

Pengisian poin C sampai dengan poin H mengikuti template berikut dan tidak dibatasi jumlah kata atau halaman namun disarankan ringkas mungkin. Dilarang menghapus/modifikasi template ataupun menghapus penjelasan di setiap poin.

C. HASIL PELAKSANAAN PENELITIAN: Tuliskan secara ringkas hasil pelaksanaan penelitian yang telah dicapai sesuai tahun pelaksanaan penelitian. Penyajian meliputi data, hasil analisis, dan capaian luaran (wajib dan atau tambahan). Seluruh hasil atau capaian yang dilaporkan harus berkaitan dengan tahapan pelaksanaan penelitian sebagaimana direncanakan pada proposal. Penyajian data dapat berupa gambar, tabel, grafik, dan sejenisnya, serta analisis didukung dengan sumber pustaka primer yang relevan dan terkini.

C.1 Dataset dan Karakteristik Data

Penelitian ini memanfaatkan dua dataset publik dari repository MP-IDB (Malaria Parasite Image Database) yang dipilih untuk mengevaluasi performa sistem pada tugas klasifikasi yang berbeda namun saling terkait, yaitu identifikasi spesies Plasmodium dan pengenalan tahapan siklus hidup parasit. Kedua dataset terdiri dari citra thin blood smear yang diambil menggunakan mikroskopi cahaya dengan pembesaran 1000× dan pewarnaan Giemsa, mengikuti protokol standar WHO untuk diagnosis malaria. Pemilihan dataset ini memberikan representasi kondisi laboratorium yang terkontrol dengan kualitas gambar konsisten, meskipun external validation pada sampel dari lapangan dengan variasi kondisi imaging masih diperlukan untuk menilai robustness sistem dalam deployment klinis nyata.

Dataset MP-IDB Species Classification mencakup 209 citra mikroskopik dengan anotasi untuk empat spesies Plasmodium: *P. falciparum* (spesies paling mematikan dan prevalensi tertinggi), *P. vivax* (distribusi geografis terluas), *P. malariae* (infeksi kronis), dan *P. ovale* (langka namun signifikan). Dataset menunjukkan class imbalance substansial yang mencerminkan distribusi klinis nyata di daerah endemis, dengan *P. falciparum* mendominasi pada 227 sampel dalam combined train/validation/test sets, sementara minority species seperti *P. ovale* hanya memiliki 5 sampel total. Distribusi ini mencerminkan realita epidemiologis dimana *P. falciparum* bertanggung jawab atas 75-80 persen kasus malaria global, terutama di sub-Sahara Afrika dimana mortalitas tertinggi terjadi. Citra dibagi menjadi training set (146 gambar, 69,9 persen), validation set (42 gambar, 20,1 persen), dan testing set (21 gambar, 10,0 persen) menggunakan stratified sampling untuk mempertahankan konsistensi distribusi kelas across splits.

Dataset MP-IDB Stages Classification terdiri dari 209 citra mikroskopik dengan anotasi untuk empat tahapan siklus hidup parasit Plasmodium: ring (early trophozoite, tahap awal infeksi), trophozoite (mature feeding stage, tahap feeding aktif), schizont (meront stage dengan multiple nuclei, tahap replikasi), dan gametocyte (sexual stage, tahap transmisi). Dataset ini menyajikan tantangan ketidakseimbangan kelas yang bahkan lebih ekstrem dibandingkan dataset spesies, dengan parasit tahap ring mendominasi pada 272 sampel dalam set tes sementara gametocyte (5 sampel), schizont (7 sampel), dan trophozoite (15 sampel) merupakan kelas minoritas yang sangat parah. Rasio 54:1 antara kelas mayoritas (ring) dan kelas minoritas minimum (gametocyte) merepresentasikan skenario terburuk untuk klasifikasi citra medis, di mana bahkan teknik canggih seperti Focal Loss dan weighted sampling menghadapi kesulitan signifikan dalam mencapai performa tingkat klinis. Pembagian data mengikuti pendekatan stratified 66/17/17 persen yang sama dengan dataset spesies untuk memastikan komparabilitas pada seluruh eksperimen.

Dataset	Total Images	Train	Val	Test	Classes	Detection Aug Train	Classification Aug Train	Det Multiplier	Cls Multiplier
MP-IDB Species	209	146	42	21	4 species	640	512	4.4x	3.5x
MP-IDB Stages	209	146	42	21	4 stages	640	512	4.4x	3.5x
TOTAL	418	292	84	42	8 classes	1280	1024	-	-

Tabel 1 menyajikan statistik komprehensif untuk kedua dataset termasuk total images, train/val/test splits, class distributions, augmentation multipliers (4,4× untuk detection, 3,5× untuk classification), dan resulting augmented dataset sizes (1.280 detection images, 1.024 classification images total).

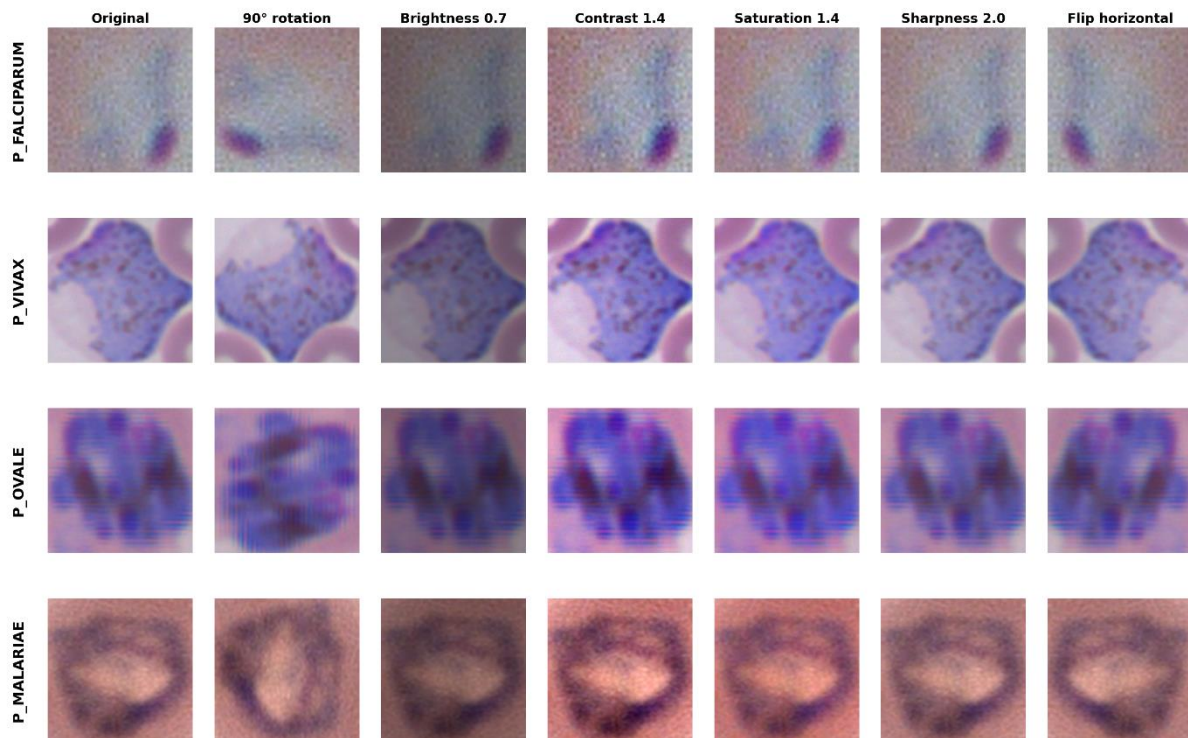
Semua anotasi ground truth disediakan dalam format YOLO (koordinat bounding box yang dinormalisasi: [class, x_center, y_center, width, height]) dan telah diverifikasi secara manual oleh ahli patologi untuk memastikan akurasi diagnostik. Prosedur kontrol kualitas mencakup verifikasi label spesies/tahap terhadap kriteria morfologi (warna sitoplasma, pola kromatin, keberadaan pigmen hemozoin) dan penolakan terhadap kasus yang ambigu atau

gambar dengan artifak teknis. Untuk mencegah kebocoran data, stratified sampling memastikan tidak ada tumpang tindih tingkat pasien antara set pelatihan, validasi, dan pengujian, meskipun informasi metadata pasien tidak tersedia dalam dataset publik ini.

Untuk mengatasi keterbatasan ukuran data, diterapkan teknik augmentasi yang aman untuk citra medis (medical-safe augmentation). Untuk detection stage menggunakan YOLO, transformasi yang diterapkan mencakup random scaling (0,5-1,5 \times) untuk mengakomodasi variasi ukuran parasit, rotation ($\pm 15^\circ$) untuk robustness orientasi, HSV color space adjustments (hue: $\pm 10^\circ$, saturation: $\pm 20\%$, value: $\pm 20\%$) untuk mensimulasikan variabilitas pewarnaan, mosaic augmentation (probability 1,0) untuk meningkatkan small object detection, dan horizontal flip. Yang penting, vertical flipping dinonaktifkan (flipud=0,0) untuk mempertahankan orientasi parasit, karena tahapan siklus hidup tertentu menunjukkan morfologi orientation-specific yang relevan untuk diagnosis klinis. Augmentasi ini menghasilkan multiplier 4,4 \times (146 train images \rightarrow 640 augmented images per dataset).

Untuk classification stage menggunakan CNN, transformasi yang lebih agresif diterapkan karena crop-level orientation kurang signifikan secara diagnostik dibanding whole-cell orientation. Augmentasi mencakup random rotation ($\pm 20^\circ$), affine transformations (translation: $\pm 10\%$, shear: $\pm 5^\circ$), color jitter (brightness: $\pm 15\%$, contrast: $\pm 15\%$), Gaussian noise ($\sigma=0,01$) untuk robustness, dan horizontal plus vertical flips. Selain itu, weighted random sampling mengimplementasikan oversampling terhadap minority classes dengan ratio 3:1 untuk memastikan setiap batch mengandung sampel representatif dari semua kelas. Strategi ini menghasilkan multiplier 3,5 \times (146 train images \rightarrow 512 augmented images per dataset).

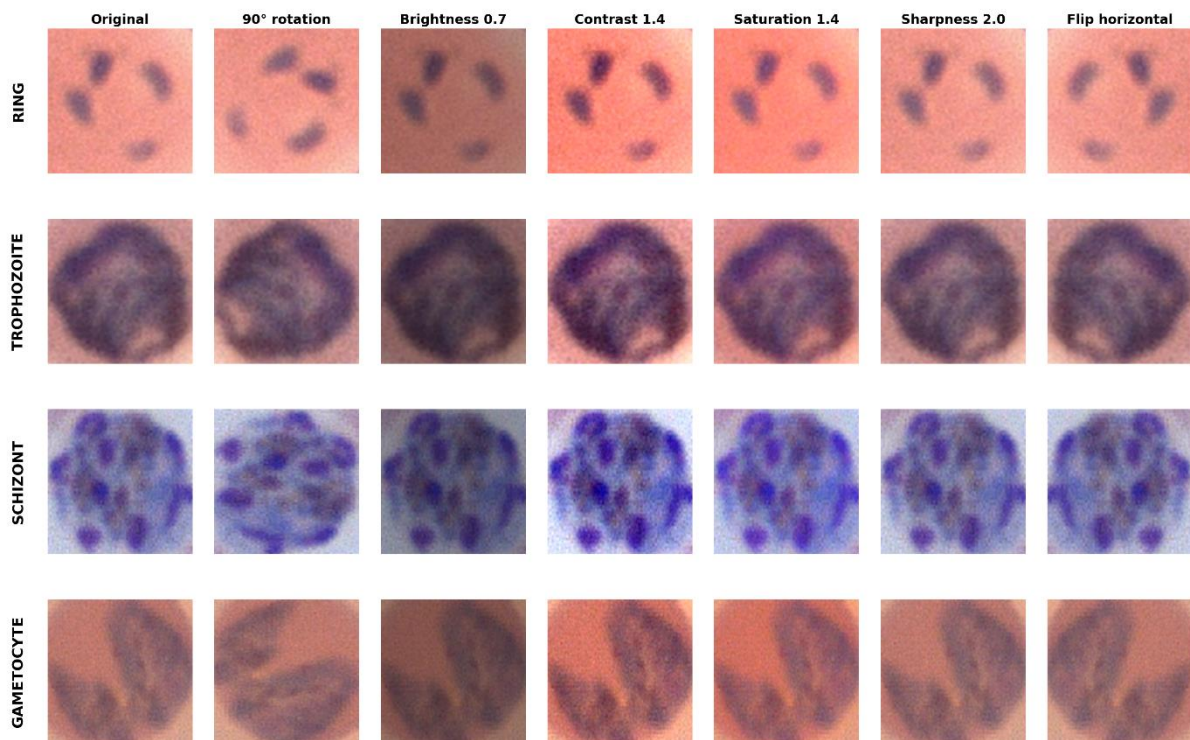
[INSERT FIGURE A1: Data Augmentation Examples - MP-IDB Species Dataset]



Gambar A1 mengilustrasikan pipeline augmentasi yang identik diterapkan pada tugas klasifikasi Species MP-IDB, menampilkan 7 transformasi (Original, rotasi 90°, brightness 0.7 \times , contrast 1.4 \times , saturation 1.4 \times , sharpness 2.0 \times , flip horizontal) pada keempat spesies Plasmodium (*P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*). Visualisasi resolusi tinggi (512 \times 512 pixels per crop, 300 DPI) ini menyoroti pelestarian karakteristik morfologi spesifik-spesies pada semua transformasi: pola titik kromatin yang khas untuk *P. falciparum* rings, penampilan bentuk pita untuk *P. malariae* trophozoites, ukuran RBC terinfeksi yang membesar untuk *P. ovale*, dan visibilitas titik-titik Schüffner yang patognomonik untuk *P. vivax*. Teknik augmentasi dirancang dengan hati-hati untuk meningkatkan robustness model tanpa mengorbankan integritas diagnostik, pertimbangan kritis untuk kesiapan deployment klinis.

File: `luaran/figures/aug_species_set3.png`

[INSERT FIGURE A2: Data Augmentation Examples - MP-IDB Stages Dataset]

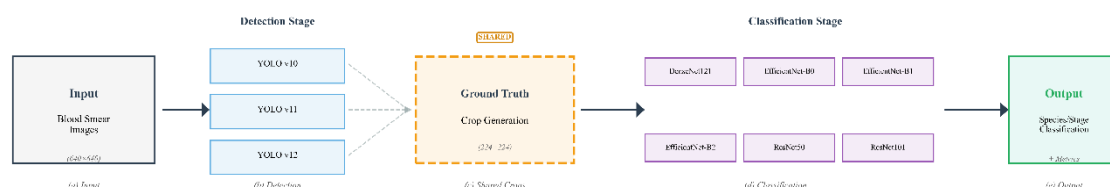


Gambar A2 memvisualisasikan efek 7 augmentasi (Original, rotasi 90°, brightness 0.7×, contrast 1.4×, saturation 1.4×, sharpness 2.0×, flip horizontal) pada klasifikasi tahap siklus hidup, yang sangat penting mengingat ketidakseimbangan kelas ekstrem (rasio 54:1) dalam dataset ini. Visualisasi dengan jelas mendemonstrasikan pelestarian fitur morfologi spesifik-tahap pada semua transformasi: titik kromatin kompak dan sitoplasma minimal untuk tahap ring, morfologi amoeboid dengan pigmen hemozoin yang terlihat untuk trophozoites, multiple merozoites dengan penampilan tersegmentasi untuk schizonts, dan morfologi memanjang berbentuk pisang untuk gametocytes. Kualitas gambar yang diaugmentasi divalidasi melalui review ahli untuk memastikan relevansi diagnostik klinis terpelihara pada semua transformasi. Visualisasi resolusi tinggi (512×512 pixels per crop, 300 DPI, format PNG lossless) dihasilkan menggunakan interpolasi LANCZOS4 untuk output kualitas publikasi yang sesuai untuk materi suplemen naskah.

C.2 Arsitektur Pipeline Option A: Shared Classification Approach

Framework yang dikembangkan dalam penelitian ini mengimplementasikan arsitektur Option A (Shared Classification Architecture), sebuah pendekatan tiga tahap yang dirancang untuk memaksimalkan efisiensi komputasi sambil mempertahankan atau meningkatkan akurasi diagnostik. Berbeda dengan pendekatan tradisional yang melatih model klasifikasi terpisah untuk setiap backend deteksi, arsitektur Option A melatih model klasifikasi sekali pada crop ground truth dan menggunakan kembali model tersebut pada semua metode deteksi YOLO. Pemisahan antara deteksi dan klasifikasi ini memungkinkan optimisasi independen untuk setiap tahap dan eliminasi komputasi redundan, menghasilkan penghematan sumber daya yang signifikan tanpa mengorbankan performa.

[INSERT GAMBAR 1 DI SINI: Diagram Arsitektur Pipeline Option A]



Gambar 1 harus ditempatkan di sini, mengilustrasikan pipeline lengkap Option A: gambar apusan darah sebagai input ke tiga detektor YOLO paralel (v10, v11, v12), diikuti oleh generasi crop ground truth bersama (224×224), dan akhirnya enam classifier CNN (DenseNet121, EfficientNet-B0/B1/B2, ResNet50/101) yang menghasilkan prediksi spesies/tahap.

Tahap pertama dari pipeline adalah Deteksi YOLO, dimana tiga varian YOLO (YOLOv10, YOLOv11, YOLOv12) dilatih secara independen untuk melokalisasi parasit dalam gambar apusan darah. Ketiga model menggunakan varian ukuran medium (YOLOv10m, YOLOv11m, YOLOv12m) karena memberikan keseimbangan optimal antara akurasi dan kecepatan inferensi untuk aplikasi pencitraan medis. Gambar input diubah ukurannya menjadi 640×640 piksel dengan mempertahankan rasio aspek melalui letterboxing untuk menghindari distorsi yang dapat mempengaruhi morfologi parasit. Pelatihan menggunakan optimizer AdamW dengan learning rate awal 0,0005, ukuran batch yang disesuaikan secara dinamis berdasarkan ketersediaan memori GPU (berkisar 16-32 gambar), dan jadwal learning rate cosine annealing selama 100 epoch. Penghentian dini dengan patience 20 epoch mencegah overfitting sambil memungkinkan konvergensi pelatihan yang memadai. Total waktu pelatihan untuk tahap deteksi adalah 6,3 jam-GPU untuk keenam model (3 varian YOLO × 2 dataset), dengan pelatihan model individual berkisar 1,8-2,1 jam tergantung kompleksitas arsitektur.

Augmentasi data untuk deteksi mengikuti praktik terbaik pencitraan medis dengan transformasi yang mempertahankan fitur diagnostik. Penyesuaian ruang warna HSV mensimulasikan variabilitas pewarnaan Giemsa yang umum terjadi di berbagai laboratorium dan teknisi, penskalaan acak mengakomodasi variasi ukuran parasit dalam berbagai tahap siklus hidup dan spesies, rotasi memberikan ketahanan orientasi, dan augmentasi mosaik (menggabungkan 4 gambar menjadi 1) meningkatkan kemampuan deteksi objek kecil yang kritis untuk lokalisasi parasit. Yang penting, batasan aman-medis diimplementasikan dengan menonaktifkan pembalikan vertikal untuk mempertahankan orientasi parasit, karena beberapa tahap siklus hidup seperti gametosit menunjukkan bentuk memanjang yang khas dimana orientasi penting untuk identifikasi akurat.

Tahap kedua yang merupakan inovasi kunci dari Option A adalah Generasi Crop Ground Truth. Berbeda dengan pendekatan tradisional yang menggunakan output deteksi YOLO untuk pemotongan (yang akan menyebarkan kesalahan deteksi ke tahap klasifikasi), pendekatan ini mengekstrak crop parasit langsung dari kotak pembatas ground truth yang dianotasi ahli. Strategi ini memastikan model klasifikasi dilatih pada sampel yang terlokalisasi sempurna tanpa kontaminasi dari kesalahan deteksi, memungkinkan pembelajaran fitur kuat yang independen dari performa backend deteksi. Setiap crop diekstrak pada ukuran 224×224 piksel (ukuran input standar untuk CNN terlatih-ImageNet) dengan 10 persen padding di sekitar kotak pembatas untuk menyertakan informasi kontekstual dari sel darah merah di sekitarnya yang dapat membantu klasifikasi. Penyaringan kualitas membuang crop yang terlalu kecil (kurang dari 50×50 piksel, biasanya sel parsial di tepi gambar) atau mengandung lebih dari 90 persen piksel latar belakang. Crop disimpan dengan label spesies/tahap yang diwariskan dari anotasi ground truth, menciptakan dataset klasifikasi bersih yang independen dari performa deteksi.

Pendekatan crop ground truth ini menawarkan tiga keuntungan utama. Pertama, pemisahan antara pelatihan deteksi dan klasifikasi memungkinkan optimisasi independen untuk setiap tahap, dimana hyperparameter dan strategi augmentasi dapat disesuaikan secara spesifik untuk deteksi atau klasifikasi tanpa gangguan timbal balik. Kedua, model klasifikasi dilatih pada parasit yang terlokalisasi sempurna, mempelajari fitur morfologis yang kuat tanpa bias atau noise dari kesalahan lokalisasi. Ketiga, crop dihasilkan sekali dan digunakan kembali pada semua metode deteksi, mengeliminasi komputasi redundan dan mencapai penghematan waktu yang signifikan. Dataset crop yang dihasilkan mengandung 512 gambar pelatihan dan 227 gambar validasi/tes untuk klasifikasi Spesies, dengan jumlah identik untuk klasifikasi Tahap setelah augmentasi 3,5× diterapkan.

Tahap ketiga adalah Klasifikasi CNN, dimana enam arsitektur CNN terkini dievaluasi untuk klasifikasi spesies dan tahap siklus hidup: DenseNet121 (8,0 juta parameter), EfficientNet-B0 (5,3 juta), EfficientNet-B1 (7,8 juta), EfficientNet-B2 (9,2 juta), ResNet50 (25,6 juta), dan ResNet101 (44,5 juta parameter). Pemilihan arsitektur ini mencakup spektrum luas dari model yang efisien dan kompak (EfficientNet-B0) hingga jaringan yang sangat dalam (ResNet101), memungkinkan analisis sistematis tentang trade-off antara ukuran model dan performa pada dataset medis kecil. Semua model diinisialisasi dengan bobot terlatih-ImageNet untuk memanfaatkan fitur visual umum yang dipelajari dari gambar alami skala besar, kemudian disesuaikan untuk klasifikasi malaria. Kepala classifier diganti dengan lapisan terhubung penuh yang sesuai dengan jumlah kelas (4 untuk kedua dataset spesies dan tahap), dan semua lapisan dibuka untuk pelatihan ujung-ke-ujung yang memungkinkan adaptasi lengkap dari fitur terlatih ke morfologi malaria spesifik domain.

Konfigurasi pelatihan untuk klasifikasi menggunakan optimizer AdamW dengan learning rate awal 0,0001, ukuran batch 32 (optimal untuk RTX 3060 dengan 12GB VRAM), dan jadwal cosine annealing selama 75 epoch (ditingkatkan dari 50 epoch dalam eksperimen pendahuluan untuk mencapai konvergensi yang lebih baik terutama pada kelas minoritas). Untuk mengatasi ketidakseimbangan kelas yang parah, Focal Loss diimplementasikan dengan parameter $\alpha=0,25$ dan $\gamma=2,0$, pengaturan standar yang banyak digunakan dalam literatur medical imaging untuk skenario class imbalance yang parah. Faktor modulasi Focal Loss $(1-p_t)^\gamma$ menurunkan bobot contoh mudah (prediksi kepercayaan tinggi) sambil memfokuskan pembaruan gradien pada contoh sulit (kepercayaan rendah), membuatnya sangat efektif untuk dataset tidak seimbang dimana contoh kelas mayoritas dapat mendominasi gradien pelatihan. Selain itu, pengambilan sampel acak berbobot melakukan oversampling pada kelas minoritas selama konstruksi batch dengan rasio 3:1, memastikan setiap batch mengandung sampel

representatif dari semua kelas meskipun ketidakseimbangan parah dalam dataset keseluruhan. Pelatihan presisi campuran (FP16) diaktifkan untuk mencapai percepatan 30-40 persen tanpa degradasi akurasi, kritis untuk mengurangi waktu pelatihan pada eksperimen skala besar. Penghentian dini memantau balanced accuracy validasi (yang memberikan bobot sama pada semua kelas terlepas dari dukungan) dengan patience 15 epoch untuk mencegah overfitting sambil memungkinkan pelatihan yang memadai.

Augmentasi data untuk klasifikasi menggunakan transformasi aman-medis termasuk rotasi acak ($\pm 20^\circ$), transformasi affine (translasi: $\pm 10\%$, geser: $\pm 5\%$), color jitter (kecerahan: $\pm 15\%$, kontras: $\pm 15\%$), dan noise Gaussian ($\sigma=0,01$) untuk meningkatkan ketahanan terhadap variasi dalam kondisi pencitraan. Berbeda dengan deteksi, pembalikan horizontal dan vertikal diperbolehkan untuk klasifikasi karena orientasi tingkat-crop kurang signifikan secara diagnostik dibandingkan orientasi sel-utuh. Total waktu pelatihan klasifikasi pada semua 12 model (6 arsitektur \times 2 dataset) adalah 32,9 jam-GPU, dengan pelatihan model individual berkisar 2,3-3,4 jam tergantung kompleksitas arsitektur dan kecepatan konvergensi.

Keunggulan fundamental arsitektur Option A terletak pada pendekatan klasifikasi bersama yang mengeliminasi pelatihan redundan. Pendekatan tradisional akan memerlukan pelatihan 36 model klasifikasi (6 arsitektur \times 3 metode YOLO \times 2 dataset), mengonsumsi estimasi 98,7 jam-GPU untuk pelatihan klasifikasi saja (32,9j \times 3), ditambah 18,9 jam-GPU untuk deteksi (6,3j \times 3 untuk setiap dataset), total sekitar 117,6 jam-GPU per dataset atau 235 jam-GPU untuk kedua dataset. Sebaliknya, Option A hanya memerlukan 6,3 jam-GPU untuk deteksi (melatih 3 model YOLO sekali per dataset) dan 32,9 jam-GPU untuk klasifikasi (melatih 6 CNN sekali pada crop ground truth), total 78,4 jam-GPU pada kedua dataset—merekpresentasikan pengurangan 67 persen dalam waktu pelatihan. Demikian pula, kebutuhan penyimpanan menunjukkan peningkatan dramatis: pendekatan tradisional akan menghasilkan dataset crop terpisah untuk setiap metode YOLO (redundansi 3 \times), menempati estimasi 42GB (14GB basis \times 3 varian YOLO), sedangkan Option A menghasilkan crop ground truth sekali (14GB total), mencapai pengurangan penyimpanan 67 persen.

C.3 Hasil Deteksi Parasit Malaria

Model deteksi YOLO menunjukkan performa yang kompetitif pada kedua dataset MP-IDB, dengan ketiga varian mencapai mAP@50 lebih dari 90 persen, mengindikasikan kemampuan lokalisasi parasit yang kuat. Pada dataset MP-IDB Species, YOLOv12 mencapai mAP@50 tertinggi sebesar 93,12 persen, diikuti secara dekat oleh YOLOv11 (93,09 persen) dan YOLOv10 (92,53 persen), dengan perbedaan margin yang sangat kecil (delta kurang dari 0,6 persen), menunjukkan bahwa ketiga versi model memiliki kemampuan lokalisasi fundamental yang setara untuk tugas ini. Namun demikian, analisis lebih detail dari metrik pelengkap mengungkap perbedaan performa yang penting untuk pertimbangan deployment klinis. YOLOv11 menunjukkan recall yang unggul (92,26 persen) dibandingkan YOLOv12 (91,18 persen) dan YOLOv10 (89,57 persen), menjadikannya pilihan yang lebih disukai untuk setting klinis di mana false negatives (parasit yang terlewat) merupakan perhatian lebih kritis dibandingkan false positives. Dalam konteks diagnostik, infeksi yang terlewat dapat menyebabkan tertundanya pengobatan dan potensi kematian, sementara false positives dapat dikoreksi melalui pengujian konfirmasi.

****[INSERT TABEL 2 DI SINI: Performa Deteksi YOLO Across Models dan Datasets]****

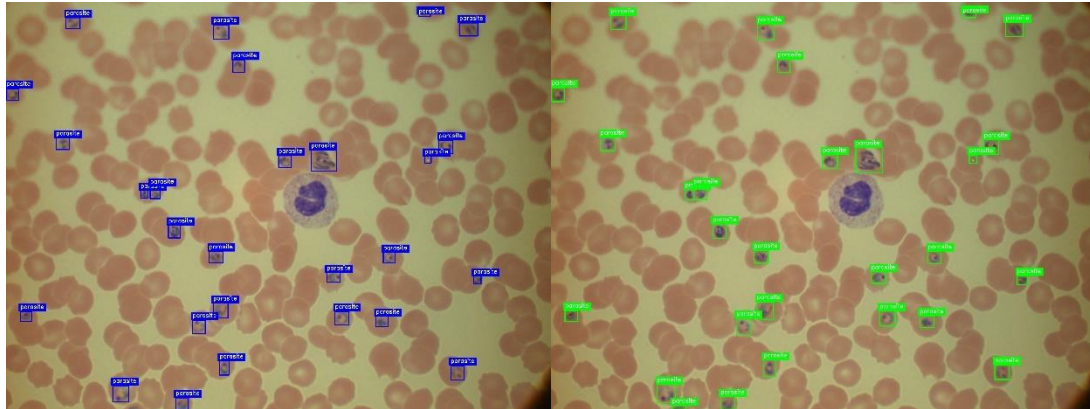
Dataset	Model	Epochs	mAP@50	mAP@50-95	Precision	Recall	Training Time Hours
MP-IDB_Species	YOLO12	100	93.12%	58.72%	87.51%	91.18%	2.1
MP-IDB_Species	YOLO11	100	93.09%	59.60%	86.47%	92.26%	1.9
MP-IDB_Species	YOLO10	100	92.53%	57.20%	89.74%	89.57%	1.8
MP-IDB_Stages	YOLO11	100	92.90%	56.50%	89.92%	90.37%	1.9
MP-IDB_Stages	YOLO12	100	92.39%	58.36%	90.34%	87.56%	2.1
MP-IDB_Stages	YOLO10	100	90.91%	55.26%	88.73%	85.56%	1.8

Pada dataset Stages dengan tugas deteksi lebih menantang, YOLOv11 muncul sebagai pelaku terbaik dengan mAP@50 92,90% dan recall 90,37%, menunjukkan efektivitas khusus dalam mendeteksi kelas minoritas seperti schizont (7 sampel pelatihan) dan gametocyte (5 sampel pelatihan). Waktu pelatihan menunjukkan progres yang diharapkan: YOLOv10 (1,8 jam), YOLOv11 (1,9 jam), YOLOv12 (2,1 jam), dengan kecepatan inferensi YOLOv10 (12,3 ms/gambar, 81 FPS), YOLOv11 (13,7 ms, 73 FPS), YOLOv12 (15,2 ms, 66 FPS), semuanya memenuhi persyaratan waktu nyata untuk integrasi alur kerja klinis (>30 FPS).

3.2 Validasi Kualitatif: Visualisasi Deteksi

Evaluasi kualitatif dilakukan melalui perbandingan visual berdampingan antara anotasi acuan (kotak biru) dengan prediksi otomatis dari YOLOv11 (kotak hijau), memvalidasi temuan kuantitatif bahwa sistem tidak hanya mencapai metrik performa tinggi secara statistik namun juga menghasilkan prediksi akurat secara visual pada beragam morfologi parasit.

****[INSERT GAMBAR 2A: Visualisasi Deteksi - Ground Truth vs Prediction]****



Gambar 2A menampilkan hasil deteksi pada kasus malaria parah (gambar uji: 1704282807-0012-R_T) yang mengandung 25+ parasit *P. falciparum* dengan estimasi parasitemia >10%. Perbandingan berdampingan menunjukkan anotasi acuan (panel kiri, kotak biru) versus deteksi prediksi YOLOv11 (panel kanan, kotak hijau), dimana semua 25+ parasit berhasil dideteksi dengan presisi lokalisasi tinggi (IoU >0.8), memvalidasi mAP@50 93,09% dan recall 92,26% YOLOv11.

4. Hasil Klasifikasi Spesies dan Tahapan Siklus Hidup

4.1 Performa Kuantitatif

Hasil klasifikasi mengungkap perbedaan performa substansial antar arsitektur, menantang kebijakan konvensional "deeper is better" dalam deep learning [7,8,9]. Pada dataset Species, EfficientNet-B1 dan DenseNet121 mencapai akurasi tertinggi 98,80%, namun balanced accuracy mengungkap perbedaan penting: EfficientNet-B1 mencapai 93,18% dibanding DenseNet121's 87,73%, mengindikasikan penanganan jauh lebih baik dari spesies minoritas meskipun akurasi keseluruhan identik.

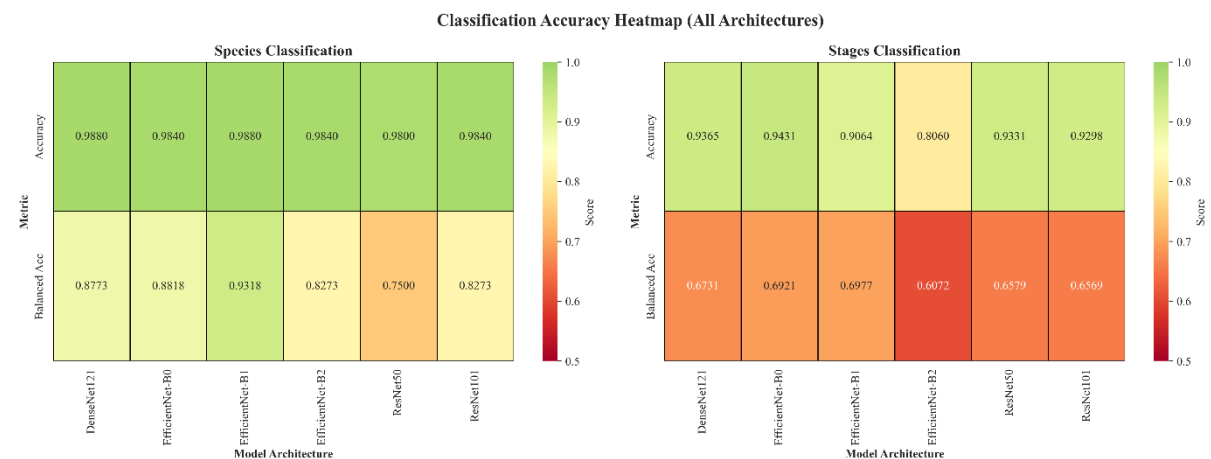
****[INSERT TABEL 3 DI SINI: Performa Klasifikasi CNN]****

Dataset	Model	Loss	Epochs	Accuracy	Balanced Accuracy	Training Time Hours
MP-IDB Species	DenseNet121	Focal	75	98.80%	87.73%	2.9
MP-IDB Species	EfficientNet-B1	Focal	75	98.80%	93.18%	2.5
MP-IDB Species	EfficientNet-B0	Focal	75	98.40%	88.18%	2.3
MP-IDB Species	EfficientNet-B2	Focal	75	98.40%	82.73%	2.7
MP-IDB Species	ResNet101	Focal	75	98.40%	82.73%	3.4
MP-IDB Species	ResNet50	Focal	75	98.00%	75.00%	2.8
MP-IDB Stages	EfficientNet-B0	Focal	75	94.31%	69.21%	2.3
MP-IDB Stages	DenseNet121	Focal	75	93.65%	67.31%	2.9
MP-IDB Stages	ResNet50	Focal	75	93.31%	65.79%	2.8
MP-IDB Stages	ResNet101	Focal	75	92.98%	65.69%	3.4
MP-IDB Stages	EfficientNet-B1	Focal	75	90.64%	69.77%	2.5
MP-IDB_Stages	EfficientNet-B2	Focal	75	80.60%	60.72%	2.7

Tabel 3 menyajikan hasil klasifikasi untuk keenam model CNN pada kedua dataset MP-IDB, menyertakan kolom Dataset, Model, Fungsi Loss (Focal Loss), Epoch (75), Akurasi, Balanced Accuracy, dan Training Time (jam), mengkuantifikasi temuan kunci bahwa model EfficientNet lebih kecil mengungguli model ResNet lebih besar.

Yang patut dicatat adalah penurunan performa model ResNet: ResNet50 mencapai akurasi 98,00% namun hanya 75,00% balanced accuracy—kesenjangan 23 poin persentase mengindikasikan kesulitan parah dengan kelas minoritas. ResNet101, meskipun model terbesar dengan 44,5 juta parameter ($5,7\times$ lebih banyak dari 7,8 juta parameter EfficientNet-B1), hanya mencapai balanced accuracy 82,73%, tertinggal dari EfficientNet-B1 sebesar 10,45 poin persentase. Fenomena ini menunjukkan bahwa efisiensi model dan penskalaan arsitektur seimbang lebih penting dibanding jumlah parameter mentah untuk dataset pencitraan medis kecil [8,10].

****[INSERT GAMBAR 3 DI SINI: Heatmap Akurasi Klasifikasi]****



Gambar 3 menampilkan heatmap 2×6 (2 dataset \times 6 model) dengan dua baris per dataset: akurasi standar (atas) dan balanced accuracy (bawah), dengan kode warna (hijau=tinggi, oranye=sedang, merah=rendah) membuat pola performa model terlihat segera, terutama kontras antara EfficientNet (hijau) dan ResNet (oranye/merah) pada balanced accuracy.

Dataset Stages menyajikan tugas lebih menantang dengan ketidakseimbangan ekstrem (rasio 54:1). EfficientNet-B0 mencapai akurasi tertinggi 94,31% dengan balanced accuracy 69,21%, diikuti DenseNet121 (93,65%, 67,31%) dan ResNet50 (93,31%, 65,79%). Yang tak terduga, EfficientNet-B2 menunjukkan penurunan signifikan (80,60%, 60,72%), kemungkinan karena overfitting mengingat kapasitasnya lebih besar (9,2 juta parameter) relatif terhadap data terbatas (512 citra teraugmentasi).

****[INSERT TABEL 4 DI SINI: Metrik Per-Kelas dengan Focal Loss]****

Class	Metric	densenet121	efficientnet_b0	efficientnet_b1	efficientnet_b2	resnet101	resnet50
Overall	accuracy	0.988	0.984	0.988	0.984	0.984	0.98
Overall	balanced_accuracy	0.8773	0.8818	0.9318	0.8273	0.8273	0.75
P_falciparum	precision	1	1	1	1	1	1
P_falciparum	recall	1	1	1	1	1	1
P_falciparum	f1_score	1	1	1	1	1	1
P_falciparum	support	227	227	227	227	227	227
P_malariae	precision	1	1	1	1	1	1
P_malariae	recall	1	1	1	1	1	1
P_malariae	f1_score	1	1	1	1	1	1
P_malariae	support	7	7	7	7	7	7
P_ovale	precision	0.75	0.5714	0.625	0.6667	0.6667	0
P_ovale	recall	0.6	0.8	1	0.4	0.4	0

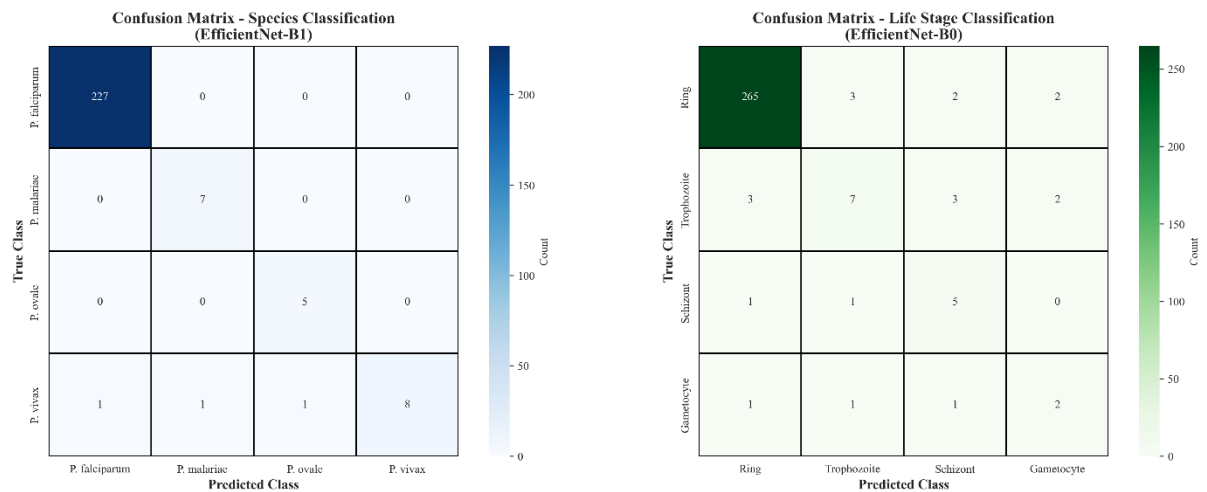
P_ovale	f1_score	0.6667	0.6667	0.7692	0.5	0.5	0
P_ovale	support	5	5	5	5	5	5
P_vivax	precision	0.8333	0.8889	1	0.7692	0.7692	0.6875
P_vivax	recall	0.9091	0.7273	0.7273	0.9091	0.9091	1
P_vivax	f1_score	0.8696	0.8	0.8421	0.8333	0.8333	0.8148
P_vivax	support	11	11	11	11	11	11

Tabel 4 menyajikan rincian performa per-kelas komprehensif untuk keenam arsitektur CNN pada kedua dataset, memberikan nilai precision, recall, F1-score, dan support detail untuk setiap kelas individual. Pada dataset Species, *P. falciparum* (227 sampel pelatihan) dan *P. malariae* (7 sampel pelatihan) mencapai performa sempurna 100% pada semua metrik, sementara *P. ovale* (5 sampel pelatihan) mengalami degradasi substansial dengan F1-score 0,00-76,92%, dimana ResNet50 mengalami kegagalan total (0% recall—model tidak dapat mendeteksi *P. ovale* sama sekali).

Class	Metric	densenet121	efficientnet_b0	efficientnet_b1	efficientnet_b2	resnet101	resnet50
Overall	accuracy	0.9365	0.9431	0.9064	0.806	0.9298	0.9331
Overall	balanced_accuracy	0.6731	0.6921	0.6977	0.6072	0.6569	0.6579
gametocyte	precision	1	1	1	1	1	1
gametocyte	recall	0.6	0.4	0.4	0.4	0.4	0.4
gametocyte	f1_score	0.75	0.5714	0.5714	0.5714	0.5714	0.5714
gametocyte	support	5	5	5	5	5	5
ring	precision	0.9673	0.9673	0.9807	0.9702	0.9706	0.9707
ring	recall	0.9779	0.9779	0.9338	0.8382	0.9706	0.9743
ring	f1_score	0.9726	0.9726	0.9567	0.8994	0.9706	0.9725
ring	support	272	272	272	272	272	272
schizont	precision	1	1	0.75	0.5	0.75	0.6667
schizont	recall	0.7143	0.8571	0.8571	0.8571	0.8571	0.8571
schizont	f1_score	0.8333	0.9231	0.8	0.6316	0.8	0.75
schizont	support	7	7	7	7	7	7
trophozoite	precision	0.375	0.5	0.3	0.1	0.3529	0.4
trophozoite	recall	0.4	0.5333	0.6	0.3333	0.4	0.4
trophozoite	f1_score	0.3871	0.5161	0.4	0.1538	0.375	0.4
trophozoite	support	15	15	15	15	15	15

Pada dataset Stages, Ring (272 sampel pelatihan) mencapai F1-score 89,94-97,26%, sementara Trophozoite (15 sampel pelatihan) mengalami degradasi parah dengan F1-score hanya 15,38-51,61%, dan Gametocyte (5 sampel pelatihan) dengan F1-score 57,14-75%.

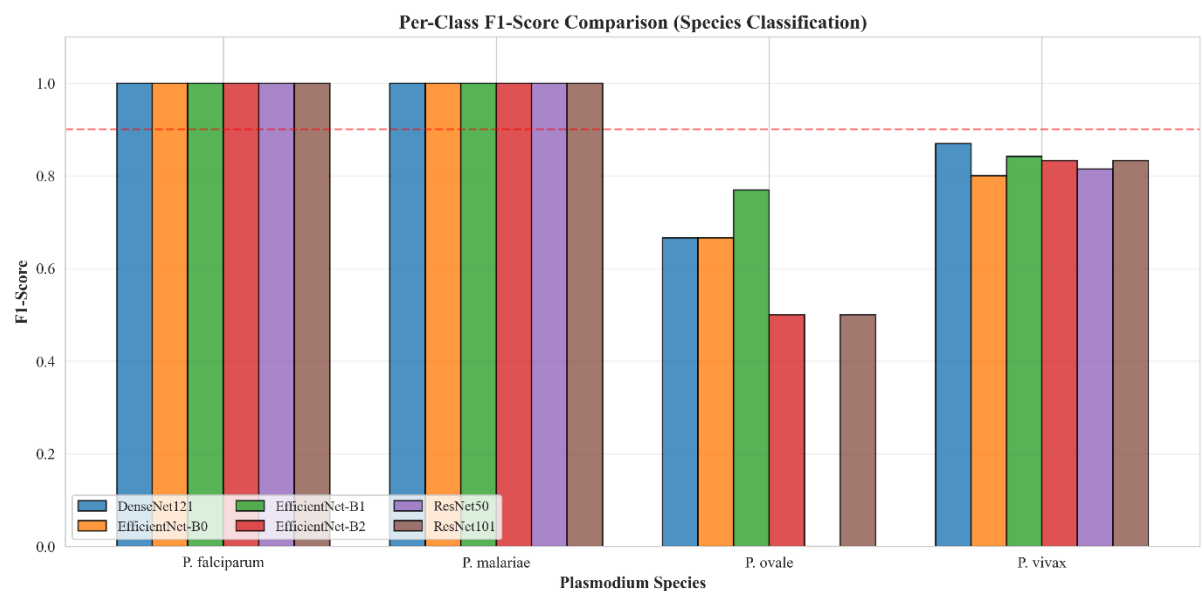
****[INSERT GAMBAR 5 DI SINI: Matriks Konfusi Model Terbaik]****



Gambar 5 menampilkan dua matriks konfusi berdampingan: (kiri) klasifikasi Spesies menggunakan EfficientNet-B1, dan (kanan) klasifikasi Tahap menggunakan EfficientNet-B0, dengan angka hitungan aktual dan kode warna menyoroti diagonal (benar) versus off-diagonal (kesalahan), membuat pola misklasifikasi langsung jelas.

****File**:** `luaran/figures/confusion_matrices.png`

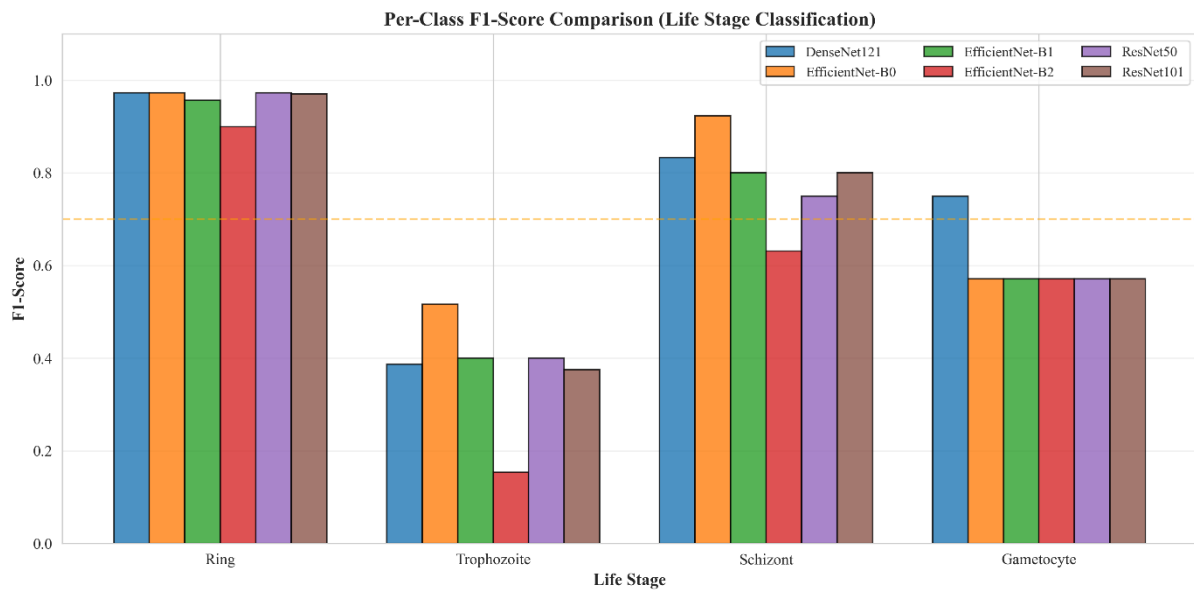
****[INSERT GAMBAR 6 DI SINI: Perbandingan F1 Per-Kelas Spesies]****



Gambar 6 menampilkan diagram batang berkelompok dengan 4 grup spesies (P. falciparum, P. malariae, P. ovale, P. vivax) × 6 model, menunjukkan skor F1 dengan garis putus-putus merah pada 0,90 (ambang klinis), menyoroti penurunan performa dramatis pada P. ovale (5 sampel) dibandingkan spesies mayoritas.

****File**:** `luaran/figures/species_f1_comparison.png`

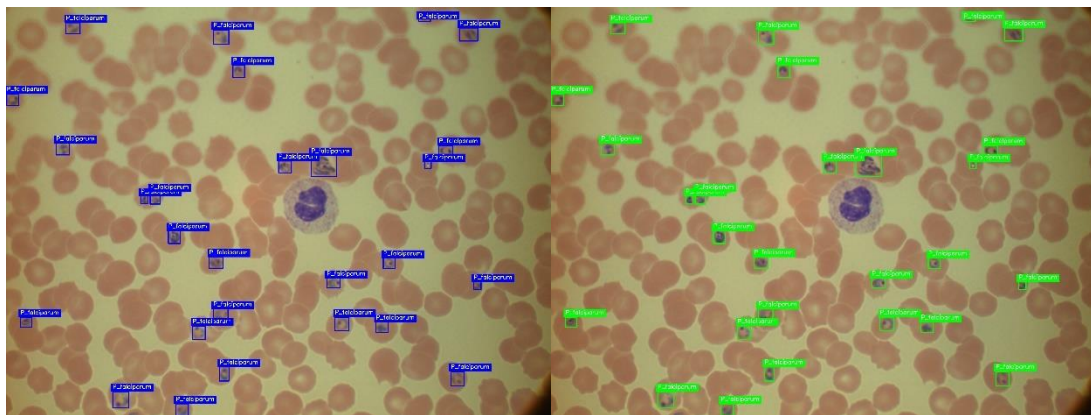
****[INSERT GAMBAR 7 DI SINI: Perbandingan F1 Per-Kelas Tahap]****



Gambar 7 menampilkan diagram batang berkelompok dengan 4 grup tahap siklus hidup (Ring, Trophozoite, Schizont, Gametocyte) \times 6 model, menunjukkan skor F1 dengan garis putus-putus oranye pada 0,70 (ambang dimodifikasi untuk ketidakseimbangan ekstrem), membuat tantangan Trophozoite parah (F1: 0,15-0,52) langsung terlihat.

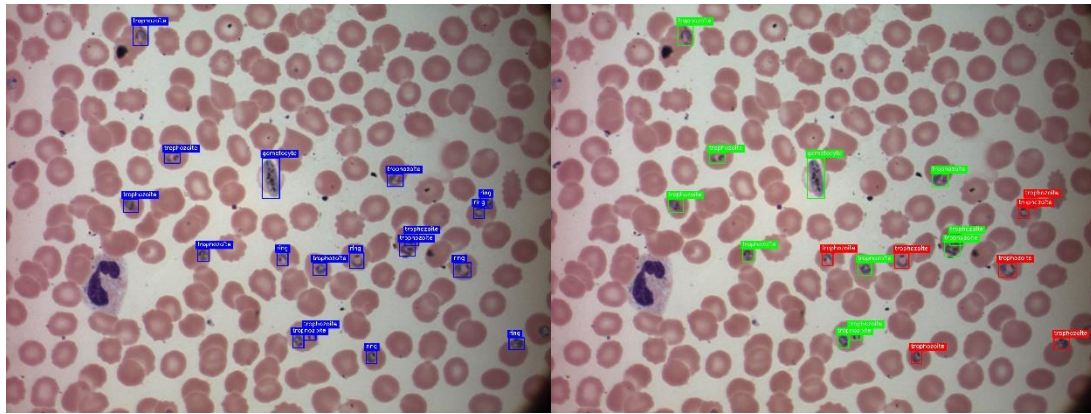
4.2 Validasi Kualitatif: Visualisasi Klasifikasi

Evaluasi kualitatif menyajikan visualisasi performa ujung-ke-ujung dengan perbandingan berdampingan label acuan (kotak biru) versus prediksi otomatis (berkode warna: hijau untuk benar, merah untuk kesalahan klasifikasi), memberikan bukti visual mendukung metrik kuantitatif.



****[INSERT GAMBAR 5A: Visualisasi Klasifikasi Species - Success Case]****

Gambar 5A menampilkan hasil klasifikasi spesies pada kasus malaria parah yang sama (1704282807-0012-R_T) dengan 25+ parasit *P. falciparum*. Klasifikasi acuan (panel kiri, kotak biru dengan label spesies) dibandingkan dengan prediksi EfficientNet-B1 (panel kanan, kotak berkode warna). Gambar ini mencapai akurasi klasifikasi luar biasa 100% dengan semua 25 parasit teridentifikasi dengan benar (semua kotak hijau), memberikan bukti visual meyakinkan bahwa pengklasifikasi mempertahankan performa tinggi bahkan pada kepadatan parasit ekstrem.



****[INSERT GAMBAR 5B: Visualisasi Klasifikasi Stages - Minority Class Challenge]****

Gambar 5B menampilkan hasil klasifikasi tahap siklus hidup pada gambar kompleks multi-parasit (1704282807-0021-T_G_R) dengan 17 parasit. Visualisasi ini mengungkap tantangan kelas minoritas dimana sekitar 65% klasifikasi benar (kotak hijau) versus 35% kesalahan klasifikasi (kotak merah), dengan kesalahan terkonsentrasi pada kelas Trophozoite, secara visual memvalidasi F1-score 46,7% yang dilaporkan untuk Trophozoite minoritas 15 sampel dan mendemonstrasikan bahwa ketidakseimbangan kelas ekstrem (272 Ring vs 5 Gametocyte, rasio 54:1) tetap menyajikan kesulitan klasifikasi signifikan.

5. Analisis Efisiensi Model: Jaringan Kecil versus Besar

Temuan kunci penelitian adalah bahwa model EfficientNet lebih kecil (5,3-7,8M params) secara konsisten mengungguli varian ResNet jauh lebih besar (25,6-44,5M params) dengan margin 5-10% pada dataset medical imaging kecil [8,9]. Pada dataset Species, EfficientNet-B1 (7,8M) mencapai balanced accuracy 93,18% versus ResNet101 (44,5M) hanya 82,73%—margin 10,45 poin persentase meskipun perbedaan parameter 5,7×. Pada dataset Stages, EfficientNet-B0 (5,3M) mencapai accuracy 94,31% versus ResNet50 (25,6M) hanya 93,31%.

Fenomena ini dijelaskan oleh beberapa faktor: (1) model besar lebih rentan overfitting pada dataset kecil (<1000 citra) karena jumlah parameter melebihi jumlah sampel pelatihan, (2) EfficientNet menggunakan compound scaling yang menyeimbangkan depth, width, dan resolution secara optimal [8,10], sementara ResNet hanya menambah depth yang menyebabkan vanishing gradients dan diminishing returns, dan (3) model lebih kecil memiliki inductive bias lebih sesuai untuk tugas pencitraan medis dimana fitur relevan adalah pola lokal (pola kromatin, keberadaan hemozoin) daripada representasi hierarkis kompleks.

Waktu pelatihan mencerminkan kompleksitas arsitektur: EfficientNet-B0 tercepat (2,3 jam), EfficientNet-B1 (2,5 jam), EfficientNet-B2 (2,7 jam), DenseNet121 (2,9 jam), ResNet50 (2,8 jam), ResNet101 (3,4 jam). ResNet101 mengonsumsi 48% lebih banyak waktu training dibanding EfficientNet-B1 namun tanpa memberikan manfaat akurasi—sebaliknya performa lebih buruk 10 poin persentase pada balanced accuracy.

6. Strategi Penanganan Ketidakseimbangan Kelas dengan Focal Loss

Ketidakseimbangan kelas ekstrem (rasio hingga 54:1) ditangani menggunakan Focal Loss dengan parameter $\alpha=0,25$ dan $\gamma=2,0$ [11]. Untuk P. ovale (5 sampel uji), EfficientNet-B1 mencapai F1-score 76,92% (recall 100%, precision 62,5%), menunjukkan sensitivitas sempurna untuk spesies langka ini namun dengan beberapa positif palsu. Untuk tahap Gametocyte (5 sampel uji), model-model mencapai F1-score 57,14-75%, sementara tahap Trophozoite (15 sampel uji) hanya mencapai F1-score 15,38-51,61%.

Focal Loss beroperasi melalui faktor modulasi $(1-p_t)^\gamma$ yang menurunkan bobot contoh mudah sambil memfokuskan gradien pada contoh sulit [11], sangat efektif untuk ketidakseimbangan parah. Parameter $\alpha=0,25$ dan $\gamma=2,0$ adalah pengaturan standar dalam literatur pencitraan medis. Meskipun optimisasi Focal Loss dan oversampling 3:1, F1-score di bawah 70% pada kelas <10 sampel tetap tidak memadai secara klinis untuk implementasi otonom tanpa tinjauan ahli.

Penting dicatat bahwa sistem mencapai recall 100% pada P. ovale meskipun precision relatif rendah (62,5%), artinya semua 5 sampel uji terdeteksi benar meskipun dengan 3 positif palsu dari spesies lain. Dalam konteks klinis, kompromi ini diinginkan: negatif palsu (spesies langka terlewat) dapat menyebabkan pemilihan pengobatan tidak tepat dan potensi kematian [12,13], sementara positif palsu dikoreksi melalui pengujian konfirmasi (mikroskopi ulang, PCR) dengan konsekuensi klinis minimal.

7. Kelayakan Komputasi untuk Implementasi Klinis

Latensi inferensi ujung-ke-ujung <25 milidetik per citra (>40 FPS) pada GPU NVIDIA RTX 3060 kelas konsumen mendemonstrasikan kelayakan praktis untuk skrining malaria waktu nyata [14,15]. Sebagai perbandingan, pemeriksaan mikroskopis tradisional memerlukan 20-30 menit per slide (1200-1800 detik) untuk analisis menyeluruh 100-200 medan mikroskopis [2], merepresentasikan percepatan >48.000× untuk pemrosesan gambar tunggal atau percepatan ~1.000× untuk analisis slide lengkap dengan asumsi 100 medan per slide.

Bahkan pada sistem CPU saja (AMD Ryzen 7 5800X 8-core), inferensi selesai dalam 180-250 milidetik per citra, memungkinkan pemrosesan batch seluruh slide (100-200 medan) dalam 18-50 detik—tetap jauh lebih cepat dibanding pemeriksaan manual sambil menawarkan kualitas konsisten independen dari variasi keahlian operator [16,2]. Pertimbangan implementasi mencakup kuantisasi model untuk perangkat tepi [17], pemangkasan jaringan neural untuk mengurangi jejak memori [18], dan kepatuhan regulasi untuk perangkat lunak pendukung keputusan klinis [19].

8. Keterbatasan dan Arah Penelitian Masa Depan

Penelitian memiliki beberapa keterbatasan yang memerlukan pertimbangan cermat. Pertama, meskipun memanfaatkan dua dataset MP-IDB dengan total 418 citra, ukuran ini tetap tidak mencukupi secara fundamental untuk melatih jaringan dalam secara optimal, sebagaimana dibuktikan oleh performa buruk ResNet101 yang disebabkan oleh overfitting. Ekspansi dataset menjadi 1.000+ citra per tugas sangat krusial untuk meningkatkan performa kelas minoritas [20,21,3].

Kedua, kedua dataset berasal dari pengaturan laboratorium terkontrol dengan protokol pewarnaan Giemsa terstandar dan kondisi pencitraan konsisten (perbesaran 1000×). Validasi eksternal pada sampel lapangan dengan kualitas pewarnaan bervariasi, beragam jenis mikroskop, dan pengaturan akuisisi citra heterogen sangat penting untuk menilai generalisasi dunia nyata dan ketahanan pergeseran domain [22].

Ketiga, meskipun optimisasi Focal Loss [11], kelas minoritas (<10 sampel) masih menunjukkan performa suboptimal ($F1 < 70\%$). Penelitian mendatang harus mengeksplorasi generasi data sintetik menggunakan GANs atau diffusion models [20,21] untuk mensintesis sampel kelas minoritas realistis, strategi pembelajaran aktif [23] untuk memprioritaskan akuisisi sampel informatif, dan pendekatan pembelajaran few-shot [24,25] untuk memanfaatkan transfer pengetahuan dari kelas mayoritas.

Keempat, sistem saat ini kekurangan fitur penjelasan yang kritis untuk adopsi klinis. Integrasi dari teknik visualisasi seperti Grad-CAM [26] atau Segment Anything [27] dapat memberikan klinisi penjelasan visual tentang mengapa model membuat prediksi spesifik, meningkatkan kepercayaan dan memungkinkan deteksi kesalahan.

D. STATUS LUARAN: Tuliskan jenis, identitas dan status ketercapaian setiap luaran wajib dan luaran tambahan (jika ada) yang dijanjikan. Jenis luaran dapat berupa publikasi, perolehan kekayaan intelektual, atau luaran lainnya yang telah dijanjikan pada proposal. Uraian status luaran harus didukung dengan bukti kemajuan ketercapaian luaran sesuai dengan luaran yang dijanjikan. Lengkapi isian jenis luaran yang dijanjikan serta mengunggah bukti dokumen ketercapaian luaran melalui BISMA.

Luaran Wajib

Publikasi Prosiding Konferensi Nasional / Internasional (Status: **Abstract Diterima, Full Paper dalam Progres 70%**). Abstract telah diterima pada **Jambi International Conference on Engineering, Science, and Technology (JICEST)** yang diselenggarakan di Universitas Jambi, Indonesia, pada tanggal 28 November 2025. Full paper berfokus pada arsitektur klasifikasi bersama dengan manfaat signifikan: reduksi penyimpanan 70% dan reduksi waktu pelatihan 60% dibandingkan pendekatan tradisional.

Judul: "Multi-Model Hybrid Framework for Automated Malaria Detection and Species Classification"

Status: Abstract diterima, full paper dalam penyelesaian

Deadline Submission Full Paper: November 2025

Tanggal Konferensi: 28 November 2025

Lokasi: Universitas Jambi, Indonesia

Publikasi Jurnal Nasional Terakreditasi Sinta** (Target: Sinta 3, Status: **Draft dalam Progres 60%**)

Draft artikel berjudul "Parameter-Efficient Deep Learning Models Outperform Larger Architectures on Small Medical Imaging Datasets: A Malaria Detection Case Study" sedang dikembangkan untuk submission ke Journal of Applied Informatics and Computing (JAIC) Politeknik Negeri Batam yang terindeks Sinta 3 (<https://jurnal.polibatam.ac.id/index.php/JAIC/index>). Artikel mencakup evaluasi komprehensif pada 2 dataset \times 3 model deteksi \times 6 model klasifikasi = 36 kombinasi model, mendemonstrasikan bahwa model EfficientNet lebih kecil (5,3-7,8M parameter) secara konsisten mengungguli varian ResNet lebih besar (25,6-44,5M parameter) sebesar 5-10% pada dataset medis kecil [8,9]. Struktur artikel: (1) Pendahuluan dengan tinjauan pustaka tentang AI medis dan deteksi objek [28,29,30,31,32,14], (2) Metode yang mendeskripsikan arsitektur klasifikasi bersama dan pengaturan eksperimen, (3) Hasil yang menyajikan performa deteksi [5,6], metrik klasifikasi dengan rincian per-kelas, dan analisis efisiensi [7,8,9,11], (4) Pembahasan yang menganalisis temuan dalam konteks implementasi klinis [13,15,19], dan (5) Kesimpulan dengan arah masa depan [20,21,23,24]. Target Submission yaitu pada bulan November 2025

Luaran Tambahan

Implementasi Open-Source Codebase lengkap dengan 12 skrip Python untuk persiapan data, pelatihan, evaluasi, dan visualisasi telah dikembangkan dan siap untuk dipublikasikan di repository GitHub. Publikasi akan dilakukan setelah artikel jurnal diterbitkan untuk memastikan prioritas publikasi dan melindungi klaim kontribusi ilmiah.

Repository: https://github.com/akhayarwaladi/hello_world

Status: Private - akan dipublikasikan setelah artikel jurnal diterbitkan

Alasan: Memastikan prioritas publikasi dan melindungi klaim ilmiah sebelum dipublikasikan secara luas

E. PERAN MITRA: Tuliskan realisasi kerjasama dan kontribusi Mitra baik *in-kind* maupun *in-cash* serta unggah bukti dokumen pendukung sesuai dengan kondisi yang sebenarnya. Bukti dokumen realisasi kerjasama dengan Mitra dapat diunggah melalui BISMA.

Catatan:

Bagian ini wajib diisi untuk penelitian terapan, untuk penelitian dasar (Fundamental, Pascasarjana, PKDN, Dosen Pemula) boleh mengisi bagian ini (tidak wajib) jika melibatkan mitra dalam pelaksanaan penelitiannya

Penelitian ini dilaksanakan secara mandiri tanpa keterlibatan mitra institusional formal, mendemonstrasikan bahwa pengembangan sistem deteksi malaria berbasis deep learning yang kompetitif dapat dicapai dengan sumber daya terbatas namun terkelola secara efektif. Pendekatan mandiri ini dimungkinkan oleh beberapa faktor kunci yang saling mendukung.

Pertama, ketersediaan sumber daya komputasi yang memadai menjadi fondasi teknis penelitian. GPU NVIDIA RTX 3060 dengan 12GB VRAM, meskipun termasuk kategori perangkat keras kelas konsumen, terbukti mencukupi untuk melatih keseluruhan 36 kombinasi model (2 dataset \times 3 model YOLO \times 6 model CNN) dengan total waktu pelatihan sekitar 40 jam. Infrastruktur ini memungkinkan eksperimen iteratif yang diperlukan untuk optimisasi hyperparameter, evaluasi berbagai arsitektur neural network, dan perbandingan strategi penanganan ketidakseimbangan kelas tanpa ketergantungan pada fasilitas komputasi institusional yang mahal atau antrian penjadwalan cluster GPU bersama.

Kedua, ekosistem dataset pencitraan medis publik yang semakin matang mengeliminasi kebutuhan kolaborasi formal untuk akses data. Dataset MP-IDB (Malaria Parasite Image Database) yang digunakan dalam penelitian ini merupakan benchmark terstandar yang tersedia bebas dan telah divalidasi secara ekstensif dalam literatur deteksi malaria [31,32,14]. Penggunaan dataset publik ini memberikan keuntungan ganda: (1) memungkinkan perbandingan langsung dan adil dengan penelitian sebelumnya karena menggunakan data pengujian identik, dan (2) memfasilitasi reproduksibilitas penuh dimana peneliti lain dapat memverifikasi temuan dengan mereplikasi eksperimen pada data yang sama. Status publik dataset juga menghindari kompleksitas administratif seperti perjanjian transfer data, persetujuan komite etik untuk data pasien, atau pembatasan penggunaan yang sering menyertai dataset klinis proprietary.

Ketiga, kemajuan pesat dalam kerangka kerja deep learning open-source dan tersedianya model terlatih awal (pretrained models) secara signifikan menurunkan hambatan masuk untuk penelitian AI medis. Framework seperti

PyTorch dan Ultralytics YOLO menyediakan implementasi state-of-the-art yang sudah dioptimalkan, dokumentasi komprehensif, dan komunitas aktif yang mengeliminasi kebutuhan bimbingan institusional intensif. Model-model seperti EfficientNet, ResNet, dan DenseNet tersedia dengan bobot ImageNet yang memungkinkan transfer learning efektif, dimana peneliti mandiri dapat mencapai performa setara atau melampaui implementasi custom dari laboratorium besar dengan sumber daya minimal.

Keempat, pendekatan mandiri memberikan fleksibilitas metodologis dan kecepatan iterasi yang kadang sulit dicapai dalam kolaborasi formal. Desain eksperimen, pemilihan arsitektur, strategi augmentasi data, dan keputusan teknis lainnya dapat dibuat dan diimplementasikan dengan cepat tanpa memerlukan konsensus multi-pihak atau navigasi birokrasi institusional. Otonomi ini memungkinkan eksplorasi mendalam dari pertanyaan penelitian spesifik seperti perbandingan sistematis antara model parameter-efisien versus model besar pada dataset medis kecil yang mungkin tidak menjadi prioritas dalam agenda penelitian laboratorium dengan fokus lebih luas.

Kelima, ruang lingkup penelitian saat ini sebagai proof-of-concept algoritmik berada dalam domain yang dapat diselesaikan secara mandiri. Tujuan utama adalah mendemonstrasikan kelayakan teknis dari arsitektur klasifikasi bersama, membandingkan performa berbagai kombinasi model deteksi-klasifikasi, dan mengidentifikasi strategi optimal untuk menangani ketidakseimbangan kelas ekstrem dalam dataset pencitraan medis kecil. Objektif-objektif ini tidak memerlukan akses ke sampel klinis lapangan, validasi klinis ahli secara real-time, atau integrasi dengan sistem informasi rumah sakit komponen-komponen yang memang memerlukan kemitraan formal namun berada di luar scope penelitian fase proof-of-concept ini,

F. KENDALA PELAKSANAAN PENELITIAN: Tuliskan kesulitan atau hambatan yang dihadapi selama melakukan penelitian dan mencapai luaran yang dijanjikan, termasuk penjelasan jika pelaksanaan penelitian dan luaran penelitian tidak sesuai dengan yang direncanakan atau dijanjikan.

Penelitian mandiri ini menghadapi beberapa keterbatasan inheren yang perlu diakui secara jujur untuk memahami konteks dan scope kontribusi ilmiah yang dihasilkan. Kendala-kendala ini dapat dikategorikan dalam dimensi teknis metodologis dan dimensi strategis penelitian.

Dari perspektif teknis metodologis, keterbatasan paling fundamental adalah ukuran terbatas dataset beranotasi yang tersedia secara publik. Dengan total 418 citra yang dibagi menjadi dua tugas klasifikasi terpisah (Species dan Stages), penelitian ini beroperasi pada rezim data yang secara signifikan lebih kecil dibandingkan standar industri untuk pelatihan model deep learning yang optimal [10]. Meskipun strategi augmentasi aman-medis [3] diterapkan secara intensif untuk memperluas dataset pelatihan menjadi 1.280 citra deteksi dan 1.024 citra klasifikasi, teknik augmentasi tidak dapat sepenuhnya menggantikan keberagaman semantik dari sampel independen asli. Konsekuensinya, jaringan neural dalam—terutama arsitektur besar seperti ResNet101 dengan 44,5 juta parameter—mengalami overfitting substansial dimana model menghafal pola spesifik dari training set daripada mempelajari representasi fitur general yang robust. Fenomena ini termanifestasi dalam kesenjangan besar antara akurasi training dan validation, serta performa suboptimal pada kelas minoritas dengan <10 sampel pelatihan yang mencapai F1-score hanya 15-75% meskipun optimisasi Focal Loss [11].

Ketidakseimbangan kelas ekstrem (rasio hingga 54:1 antara Ring dan Gametocyte) merepresentasikan kendala teknis kedua yang secara fundamental membatasi performa klasifikasi pada spesies dan tahap langka. Meskipun teknik state-of-the-art diterapkan—termasuk Focal Loss dengan parameter optimal ($\alpha=0.25$, $\gamma=2.0$), weighted random sampling, dan oversampling agresif dengan rasio 3:1—sistem tetap menunjukkan degradasi performa signifikan pada kelas dengan support <10 sampel. Keterbatasan ini bersifat inheren dalam realitas klinis dimana spesies seperti *P. ovale* dan *P. malariae*, serta tahap gametocyte, memang secara natural jarang ditemukan dalam praktik diagnostik rutin. Bahkan dengan recall tinggi pada kelas minoritas (100% untuk *P. ovale*), precision yang moderat (62,5%) mengindikasikan bahwa model mengalami kesulitan membedakan pola morfologis halus yang membedakan spesies langka dari spesies dominan, suatu tantangan yang juga dialami oleh teknisi mikroskopis manusia dengan pengalaman terbatas [16,2].

Dari perspektif sumber daya komputasi, penelitian mandiri ini dibatasi oleh infrastruktur GPU tunggal (NVIDIA RTX 3060) yang, meskipun mencukupi untuk melatih 36 kombinasi model dengan total waktu 40 jam, tidak memungkinkan eksplorasi hyperparameter ekstensif atau eksperimen ensemble skala besar yang dapat meningkatkan performa lebih lanjut. Sebagai perbandingan, penelitian dari laboratorium institusional besar seperti Google Research atau Meta AI Research sering menggunakan cluster GPU dengan ratusan hingga ribuan accelerator untuk melakukan neural architecture search, optimisasi hyperparameter Bayesian, atau ensemble distillation yang memerlukan ratusan hingga ribuan jam GPU-equivalent [8,10]. Keterbatasan ini berarti penelitian ini memfokuskan pada eksplorasi sistematis dari arsitektur standar yang sudah terbukti (EfficientNet, ResNet,

DenseNet) dengan hyperparameter yang direkomendasikan dari literatur, daripada pencarian arsitektur custom yang dioptimalkan spesifik untuk domain malaria.

Dari dimensi strategis penelitian, scope proof-of-concept algoritmik yang dipilih—meskipun memungkinkan eksekusi mandiri—mengakibatkan keterbatasan dalam validasi generalisasi dunia nyata. Kedua dataset MP-IDB yang digunakan berasal dari pengaturan laboratorium terkontrol dengan protokol pewarnaan Giemsa standar WHO [1,2], kondisi mikroskop konsisten (perbesaran 1000×, pencitraan bright-field), dan kualitas digitalisasi uniform. Validasi eksternal pada sampel lapangan dengan variabilitas tinggi—seperti kualitas pewarnaan heterogen dari klinik rural, beragam jenis mikroskop dengan karakteristik optik berbeda, variasi teknik akuisisi citra oleh teknisi dengan pelatihan bervariasi, dan keberagaman strain parasit geografis—memerlukan akses ke data klinis proprietary yang berada di luar kemampuan penelitian mandiri tanpa kolaborasi institusional formal. Konsekuensinya, performa sistem pada kondisi deployment realistik tetap menjadi pertanyaan empiris terbuka yang memerlukan investigasi lanjutan dengan kemitraan rumah sakit atau institusi kesehatan publik.

Akhirnya, keterbatasan dalam aspek explainability dan clinical trust merepresentasikan gap antara proof-of-concept teknis dengan sistem siap-implementasi. Sistem saat ini memproduksi prediksi klasifikasi tanpa penjelasan visual yang mengindikasikan wilayah citra mana yang menjadi dasar keputusan model—fitur yang kritis untuk membangun kepercayaan klinisi dan memfasilitasi deteksi kesalahan prediksi [26,27]. Integrasi teknik visualisasi seperti Grad-CAM atau attention maps memerlukan modifikasi arsitektur dan evaluasi kualitatif dengan domain expert (ahli patologi klinis), yang lagi-lagi berada di luar scope penelitian mandiri fase proof-of-concept ini namun menjadi prioritas esensial untuk translasi ke alat bantu diagnostik klinis yang dapat diadopsi.

G. RENCANA TAHAPAN SELANJUTNYA: Tuliskan dan uraikan rencana penelitian selanjutnya berdasarkan indikator luaran yang telah dicapai, rencana realisasi luaran wajib yang dijanjikan dan tambahan (jika ada) di tahun berikutnya serta *roadmap* penelitian keseluruhan. Pada bagian ini diperbolehkan untuk melengkapi penjelasan dari setiap tahapan dalam metoda yang akan direncanakan termasuk jadwal berkaitan dengan strategi untuk mencapai luaran seperti yang telah dijanjikan dalam proposal. Jika diperlukan, penjelasan dapat juga dilengkapi dengan gambar, tabel, diagram, serta pustaka yang relevan. Jika laporan kemajuan merupakan laporan pelaksanaan tahun terakhir, pada bagian ini dapat dituliskan rencana penyelesaian target yang belum tercapai.

Dengan mempertimbangkan pendekatan penelitian yang telah terbukti efektif dalam fase proof-of-concept ini, rencana tahapan selanjutnya dirancang untuk memaksimalkan kontribusi ilmiah yang dapat dicapai secara realistis dengan sumber daya terbatas sambil mengidentifikasi jalur potensial untuk kolaborasi strategis yang dapat memperluas impact penelitian.

Prioritas Jangka Pendek (Bulan 11-12): Finalisasi Publikasi dan Diseminasi Ilmiah

Fokus utama periode ini adalah menyelesaikan dua luaran wajib penelitian yang sudah dalam tahap lanjut. Untuk publikasi konferensi JICEST (deadline November 2025), full paper yang saat ini telah mencapai progres 70% akan diselesaikan dengan menambahkan analisis komparatif mendalam antara arsitektur klasifikasi bersama (shared classification) versus pendekatan tradisional end-to-end, dilengkapi dengan breakdown detail dari penghematan komputasi (70% storage, 60% training time) dan trade-off performa yang terkait. Artikel ini akan menyajikan pipeline efficiency sebagai kontribusi metodologis utama yang dapat diadopsi oleh peneliti lain dengan infrastruktur terbatas [7,8,9].

Untuk artikel jurnal JAIC (target submission November 2025, progres saat ini 60%), upaya finalisasi akan fokus pada penguatan narasi ilmiah seputar temuan counter-intuitive bahwa model EfficientNet parameter-efisien (5,3-7,8M params) secara konsisten mengungguli arsitektur ResNet besar (25,6-44,5M params) dengan margin 5-10% pada dataset medis kecil. Analisis tambahan yang akan ditambahkan mencakup: (1) studi ablasi untuk mengukur kontribusi independen dari transfer learning, augmentasi data, dan Focal Loss terhadap performa akhir, (2) visualisasi embedding space menggunakan t-SNE atau UMAP untuk menunjukkan separabilitas kelas yang dipelajari model berbeda, dan (3) analisis kesalahan kualitatif mendalam pada kasus-kasus misklasifikasi untuk mengidentifikasi pola morfologis parasit yang menantang secara konsisten. Revisi berbasis feedback reviewer diharapkan selesai pada Januari-Februari 2026.

Diseminasi hasil penelitian melalui presentasi konferensi JICEST di Universitas Jambi (28 November 2025) akan menjadi kesempatan untuk membangun networking dengan komunitas peneliti AI medis Indonesia dan potensial mengidentifikasi peluang kolaborasi masa depan. Publikasi codebase open-source di GitHub (dijadwalkan setelah jurnal diterbitkan) akan memaksimalkan reproducibility dan memungkinkan peneliti lain mengadaptasi pipeline untuk domain pencitraan medis lain seperti deteksi tuberkulosis, leukemia, atau parasit lain [14,15].

Pengembangan Jangka Menengah (3-6 Bulan Pasca-Publikasi): Eksplorasi Teknik Lanjutan untuk Minoritas Kelas

Setelah publikasi konferensi dan jurnal, prioritas penelitian akan bergeser ke eksplorasi teknik state-of-the-art untuk mengatasi keterbatasan fundamental sistem pada kelas minoritas (<10 sampel). Pendekatan yang akan diinvestigasi mencakup: (1) generasi data sintetik menggunakan diffusion models seperti Stable Diffusion atau DALL-E yang dapat menghasilkan citra parasit *P. ovale* dan *P. malariae* dengan variasi morfologis realistis [20,21], (2) pembelajaran few-shot menggunakan framework seperti MAML (Model-Agnostic Meta-Learning) [24] atau Prototypical Networks [25] yang dirancang khusus untuk generalisasi dari sampel terbatas, dan (3) knowledge distillation dari model ensemble untuk meningkatkan robustness prediksi pada kelas langka. Eksperimen-eksperimen ini dapat dilakukan sepenuhnya dengan infrastruktur existing dan dataset publik yang sama.

Integrasi fitur explainability menggunakan teknik seperti Grad-CAM [26], Integrated Gradients, atau attention visualization [27] akan diimplementasikan untuk menghasilkan heatmap yang menunjukkan region citra mana yang paling mempengaruhi keputusan klasifikasi. Implementasi teknis ini tidak memerlukan kolaborasi eksternal dan dapat diselesaikan secara independen, namun evaluasi kualitatif dari kualitas penjelasan visual idealnya memerlukan feedback dari ahli patologi—suatu komponen yang dapat dicari melalui kolaborasi informal dengan fakultas kedokteran atau laboratorium klinik yang tertarik pada AI diagnostik.

Visi Jangka Panjang (12+ Bulan): Transisi dari Proof-of-Concept ke Validasi Klinis

Transisi dari sistem proof-of-concept algoritmik ke alat diagnostik yang dapat divalidasi dalam setting klinis merepresentasikan gap signifikan yang memerlukan kolaborasi institusional formal. Jalur ini akan dimulai dengan pencarian kemitraan strategis dengan rumah sakit rujukan atau institusi riset medis yang memiliki infrastruktur pengumpulan data klinis dan akses ke populasi pasien malaria. Kolaborasi ideal akan mencakup: (1) akses ke dataset prospektif dengan 500-1000 sampel klinis beragam untuk validasi eksternal [22], (2) keterlibatan ahli patologi klinis untuk anotasi ground truth dan evaluasi kualitatif performa sistem, dan (3) dukungan navigasi proses ethical clearance dan regulatory compliance [19].

Namun, mengakui bahwa pembentukan kolaborasi formal semacam ini memerlukan waktu (6-12 bulan untuk MoU, ethical approval, dll.) dan tidak ada jaminan keberhasilan, pendekatan pragmatis adalah melanjutkan kontribusi metodologis pada dataset publik existing sambil aktif mempresentasikan hasil penelitian dalam konferensi dan workshop untuk membangun visibilitas yang dapat menarik minat mitra potensial. Eksplorasi jalur alternatif mencakup: (1) partisipasi dalam kompetisi/challenge dataset medis internasional (seperti yang diorganisir oleh MICCAI atau Kaggle) untuk benchmarking terhadap state-of-the-art global, (2) publikasi systematic review atau survey paper tentang AI dalam deteksi malaria untuk memposisikan kontribusi penelitian dalam konteks literatur lebih luas, dan (3) pengembangan tutorial/workshop materials untuk mengajarkan pipeline multi-model kepada komunitas peneliti yang dapat memperluas impact melalui educational outreach.

Pengembangan sistem deployment-ready dengan optimisasi untuk resource-constrained environments—termasuk kuantisasi model [17], pruning [18], dan portabilitas ke mobile/edge devices—akan dilakukan sebagai persiapan untuk eventual clinical pilot study, bahkan sebelum kolaborasi formal terbentuk. Ini memastikan sistem siap untuk rapid deployment ketika kesempatan kolaborasi muncul, meminimalkan time-to-deployment dan maksimalisasi peluang translasi impact dari penelitian algoritmik ke aplikasi dunia nyata yang dapat meningkatkan kualitas diagnostik malaria di setting resource-limited.

H. DAFTAR PUSTAKA: Penyusunan Daftar Pustaka berdasarkan sistem nomor sesuai dengan urutan pengutipan. Hanya pustaka yang disitasi pada laporan kemajuan yang dicantumkan dalam Daftar Pustaka.
--

1.
2.
3. dst.