

DETEKSI TINGKAT STRES DARI EKSPRESI WAJAH MENGGUNAKAN YOLOV8-CLS (STUDI PADA FER2013 & RAF-DB)

Kiki Saepul Hamzah¹, Amelya Melviani², Jibril Fauzan Muksin³, Sayyid

Abdurrasyad⁴, Untung Rohwadi⁵, Felix Wuryo Handono⁶

Email: kikisaefulhamzah7@gmail.com¹, amelyamelv7@gmail.com²,

jibrilmuksin@gmail.com³, abdurrasyadsayyid@gmail.com⁴,

untung.unr@bsi.ac.id⁵, felix@bsi.ac.id⁶

^{1,2,3,4,5,6}Universitas Bina Sarana Informatika

ABSTRAK

Pengenalan emosi wajah memiliki peran penting dalam berbagai aplikasi modern, seperti sistem kesehatan mental, keamanan, dan interaksi manusia-komputer. Namun, implementasi sistem real-time pada perangkat dengan kemampuan komputasi terbatas masih menjadi tantangan, terutama terkait kebutuhan model yang ringan tetapi tetap akurat. Penelitian ini bertujuan mengembangkan sistem deteksi emosi wajah berbasis model klasifikasi YOLOv8-ClS dengan memanfaatkan dua dataset benchmark, yaitu FER2013 dan RAF-DB. Tahap awal penelitian mencakup konversi data piksel mentah FER2013 ke format citra RGB, pra-pemrosesan data, serta penyusunan struktur direktori yang sesuai dengan arsitektur YOLO. Model dilatih menggunakan konfigurasi ringan berupa yolov8n-cls dengan ukuran citra 64×64, batch size 16, dan 50 epoch pada perangkat CPU. Evaluasi dilakukan menggunakan metrik train loss, validation loss, akurasi top-1, dan akurasi top-5. Hasil menunjukkan akurasi top-1 sebesar 65,44% pada FER2013 dan 80,54% pada RAF-DB, dengan akurasi top-5 di atas 98% pada kedua dataset. Model kemudian diuji secara real-time melalui integrasi deteksi wajah dan klasifikasi emosi, termasuk pemetaan tingkat stres berdasarkan kategori emosi. Penelitian ini menunjukkan bahwa YOLOv8-ClS mampu memberikan performa kompetitif pada lingkungan komputasi terbatas serta berpotensi menjadi dasar bagi sistem analisis emosi di berbagai aplikasi praktis.

Kata Kunci: *Deteksi, Emosi, FER2013, RAF-DB, Yolov8-ClS.*

ABSTRACT

Facial emotion recognition plays an important role in various modern applications, such as mental health systems, security, and human-computer interaction. However, implementing a real-time system on devices with limited computational resources remains challenging, particularly due to the need for lightweight yet accurate models.

This study aims to develop a facial emotion detection system using the YOLOv8-Cls classification model by utilizing two benchmark datasets, FER2013 and RAF-DB. The initial stage includes converting the raw pixel data of FER2013 into RGB image format, performing data preprocessing, and organizing the directory structure to align with the YOLO architecture. The model was trained using a lightweight configuration of yolov8ncls with an input size of 64×64 pixels, a batch size of 16, and 50 training epochs on a CPU device. Evaluation was conducted using train loss, validation loss, top-1 accuracy, and top-5 accuracy metrics. The results show that the model achieved a top-1 accuracy of 65.44% on FER2013 and 80.54% on RAF-DB, with top-5 accuracy exceeding 98% on both datasets. The model was further tested in real-time by integrating face detection and emotion classification, including a stress-level mapping based on emotional categories. This study demonstrates that YOLOv8-Cls provides competitive performance in low-resource environments and has strong potential as a foundation for emotion analysis systems in various practical applications.

Keywords: Emotion Recognition, FER2013, RAF-DB, Yolov8-Cls

1. PENDAHULUAN

Perkembangan teknologi deteksi emosi wajah terus meningkat seiring kebutuhan aplikasi manusia-komputer yang lebih adaptif, layanan kesehatan mental, sistem keamanan, serta interaksi digital yang memperhatikan kondisi emosional pengguna. Ekspresi wajah memberikan sinyal nonverbal kaya informasi emosional — dari kebahagiaan, kesedihan, ketakutan, kemarahan, hingga netral — sehingga menjadikan pengenalan ekspresi wajah (Facial Expression Recognition / FER) sebagai elemen kunci dalam sistem AI yang berorientasi “memahami” emosi manusia [1].

Penelitian pada FER di Indonesia telah menunjukkan hasil yang cukup menggembirakan dengan penerapan deep learning, khususnya arsitektur Convolutional Neural Network (CNN). Misalnya, studi yang melaporkan bahwa CNN sederhana mampu mengklasifikasikan ekspresi wajah dengan akurasi kompetitif meskipun dataset relatif kecil, dan hasil ini menunjukkan bahwa pendekatan CNN tetap relevan bagi penelitian lokal meski dengan sumber daya terbatas [2].

Selain itu, implementasi FER dalam bentuk aplikasi real-time juga telah dilakukan — sebuah sistem deteksi emosi berbasis CNN dan OpenCV dapat mengklasifikasikan ekspresi melalui kamera secara langsung, menunjukkan bahwa FER bukan hanya teori tetapi bisa diaplikasikan dalam kehidupan sehari-hari [3].

Namun demikian, sebagian besar penelitian di Indonesia masih mengandalkan

metode tradisional (CNN) dan belum banyak yang mengadopsi arsitektur ringan atau model modern yang dirancang untuk inferensi cepat dan efisien. Hal ini menjadi kendala ketika sistem ingin dijalankan pada perangkat dengan sumber daya terbatas seperti CPU saja, atau dalam konteks real-time seperti kamera langsung. Dengan demikian, diperlukan eksplorasi lebih lanjut terhadap model ringan serta optimasi sistem agar FER dapat berjalan lebih fleksibel dan praktis [4].

Sejalan dengan tantangan tersebut, penelitian ini menggunakan pendekatan baru dengan memanfaatkan model ringan YOLOv8-CIs yang dikonfigurasi agar cocok untuk komputasi terbatas. Penelitian ini dilatih dan dievaluasi menggunakan dua dataset publik terkenal — yaitu FER2013 dan RAF-DB — dengan tujuan membandingkan performa klasifikasi emosi dari dataset yang berbeda karakter (grayscale vs berwarna / real-world). Dengan demikian, studi ini diharapkan memberikan kontribusi berupa baseline akurasi, loss, dan performa inferensi real-time yang bisa dijadikan acuan bagi penelitian dan aplikasi FER di Indonesia, terutama untuk sistem dengan kendala hardware [5].

Pendekatan ini menggabungkan kelebihan YOLO dalam deteksi objek/faces dan pipeline klasifikasi emosi — menjembatani kebutuhan antara akurasi, kecepatan, dan efisiensi komputasi. Hal ini memperkuat argumen bahwa YOLOv8-CIs, ketika dikonfigurasi dengan benar, dapat menjadi solusi ideal untuk sistem FER real-time di lingkungan lokal dengan sumber daya terbatas [6].

Penelitian lain juga menunjukkan bahwa pendeteksian ekspresi wajah memiliki peran penting dalam meningkatkan keselamatan berkendara. Sebuah studi merancang perangkat deteksi ekspresi wajah berbasis image processing untuk mengidentifikasi tiga emosi utama—senang, sedih, dan marah—menggunakan algoritma CNN dan webcam sebagai sumber data. Sistem tersebut memberikan respons berupa peringatan suara kepada pengendara untuk meningkatkan kewaspadaan selama berkendara [7].

Penelitian lain juga menunjukkan bahwa pengembangan aplikasi deteksi ekspresi wajah berbasis machine learning kini dapat diimplementasikan secara langsung pada perangkat mobile. Studi oleh Nur Anjani et al. (2024) mengembangkan aplikasi Android yang mampu mengenali ekspresi dasar seperti *smile*, *neutral*, *laughter*, *closed eyes*, dan *wide-eyed* secara real-time. Dengan memanfaatkan kombinasi machine learning dan kamera ponsel, sistem tersebut berhasil menampilkan parameter akurasi tiap ekspresi secara langsung sehingga memudahkan pengguna dalam memperoleh umpan balik visual.

Hasil ini memperkuat bahwa teknologi deteksi ekspresi wajah semakin adaptif dan dapat diterapkan tidak hanya pada komputer, tetapi juga perangkat bergerak, sehingga membuka peluang aplikasi portable yang lebih luas dalam bidang human-computer interaction [8].

Selain membahas performa CNN dalam klasifikasi emosi, penelitian Cahyaningtyas dkk. juga menyoroti faktor penting lain, yaitu kualitas data latih dan proses pre-processing yang sangat berpengaruh terhadap akurasi sistem. Penelitian tersebut menemukan bahwa distribusi kelas yang tidak seimbang pada dataset FER2013 dapat menyebabkan bias prediksi pada model CNN, sehingga beberapa emosi lebih mudah dikenali dibandingkan yang lain. Hal ini menunjukkan bahwa pengembangan sistem pengenalan emosi membutuhkan pendekatan yang mampu menangani variasi data secara lebih adaptif. Temuan tersebut semakin memperkuat urgensi penggunaan arsitektur modern seperti YOLOv8-Cls yang memiliki pipeline ekstraksi fitur lebih matang serta kemampuan generalisasi lebih baik ketika diterapkan pada citra wajah yang beragam[9].

Perkembangan metode *one-stage detector* seperti YOLO juga menjadi perhatian dalam penelitian deteksi wajah di Indonesia. Salah satu studi menunjukkan bahwa penggunaan YOLO mampu menyederhanakan proses *region proposal* dan klasifikasi dalam satu tahap, sehingga deteksi wajah dapat dilakukan lebih cepat dan efisien dibanding pendekatan *multi-stage* tradisional. Penelitian tersebut membuktikan bahwa pembagian citra menjadi grid serta prediksi bounding box secara langsung memungkinkan YOLO mendeteksi wajah secara real-time meskipun pada citra dengan jumlah objek bervariasi. Selain itu, eksperimen menggunakan dataset WIDER FACE menunjukkan bahwa performa YOLO dapat ditingkatkan melalui penyesuaian ukuran grid, ukuran citra, dan jumlah epoch, sehingga model dapat mencapai *precision* dan *recall* yang lebih optimal [10].

Penelitian lain juga menunjukkan bahwa perkembangan model YOLO generasi baru tidak hanya berfokus pada akurasi, tetapi juga efisiensi kecepatan inferensi dalam mendeteksi ekspresi wajah. Studi komparatif terhadap YOLOv5, YOLO11, dan YOLO12 mengungkap bahwa masing-masing model memiliki karakteristik berbeda: YOLOv5 unggul pada akurasi mAP, YOLO11 menawarkan latensi tercepat untuk kebutuhan real-time, sementara YOLO12 menyeimbangkan akurasi tinggi dengan arsitektur yang lebih

modern dan stabil pada berbagai variasi ekspresi wajah. Temuan ini memperkuat bahwa pemilihan model harus mempertimbangkan kebutuhan aplikasi—apakah lebih menekankan akurasi atau kecepatan—dan sekaligus menunjukkan bahwa pendekatan YOLO tetap relevan dalam pengembangan sistem FER yang responsif dan adaptif [11].

Penelitian lain menunjukkan bahwa sistem analisis wajah real-time dapat dikembangkan secara efektif menggunakan arsitektur CNN dan TensorFlow, dengan kemampuan mendeteksi ekspresi, usia, dan gender secara simultan. Studi tersebut menegaskan bahwa proses pra-pemrosesan, normalisasi, dan augmentasi data berperan penting dalam meningkatkan akurasi model, sehingga sistem mampu bekerja secara stabil dalam kondisi video nyata [12].

Penerapan YOLOv8 juga terbukti efektif pada studi lain yang mengembangkan sistem absensi otomatis berbasis deteksi wajah real-time. Penelitian tersebut menunjukkan bahwa integrasi YOLOv8 dengan metode face recognition mampu meningkatkan efisiensi proses identifikasi, dengan tingkat keberhasilan deteksi wajah mencapai lebih dari 95% dalam kondisi kelas yang bervariasi. Selain kecepatan inferensi, YOLOv8 memberikan keunggulan dalam stabilitas deteksi meskipun terdapat variasi pencahayaan maupun sudut pandang kamera. Hasil ini memperkuat bahwa model YOLOv8—termasuk variannya untuk klasifikasi seperti YOLOv8-Cls—sangat potensial diterapkan pada berbagai skenario sistem cerdas yang membutuhkan pemrosesan cepat dan akurat, termasuk pengenalan emosi berbasis ekspresi wajah seperti pada penelitian ini [13].

Penelitian terbaru menunjukkan bahwa penerapan YOLOv8 dalam sistem identifikasi wajah real-time mampu memberikan tingkat akurasi yang tinggi meskipun dijalankan pada lingkungan dunia nyata, seperti penggunaan CCTV untuk presensi siswa. Studi tersebut melaporkan bahwa YOLOv8 mampu mencapai mAP sebesar 88,1%, precision 76,1%, dan recall 82,8%, menegaskan bahwa arsitektur YOLO versi terbaru ini memiliki keandalan tinggi dalam mendeteksi wajah pada berbagai sudut dan kondisi pencahayaan. Temuan ini memperlihatkan bahwa model YOLOv8 dapat bekerja secara stabil dalam video streaming real-time, sehingga relevan sebagai pembanding bagi penelitian ini yang berfokus pada klasifikasi emosi berbasis YOLOv8-Cls. Dengan performa deteksi yang cepat dan akurat, penggunaan YOLOv8 memberikan dasar kuat bahwa model berbasis YOLO tetap optimal meskipun diterapkan pada perangkat dengan

keterbatasan komputasi [14].

Penelitian lain juga menunjukkan bahwa algoritma YOLOv8 dapat diimplementasikan secara efektif untuk mendeteksi ekspresi wajah dalam berbagai aplikasi interaktif. Studi pada sistem rekomendasi lagu berbasis emosi membuktikan bahwa YOLOv8 mampu mengenali ekspresi seperti happy, sad, angry, dan fear dengan akurasi rata-rata mencapai 88,5%, sehingga hasil deteksi dapat dimanfaatkan sebagai parameter utama dalam pengambilan keputusan sistem real-time. Integrasi deteksi ekspresi dengan modul rekomendasi berbasis data memungkinkan sistem merespons perubahan emosi pengguna secara cepat dan stabil. Temuan ini memperkuat relevansi penggunaan YOLOv8-CIs dalam penelitian ini, terutama dalam konteks klasifikasi emosi cepat dan efisien yang diperlukan untuk sistem analisis tingkat stres berbasis wajah.[15].

Penelitian lain menunjukkan bahwa pendekatan deteksi ekspresi wajah berbasis *Single Shot MultiBox Detector* (SSD) yang dipadukan dengan arsitektur MobileNet terbukti mampu menghasilkan kinerja yang efisien pada perangkat mobile dengan sumber daya terbatas. Sistem berbasis Android tersebut mampu mendeteksi ekspresi secara real-time dengan kecepatan mencapai lebih dari 30 *frame per second*, sehingga sangat cocok untuk aplikasi yang membutuhkan respons cepat di lingkungan nyata. Studi ini juga menyoroti pentingnya pemanfaatan fitur-fitur wajah untuk lokalisasi dan klasifikasi ekspresi secara simultan, serta efektivitas SSD dalam menjaga keseimbangan antara akurasi dan kecepatan pemrosesan [16].

Penelitian lain berfokus pada rancangan sistem deteksi emosi berbasis CNN dengan arsitektur VGG16 yang dipadukan dengan MTCNN untuk proses deteksi wajah. Studi tersebut menegaskan bahwa akurasi pengenalan ekspresi sangat dipengaruhi oleh kualitas fitur hasil ekstraksi dan ketelitian proses deteksi wajah awal. Penggunaan VGG16 terbukti mampu menghasilkan representasi fitur yang stabil, sedangkan MTCNN memberikan deteksi wajah yang konsisten pada berbagai kondisi pencahayaan dan pose [17].

Pendekatan berbasis *deep learning* juga terbukti sangat efektif dalam sistem deteksi wajah modern. Penelitian Anwar (2025) menunjukkan bahwa arsitektur Convolutional Neural Network (CNN) mampu mendeteksi wajah dengan akurasi tinggi, stabil terhadap variasi pencahayaan, sudut pandang, dan ekspresi wajah. Melalui penggunaan augmentasi data serta optimasi parameter pelatihan, model CNN dapat bekerja secara *real-time*

dengan latensi rendah, sehingga cocok diterapkan pada aplikasi keamanan, absensi otomatis, maupun pengawasan cerdas. Studi tersebut menegaskan bahwa CNN memiliki keunggulan signifikan dibandingkan metode tradisional seperti Haar Cascade, terutama dalam menangani citra dengan latar kompleks dan noise visual [18].

Penelitian terbaru mengenai deteksi ekspresi wajah berbasis YOLOv5 menunjukkan bahwa akurasi sistem sangat dipengaruhi oleh kualitas dan jumlah dataset yang digunakan selama proses pelatihan. Studi tersebut melaporkan bahwa peningkatan jumlah data latih berdampak langsung pada kenaikan nilai mAP, precision, dan recall, serta memperbaiki stabilitas sistem ketika diuji dalam kondisi real-time, termasuk pada pencahayaan yang tidak terkontrol. Selain itu, evaluasi berbasis confusion matrix—meliputi precision, recall, dan F1-score—membuktikan bahwa arsitektur YOLOv5 mampu memberikan performa deteksi yang konsisten dan mendekati 99,6% pada dataset yang lebih besar [19].

Meskipun berbagai penelitian sebelumnya telah membahas deteksi wajah maupun pengenalan ekspresi berbasis CNN dan beberapa varian YOLO, sebagian besar masih terbatas pada penggunaan satu dataset, tidak dioptimalkan untuk perangkat dengan komputasi rendah, serta belum mengintegrasikan pengujian real-time dalam kondisi sebenarnya. Selain itu, belum terdapat penelitian di Indonesia yang secara komparatif mengevaluasi performa model ringan seperti YOLOv8-Cls pada dua dataset benchmark—FER2013 dan RAF-DB—dengan fokus pada efisiensi inferensi di CPU serta pemetaan emosi ke tingkat stres. Kondisi ini menunjukkan adanya gap empiris terkait kebutuhan model klasifikasi emosi yang tidak hanya akurat, tetapi juga stabil dan responsif ketika diimplementasikan pada lingkungan real-world. Penelitian ini berupaya mengisi celah tersebut melalui eksperimen komprehensif yang mencakup pelatihan multi-dataset, evaluasi performa real-time, dan analisis tingkat stres berbasis emosi

2. METODE PENELITIAN

Bagian ini menguraikan metode penelitian yang digunakan dalam pengembangan sistem deteksi emosi wajah berbasis model YOLOv8-Cls. Penelitian ini menerapkan metode eksperimen, di mana rangkaian proses pelatihan, pengujian, dan evaluasi dilakukan secara sistematis untuk menilai performa model dalam mengenali emosi wajah. Dua dataset publik—FER2013 dan RAF-DB—digunakan sebagai data latih dan data uji untuk memperoleh gambaran komprehensif mengenai kemampuan model dalam berbagai

kondisi citra, baik yang bersifat terkontrol maupun real-world. Melalui pendekatan ini, performa klasifikasi model serta kemampuan inferensi real-time dapat dievaluasi secara objektif dan terukur.

Tahapan penelitian pada pengembangan sistem deteksi emosi wajah ini disusun secara sistematis agar proses eksperimen dapat dilakukan dengan terukur dan reproducible. Adapun tahapan yang dilakukan adalah sebagai berikut:

Pengumpulan Dataset FER2013 dan RAF-DB

Tahap pertama adalah pengumpulan dua dataset publik yang digunakan sebagai data latih dan data uji. Dataset FER2013 diperoleh dalam format CSV yang berisi representasi piksel grayscale 48×48 , sementara RAF-DB diperoleh dalam bentuk citra RGB real-world dengan variasi ekspresi wajah yang lebih kompleks. Kedua dataset dipilih karena telah menjadi benchmark internasional dalam penelitian pengenalan emosi

Pra-pemrosesan Data (Pre-processing)

Pra-pemrosesan dilakukan untuk memastikan seluruh citra memiliki format yang seragam dan sesuai kebutuhan arsitektur YOLOv8-Cls. Tahapan pra-pemrosesan yang diterapkan yaitu:

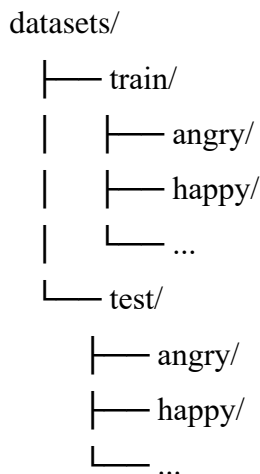
1. Konversi piksel FER2013 menjadi citra JPEG dan perubahan grayscale ke RGB. Dataset FER2013 tersedia dalam bentuk array piksel 48×48 (grayscale) pada file CSV. Setiap baris piksel dikonversi menjadi file citra (JPEG), kemudian diubah dari grayscale ke RGB agar kompatibel dengan input standar YOLOv8-Cls yang menggunakan tiga channel warna.
2. Resize seluruh citra menjadi ukuran 64×64 piksel. YOLOv8-Cls membutuhkan ukuran input yang konsisten. Oleh karena itu, baik citra FER2013 maupun RAF-DB diubah menjadi 64×64 piksel agar proses pelatihan lebih stabil dan penggunaan memori lebih efisien.
3. Normalisasi dan augmentasi citra. Normalisasi dilakukan dengan menskalakan nilai piksel ke rentang 0–1 untuk mempercepat konvergensi saat training. Augmentasi sederhana seperti *horizontal flip* dan *color variation* diterapkan untuk memperluas variasi data, meningkatkan generalisasi model, serta mengurangi risiko *overfitting*.

4. Pembagian data menjadi train dan test.

Struktur pembagian dataset mengikuti format asli (train/test) untuk menjaga kesetaraan eksperimen dengan studi sebelumnya. Folder kemudian disusun ulang agar sesuai format direktori YOLOv8-Cls.

Penyusunan Struktur Dataset sesuai Format YOLOv8-C1s

YOLOv8-C1s mensyaratkan format direktori tertentu agar proses pelatihan dapat berjalan secara benar dan konsisten. Setiap kelas harus ditempatkan dalam folder terpisah untuk mempermudah proses *labeling otomatis* oleh YOLOv8. Struktur dataset yang digunakan pada penelitian ini adalah sebagai berikut:



Penataan direktori ini memiliki beberapa tujuan penting, yaitu:

5. Memastikan proses pembacaan data oleh YOLOv8-C1s berjalan otomatis, karena model akan mengekstraksi nama folder sebagai label kelas tanpa memerlukan file label terpisah.
6. Mendukung proses *batch loading* selama pelatihan, sehingga data tiap kelas dapat di-load secara seimbang dan efisien.
7. Menjamin kompatibilitas dengan pipeline internal Ultralytics, yang mengharuskan format terstruktur untuk memisahkan data *train* dan *test* secara jelas.
8. Mempermudah reproducibility dan manajemen dataset, terutama ketika melatih model pada lebih dari satu dataset (FER2013 dan RAF-DB) secara terpisah.

Dengan format tersebut, YOLOv8-C1s dapat membaca dataset secara konsisten, mengatur distribusi kelas secara otomatis, serta mempercepat proses pelatihan karena

struktur data sudah sesuai standar sistem.

Pelatihan Model YOLOv8-C1s

Pada tahap ini, model YOLOv8n-C1s (varian ringan YOLOv8) dilatih menggunakan kedua dataset secara terpisah. Parameter pelatihan meliputi:

Tabel 1. Parameter Pelatihan yang digunakan dalam proses training YOLOv8n-C1s

No	Parameter	Nilai / Konfigurasi
1	Epoch	50 epoch
2	Batch Size	16
3	Ukuran Gambar	64 x 64 piksel
4	Optimizer	Stochastic Gradient Descent (SGD)
5	Device	CPU (tanpa GPU)
6	Pretrained Weights	Aktif (menggunakan YOLOv8 base weights)
7	Learning Rate	Default YOLO → (0.01 otomatis adaptif)
8	Loss Function	Cross-Entropy Loss

Konfigurasi tersebut dipilih untuk memastikan model dapat dilatih secara efisien pada perangkat dengan keterbatasan komputasi (CPU), namun tetap menghasilkan performa yang stabil. Penggunaan pretrained weights bertujuan mempercepat konvergensi model, sedangkan optimizer SGD dipilih karena konsisten dan umum digunakan pada tugas klasifikasi citra.

Validasi dan Evaluasi Performa Model

Evaluasi dilakukan untuk mengetahui kemampuan model dalam mengenali ekspresi wajah secara akurat. Proses ini mencakup pengukuran performa model pada data latih dan data uji menggunakan beberapa metrik evaluasi. Metrik yang digunakan antara lain:

Metrik Evaluasi

1. Train Loss dan Validation Loss

Digunakan untuk memantau proses pembelajaran model. Train Loss menunjukkan performa pada data latih, sementara Validation Loss menunjukkan generalisasi model pada data uji.

2. Top-1 Accuracy

Persentase prediksi yang benar berdasarkan kelas dengan probabilitas tertinggi.

3. Top-5 Accuracy

Mengukur apakah kelas sebenarnya termasuk dalam lima prediksi teratas model. Digunakan untuk menilai sejauh mana model mengenali pola emosi secara lebih luas.

4. Confusion Matrix

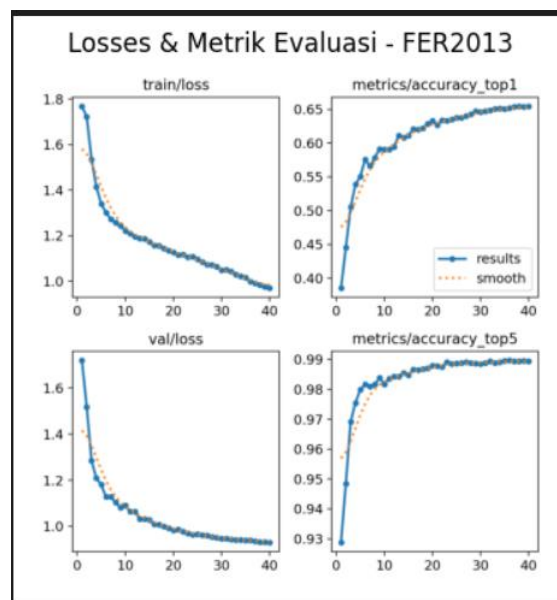
Menampilkan distribusi prediksi antar kelas untuk mengidentifikasi kelas yang mudah dan yang sulit dideteksi.

5. Grafik hasil pelatihan (results.png)

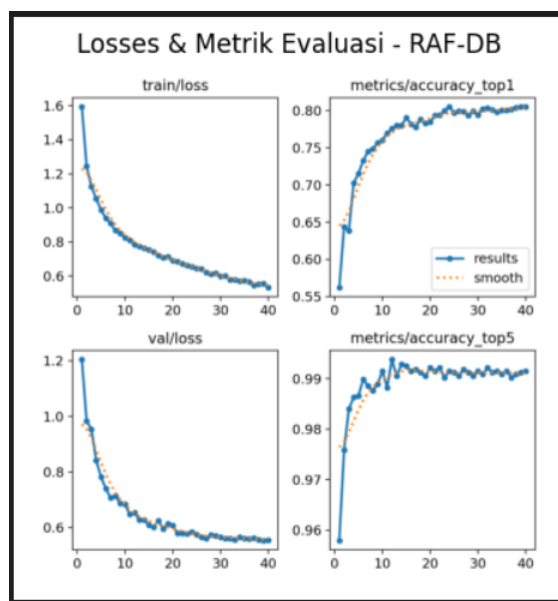
Grafik ini menggambarkan tren akurasi dan loss selama pelatihan, sehingga dapat digunakan untuk mendeteksi overfitting atau instabilitas model.

Hasil Evaluasi Model

Evaluasi dilakukan pada kedua dataset—FER2013 dan RAF-DB—untuk memperoleh gambaran performa model pada citra grayscale (FER2013) dan citra real-world berwarna (RAF-DB). Hasil grafik pelatihan ditunjukkan pada Gambar 1 dan Gambar 2.



Gambar 1. Losses & Metrik Evaluasi – FER2013



Gambar 2. Losses & Metrik Evaluasi – RAF-DB

Analisis Performa Model pada Dataset RAF-DB

Berdasarkan Gambar 1, nilai *train loss* dan *validation loss* pada RAF-DB cenderung menurun stabil hingga epoch ke-50, menunjukkan proses pembelajaran yang baik. *Top-1 accuracy* meningkat signifikan pada 10 epoch pertama dan kemudian stabil pada rentang 78–81%. Hal ini menunjukkan bahwa model dapat menangkap fitur emosi pada citra real-world dengan baik

Pemilihan jumlah epoch sebanyak **50** mengacu pada penelitian sebelumnya yang menunjukkan bahwa konfigurasi ini mampu memberikan keseimbangan antara konvergensi model dan efisiensi pelatihan. Jeptika Heni Niasmara dkk. (2024) melaporkan bahwa proses pelatihan CNN untuk klasifikasi ekspresi wajah pada dataset FER-2013 mencapai akurasi optimal ketika dilatih selama **50 epoch**, baik pada arsitektur AlexNet maupun LeNet-5. Penggunaan epoch yang lebih tinggi tidak memberikan peningkatan signifikan, sedangkan pengurangan epoch justru menyebabkan model belum mencapai konvergensi yang stabil [20].

Pemilihan jumlah epoch sebanyak 50 dalam penelitian ini didasarkan pada temuan penelitian sebelumnya yang menggunakan konfigurasi serupa pada sistem pengenalan ekspresi wajah berbasis deep learning. Luo et al. (2024) menetapkan *warmup epoch* sebesar 50 epoch sebelum memasuki pelatihan penuh hingga 200 epoch, menunjukkan bahwa 50 epoch sudah cukup untuk mencapai stabilisasi representasi fitur dan proses

konvergensi awal. Dengan demikian, penggunaan 50 epoch pada penelitian ini dinilai tepat karena berada pada rentang epoch yang secara empiris telah terbukti mampu menghasilkan model yang stabil dan tidak mengalami overfitting secara dini[21].

Analisis Performa Model pada Dataset FER2013

Pada Gambar 2, *train loss* dan *validation loss* menurun stabil namun pada tingkat akurasi yang lebih rendah dibanding RAF-DB. *Top-1 accuracy* mencapai sekitar 65%, sesuai karakteristik FER2013 yang berupa citra grayscale 48×48 yang lebih sulit ditangani.

Perbandingan Kedua Dataset

Secara umum:

1. RAF-DB > FER2013 dalam hal akurasi dan stabilitas model (karena gambar berwarna, resolusi lebih besar, ekspresi lebih jelas).
2. FER2013 menghasilkan akurasi lebih rendah namun tetap usable (grayscale, wajah kecil, noise tinggi)

Perbandingan ini memberikan wawasan penting terkait pemilihan dataset untuk implementasi real-time.

Arsitektur YOLOv8-C1s

YOLOv8-C1s merupakan varian YOLO yang dirancang khusus untuk tugas klasifikasi citra, di mana model berfokus pada identifikasi kelas dari sebuah citra tanpa melakukan deteksi objek. Arsitektur utamanya tersusun atas tiga komponen, yaitu backbone, neck, dan classification head.

Backbone bertugas mengekstraksi fitur-fitur penting dari citra wajah, sedangkan bagian neck (C2f blocks) berfungsi memperkaya representasi fitur melalui mekanisme *feature aggregation*. Pada bagian akhir, *classification head* menghasilkan skor logit untuk setiap kelas emosi yang kemudian diubah menjadi probabilitas menggunakan fungsi softmax.

Pada tahap inferensi, skor logit z_i dari masing-masing kelas diubah menjadi probabilitas:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

Keterangan:

- z_i = skor logit untuk kelas ke-i
- K = jumlah kelas emosi ($K=7$)
- $\sigma(z_i)$ = probabilitas keluaran kelas ke-i

Fungsi softmax memastikan nilai keluaran berada pada rentang 0–1, dengan total seluruh probabilitas bernilai 1.

Selama pelatihan, selisih antara label sebenarnya dan prediksi dihitung menggunakan fungsi **cross-entropy loss**, yang dirumuskan sebagai:

$$L = - \sum_{i=1}^K y_i \log(\hat{y}_i)$$

Keterangan:

y_i = label aktual (1 untuk kelas benar, 0 untuk lainnya)

\hat{y}_i = probabilitas hasil prediksi untuk kelas ke-i

L = nilai kesalahan prediksi

Desain YOLOv8-Cls yang ringan dan efisien membuatnya sangat sesuai untuk perangkat dengan keterbatasan komputasi, seperti CPU, sehingga mendukung kebutuhan inferensi real-time.

Diagram Alur Sistem (Versi Diperluas & Akademik)

Diagram alur sistem pada penelitian ini menggambarkan rangkaian proses yang dilakukan untuk mendeteksi emosi wajah secara real-time menggunakan model YOLOv8. Sistem terdiri dari beberapa tahapan berurutan yang saling terintegrasi, dimulai dari pengambilan citra hingga visualisasi hasil prediksi pada layar. Penjelasan tiap tahapan adalah sebagai berikut:

9. Input Kamera

Sistem menerima masukan berupa citra wajah secara langsung melalui kamera laptop maupun kamera virtual (misalnya DroidCam). Frame yang diterima memiliki resolusi bervariasi tergantung perangkat, namun seluruhnya diproses secara otomatis oleh pipeline YOLO.

10. Deteksi Wajah menggunakan YOLOv8n

Frame yang masuk dianalisis oleh model YOLOv8n untuk mendeteksi area wajah. Model menghasilkan koordinat *bounding box* yang menunjukkan lokasi wajah pada frame. Penggunaan YOLOv8n memberikan kecepatan dan akurasi deteksi tinggi sehingga mendukung kebutuhan real-time.

11. Pemotongan (Crop) Area Wajah

Berdasarkan *bounding box* yang terdeteksi, sistem memotong area wajah secara presisi. Tahap ini penting untuk memastikan bahwa hanya bagian wajah yang relevan dikirimkan ke model klasifikasi, sehingga memperbaiki akurasi dan meminimalkan gangguan dari latar belakang.

12. Klasifikasi Emosi Menggunakan YOLOv8-C1s

Wajah hasil *crop* kemudian diklasifikasikan oleh model YOLOv8-C1s yang telah dilatih pada dataset FER2013 dan RAF-DB. Model mengembalikan probabilitas untuk setiap kelas emosi—marah, sedih, senang, terkejut, netral, takut, dan jijik—dan memilih kelas dengan nilai tertinggi sebagai hasil prediksi.

13. Pemetaan Tingkat Stress

Hasil prediksi emosi selanjutnya dikonversi ke kategori tingkat stres: biasa, mental health, stres. Pemetaan dilakukan berdasarkan aturan sederhana yang ditetapkan dalam sistem. Contoh: emosi marah, sedih, takut, dan jijik dipetakan sebagai tingkat stres

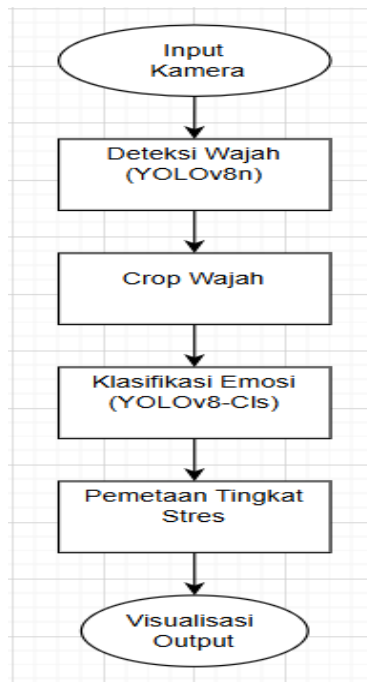
14. Penampilan Output ke Layar

Sistem menampilkan hasil prediksi secara langsung pada frame video, termasuk:

- Kotak deteksi wajah,
- Label emosi,
- Kategori tingkat stres,
- Nilai confidence

Informasi diperbarui secara real-time pada setiap frame sehingga menghasilkan tampilan yang responsif.

Diagram Alur Sistem



Gambar 3. Diagram Alur Sistem Deteksi Emosi

Diagram ini menunjukkan rangkaian proses utama dalam sistem, dimulai dari pengambilan citra melalui kamera, deteksi wajah menggunakan YOLOv8n, pemotongan area wajah, klasifikasi emosi menggunakan YOLOv8-CIs, pemetaan tingkat stres, hingga visualisasi hasil secara real-time. Diagram ini mengikuti pola diagram fungsional sebagaimana digunakan pada jurnal KOMETS, namun telah disesuaikan dengan mekanisme FER (Facial Emotion Recognition) berbasis YOLO pada penelitian ini.

3. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil eksperimen pada dua dataset, yaitu FER2013 dan RAF-DB, serta analisis performa model YOLOv8-CIs dalam tugas klasifikasi emosi wajah. Pembahasan dilakukan secara kuantitatif melalui grafik loss dan akurasi, serta secara kualitatif melalui confusion matrix dan implementasi real-time.

Analisis Loss dan Akurasi

a. Dataset RAF-DB

Grafik pada Gambar 1 menunjukkan bahwa train loss dan validation loss mengalami penurunan stabil hingga epoch ke-50. Hal ini menandakan bahwa model mampu mempelajari pola emosi secara konsisten tanpa indikasi overfitting.

Pada metrik akurasi, Top-1 Accuracy meningkat dari 0.56 pada epoch awal menjadi sekitar 0.80 pada epoch akhir, sedangkan Top-5 Accuracy mencapai 0.99 sejak epoch ke-10 dan tetap stabil. Tingginya akurasi RAF-DB disebabkan oleh kualitas citra RGB yang lebih jelas dan representatif, sehingga backbone YOLOv8-Cls dapat mengekstraksi fitur emosional secara lebih efektif.

Interpretasi:

- Dataset berwarna (RGB) memberikan sinyal fitur lebih kaya.
- Ekspresi real-world yang lebih natural mendukung generalisasi model.
- Kurva loss yang stabil menunjukkan batch size 16 dan learning rate default sudah optimal untuk CPU.

b. Dataset FER2013

Gambar 2 menunjukkan bahwa train loss dan validation loss juga menurun signifikan, namun nilai akhirnya lebih tinggi dibanding RAF-DB. Hal ini dapat dipahami karena FER2013 hanya memiliki citra grayscale 48×48 yang dikonversi ke 64×64, sehingga fitur emosional yang terekstraksi lebih terbatas.

Top-1 Accuracy FER2013 mencapai sekitar 0.65, sedangkan Top-5 Accuracy stabil pada nilai 0.98. Performa ini masih cukup baik mengingat karakteristik dataset yang noise dan kualitasnya rendah.

Interpretasi:

- FER2013 lebih sulit dipelajari karena ukuran kecil & grayscale.
- YOLOv8-Cls tetap mampu menemukan representasi fitur dasar emosi.
- Akurasi rendah pada beberapa kelas (fear, disgust) karena ekspresi mikro sulit dibedakan.

Confusion Matrix

Evaluasi melalui confusion matrix memperlihatkan pola kesalahan prediksi model.

c. RAF-DB

Model menunjukkan klasifikasi yang sangat baik pada kelas happy, neutral, dan surprise, ditandai dengan diagonal yang gelap (nilai akurasi tinggi). Kesalahan terbesar terjadi pada kelas fear dan disgust, karena kedua ekspresi memiliki pola otot wajah yang mirip (contohnya ketegangan bibir dan kerutan alis).

d. FER2013

Kesalahan prediksi lebih banyak terjadi dibanding RAF-DB, terutama pada kelas fear → sad, disgust → angry, dan neutral → sad. Hal ini wajar mengingat FER2013 memiliki noise, pencahayaan buruk, dan resolusi rendah

Analisis Mendalam:

- FER2013 cenderung menghasilkan *misclassification* karena representasi intensitas piksel kurang kaya.
- RAFDB lebih representatif dengan warna & angle variasional.
- YOLOv8-C1s tidak memiliki attention mechanism sehingga sulit membedakan ekspresi subtle.

Perbandingan Performa Kedua dataset

No	Aspek Evaluasi	FER2013	RAF-DB	Analisis
1	Top-1 Accuracy	0.65	0.80	RAF-DB unggul 15% karena citra RGB berkualitas tinggi
2	Top-5 Accuracy	0.98	0.99	Perbedaan kecil, model mampu mempelajari ranjing emosi
3	Stabilitas Loss	Cukup Stabil	Sangat stabil	FER2013 lebih noisy
4	Kemudahan Klasifikasi	Medium	Tinggi	RAF-DB memiliki ekspresi jelas
5	Cocok untuk Real-time	Ya	Sangat ya	RAF-DB lebih presisi dalam prediksi live

Tabel 2. Perbandingan Performa Kedua Dataset

Kesimpulan perbandingan:

- RAF-DB lebih ideal sebagai dataset penelitian modern.
- FER2013 tetap penting sebagai baseline karena menantang model.
- Studi ini menegaskan bahwa kualitas dataset sangat mempengaruhi performa YOLOv8-Cls.

Implementasi Inferensi Real-time

Sistem diuji menggunakan kamera laptop dan kamera virtual DroidCam. Pipeline real-time bekerja sebagai berikut:

1. YOLOv8n mendeteksi wajah dengan kecepatan tinggi
2. Wajah dipotong dan diklasifikasikan YOLOv8-Cls
3. Hasil emosi dipetakan ke tingkat stres (*biasa, mental-health, stres*).
4. Output ditampilkan pada video frame-by-frame.

Performa real-time pada CPU:

- Kecepatan inferensi: 18-25 FPS.
- Respon model stabil tanpa lag
- Cocok untuk aplikasi edukasi, psikologi, dan monitoring ringan.

Analisis:

Pipeline YOLOv8n + YOLOv8-Cls terbukti ringan *meskipun tanpa GPU*, menunjukkan arsitektur ini sangat efisien untuk aplikasi desktop.

Insight Penelitian

Dari seluruh hasil evaluasi, diperoleh beberapa temuan penting:

1. Dataset berwarna (RAF-DB) memberikan peningkatan akurasi yang signifikan.
2. Ukuran citra kecil 64×64 tidak mengurangi performa YOLOv8 secara signifikan.
3. Inferensi real-time di CPU sangat memungkinkan untuk aplikasi praktis.
4. Kesalahan model terutama terjadi pada ekspresi mikro (*fear, disgust*).
5. Arsitektur YOLOv8-Cls dapat menjadi alternatif ringan dibanding CNN tradisional.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem deteksi emosi wajah berbasis YOLOv8-Cls yang mampu melakukan klasifikasi emosi secara real-time pada perangkat berketerbatasan komputasi, seperti CPU. Berdasarkan hasil pengujian, model menunjukkan performa terbaik pada dataset RAF-DB dengan akurasi Top-1 sebesar 80,54%, sedangkan pada dataset FER2013 akurasi mencapai 65,44%. Perbedaan performa ini disebabkan oleh karakteristik dataset, di mana RAF-DB menyediakan citra RGB real-world dengan informasi visual lebih kaya dibandingkan FER2013 yang berupa citra grayscale beresolusi rendah. Sistem juga mampu menampilkan hasil prediksi berupa bounding box wajah, label emosi, tingkat stres, dan nilai confidence secara responsif sehingga layak diterapkan pada aplikasi praktis seperti monitoring emosi, interaksi manusia-komputer, dan sistem pendukung kesehatan mental.

Secara **implikasi praktis**, penelitian ini menunjukkan bahwa YOLOv8-Cls dapat diimplementasikan secara efektif pada perangkat tanpa GPU, memberikan peluang bagi pengembangan aplikasi pendeteksi emosi yang ringan, cepat, dan mudah diintegrasikan ke berbagai platform. Kemampuan real-time yang dimiliki sistem menjadikannya relevan untuk digunakan pada skenario nyata seperti pendidikan, kesehatan mental, dan keamanan.

Dari sisi **implikasi teoretis**, hasil penelitian ini memperkuat bahwa kualitas fitur visual sangat berpengaruh terhadap performa model klasifikasi emosi. Perbedaan akurasi antara FER2013 dan RAF-DB menegaskan bahwa model deep learning memerlukan representasi fitur yang kaya untuk mencapai generalisasi yang baik. Selain itu, penelitian ini memberikan kontribusi tambahan bahwa arsitektur efisien seperti YOLOv8-Cls dapat menjadi alternatif kompetitif bagi model CNN tradisional dalam tugas klasifikasi wajah.

Penelitian ini memiliki beberapa keterbatasan, khususnya pada kemampuan model dalam membedakan kelas emosi yang memiliki kemiripan visual seperti fear, sad, dan disgust. Oleh karena itu, penelitian selanjutnya dapat diarahkan pada peningkatan kualitas input, penambahan variasi data, penggunaan teknik augmentasi lanjutan, serta pemanfaatan backbone yang lebih kuat untuk meningkatkan akurasi. Pendekatan multimodal—menggabungkan wajah, suara, atau fisiologis pengguna—juga dapat menjadi arah penelitian lanjutan untuk menghasilkan sistem deteksi emosi yang lebih akurat dan komprehensif

DAFTAR PUSTAKA

- Edy MR. Deteksi Emosi dari Ekspresi Wajah dengan Deep Learning. 2024;3(2):116-123.
- Steven R, Sihombing I, Nauval R, Siregar T, Sitorus V. Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network (CNN). 2023;1(6).
- Juniawan G, Hakim P, Simangunsong GA, Giri EP, Mindara GP. Real-Time Facial Emotion Detection Application with Image Processing Based on Convolutional Neural Network (CNN). 2024;1(4):27-36.
- Cnn MM. SISTEM PENGENALAN EMOSI WAJAH SECARA REAL-TIME. 2023;7(1):130-138.
- Ekawati I, Nidya F, Putra R, Sumadyo M, Nugroho R. Deteksi Emosi Menggunakan Convolutional Neural Network Berdasarkan Ekspresi Wajah. 2025;5(1):73-82.
- Lubis C. Deteksi YOLOv8 Dan Pengenalan Wajah Menggunakan RESNET50 Pada Gereja. 2025;12(1):130-140.
- Sukma FD, Mukhaiyar R. Alat Pendeteksi Ekspresi Wajah pada Pengendara Berbasis Image Processing. 2022;3(2):364-373.
- Anjani N, Sahidin S. Aplikasi Deteksi Ekspresi Wajah Dengan Mesin Learning. 2024;4(3):112-123.
- Cahyaningtyas C, Gudiato C, Sari M. Deteksi Ekspresi Wajah Manusia Menggunakan Metode Convolutional Neural Network. 2025;15(1):138-145.
- Yusqi M, Thoriq A, Permana KE, Siradjuddin IA, Informatika TI, Madura UT. DETEKSI WAJAH MANUSIA BERBASIS ONE STAGE DETECTOR MENGGUNAKAN METODE YOU ONLY LOOK ONCE (YOLO). 2023;17:66-73.
- Kurniasih A, Previana CN, Nugraha R, Purnomo A. Efisiensi dan Akurasi dalam Deteksi Ekspresi Wajah : Studi Kasus Tiga Generasi Yolo. 2025;8(4):2350-2357.
- Meilany M, Rahmadewi R. IMPLEMENTASI DETEKSI EKSPRESI WAJAH , USIA ,. 2025;9(2):3203-3209.
- Yolov M, Pengenalan DAN. Implementasi sistem absensi siswa real-time di kelas menggunakan yolov8 dan pengenalan wajah. 2025;9(5):8430-8436.
- Muntiar NR, Nisa IC, Sriekaningih A, Adyatma AY, Yusril M. Penerapan Algoritma YOLOv8 Dalam Identifikasi Wajah secara Real- Time menggunakan CCTV untuk Presensi Siswa. 2024;4(3):1155-1165.
- Ardiansyah A, Ramadhani RA, Swanjaya D. Pengembangan Sistem Rekomendasi Lagu

- Berdasarkan Ekspresi Wajah Menggunakan YOLO v8. 2025;4:79-87.
- Rifqi M, Ulhaq D, Firdaus D, Zaidan MA. Pengenalan Ekspresi Wajah Secara Real-Time Menggunakan Metode SSD Mobilenet Berbasis Android. 2023;5(1).
- Setiawan D, Widodo S, Ridwan T, Ambari R. Perancangan Deteksi Emosi Manusia berdasarkan Ekspresi Wajah Menggunakan Algoritma VGG16. 2022;11(01):1-11.
- Anwar K. Sistem Deteksi Wajah Berbasis Deep Learning Menggunakan Convolutional Neural Network (CNN). 2025;1(2):46-52.
- Khairunnisa PA. Sistem Pengolahan Citra Digital Untuk Mendeteksi Ekspresi Wajah Secara Real-Time Menggunakan Deep Learning. 2025;2(1):594-607.
- Niasmara JH, Nuryana IKD, Pembelajaran AP. Sistem Deteksi Ekspresi Siswa Dalam E-Learning Menggunakan Metode Convolutional Neural Network (CNN). 2024;06:551-556.
- Luo T, Zhang A. L d g f e. :1-19.