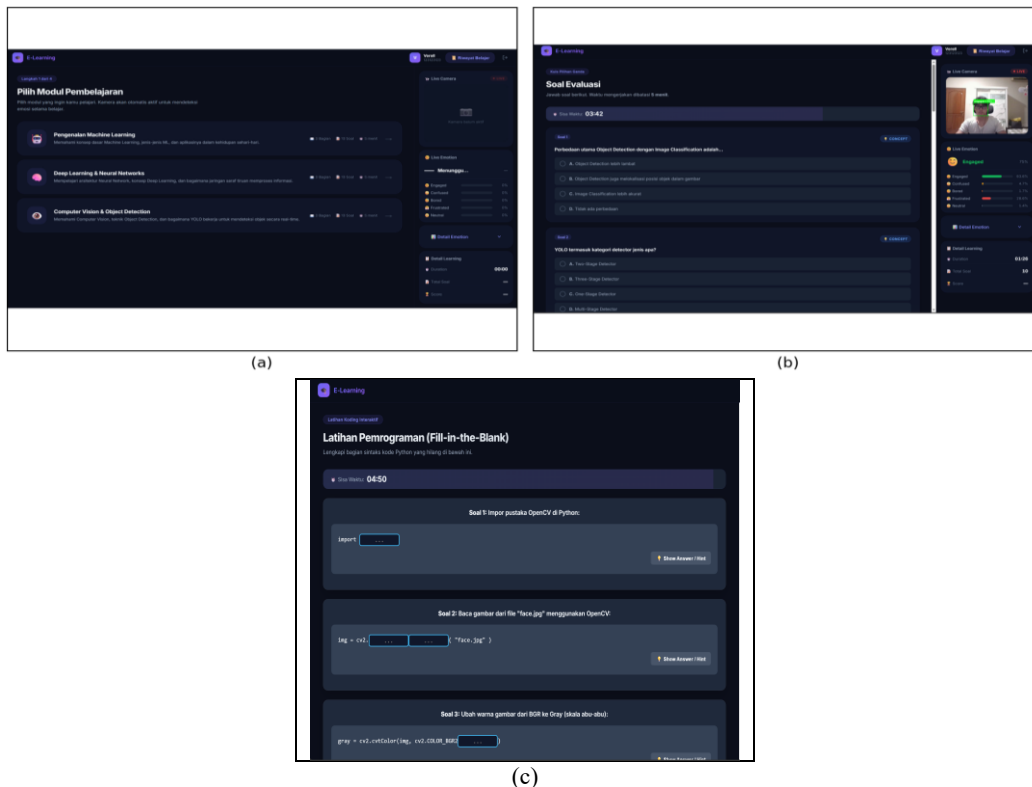
Pengurangan prediksi berdurasi singkat terlihat dari jumlah segmen yang memiliki panjang kurang dari delapan frame. Jumlahnya menurun dari 371 menjadi 38 segmen atau sebesar 89,76%. Persentase segmen singkat juga turun dari 70,94% menjadi 31,67%, dengan selisih sebesar 39,27 poin persentase. Sementara itu, median panjang segmen meningkat dari tiga menjadi 23 frame dan median durasi segmen meningkat dari 0,10 menjadi 0,77 detik. Hasil ini menunjukkan bahwa label setelah pasca-pemrosesan cenderung bertahan lebih lama sehingga rangkaian prediksi menjadi lebih stabil dan lebih mudah dibaca pada visualisasi timeline. Temuan tersebut sesuai dengan analisis dalam laporan tugas akhir. Setelah pasca-pemrosesan, sebanyak 105 frame atau 2,16% dari keseluruhan data memperoleh status operasional neutral. Status tersebut bukan merupakan kelas kelima yang dipelajari oleh YOLOv13n, melainkan digunakan ketika jumlah prediksi dalam sliding window belum mencukupi atau ketika rasio voting dan rata-rata confidence belum memenuhi ambang minimum.

Secara keseluruhan, pasca-pemrosesan berhasil meningkatkan kestabilan penyajian hasil prediksi dengan mengurangi pergantian label dan segmen berdurasi singkat. Meskipun demikian, hasil tersebut tidak menunjukkan peningkatan precision, recall, F1-score, atau mAP dasar YOLOv13n. Apabila salah klasifikasi terjadi secara konsisten pada beberapa frame, mekanisme majority voting masih dapat mempertahankan label yang kurang tepat sebagai keluaran yang stabil. Oleh karena itu, pasca-pemrosesan berfungsi untuk meningkatkan keterbacaan temporal hasil deteksi, bukan untuk memperbaiki akurasi dasar model.

3.4. Implementasi Model Aplikasi

Aplikasi berhasil mengintegrasikan identitas pengguna, tiga modul pembelajaran, kamera, deteksi *real-time*, materi, kuis pilihan ganda, latihan kode, perhitungan skor, penyimpanan sesi, dashboard, dan riwayat belajar. Dashboard *E-Learning* menampilkan tiga kartu modul beserta jumlah bagian materi, jumlah soal, estimasi waktu, panel kamera, status emosi saat ini, distribusi ekspresi, durasi sesi, dan skor. Pada mode kuis, sistem menyajikan 10 soal pilihan ganda,

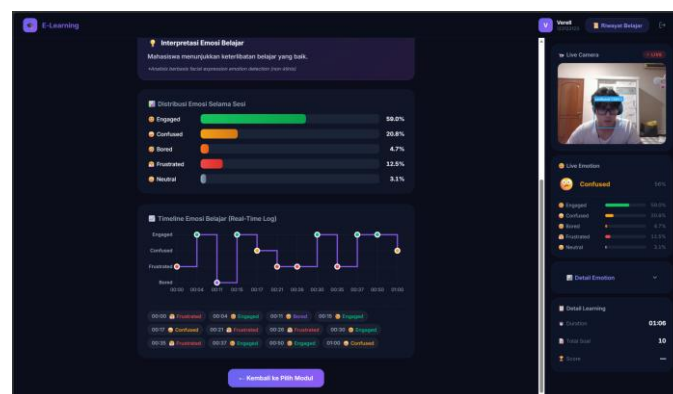
indikator jawaban, penghitung waktu lima menit, serta deteksi ekspresi yang tetap berjalan selama evaluasi. Antarmuka dashboard *E-Learning* dan kuis ditunjukkan pada Gambar 4 berikut:



Gambar 4. (a) Dashboard E-Learning Pada Halaman Pemilihan Modul; (b) Antarmuka Kuis Pilihan Ganda; (c) Antarmuka Latihan Kode Interaktif Berbasis *Fill-In-The-Blank*.

Selain kuis pilihan ganda, aplikasi menyediakan lima latihan kode berbasis Python dalam bentuk *fill-in-the-blank*. Pada aktivitas ini, mahasiswa melengkapi bagian sintaks atau nilai yang dikosongkan pada potongan program berdasarkan materi yang telah dipelajari. Pemeriksaan jawaban dilakukan terhadap jawaban acuan yang telah ditentukan dan belum mengeksekusi program pada lingkungan *sandbox*. Webcam dan proses deteksi ekspresi tetap aktif selama mahasiswa mengerjakan latihan kode, sehingga hasil prediksi selama aktivitas tersebut menjadi bagian dari riwayat ekspresi pada sesi pembelajaran.

Setelah sesi selesai, halaman hasil menampilkan skor total, durasi, ekspresi dominan, distribusi engaged, confused, bored, frustrated, serta neutral, dan timeline perubahan hasil deteksi. Implementasi dashboard hasil sesi ditunjukkan pada Gambar 5. Penyajian tersebut sejalan dengan tujuan learning analytics dashboard untuk menggabungkan indikator belajar dalam tampilan yang dapat dibaca pengguna [15]-[18].



Gambar 5. Implementasi Distribusi Dashboard Deteksi Riwayat Ekspresi

Pada implementasi penelitian ini, hasil deteksi ekspresi dirangkum pada tingkat sesi pembelajaran dan belum dianalisis secara terpisah berdasarkan jenis aktivitas, seperti membaca materi, mengerjakan kuis pilihan ganda, atau mengerjakan latihan kode. Penelitian juga belum mengelompokkan soal berdasarkan tingkat kesulitan secara formal. Oleh karena itu, hasil penelitian ini belum dapat digunakan untuk menyimpulkan bahwa *confusion* meningkat ketika mahasiswa mengerjakan latihan kode tertentu, bahwa *frustration* muncul akibat kegagalan menyelesaikan suatu program, atau bahwa *engagement* berbeda antara aktivitas membaca dan latihan kode. Konteks pembelajaran pemrograman dalam penelitian ini berfungsi sebagai lingkungan implementasi sistem, sedangkan hubungan antara jenis aktivitas pembelajaran dan perubahan ekspresi afektif masih memerlukan analisis khusus pada penelitian selanjutnya.

3.5. Pengujian Fungsional Aplikasi dengan *Black-box Testing*

Pengujian *black-box* dilakukan untuk menilai kesesuaian fungsi aplikasi terhadap keluaran yang dirancang tanpa mengevaluasi struktur internal program. Sebanyak 27 skenario mencakup validasi identitas, navigasi, aktivasi dan kegagalan kamera, pemilihan wajah, pasca-pemrosesan prediksi, penyajian materi, kuis, latihan kode, penyimpanan hasil, riwayat pengguna, penanganan respons kesalahan, dan *logout*. Seluruh skenario menghasilkan keluaran sesuai dengan hasil yang diharapkan. Dengan demikian, tingkat keberhasilan 100% pada pengujian ini menunjukkan bahwa fungsi aplikasi sesuai spesifikasi pada skenario yang diuji dan tidak menunjukkan tingkat akurasi model maupun efektivitas terhadap proses pembelajaran mahasiswa.

Tabel 7. Rekapitulasi Kelompok Skenario Pengujian Fungsional

Kelompok Pengujian	Nomor Skenario	Jumlah	Status	Cakupan Utama
Identitas dan navigasi	1-5	5	Berhasil	Validasi identitas, sesi, daftar dan pemilihan modul
Kamera, deteksi, dan stabilisasi	6-12	7	Berhasil	Aktivasi kamera, wajah utama, filter ukuran, sliding window
Materi dan evaluasi	13-20	8	Berhasil	Materi, durasi baca, kuis, latihan kode, batas waktu
Hasil, penyimpanan, dan riwayat	21-27	7	Berhasil	Dashboard, penghentian kamera, riwayat, 401/404, logout
Total Keseluruhan	1-27	27	Berhasil	Seluruh keluaran sesuai hasil yang diharapkan

Rekapitulasi pada Tabel 7 memperlihatkan bahwa pengujian mencakup alur normal dan kondisi tidak valid. Tingkat keberhasilan 100% menyatakan kesesuaian fungsi terhadap rancangan pada skenario yang diuji, bukan jaminan bahwa klasifikasi ekspresi selalu tepat. Pengujian kamera tetap menemukan kesalahan kelas pada variasi pose dan gestur, sehingga evaluasi fungsional aplikasi harus dibaca terpisah dari evaluasi performa model.

Integrasi deteksi afektif dan visualisasi real-time memiliki kesamaan arah dengan sistem pemantauan emosi pembelajar jarak jauh yang menyajikan hasil deteksi secara visual [16]. Namun, rancangan dashboard yang berorientasi pada mahasiswa perlu memastikan bahwa informasi mudah dipahami, relevan terhadap tujuan belajar, dan tidak menimbulkan interpretasi berlebihan; kebutuhan tersebut juga ditekankan dalam pendekatan partisipatif terhadap dashboard mahasiswa [28]. Pada pembelajaran pemrograman, sistem penilaian otomatis [29] dan intervensi learning analytics [30] menunjukkan potensi integrasi data aktivitas untuk mendukung umpan balik, tetapi manfaat pedagogis tetap perlu diuji melalui studi pengguna dan desain eksperimen di kelas.

3.6. Diskusi

Perbandingan dengan penelitian terdahulu perlu dilihat dari konteks, metode, dan keluaran sistem. Studi FER pada pembelajaran sains dan kelas daring menekankan pemanfaatan ekspresi untuk menelusuri engagement [3], [4], sedangkan analisis multimodal pada aktivitas pemrograman dan penilaian engagement menggunakan lebih dari satu sumber data [5], [6]. Pendekatan tersebut memberikan konteks afektif yang lebih luas, tetapi membutuhkan sensor, sinkronisasi data, atau proses komputasi yang lebih kompleks dibandingkan sistem berbasis

webcam tunggal. Sistem dashboard afektif dan visualisasi emosi real-time juga telah dikembangkan [19], [16], tetapi belum menggabungkan YOLOv13n, materi e-learning, kuis pemrograman, pasca-pemrosesan temporal, skor, dan timeline dalam satu alur aplikasi.

Nilai evaluasi yang tinggi pada penelitian ini juga perlu ditafsirkan dengan mempertimbangkan ukuran data pengujian dan mekanisme pembagian dataset. Data pengujian hanya terdiri atas 173 citra, sedangkan pembagian data dilakukan pada tingkat citra dan belum menggunakan identitas subjek sebagai kelompok pembagian. Dengan demikian, kemungkinan *subject leakage* belum dapat dikesampingkan dan performa yang diperoleh dapat lebih optimistis dibandingkan evaluasi *subject-independent*. Oleh karena itu, penelitian ini tidak menggunakan nilai mAP yang lebih tinggi daripada beberapa penelitian sebelumnya sebagai bukti bahwa YOLOv13n secara umum lebih unggul. Evaluasi lanjutan dengan pemisahan subjek yang ketat dan jumlah pengguna yang lebih besar diperlukan untuk menilai kemampuan generalisasi model.

Pada kelompok object detection, YOLOv5 telah digunakan untuk tiga kelas emosi mahasiswa [10], YOLOv8 telah dibandingkan pada ekspresi wajah pembelajaran daring [11], dan varian YOLO ringan menunjukkan potensi untuk FER [12]. YOLOv13 sendiri mula-mula dievaluasi pada deteksi objek umum [13]. Pada dataset pengujian penelitian ini, YOLOv13n menghasilkan metrik yang lebih tinggi daripada angka yang dilaporkan pada beberapa studi YOLO sebelumnya, tetapi perbandingan lintas penelitian tersebut tidak membuktikan keunggulan absolut karena dataset, jumlah kelas, pembagian data, perangkat, dan skenario evaluasi berbeda. Kontribusi penelitian ini terletak pada evaluasi varian Nano untuk empat kategori emosi belajar dan integrasinya dengan UI e-learning, kuis pemrograman, pasca-pemrosesan temporal, serta dashboard dalam satu prototipe.

Integrasi dengan materi, kuis, dan latihan kode menunjukkan bahwa model dapat dijalankan bersamaan dengan beberapa bentuk aktivitas pada aplikasi pembelajaran. Namun, penelitian ini belum mengevaluasi perbedaan distribusi ekspresi berdasarkan jenis aktivitas secara terpisah. Dengan demikian, belum dapat disimpulkan bahwa aktivitas latihan kode menghasilkan *confusion* atau *frustration* yang lebih tinggi dibandingkan membaca materi atau mengerjakan kuis. Penelitian ini juga belum mengevaluasi hubungan antara hasil deteksi dan tingkat kesulitan soal, keberhasilan menyelesaikan kode, maupun kejadian *debugging*. Kontribusi pada konteks pembelajaran pemrograman saat ini terletak pada integrasi teknis deteksi ekspresi dengan aktivitas pembelajaran berbasis materi dan latihan kode, sedangkan hubungan antara karakteristik aktivitas pemrograman dan dinamika ekspresi afektif masih menjadi ruang penelitian lanjutan.

Pengujian langsung memperlihatkan bahwa wajah tetap dapat terdeteksi ketika klasifikasi ekspresinya kurang tepat. Kepala miring dapat menghasilkan prediksi bored, sedangkan gestur memegang kepala dapat menghasilkan prediksi frustrated, termasuk ketika pengguna tersenyum. Hal ini menunjukkan bahwa keberhasilan lokalisasi wajah tidak selalu diikuti ketepatan klasifikasi kelas. Dataset hard samples membantu mendekatkan kondisi pelatihan dengan implementasi, tetapi jumlah pengguna dan variasi visual masih terbatas. Selain itu, identitas pengguna belum diverifikasi melalui basis data akademik dan riwayat masih disimpan dalam berkas JSON. Selain itu, confusion dan frustration dapat berkembang sebagai dinamika nonlinier selama aktivitas belajar [9], sehingga ringkasan distribusi kelas tidak boleh diperlakukan sebagai representasi utuh proses kognitif-afektif mahasiswa. Karena itu, sistem diposisikan sebagai prototipe indikator visual pendukung, bukan alat diagnosis, penentu intervensi otomatis, atau bukti hubungan sebab-akibat antara ekspresi dan skor belajar.

Penelitian ini belum melakukan studi pengguna, evaluasi *usability*, maupun eksperimen terkontrol untuk mengukur pengaruh penggunaan dashboard terhadap pemahaman, keterlibatan, atau hasil belajar mahasiswa. Oleh karena itu, keberhasilan integrasi sistem dan penyajian learning analytics tidak dapat diartikan sebagai bukti bahwa dashboard meningkatkan proses pembelajaran. Manfaat pedagogis sistem pada penelitian ini masih berupa potensi sebagai informasi pendukung dan memerlukan evaluasi lebih lanjut melalui studi pengguna serta eksperimen pembelajaran pada kondisi kelas nyata.

4. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan sistem *Facial Emotion Recognition* berbasis YOLOv13n untuk mendeteksi indikator ekspresi visual yang diklasifikasikan ke dalam empat kategori afektif, yaitu *engagement*, *confusion*, *boredom*, dan *frustration*, pada sesi pembelajaran berbasis web. Model di-*fine-tuning* menggunakan 1.660 citra dan dievaluasi pada 173 citra pengujian dengan memperoleh *precision* 0,994, *recall* 0,982, F1-score 0,988, mAP@0.5 0,994, mAP@0.5:0.95 0,982, serta waktu inferensi 6,2 ms per citra. Namun, pembagian dataset masih dilakukan pada tingkat citra dan belum berbasis identitas subjek, sehingga terdapat risiko *subject leakage* dan hasil evaluasi belum menunjukkan generalisasi terhadap subjek yang sepenuhnya baru.

Pasca-pemrosesan temporal menurunkan perubahan label dari 522 menjadi 119 atau sebesar 77,20%. Hasil tersebut menunjukkan peningkatan kestabilan temporal dan keterbacaan rangkaian prediksi, bukan peningkatan akurasi dasar model. Pengujian *black-box* terhadap 27 skenario juga menunjukkan seluruh fungsi aplikasi berjalan sesuai rancangan, tetapi keberhasilan tersebut hanya menunjukkan kesesuaian fungsi aplikasi dan bukan ketepatan klasifikasi ekspresi maupun efektivitas sistem terhadap pembelajaran.

Sistem berhasil mengintegrasikan pembacaan materi, kuis pilihan ganda, latihan kode Python terstruktur, webcam, skor, durasi, distribusi ekspresi, ekspresi dominan, *timeline*, dan riwayat sesi. Keluaran *engaged*, *confused*, *bored*, dan *frustrated* digunakan sebagai indikator ekspresi afektif berbasis visual dan bukan sebagai pengukuran langsung terhadap kondisi emosi atau psikologis mahasiswa. Penelitian ini juga belum menganalisis hasil deteksi secara terpisah antara aktivitas membaca materi, mengerjakan kuis, dan latihan kode, serta belum mengevaluasi hubungan ekspresi dengan tingkat kesulitan soal, proses *debugging*, keberhasilan penyelesaian kode, maupun hasil belajar. Oleh karena itu, manfaat pedagogis sistem masih diposisikan sebagai potensi.

Penelitian selanjutnya disarankan memperluas dataset dengan lebih banyak subjek dan variasi kondisi visual serta menerapkan *subject-independent split* untuk mengurangi risiko *subject leakage*. Proses anotasi juga perlu melibatkan lebih dari satu penilai dan, apabila memungkinkan, dilengkapi laporan diri pengguna. Analisis hasil deteksi sebaiknya dipisahkan berdasarkan jenis aktivitas pembelajaran dan karakteristik latihan pemrograman, seperti tingkat kesulitan, kesalahan eksekusi, proses *debugging*, dan waktu penyelesaian.

Selain itu, studi pengguna dan eksperimen pembelajaran perlu dilakukan untuk mengevaluasi *usability*, pemahaman terhadap dashboard, serta potensi pengaruh sistem terhadap keterlibatan dan hasil belajar mahasiswa. Penyimpanan berbasis JSON juga dapat dikembangkan menjadi basis data dengan autentikasi, kontrol akses, pencadangan, dan perlindungan privasi, dengan tetap menegaskan bahwa hasil deteksi merupakan indikator visual pendukung dan bukan diagnosis psikologis.

DAFTAR PUSTAKA

- [1] Z. Atiq and R. Batra, "How Do First-Year Engineering Students' Emotions Change while Working on Programming Problems?," *ACM Transactions on Computing Education*, vol. 24, no. 2, pp. 1-30, 2024, doi: 10.1145/3643865.
- [2] R. Pekrun, "Control-Value Theory: From Achievement Emotion to a General Theory of Human Emotions," *Educ. Psychol. Rev.*, vol. 36, no. 3, pp. 1-36, 2024, doi: 10.1007/s10648-024-09909-7.
- [3] X. Tang, *et al.*, "Facial Expression Recognition for Probing Students' Emotional Engagement in Science Learning," *J. Sci. Educ. Technol.*, vol. 34, no. 1, pp. 13–30, 2025, doi: 10.1007/s10956-024-10143-7.
- [4] X. Hu and J. Gao, "Facial expression recognition reveals students' engagement in online class: Correlations with six engagement measurements," *PLoS One*, vol. 20, no. 10, 2025, doi: 10.1371/journal.pone.0334232.

- [5] K. Mangaroska, K. Sharma, D. Gašević, and M. Giannakos, "Exploring students' cognitive and affective states during problem solving through multimodal data: Lessons learned from a programming activity," *J. Comput. Assist. Learn.*, vol. 38, no. 1, pp. 40–59, 2022, doi: 10.1111/jcal.12590.
- [6] L. Yan, X. Wu, and Y. Wang, "Student engagement assessment using multimodal deep learning," *PLoS One*, vol. 20, no. 6, pp. e0325377, 2025, [Online]. Available: 10.1371/journal.pone.0325377.
- [7] A. Gupta, A. D'Cunha, K. Awasthi, and V. Balasubramanian, "DAiSEE: Towards User Engagement Recognition in the Wild," vol. 14, no. 8, pp. 1-12, 2022, doi: 10.48550/arXiv.1609.01885.
- [8] Y. Xu, Y. Li, Y. Chen, H. Bao, and Y. Zheng, "Spontaneous visual database for detecting learning-centered emotions during online learning," *Image Vis. Comput.*, vol. 136, p. 104739, 2023, doi: 10.1016/j.imavis.2023.104739.
- [9] E. B. Cloude, *et al.*, "Exploring Confusion and Frustration as Non-linear Dynamical Systems," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, 2024, pp. 241–252. doi: 10.1145/3636555.3636875.
- [10] A. K. Ulandari, F. Bimantoro, and I. G. P. S. Wijaya, "Real Time Student Emotion Detection using YOLOv5," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 222–231, 2024, doi: 10.29408/edumatic.v8i1.25726.
- [11] D. Tresnawati, S. Nurhidayanti, and N. Lestari, "A Comparison of YOLOv8 Series Performance in Student Facial Expressions Detection on Online Learning," *Jurnal Online Informatika*, vol. 10, no. 1, pp. 93–104, 2025, doi: 10.15575/join.v10i1.1390.
- [12] A. I. Aulia, A. J. Hutapea, A. M. Siregar, and Surjandy, "Lightweight YOLO Models for Robust Facial Expression Detection," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 19, no. 2, 2026, doi: 10.24036/jtip.v19i2.1120.
- [13] M. Lei *et al.*, "YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception," pp. 1-12, 2025, doi: 10.48550/arXiv.2506.17733.
- [14] A. S. Kusuma, A. I. Pradana, and B. W. Pamekas, "Pengembangan Sistem Perhitungan Jumlah Kendaraan Berdasarkan Jenis Kendaraan Menggunakan Algoritma Yolo Secara Realtime," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 7, no. 2, pp. 166–179, 2024, doi: 10.36080/skanika.v7i2.3201.
- [15] M. N. Hasnine, *et al.*, "A Real-Time Learning Analytics Dashboard for Automatic Detection of Online Learners' Affective States," *Sensors*, vol. 23, no. 9, pp. 1-15, 2023, doi: 10.3390/s23094243.
- [16] A. Harb *et al.*, "Diverse distant-students deep emotion recognition and visualization," *Computers and Electrical Engineering*, vol. 111, 2023, doi: 10.1016/j.compeleceng.2023.108963.
- [17] A. B. Amjoud and M. Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," *IEEE Access*, vol. 11, pp. 35479–35516, 2023, doi: 10.1109/ACCESS.2023.3266093.
- [18] Y. ELSayed, A. ELSayed, and M. A. Abdou, "An automatic improved facial expression recognition for masked faces," *Neural Comput. Appl.*, vol. 35, no. 20, pp. 14963–14972, 2023, doi: 10.1007/s00521-023-08498-w.
- [19] I. Dominguez-Catena, D. Paternain, and M. Galar, "Metrics for Dataset Demographic Bias: A Case Study on Facial Expression Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, pp. 5209–5226, 2024, doi: 10.1109/TPAMI.2024.3361979.
- [20] M. Lukac, G. Zhambulova, K. Abdiyeva, and M. Lewis, "Study on emotion recognition bias in different regional groups," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-34932-z.
- [21] L. D. Vreugd, A. V. Leeuwen, R. Jansen, and M. V. D. Schaaf, "Learning Analytics Dashboard Design and Evaluation to Support Student Self-Regulation of Study Behaviour," *Journal of Learning Analytics*, vol. 11, no. 3, pp. 249–262, 2024, doi: 10.18608/jla.2024.8529.

- [22] L. Paulsen and E. Lindsay, "Learning analytics dashboards are increasingly becoming about learning and not just analytics - A systematic review," *Educ. Inf. Technol. (Dordr.)*, vol. 29, no. 11, pp. 14279–14308, 2024, doi: 10.1007/s10639-023-12401-4.
- [23] A. Kaveri, A. Silvola, and H. Muukkonen, "Supporting Student Agency with a Student-Facing Learning Analytics Dashboard: Perceptions of an Interdisciplinary Development Team," *Journal of Learning Analytics*, vol. 10, no. 2, pp. 85–99, 2023, doi: 10.18608/jla.2023.7729.
- [24] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 2, pp. 331–339, 2024, doi: 10.14569/IJACSA.2024.0150235.
- [25] Z. Wang, A. Li, Z. Li, and Q. Dou, "Efficient Transferability Assessment for Selection of Pre-Trained Detectors," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 1679–1689, doi: 10.1109/WACV57701.2024.00170.
- [26] A. Psaroudakis and D. Kollias, "MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022, pp. 2366–2374, doi: 10.1109/CVPRW56347.2022.00264.
- [27] D. Felicio, J. Simao, and N. Datia, "Rapitest: Continuous black-box testing of restful web apis," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 537–545. doi: 10.1016/j.procs.2023.01.322.
- [28] C. Herodotou *et al.*, "A Participatory Approach to Designing a Student-Facing Dashboard for Online and Distance Education," *Journal of Learning Analytics*, vol. 12, no. 2, pp. 158–174, 2025, doi: 10.18608/jla.2025.8481.
- [29] A. L. C. Barczak, A. Mathrani, B. Han, and N. H. Reyes, "Automated assessment system for programming courses: a case study for teaching data structures and algorithms," *Educational Technology Research and Development*, vol. 71, no. 6, pp. 2365–2388, 2023, doi: 10.1007/s11423-023-10277-2.
- [30] P. Utamachant, C. Anutariya, and S. Pongnumkul, "i-Ntervene: applying an evidence-based learning analytics intervention to support computer programming instruction," *Smart Learning Environments*, vol. 10, no. 1, pp. 1–30, 2023, doi: 10.1186/s40561-023-00257-7.