

RANCANG BANGUN SISTEM PENGENALAN EKSPRESI BAYI BERBASIS YOLOV11 UNTUK PEMANTAUAN JARAK JAUH

Isa Mahfudi, Bylliano Helmiananda, Dianthy Marya,
Mila Kusumawardani, Chandrasena Setiadi, Hadiwiyatno

Teknik Elektro, Politeknik Negeri Malang

Jl. Soekarno Hatta No.9, Jatimulyo, Kec. Lowokwaru, Kota Malang, Jawa Timur

Isa_mahfudi@polinema.ac.id

ABSTRAK

Perkembangan teknologi dan meningkatnya jumlah keluarga dengan kedua orang tua bekerja menimbulkan tantangan baru dalam pemantauan ekspresi emosional bayi, terutama karena bayi belum mampu berkomunikasi secara verbal dan bergantung pada ekspresi wajah sebagai bentuk utama komunikasi. Sistem pemantauan bayi yang ada umumnya masih terbatas pada deteksi suara atau *video streaming* tanpa analisis otomatis terhadap ekspresi wajah. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem pengenalan ekspresi bayi berbasis YOLOv11 yang diintegrasikan dengan Raspberry Pi 5 sebagai perangkat *edge computing* serta *Firebase Realtime Database* untuk mendukung pemantauan jarak jauh dan notifikasi real-time kepada orang tua. Model dilatih menggunakan dataset tiga kelas ekspresi yakni tidur, senyum, dan menangis. Konfigurasi terbaik diperoleh pada *batch size* 16, *learning rate* 0.001, dan 20 *epoch*. Hasil pengujian menunjukkan akurasi tertinggi sebesar 91,2% pada kelas tidur, dengan nilai *precision* 0.85, *recall* 0.83, dan *mAP@50* 0.87. Pengujian notifikasi menunjukkan sistem mampu mengirimkan data dengan latensi rata-rata 1,24 detik pada jaringan Wi-Fi dan 3,45 detik pada jaringan seluler 4G. Hasil ini membuktikan bahwa sistem mampu mendeteksi ekspresi bayi secara akurat dan mengirimkan notifikasi secara cepat dan andal melalui koneksi internet, sehingga mendukung pemantauan bayi jarak jauh secara efisien dan aman.

Kata kunci : Ekspresi Bayi, Firebase, Notifikasi Real-time, Pemantauan Jarak Jauh, YOLOv11.

1. PENDAHULUAN

Perkembangan teknologi dan dinamika sosial, terutama meningkatnya jumlah keluarga dengan kedua orang tua bekerja, telah menimbulkan tantangan baru dalam pengasuhan anak. Berdasarkan data dari Biro Statistik Tenaga Kerja Amerika, sekitar 49,6% keluarga menikah, memiliki kedua pasangan yang sama-sama bekerja[1]. Kondisi ini berdampak pada terbatasnya waktu orang tua untuk melakukan pengawasan terhadap ekspresi emosional bayi secara langsung, terutama pada usia dini ketika bayi sangat bergantung pada sinyal non-verbal untuk berkomunikasi[2]. Pada fase perkembangan usia 3 hingga 12 bulan, ekspresi wajah seperti tersenyum, sedih, dan menangis menjadi bentuk utama komunikasi sosial awal bayi, riset menunjukkan bayi usia 3–7 bulan sudah mampu membedakan dan memadankan ekspresi wajah dan suara[3]. Sensitivitas orang tua dalam mengenali dan merespons ekspresi tersebut terbukti memiliki korelasi positif dengan pola pengasuhan yang lebih responsif dan mendukung perkembangan sosial emosional anak[4].

Di sisi teknologi, berbagai sistem pemantauan bayi yang telah dikembangkan umumnya masih berfokus pada penggunaan sensor lingkungan, deteksi suara tangisan, atau *streaming video* tanpa kemampuan analisis otomatis terhadap ekspresi wajah secara *real-time*. Pendekatan berbasis audio sering kali rentan terhadap kebisingan lingkungan dan kesalahan klasifikasi, sedangkan pemantauan manual melalui video menuntut perhatian terus-menerus dari orang tua

dan berisiko menyebabkan keterlambatan respons terhadap kondisi bayi.

Penelitian terkini di bidang Computer Vision telah memperkenalkan sejumlah dataset ekspresi bayi serta penerapan model *Convolutional Neural Network* (CNN) dan *You Only Look Once* (YOLO) yang menjanjikan untuk deteksi ekspresi wajah secara otomatis. Namun demikian, integrasi sistem pengenalan ekspresi bayi secara real-time dengan kemampuan pemantauan jarak jauh dan notifikasi otomatis kepada orang tua masih sangat terbatas[5][6][7][8]. Kondisi ini menjadi peluang pengembangan sistem berbasis YOLOv11 yang tidak hanya mampu mengenali ekspresi bayi secara cepat dan akurat, tetapi juga dapat diimplementasikan dalam ekosistem IoT untuk mendukung pemantauan jarak jauh secara efisien.

Kemajuan pada model deteksi YOLO telah membuka peluang besar untuk meningkatkan kecerdasan sistem pemantauan bayi. Versi terbarunya, YOLOv11, merupakan pengembangan arsitektur yang menawarkan peningkatan signifikan pada akurasi deteksi objek kecil, efisiensi komputasi, dan kemampuan generalisasi lintas domain[9][10]. Karakteristik ini sangat relevan dengan kebutuhan pengenalan ekspresi wajah bayi, yang umumnya memiliki ukuran wajah kecil di bidang pandang kamera, pose yang dinamis, sering mengalami *occlusion* (misalnya tertutup selimut atau tangan), serta berada dalam kondisi pencahayaan ruangan yang bervariasi.

Dengan dukungan perangkat *edge computing* seperti Raspberry Pi, proses inferensi dapat dilakukan langsung di perangkat lokal (*on-device inference*). Pendekatan ini menjaga latensi sistem tetap rendah dan menghemat bandwidth, karena proses deteksi utama tidak bergantung pada server eksternal. Selain efisiensi, pendekatan ini juga mendukung keamanan dan privasi data, mengingat informasi yang diproses berupa citra wajah bayi. Prinsip *privacy-by-design* perlu diterapkan melalui inferensi lokal, penyimpanan data terenkripsi, kontrol akses berbasis peran, serta opsi anonimisasi visual untuk menjaga kerahasiaan data pengguna[11].

Dari sisi arsitektur, sistem pengenalan ekspresi bayi yang ideal tidak hanya mengandalkan deteksi visual, tetapi juga harus andal dan berkelanjutan. Hal ini mencakup proses mulai dari akuisisi citra, prapemrosesan, deteksi wajah, klasifikasi ekspresi, hingga manajemen notifikasi berbasis kejadian (*event-driven alerts*) dan antarmuka pemantauan jarak jauh bagi orang tua. Integrasi menyeluruh ini diharapkan mampu menghasilkan sistem yang responsif, aman, serta efisien untuk digunakan dalam lingkungan rumah tangga. Berdasarkan kebutuhan dan celah penelitian tersebut, penelitian ini mengusulkan “Rancang Bangun Sistem Pengenalan Ekspresi Bayi Berbasis YOLOv11 untuk Pemantauan Jarak Jauh.” Penelitian ini difokuskan pada pengembangan model deteksi dan klasifikasi ekspresi bayi berbasis YOLOv11 yang tangguh terhadap kondisi lingkungan rumah, implementasi arsitektur berbasis edge device berlatensi rendah, serta integrasi modul notifikasi real-time dan antarmuka pemantauan jarak jauh. Evaluasi sistem dilakukan menggunakan metrik akurasi deteksi (mAP, F1-score), latensi end-to-end, frame rate (FPS), serta tingkat kesalahan alarm (false alarm rate) untuk menunjukkan manfaat praktis sistem dalam pemantauan ekspresi bayi sehari-hari di rumah.

2. TINJAUAN PUSTAKA

Perkembangan teknologi visi komputer (*Computer Vision*) dan pembelajaran mendalam (*Deep Learning*) telah mendorong kemajuan signifikan dalam bidang pengenalan citra, termasuk penerapannya pada sistem pemantauan bayi. Berbagai penelitian sebelumnya telah dilakukan untuk mengenali ekspresi wajah, mendeteksi aktivitas, maupun menganalisis sinyal emosi bayi dengan memanfaatkan model berbasis YOLO. Kajian terhadap penelitian terdahulu diperlukan untuk memahami sejauh mana kemajuan yang telah dicapai, metode yang digunakan, serta keterbatasan yang masih terdapat pada sistem-sistem sebelumnya.

Penelitian mengenai pengenalan ekspresi wajah bayi telah dilakukan oleh Lin et al. [12] melalui pengembangan Feature Guided CNN menggunakan dataset BabyExp yang berisi lebih dari 12.000 citra bayi berusia di bawah dua tahun dengan tiga kategori

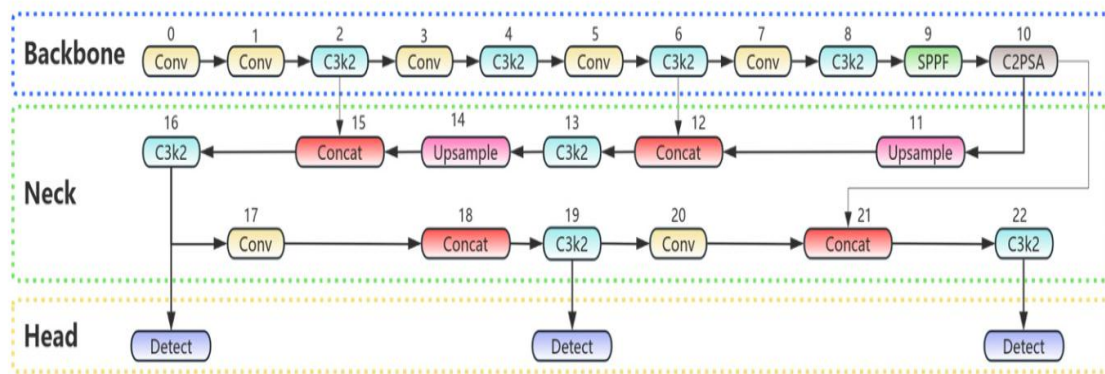
ekspresi utama, yaitu senang, sedih, dan biasa. Model ini menggunakan pendekatan konvolusional terarah (*feature-guided convolutional blocks*) untuk memperkuat deteksi fitur wajah halus. Kelebihan penelitian ini adalah tersedianya dataset khusus bayi yang valid, sedangkan kelemahannya terletak pada keterbatasan generalisasi model terhadap variasi pencahayaan dan orientasi wajah yang beragam. Sementara itu, Hammal et al. [13] mengusulkan pendekatan berbasis Facial Action Coding System (FACS) dengan menyesuaikan skema BabyFACS untuk mendeteksi pergerakan mikro pada otot wajah bayi. Metode ini unggul dalam presisi deteksi facial action units, namun memiliki kelemahan berupa kompleksitas tinggi dan ketergantungan terhadap proses pelabelan manual yang memakan waktu, sehingga kurang sesuai untuk implementasi real-time.

Selanjutnya, Laraib et al. [14] menerapkan arsitektur Deep Convolutional Neural Network (CNN) untuk mengenali tujuh ekspresi dasar (happy, sad, angry, disgust, fear, surprise, dan neutral) pada anak-anak menggunakan dataset LIRIS-CSE. Model ini mencapai akurasi 90,12% dan menunjukkan kemampuan generalisasi yang baik terhadap ekspresi anak. Akan tetapi, model tersebut belum dioptimalkan untuk ukuran wajah kecil serta karakteristik ekspresi bayi, sehingga tidak dapat langsung diterapkan pada domain bayi tanpa penyesuaian arsitektur model.

Penelitian Mekhfioui et al. [15] mengembangkan sistem identifikasi tangisan bayi berbasis Deep Learning menggunakan kombinasi CNN dan LSTM untuk membedakan tangisan bayi sehat dan sakit. Model ini unggul dalam pemrosesan audio temporal, tetapi memiliki kelemahan pada tingkat kesalahan klasifikasi akibat kebisingan lingkungan. Pendekatan ini relevan sebagai pembanding dengan sistem berbasis citra wajah yang lebih tahan terhadap gangguan suara. Lalu peneliti he et al [10] memperkenalkan penerapan YOLOv11 pada deteksi objek di lingkungan kompleks, menunjukkan peningkatan signifikan dalam akurasi deteksi objek kecil serta efisiensi pemrosesan pada perangkat dengan sumber daya terbatas. Hasil tersebut menunjukkan potensi besar YOLOv11 untuk diterapkan pada pengenalan ekspresi bayi, di mana ukuran wajah kecil dan kondisi lingkungan rumah menjadi tantangan utama.

2.1. Arsitektur YOLOv11

YOLOv11 adalah iterasi terkini dari seri YOLO yang memperkenalkan inovasi arsitektur penting seperti blok C3k2 (*Cross Stage Partial* dengan kernel 2), SPPF (*Spatial Pyramid Pooling – Fast*), dan C2PSA (*Convolutional block with Parallel Spatial Attention*) yang meningkatkan kemampuan ekstraksi fitur, terutama pada objek-objek kecil dan dalam kondisi pencahayaan atau sudut pandang yang variatif.



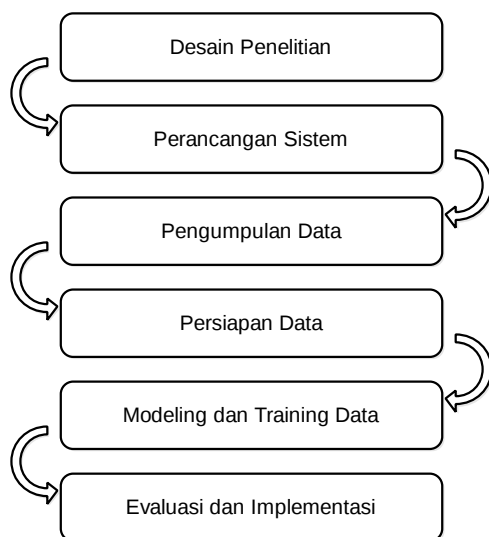
Gambar 1. Diagram Arsitektur Jaringan YOLOv11

2.2. RoboFlow

Roboflow adalah sebuah platform berbasis cloud yang dirancang untuk mengelola, melabeli, memproses, dan menyiapkan dataset citra agar siap digunakan dalam pelatihan (training) model Computer Vision seperti YOLO, TensorFlow, atau PyTorch. Platform ini mendukung seluruh siklus kerja (workflow) pengembangan model deteksi objek, mulai dari import data mentah, anotasi (labelling), augmentasi, preprocessing, hingga ekspor dataset ke berbagai format standar seperti YOLO, COCO, dan Pascal VOC[16].

3. METODE PENELITIAN

Metode penelitian ini disusun dalam beberapa tahapan yang menggambarkan alur pengembangan sistem secara menyeluruh. Tahapan tersebut dimulai dari desain penelitian hingga proses implementasi dan evaluasi sistem berbasis YOLOv11. Diagram alur tahapan metode penelitian ditunjukkan pada Gambar 2.

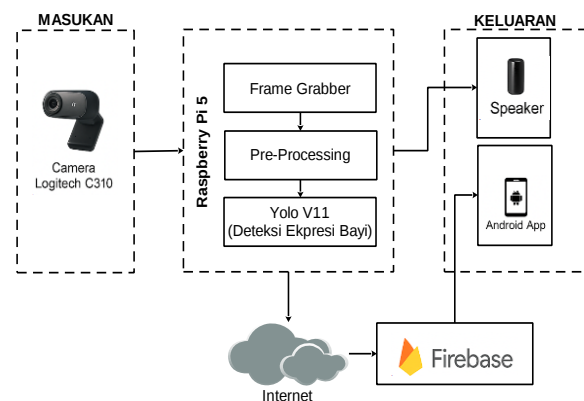


Gambar 2. Tahapan Metode Penelitian.

3.1. Perancangan Sistem

Perancangan sistem dilakukan untuk menggambarkan hubungan fungsional antara komponen input, pemrosesan, dan output pada sistem

pengenalan ekspresi bayi berbasis YOLOv11. Sistem ini mengintegrasikan kamera Logitech C310 sebagai perangkat akuisisi citra, Raspberry Pi 5 sebagai unit pemrosesan utama, serta Firebase dan aplikasi Android sebagai media komunikasi jarak jauh. Diagram blok sistem secara umum ditunjukkan pada Gambar 3.

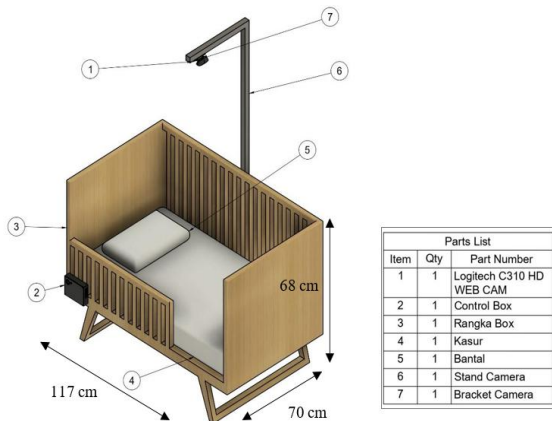


Gambar 3. Diagram blok sistem.

Sistem yang dikembangkan terdiri dari empat komponen utama, yaitu sistem masukan, pemrosesan, keluaran, dan cloud. Komponen masukan menggunakan kamera Logitech C310 untuk menangkap ekspresi wajah bayi secara real-time dan mengirimkan citra ke Raspberry Pi 5 melalui antarmuka USB. Pada sistem pemrosesan, Raspberry Pi 5 menjalankan algoritma YOLOv11 untuk mendeteksi dan mengklasifikasikan ekspresi bayi. Proses ini meliputi frame grabbing, pre-processing (seperti resizing dan normalization), serta inferensi model untuk mengenali ekspresi dasar seperti senang, netral, dan menangis. Hasil deteksi kemudian diteruskan ke sistem keluaran, yang terdiri dari speaker dan aplikasi Android. Speaker akan mengeluarkan musik penenang secara otomatis saat bayi menangis, sedangkan aplikasi Android menampilkan notifikasi ekspresi bayi kepada orang tua.

Firebase digunakan sebagai sistem cloud untuk menghubungkan Raspberry Pi dengan aplikasi Android melalui internet. Platform ini memungkinkan sinkronisasi data secara real-time dengan latensi

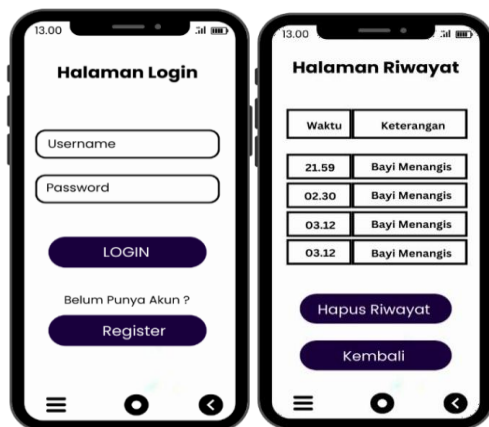
rendah, sehingga mendukung pemantauan kondisi bayi secara efisien dan jarak jauh. Perancangan dari sistem yang diusulkan ditunjukkan pada Gambar 4.



Gambar 4. Desain Perancangan sistem

3.2. Perancangan aplikasi

Perancangan aplikasi dilakukan untuk mengembangkan antarmuka yang dapat menampilkan hasil deteksi ekspresi bayi secara real-time dan memberikan notifikasi otomatis kepada pengguna. Aplikasi ini dirancang berbasis Android dengan dukungan layanan Firebase sebagai media komunikasi antara sistem deteksi di Raspberry Pi dan perangkat pengguna. Perancangan aplikasi ini ditunjukkan pada Gambar 5.



Gambar 4. Perancangan aplikasi

4. HASIL DAN PEMBAHASAN

Bab ini membahas hasil implementasi sistem pengenalan ekspresi bayi berbasis YOLOv11 yang telah dirancang dan diujikan. Pengujian dilakukan untuk mengevaluasi kinerja sistem dari sisi akurasi deteksi ekspresi, kecepatan inferensi, serta efektivitas sistem dalam memberikan notifikasi real-time melalui aplikasi Android. Hasil dari implementasi perangkat ini ditunjukkan pada Gambar 5.



Gambar 5. Hasil implementasi sistem

4.1. Hasil Deteksi Ekspresi YOLOv11

Pengujian sistem dilakukan untuk mengevaluasi kemampuan model YOLOv11 dalam mengenali ekspresi wajah bayi secara *real-time*. Pengujian dilakukan pada tiga kelas utama ekspresi bayi, yaitu tidur, senyum, dan menangis, dengan mengukur nilai *confidence* (tingkat keyakinan model terhadap hasil prediksi) serta *accuracy* (tingkat kebenaran hasil deteksi terhadap data uji).

4.2. Hasil Confidence Deteksi Ekspresi

Nilai *confidence* diperoleh langsung dari keluaran model YOLOv11 untuk setiap frame citra yang terdeteksi. Nilai ini menunjukkan seberapa yakin model terhadap prediksi kelas ekspresi yang dihasilkan. Hasil pengujian awal disajikan pada Tabel 1.

Tabel 1. Hasil Pengujian YOLOv11 untuk mengenali ekspresi bayi

No.	Ekspresi Terdeteksi	Nilai Confidence		
		Tidur	Senyum	Menangis
1		-	-	0,48
2		-	-	0.59
3		0,94	-	-
4		0.76	-	-

No.	Ekspresi Terdeteksi	Nilai Confidence		
		Tidur	Senyum	Menangis
5		-	0,71	-
6		-	0.29	-
Rata-Rata		0,94	0,71	0,48

Berdasarkan Tabel 1, terlihat bahwa nilai confidence rata-rata bervariasi antar kelas ekspresi. Nilai tertinggi terdapat pada kelas Tidur dengan rata-rata confidence sebesar 0.94, diikuti senyum (0.71) dan menangis (0.48). Nilai confidence yang tinggi menunjukkan model memiliki keyakinan lebih besar terhadap hasil prediksi pada kelas tersebut. Perbedaan ini dipengaruhi oleh karakteristik visual dari setiap ekspresi. Ekspresi tidur memiliki ciri yang relatif konstan dan mudah dikenali, seperti mata tertutup dan posisi kepala tetap. Sebaliknya, ekspresi senyum dan menangis memiliki variasi bentuk mulut, orientasi wajah, serta intensitas pencahayaan yang dapat mempengaruhi keyakinan model.

4.3. Hasil Akurasi Deteksi

Untuk mendapatkan gambaran performa model secara menyeluruh, dilakukan pengujian terhadap setiap kelas ekspresi sebanyak 10 kali percobaan. Akurasi diperoleh dengan cara :

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Seluruh Data Uji}} \times 100\% \quad (1)$$

Hasil rata-rata akurasi deteksi dari model YOLOv11 ditampilkan pada Tabel 2.

Tabel 2. Hasil pengujian akurasi

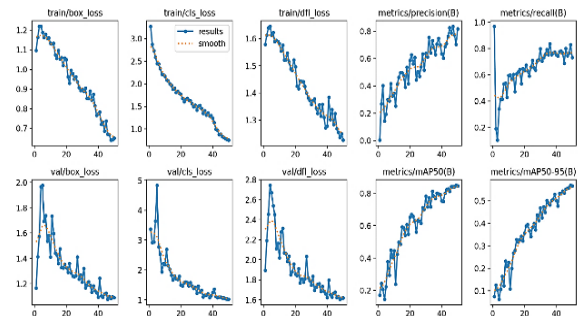
No. Pengujian	Akurasi - Menangis	Akurasi - Tidur	Akurasi - Senyum
1	55%	88%	48%
2	62%	94%	55%
3	58%	90%	50%
4	65%	95%	58%
5	54%	87%	47%
6	64%	93%	56%
7	59%	92%	52%
8	53%	89%	46%
9	63%	94%	57%
10	57%	90%	51%
Rata-rata	59.0%	91.2%	52.0%

Dari hasil pengujian pada Tabel 2, untuk ekspresi Tidur menunjukkan akurasi tertinggi sebesar 91,2%, diikuti oleh menangis (59,0%), sedangkan ekspresi Senyum memiliki akurasi terendah, yaitu 52,0%. Tingginya akurasi pada ekspresi Tidur disebabkan oleh kemiripan pola visual antar data, sehingga model lebih mudah melakukan generalisasi. Sebaliknya,

ekspresi Senyum dan Menangis cenderung memiliki variasi pose, pencahayaan, serta perbedaan bentuk wajah antar individu, yang menyebabkan penurunan performa deteksi.

4.4. Evaluasi Pelatihan Model YOLOv11

Proses pelatihan model YOLOv11 dilakukan menggunakan dataset ekspresi wajah bayi yang terdiri dari tiga kelas utama, yaitu Tidur, Senyum, dan Menangis. Dataset dibagi menjadi data latih dan data validasi dengan perbandingan 80:20. Tujuan pelatihan ini adalah untuk mendapatkan model yang memiliki akurasi tinggi dan kemampuan generalisasi yang baik terhadap variasi ekspresi bayi dalam berbagai kondisi pencahayaan dan pose wajah. Selama proses pelatihan, metrik utama yang diamati meliputi box loss, classification loss, distribution focal loss (DFL), serta metrik evaluasi seperti precision, recall, dan mean Average Precision (mAP). Visualisasi hasil pelatihan ditunjukkan pada Gambar 6.



Gambar 6. Hasil visualisasi pelatihan YOLOv11

Berdasarkan Gambar 6, memperlihatkan tren penurunan nilai *loss* dan peningkatan metrik evaluasi sepanjang proses pelatihan selama 50 epoch.

4.5. Analisis Loss Function

Tiga jenis *loss* utama yang diamati pada proses training adalah:

- Box Loss: mengukur selisih antara posisi dan ukuran *bounding box* hasil prediksi dan *ground truth*. Nilai *box loss* menurun secara konsisten dari sekitar 1.2 ke 0.7, menunjukkan peningkatan ketepatan model dalam menentukan posisi wajah bayi.
- Classification Loss: menunjukkan tingkat kesalahan dalam klasifikasi ekspresi. Penurunan dari nilai awal sekitar 3.0 menjadi di bawah 1.0 pada akhir pelatihan menunjukkan bahwa model semakin akurat dalam membedakan ketiga kelas ekspresi.
- DFL (Distribution Focal Loss): berfungsi mengoptimalkan prediksi distribusi jarak fitur objek. Nilai DFL loss juga menurun secara bertahap, mengindikasikan peningkatan stabilitas model terhadap variasi bentuk wajah bayi.

Tren penurunan *train loss* dan *validation loss* yang searah menandakan bahwa model tidak

mengalami *overfitting* yang signifikan, sehingga hasil pelatihan dapat digeneralisasikan dengan baik.

4.6. Analisis Metrik Evaluasi

Selain *loss*, performa model dievaluasi menggunakan metrik *precision*, *recall*, dan *mean Average Precision (mAP)*.

- Precision (B): meningkat secara bertahap hingga mencapai nilai sekitar 0.85 pada akhir pelatihan, menunjukkan bahwa sebagian besar prediksi model benar (minim *false positive*).
- Recall (B): mencapai nilai sekitar 0.83, yang berarti model mampu mendeteksi sebagian besar wajah bayi dengan ekspresi yang benar (minim *false negative*).
- mAP@50 (B): menunjukkan performa deteksi dengan ambang IoU 0.5 dan mencapai nilai tertinggi sekitar 0.87, menandakan tingkat akurasi deteksi yang baik.
- mAP@50–95 (B): menggambarkan performa model di berbagai ambang IoU (0.5 hingga 0.95), dengan nilai akhir sekitar 0.74, yang menandakan kemampuan generalisasi model cukup stabil untuk variasi ukuran wajah bayi.

4.7. Interpretasi Hasil Pelatihan

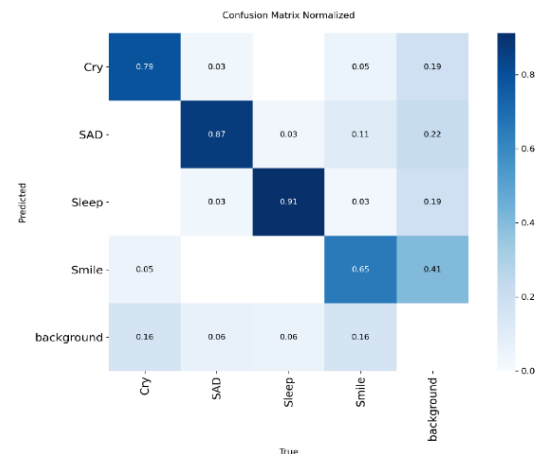
Secara keseluruhan, hasil pelatihan menunjukkan bahwa model YOLOv11 telah berhasil converge dengan tren *loss* yang menurun konsisten dan peningkatan metrik evaluasi yang stabil. Hal ini menandakan bahwa parameter model sudah optimal untuk digunakan pada tahap implementasi di Raspberry Pi 5. Dengan nilai mAP@50 sebesar 0.87 dan recall 0.83, model ini layak diimplementasikan untuk mendeteksi ekspresi bayi secara real-time, terutama karena YOLOv11 memiliki efisiensi komputasi tinggi serta kemampuan mendeteksi objek kecil seperti wajah bayi pada jarak menengah.

4.8. Analisis Confusion Matrix Model YOLOv11

Untuk mengevaluasi kemampuan klasifikasi model terhadap masing-masing kelas ekspresi bayi, dilakukan analisis menggunakan Confusion Matrix seperti ditunjukkan pada Gambar 7.

Confusion matrix menampilkan distribusi prediksi model terhadap label sebenarnya (ground truth) yang telah dinormalisasi ke dalam skala 0–1. Setiap nilai pada diagonal utama menunjukkan proporsi prediksi yang benar untuk setiap kelas, sedangkan nilai di luar diagonal menunjukkan

kesalahan klasifikasi antar kelas. Berdasarkan Gambar 7, dapat diamati bahwa:



Gambar 7. Confusion Matrix Normalized Model YOLOv11

- Kelas Sleep memiliki nilai tertinggi sebesar 0.91, menandakan tingkat deteksi yang sangat baik dan konsisten.
- Kelas SAD dan Cry juga menunjukkan performa tinggi dengan nilai 0.87 dan 0.79, menandakan model mampu mengenali ekspresi bayi dengan baik pada kondisi tersebut.
- Kelas Smile memiliki nilai 0.65, menunjukkan bahwa model masih mengalami kesulitan membedakan ekspresi senyum dari latar belakang atau kelas lain, terutama karena kemiripan pola wajah pada kondisi pencahayaan tertentu.
- Adanya nilai background pada beberapa kelas menunjukkan bahwa sebagian kecil area citra tanpa wajah bayi masih terdeteksi sebagai ekspresi tertentu.

4.9. Evaluasi Hiperparameter YOLOv11

Untuk memperoleh konfigurasi pelatihan yang optimal, dilakukan pengujian terhadap beberapa kombinasi hiperparameter utama pada model YOLOv11, yaitu jumlah epoch, batch size, dan learning rate. Tujuan dari pengujian ini adalah untuk menentukan parameter yang menghasilkan keseimbangan terbaik antara konvergensi *loss* dan peningkatan performa deteksi ekspresi bayi.

Tabel 3. Hasil Evaluasi Hiperparameter YOLOv11 untuk Pengenalan Ekspresi Bayi

Epoch	Batch Size	Learning Rate	Box Loss	mAP @0.5	Precision	Recall	F1-Score
5	16	0.001	2.060	0.013	0.355	0.030	0.02
10	16	0.001	1.922	0.086	0.611	0.074	0.11
15	16	0.001	1.825	0.152	0.577	0.121	0.20
20	16	0.001	1.756	0.229	0.602	0.176	0.32
5	16	0.01	3.126	0.000	0.080	0.001	0.00
10	16	0.01	2.809	0.000	0.140	0.005	0.00
15	16	0.01	2.831	0.008	0.053	0.043	0.04
20	16	0.01	2.793	0.012	0.081	0.061	0.05

Epoch	Batch Size	Learning Rate	Box Loss	mAP @0.5	Precision	Recall	F1-Score
5	32	0.001	3.126	0.000	0.080	0.006	0.00
10	32	0.001	2.811	0.001	0.080	0.005	0.00
15	32	0.001	2.835	0.005	0.072	0.033	0.03
20	32	0.001	2.798	0.013	0.061	0.071	0.06
5	32	0.01	3.126	0.000	0.080	0.006	0.00
10	32	0.01	2.809	0.004	0.147	0.010	0.02
15	32	0.01	2.834	0.004	0.091	0.031	0.03
20	32	0.01	2.796	0.013	0.071	0.062	0.0

Berdasarkan hasil pada Tabel 3, kombinasi batch size 16 dengan learning rate 0.001 memberikan performa paling stabil dan konvergen terhadap penurunan loss. Nilai Box Loss menunjukkan tren penurunan yang konsisten dari 2.06 menjadi 1.75 seiring peningkatan epoch, yang menandakan model semakin akurat dalam menentukan posisi bounding box. Selain itu, nilai mAP@0.5 meningkat dari 0.013 menjadi 0.229 pada konfigurasi yang sama, menunjukkan peningkatan kemampuan model dalam mengenali ekspresi wajah bayi. Nilai Precision dan Recall juga meningkat seiring bertambahnya epoch, dengan nilai F1-Score terbaik sebesar 0.32 pada epoch ke-20.

Sebaliknya, penggunaan learning rate yang lebih tinggi (0.01) menghasilkan loss yang tidak stabil dan nilai mAP mendekati nol, menandakan bahwa model gagal berkonvergensi dengan baik. Hal ini mengindikasikan bahwa learning rate terlalu besar menyebabkan overshooting pada proses pembaruan bobot. Dari hasil pengujian ini dapat disimpulkan bahwa konfigurasi terbaik diperoleh pada batch size 16, learning rate 0.001, dan epoch sebanyak 20. Parameter ini memberikan keseimbangan antara akurasi deteksi dan stabilitas proses pelatihan. Nilai ini kemudian digunakan sebagai konfigurasi akhir pada pelatihan model YOLOv11 yang diimplementasikan pada Raspberry Pi 5 untuk sistem deteksi ekspresi bayi secara real-time.

4.10. Pengujian notifikasi pada aplikasi android

Pengujian ini dilakukan untuk mengevaluasi kecepatan dan keandalan sistem dalam mengirimkan notifikasi hasil deteksi ekspresi bayi dari Raspberry Pi 5 ke aplikasi Android melalui layanan *Firestore Realtime Database*. Tujuan utama dari pengujian ini adalah untuk menilai dua parameter kinerja utama, yaitu waktu tunda (latency) pengiriman notifikasi dari saat ekspresi bayi terdeteksi hingga notifikasi diterima pada aplikasi Android, serta tingkat keberhasilan pengiriman data atau *Packet Delivery Rate* (PDR) pada berbagai kondisi jaringan internet yakni : Wi-Fi, dan jaringan seluler (4G), dengan masing-masing skenario dilakukan sebanyak sepuluh kali pengiriman data untuk menghitung rata-rata waktu tunda dan persentase keberhasilan pengiriman data.

Tabel 4. Hasil pengujian notifikasi pada aplikasi

No	Kondisi Jaringan	Rata-rata Latency (detik)	Packet Delivery Rate (%)
1	Wi-Fi	1.24	100%
2	Jaringan Seluler (4G)	3.45	95%

Berdasarkan hasil pengujian pada Tabel 4, dapat dilihat bahwa sistem pengiriman notifikasi melalui *Firestore Realtime Database* menunjukkan performa yang baik pada kedua kondisi jaringan. Pada koneksi Wi-Fi, rata-rata waktu tunda hanya sebesar 1,24 detik dengan tingkat keberhasilan pengiriman data mencapai 100%, yang menandakan bahwa setiap hasil deteksi dari Raspberry Pi 5 berhasil dikirim dan diterima oleh aplikasi Android tanpa gangguan. Hal ini menunjukkan bahwa sistem mampu memberikan notifikasi secara real-time dengan respons yang cepat ketika menggunakan jaringan dengan bandwidth tinggi dan koneksi stabil. Sementara itu, pada pengujian menggunakan jaringan seluler (4G), rata-rata waktu tunda meningkat menjadi 3,45 detik dengan tingkat keberhasilan pengiriman sebesar 95%. Walaupun terdapat sedikit peningkatan latency, notifikasi masih dapat diterima secara konsisten tanpa kehilangan data yang signifikan.

5. KESIMPULAN

Berdasarkan hasil implementasi dan pengujian sistem pengenalan ekspresi bayi berbasis YOLOv11, dapat disimpulkan bahwa model mampu mendeteksi ekspresi wajah bayi secara real-time dengan akurasi yang cukup baik, khususnya pada kelas Tidur yang mencapai 91,2% dengan nilai confidence tertinggi sebesar 0,94, sementara kelas Menangis dan Senyum memperoleh akurasi masing-masing 59,0% dan 52,0%. Selama proses pelatihan, model menunjukkan konvergensi stabil dengan penurunan loss yang konsisten serta peningkatan metrik evaluasi seperti precision (0.85), recall (0.83), dan mAP@50 (0.87), menandakan parameter yang optimal tanpa overfitting. Analisis Confusion Matrix mengonfirmasi bahwa performa terbaik terdapat pada kelas Tidur dengan nilai normalisasi 0.91, sedangkan kelas Senyum menjadi tantangan terbesar dengan nilai 0.65 karena kemiripan pola wajah dan pengaruh pencahayaan. Evaluasi hyperparameter menunjukkan konfigurasi terbaik pada batch size 16, learning rate 0.001, dan 20 epoch

yang memberikan keseimbangan antara akurasi dan stabilitas pelatihan. Selain itu, pengujian notifikasi melalui Firebase Realtime Database membuktikan bahwa sistem dapat mengirimkan hasil deteksi ke aplikasi Android dengan rata-rata latency 1,24 detik pada jaringan Wi-Fi dan 3,45 detik pada jaringan seluler (4G), serta tingkat keberhasilan pengiriman lebih dari 95%, sehingga sistem mampu memberikan notifikasi ekspresi bayi secara real-time dan andal pada berbagai kondisi jaringan.

DAFTAR PUSTAKA

- [1] S. Hodge and A. Lundeen, "Amerika Telah Menjadi Negara Pasangan Pekerja dengan Dua Penghasilan," *TAXFOUNDATION*, 2020. [Online]. Available: <https://taxfoundation.org/data/all/federal/america-has-become-nation-dual-income-working-couples/>. [Accessed: 17-Oct-2025].
- [2] L. Carnevali, A. Gui, E. J. H. Jones, and T. Farroni, "Face Processing in Early Development: A Systematic Review of Behavioral Studies and Considerations in Times of COVID-19 Pandemic," *Front. Psychol.*, vol. 13, no. February, 2022.
- [3] F. Poncet *et al.*, "A neural marker of rapid discrimination of facial expression in 3.5- and 7-month-old infants," *Front. Neurosci.*, vol. 16, no. August, pp. 1–15, 2022.
- [4] S. Nakić Radoš, "Parental Sensitivity and Responsiveness as Mediators Between Postpartum Mental Health and Bonding in Mothers and Fathers," *Front. Psychiatry*, vol. 12, no. September, pp. 1–12, 2021.
- [5] T. Abuzairi, N. Widanti, A. Kusumaningrum, and Y. Rustina, "Implementasi Convolutional Neural Network Untuk Deteksi Nyeri Bayi Melalui Citra Wajah Dengan YOLO," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 624–630, 2021.
- [6] G. J. P. Hakim *et al.*, "Real-Time Facial Emotion Detection Application with Image Processing Based on Convolutional Neural Network (CNN)," *Int. J. Electr. Eng. Math. Comput. Sci.*, vol. 1, no. 4, pp. 27–36, 2024.
- [7] A. Alshammari and M. E. Alshammari, "Emotional Facial Expression Detection using YOLOv8," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 5, pp. 16619–16623, 2024.
- [8] Y. Ihza and D. Lelono, "Face Expression Classification in Children Using CNN," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 16, no. 2, p. 159, 2022.
- [9] S. Zhang, H. Tuo, J. Hu, and Z. Jing, "Domain Adaptive YOLO for One-Stage Cross-Domain Detection," *Proc. Mach. Learn. Res.*, vol. 157, pp. 785–797, 2021.
- [10] L. He, Y. Zhou, L. Liu, and J. Ma, "Research and Application of YOLOv11-Based Object Segmentation in Intelligent Recognition at Construction Sites," *Buildings*, vol. 14, no. 12, 2024.
- [11] S. Kim, J. Park, Y. Jeong, and S. E. Lee, "Intelligent Monitoring System with Privacy Preservation Based on Edge AI," *Micromachines*, vol. 14, no. 9, 2023.
- [12] Q. Lin, R. He, and P. Jiang, "Feature guided CNN for baby's facial expression recognition," *Hindawi Complex.*, vol. 2020, pp. 0–2, 2020.
- [13] Z. Hammal, W. S. Chu, J. F. Cohn, C. Heike, and M. L. Speltz, "Automatic Action Unit Detection in Infants Using Convolutional Neural Network," *Int. Conf. Affect Comput. Intell. Interact. Work.*, vol. 176, no. 3, pp. 139–148, 2020.
- [14] U. Laraib, A. Shaukat, R. A. Khan, Z. Mustansar, M. U. Akram, and U. Asgher, "Recognition of Children's Facial Expressions Using Deep Learned Features," *Electron.*, vol. 12, no. 11, pp. 1–16, 2023.
- [15] M. Mekhfioui *et al.*, "Development of a Baby Cry Identification System Using a Raspberry Pi-Based Embedded System and Machine Learning," *Technologies*, vol. 13, no. 4, pp. 1–13, 2025.
- [16] A. H. N. Hidayah, A. R. Syafeeza, N. A. Razak, W. H. M. Saad, Y. C. Wong, and A. A. Naja, "Disease Detection of Solanaceous Crops Using Deep Learning for Robot Vision," *J. Robot. Control*, vol. 3, no. 6, pp. 790–799, 2022.