Journal of Vocational Education and Information Technology
Vol. 7, No. 2, Desember 2026, 690 - 702
E-ISSN: 2722-5305
Terakreditasi **SINTA Peringkat 4**
Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 230/E/KPT/2024

Analisis Komparatif Arsitektur CNN Modern untuk Klasifikasi Malaria dengan *Progressive Fine-Tuning* dan Grad-CAM

Muhammad Ridho Habibie[1,*], Christya Aji Putra[2], Yisti Vita Via[3]

[1,2,3] Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia

22081010246@student.upnjatim.ac.id , ajiputra@upnjatim.ac.id , yistivia.if@upnjatim.ac.id

* Korespondensi: e-mail: 22081010246@student.upnjatim.ac.id

Diterima: 12-08-2026 ; Review: 28-10-2026; Disetujui: 21-11-2026

Abstrak: Malaria merupakan penyakit endemik yang diagnosisnya masih bergantung pada pemeriksaan mikroskopis, sehingga memerlukan waktu dan keahlian pengamat. Penelitian ini membandingkan *Progressive Fine-Tuning* (PFT) dengan *feature extraction* (*backbone* dibekukan) dan *full fine-tuning* langsung pada ResNet50, EfficientNetV2-S, dan MobileNetV3-S untuk klasifikasi citra sel darah malaria. *Dataset* NIH Malaria (27.558 citra) dibagi per *slide*/pasien (StratifiedGroupKFold, *seed* 42) menjadi 20.214 citra latih (140 grup), 5.809 validasi (41 grup), dan 1.535 uji (20 grup; 170 *Parasitized* dan 1.365 *Uninfected*). Evaluasi mencakup sensitivitas, spesifisitas, *balanced accuracy*, AUC, jumlah *False Negative* (FN), waktu inferensi, dan Grad-CAM. Dibandingkan *feature extraction*, PFT meningkatkan akurasi uji dari 95,5 hingga 96,4% menjadi 97,3 hingga 98,2% (McNemar $p \leq 0{,}002$ pada ketiga model), terutama dengan menurunkan *False Positive* dari 49 hingga 50 menjadi 20 hingga 28. Sensitivitas tidak meningkat secara konsisten (88,2 hingga 95,3%). PFT tidak berbeda signifikan dari *full fine-tuning* langsung ($p = 0{,}50$ hingga 1,00) dan membutuhkan waktu pelatihan 2,4 hingga 3,4 kali lebih lama. MobileNetV3-S dengan PFT memberikan hasil numerik tertinggi, yaitu akurasi 98,2%, sensitivitas 95,3% (8 FN), dan AUC 0,991, dengan 0,94 juta parameter dan latensi 3,8 ms per citra. EfficientNetV2-S (akurasi 97,5%; sensitivitas 93,5%; 11 FN) dan ResNet50 (akurasi 97,3%; sensitivitas 88,2%; 20 FN) memerlukan sekitar 9,6 hingga 9,8 ms per citra. Hasil berasal dari satu *dataset* dan satu *seed*, dengan 170 sampel positif pada data uji, dan perbedaan antararsitektur tidak diuji secara statistik.

Kata kunci: Malaria, *Convolutional Neural Network* (CNN), *Progressive Fine-Tuning*, *Full Fine-Tuning*, Klasifikasi Citra, Grad-CAM.

***Abstract**: Malaria is an endemic disease whose diagnosis still relies on microscopic examination, which requires time and observer expertise. This study compares Progressive Fine-Tuning (PFT) with feature extraction (frozen backbone) and direct full fine-tuning on ResNet50, EfficientNetV2-S, and MobileNetV3-S for malaria blood cell image classification. The NIH Malaria dataset (27,558 images) was split by slide/patient (StratifiedGroupKFold, seed 42) into 20,214 training images (140 groups), 5,809 validation images (41 groups), and 1,535 test images (20 groups; 170 Parasitized and 1,365 Uninfected). Evaluation covered sensitivity, specificity, balanced accuracy, AUC, the number of False Negatives (FN), inference time, and Grad-CAM. Compared with feature extraction, PFT raised test accuracy from 95.5 to 96.4% to 97.3 to 98.2% (McNemar $p \leq 0.002$ for all three models), mainly by reducing False Positives from 49 to 50 down to 20 to 28. Sensitivity did not improve consistently (88.2 to 95.3%). PFT did not differ significantly from direct full fine-tuning ($p = 0.50$ to 1.00) and required 2.4 to 3.4 times longer training time. MobileNetV3-S with PFT achieved the highest numerical results: accuracy of 98.2%, sensitivity of 95.3% (8 FN), and AUC of 0.991, with 0.94 million parameters and a latency of 3.8 ms per image. EfficientNetV2-S (accuracy 97.5%; sensitivity 93.5%; 11 FN) and ResNet50 (accuracy 97.3%; sensitivity 88.2%; 20 FN) required approximately 9.6 to 9.8 ms per image. The results come from a single dataset*

Copyright@2024. LEMBAGA PENGEMBANGAN DAN INOVASI UNDHARI

Vol. 7, No. 2, Desember 2026, 690 - 702

and a single seed, with 170 positive samples in the test set, and differences between architectures were not statistically tested.

Keywords: *Malaria, Convolutional Neural Network (CNN), Progressive Fine-Tuning, Full Fine-Tuning, Image Classification, Grad-CAM.*

1. Pendahuluan

Malaria hingga kini masih menjadi salah satu penyakit menular yang mengancam kesehatan global. Menurut *World Malaria Report* 2025, penyakit ini menyebabkan sekitar 282 juta kasus infeksi dan 610.000 kematian di seluruh dunia [1]. Di Indonesia, malaria tetap menjadi masalah kesehatan masyarakat, terutama di wilayah timur seperti Papua yang menyumbang porsi terbesar beban kasus nasional. Kondisi epidemiologis ini menunjukkan bahwa sistem deteksi dini yang cepat dan akurat diperlukan untuk mendukung target eliminasi malaria global pada tahun 2030 [2]. Pemeriksaan mikroskopis sediaan apusan darah masih menjadi *gold standard* diagnosis malaria. Namun, identifikasi manual ini memakan waktu analisis yang panjang, sangat bergantung pada keahlian visual analis laboratorium, dan rentan terhadap *human error* [3], [4]. *Rapid Diagnostic Test* (RDT) di lapangan juga masih menghadapi kendala berupa sensitivitas dan spesifisitas yang berfluktuasi pada berbagai kondisi klinis [5]. Karena itu, diperlukan pendekatan diagnosis otomatis yang mampu meningkatkan kecepatan, konsistensi, dan akurasi identifikasi parasit malaria [6]. Perkembangan *Artificial Intelligence* (AI), khususnya *Deep Learning*, menawarkan solusi yang menjanjikan untuk mengotomatisasi diagnosis berbasis citra medis. Arsitektur *Convolutional Neural Network* (CNN) mampu mengekstraksi hierarki fitur visual secara mandiri, sehingga menghasilkan performa klasifikasi yang tinggi pada berbagai citra patologi [7], [8], [9]. Di antara arsitektur CNN modern, ResNet50 dikenal memiliki kemampuan ekstraksi fitur yang kuat berkat mekanisme *residual learning* dan telah dilaporkan mencapai akurasi tinggi pada klasifikasi malaria [10], [11]. EfficientNetV2-S menawarkan keseimbangan antara akurasi dan efisiensi pelatihan melalui blok Fused-MBConv, dan juga berkinerja baik dalam analisis citra darah [12], [13], [14]. Dari sisi efisiensi, MobileNetV3-S dirancang untuk perangkat berdaya komputasi rendah (*edge devices*) dengan memanfaatkan *depthwise separable convolution*, sehingga menjaga akurasi klasifikasi dengan jumlah parameter yang sangat sedikit [15], [16], [17]. Perbedaan karakteristik teknis ketiga arsitektur ini membuatnya layak dibandingkan pada citra mikroskopis malaria. Namun, performa adaptasi model CNN pada domain medis tidak hanya ditentukan oleh desain arsitektur, tetapi juga sangat bergantung pada strategi pelatihan (*fine-tuning*) yang diterapkan. *Fine-tuning* konvensional yang memperbarui seluruh lapisan jaringan sekaligus dapat mengganggu stabilitas adaptasi model dan meningkatkan risiko *overfitting* maupun *catastrophic forgetting* [18]. Literatur terdahulu juga menunjukkan bahwa tidak ada satu strategi *fine-tuning* yang selalu memberikan hasil terbaik pada semua varian arsitektur maupun karakteristik *dataset* medis [19]. Karena itu, efektivitas suatu arsitektur perlu dievaluasi bersama strategi pelatihannya untuk memperoleh keseimbangan antara performa prediktif dan kelayakan komputasi.

Sebagai salah satu alternatif, *Progressive Fine-Tuning* (PFT) mengadaptasi model secara bertahap. Pelatihan dimulai pada *classifier head* dengan *backbone* beku, lalu lapisan ekstraksi fitur dibuka secara progresif [20]. Skema ini dimaksudkan untuk menjaga representasi hasil *pre-training* dan mengurangi koreksi bobot yang terlalu besar pada awal adaptasi. PFT telah dilaporkan pada klasifikasi penyakit otak dari citra MRI 3D [20], tetapi keunggulannya dibandingkan *fine-tuning* langsung belum pasti [19]. Penerapannya pada klasifikasi malaria juga belum dibandingkan secara memadai pada berbagai arsitektur CNN modern. Penelitian terdahulu pada diagnosis malaria berbasis citra menggunakan beragam pendekatan, termasuk penjelasan model (*Explainable AI*, XAI) untuk meningkatkan kepercayaan klinis [21]. Tabel 1 merangkum studi yang paling relevan pada *dataset* NIH Malaria. Studi yang paling dekat dengan fokus penelitian ini adalah Taye dkk[22], yang membandingkan *frozen backbone*, *incremental unfreezing*, dan *full fine-tuning* pada VGG19, InceptionV3, dan Xception. Mereka melaporkan akurasi uji 0,86 hingga 0,88 untuk *incremental unfreezing*, lebih rendah daripada 0,94 hingga 0,95 untuk *full fine-tuning*. Dengan demikian, manfaat pelepasan lapisan bertahap belum terbukti dan perlu diuji pada arsitektur yang lebih modern.

Studi lain tidak membandingkan strategi pelepasan lapisan. Yebasse dkk. [23] mencapai akurasi 97,2% dibandingkan 94,49% pada *baseline* melalui strategi pelatihan yang menyorot area terinfeksi. Pamungkas dan Eljatin [24] menyetel hiperparameter pada keluarga EfficientNet B0 sampai B7 dan memperoleh akurasi terbaik 97,69% pada B5. Studi [25] membandingkan CNN kustom dengan EfficientNet-B3 yang di-*fine-tune* penuh (akurasi validasi 97,41%; AUC 0,9964). Laghari dkk. [26] membandingkan delapan CNN pralatih (VGG16, VGG19, Inception-v3, ResNet-18/34/50/101, dan Xception) yang di-*fine-tune* selama 10 *epoch* dengan data dikelompokkan per ID pasien, dan melaporkan akurasi tertinggi sekitar 89% (ResNet-50 dan ResNet-101).

Tabel 1. Perbandingan penelitian terdahulu pada klasifikasi malaria dataset NIH

Studi	Arsitektur	Strategi adaptasi	Hasil utama yang dilaporkan
Taye dkk. [22]	VGG19, InceptionV3, Xception	*Frozen*, *incremental unfreezing*, *full fine-tuning*	Akurasi uji: *frozen* 0,83/0,86/0,85; *incremental* 0,88/0,86/0,86; *full* 0,94/0,94/0,95
Yebasse dkk. [23]	Empat model, termasuk ResNet dan MobileNet	Strategi pelatihan penyorotan piksel terinfeksi	Akurasi 97,2% vs 94,49% pada *baseline*
Pamungkas & Eljatin [24]	EfficientNet B0, B1, B3, B5, B7	*Hyperparameter tuning*	Terbaik B5: akurasi 97,69%, *precision* 98,36%, *recall* 97,03%
Mohit dkk. [25]	EfficientNet-B3, SPCNN, CNN kustom	*Full fine-tuning* (300×300) + augmentasi terarah	Akurasi validasi 97,41%, AUC 0,9964
Laghari dkk. [26]	VGG16, VGG19, Inception-v3, ResNet-18/34/50/101, Xception	*Fine-tuning* 10 epoch; data dikelompokkan per ID pasien	Akurasi tertinggi ≈89% (ResNet-50, ResNet-101); AUC 0,919–0,965 (enam model)
Penelitian ini	ResNet50, EfficientNetV2-S, MobileNetV3-S	PFT vs *full fine-tuning* vs *feature extraction*; *split* per slide/pasien	Akurasi PFT 97,3–98,2%; sensitivitas 88,2–95,3%; AUC 0,987–0,992; PFT > *feature extraction* ($p \leq 0{,}002$); PFT ≈ *full fine-tuning* ($p \geq 0{,}50$)

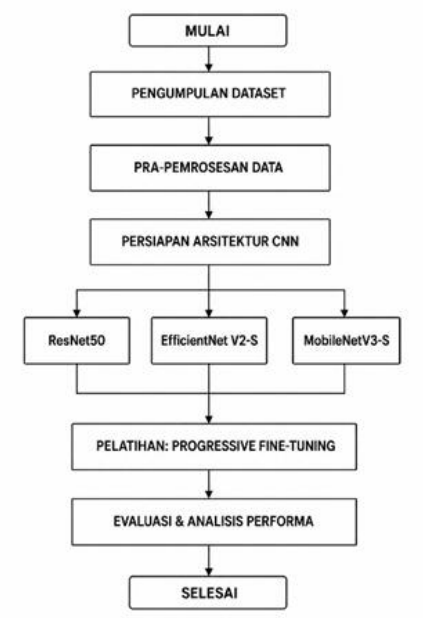
Telaah tersebut menunjukkan tiga hal. Pertama, perbandingan strategi adaptasi pada malaria baru dilaporkan pada arsitektur lama dengan hasil yang tidak mendukung pelepasan bertahap [22]. Kedua, studi lain berfokus pada perbandingan arsitektur, penyetelan hiperparameter, atau strategi pelatihan lain [23]–[26] tanpa membandingkan strategi pelepasan lapisan. Ketiga, akurasi yang dilaporkan [23]–[25] berada di kisaran 97%, sedangkan studi yang secara eksplisit mengelompokkan data per pasien melaporkan akurasi lebih rendah, sekitar 89% [26]; perbedaan protokol ini membuat perbandingan antar-studi tidak setara. Pertanyaan yang belum terjawab adalah strategi adaptasi mana yang memberi hasil setara dengan biaya terendah, serta bagaimana dampaknya pada kesalahan klinis (False Negative) dan interpretabilitas model.

Tabel 1 menunjukkan dua kesenjangan penelitian. Pertama, belum ada perbandingan terkontrol antara PFT, full fine-tuning langsung, dan feature extraction pada ResNet50, EfficientNetV2-S, dan MobileNetV3-S dengan protokol data, seed, dan callback yang sama untuk klasifikasi malaria. Kedua, studi yang melaporkan akurasi tinggi [24]–[26] tidak menyandingkan akurasi itu dengan jumlah False Negative, biaya komputasi, dan penjelasan visual dalam satu kerangka evaluasi. Evaluasi pada sel dari pasien yang sama dengan data latih juga dapat menggelembungkan performa model. Kebaruan penelitian ini ada pada perbandingan terkontrol tersebut. Akurasi studi-studi di atas berada pada kisaran yang sama dengan hasil penelitian ini, sehingga penelitian ini tidak mengklaim kebaruan pada peningkatan akurasi absolut. Untuk menjawab celah tersebut, penelitian ini membandingkan tiga strategi adaptasi, yaitu feature extraction (backbone dibekukan), full fine-tuning langsung, dan PFT, pada ResNet50, EfficientNetV2-S, dan MobileNetV3-S untuk klasifikasi citra sel darah malaria. Evaluasi memakai akurasi, sensitivitas, spesifisitas, balanced accuracy, F1-score, Area Under the Curve (AUC), jumlah False Negative, dan waktu inferensi. Uji McNemar membandingkan strategi pada data uji yang sama. Visualisasi Explainable AI berbasis Grad-CAM [22] memeriksa rasionalitas keputusan model. Penulis membagi data per slide/pasien untuk mencegah kebocoran antar-pasien. Penelitian ini menyediakan perbandingan terkontrol antara PFT, feature extraction, dan full fine-tuning langsung pada tiga arsitektur CNN modern. Evaluasinya menyatukan sensitivitas,

spesifisitas, jumlah False Negative, waktu inferensi, dan Grad-CAM dalam satu kerangka. Pembagian data per slide/pasien serta pencantuman konfigurasi dan lapisan yang dibuka pada setiap tahap memungkinkan pembaca mereproduksi eksperimen. Penulis berharap temuan ini membantu pemilihan strategi adaptasi dan arsitektur CNN untuk sistem skrining malaria berbasis AI di fasilitas layanan kesehatan dengan sumber daya terbatas.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimental kuantitatif untuk membandingkan kinerja ResNet50, EfficientNetV2-S, dan MobileNetV3-S pada klasifikasi citra mikroskopis malaria. Evaluasi dilakukan terhadap penerapan *Progressive Fine-Tuning* (PFT) dan interpretabilitas model menggunakan Grad-CAM untuk menentukan arsitektur yang paling optimal berdasarkan akurasi, efisiensi komputasi, dan transparansi keputusan.



Gambar 1 Diagram Alur Penelitian

2.1 Pengumpulan Dataset

Penelitian ini menggunakan dataset citra apusan darah tipis (*thin blood smears*) yang mencakup dua kondisi, yaitu sel yang terinfeksi malaria dan sel normal (tidak terinfeksi). Dataset ini diperoleh melalui platform repositori publik Kaggle, yang data aslinya dikurasi oleh National Institutes of Health (NIH). Secara keseluruhan, dataset ini memuat 27.558 citra sel yang telah dilabeli dan disegmentasi dari sediaan apus darah tipis dengan pewarnaan Giemsa. Jumlah data tersebut terdistribusi secara seimbang (*balanced dataset*), yang terdiri dari 13.779 citra sel terinfeksi parasit dan 13.779 citra sel normal.

2.2 Pra-Pemrosesan Data

Untuk menyeragamkan format komputasi dan meningkatkan generalisasi model, resolusi citra disesuaikan (*resizing*) menjadi 224 × 224 piksel agar kompatibel dengan standar masukan *backbone* jaringan. Guna mencegah *overfitting*, augmentasi data dinamis (rotasi acak, pergeseran, serta flip horizontal dan vertikal) diterapkan secara eksklusif pada himpunan latih untuk memperkaya variasi visual. Pembagian data dilakukan pada tingkat slide/pasien untuk mencegah kebocoran antar-pasien. ID grup diambil dari awalan nama berkas sebelum "_IMG" (mis. C100P61ThinF) dan diasumsikan merepresentasikan satu slide/pasien, menghasilkan 201 grup (65–702 citra per grup). Dengan *StratifiedGroupKFold* (*shuffle*, *seed* 42), satu dari 10 *fold* luar dijadikan data uji, dan dari sisanya 2 dari 9 *fold* dalam dijadikan data validasi. Hasilnya 20.214 citra latih (140 grup; 10.626 Parasitized, 9.588 Uninfected), 5.809 validasi (41 grup; 2.983 dan 2.826), dan 1.535 uji (20 grup; 170 dan 1.365), atau 73,4%, 21,1%, dan 5,6% dari seluruh citra.

Karena pembagian dilakukan per grup, proporsi citra tidak dapat diatur tepat, dan proporsi Parasitized berbeda antar-subset (52,6% latih, 51,4% validasi, 11,1% uji). Program memverifikasi bahwa tidak ada grup yang muncul di lebih dari satu subset. Karena itu, akurasi dilengkapi sensitivitas, spesifisitas, dan *balanced accuracy*.

2.3 Persiapan Arsitektur CNN

Penelitian ini mengimplementasikan tiga arsitektur, yaitu ResNet50, EfficientNetV2-S, dan MobileNetV3-S, dengan bobot prapelatihan (pre-trained weights) dari dataset ImageNet. Bobot ini memberi kemampuan ekstraksi fitur visual dasar sejak awal pelatihan. Penulis membuang lapisan klasifikasi bawaan setiap model dan menggantinya dengan kepala klasifikasi (classifier head) untuk membedakan sel darah terinfeksi parasit dari sel darah normal.

Kepala klasifikasi terdiri atas Global Average Pooling, Dropout 0,35, dan satu unit Dense dengan aktivasi Sigmoid, tanpa lapisan Dense tersembunyi. Global Average Pooling meringkas setiap peta fitur spasial menjadi satu nilai, dan Dropout mengurangi risiko overfitting. Unit Dense dengan Sigmoid menghasilkan probabilitas kelas Parasitized, yang menentukan label sel terinfeksi atau normal. Kepala ini menambah 2.049 (ResNet50), 1.281 (EfficientNetV2-S), dan 577 (MobileNetV3-S) parameter, sehingga total parameter model menjadi 23.589.761, 20.332.641, dan 939.697. Penulis memanggil backbone dengan training=False pada seluruh tahap, sehingga statistik Batch Normalization tidak diperbarui. Parameter skala dan offset (γ, β) tetap ikut dilatih pada lapisan yang dibuka.

2.4 Skenario dengan Progressive Fine-Tuning

PFT berjalan dalam tiga tahap: T0 (*backbone* beku, LR 1×10^{-3}), T1 (*stage* atas dibuka, LR 1×10^{-4}), dan T2 (seluruh lapisan dibuka, LR 1×10^{-5}). Pembanding *feature extraction* adalah hasil T0 dari *run* PFT yang sama (data, *seed*, dan *callback* identik), dan *full fine-tuning* langsung membuka seluruh lapisan sejak awal dengan LR 1×10^{-4}. Semua konfigurasi memakai Adam Optimizer, *Binary Cross-Entropy*, *batch* 32, maksimal 50 *epoch* per tahap, *EarlyStopping* (val_loss, *patience* 5, restore_best_weights), *ReduceLROnPlateau* (faktor 0,5, *patience* 2, LR minimum 1×10^{-7}), dan *ModelCheckpoint* (val_accuracy). Rincian lapisan yang dibuka pada Tabel 2.

Tabel 2. Lapisan dan parameter terlatih per tahap

Model	Lapisan terbuka pada T1	Param. T0	Param. T1	Param. T2
ResNet50	conv5_block1–3 (32/175 lapisan)	2.049 (0,01%)	14.978.049 (63,49%)	23.536.641 (99,77%)
EfficientNetV2-S	block5, block6, top (359/513)	1.281 (0,01%)	18.357.193 (90,28%)	20.178.769 (99,24%)
MobileNetV3-S	expanded_conv_8–10, conv_1 (38/157)	577 (0,06%)	737.065 (78,44%)	927.585 (98,71%)

2.5 Lingkungan Pengembangan

Seluruh eksperimen dijalankan pada Kaggle Notebook dengan dua GPU NVIDIA Tesla T4 (15 GB) yang tersedia, tetapi kode tidak memakai strategi multi-GPU sehingga pelatihan berjalan pada satu GPU. Lingkungan lainnya: CPU Intel Xeon 2,00 GHz (4 *core*), RAM 33,7 GB, Python 3.13.16, TensorFlow 2.21.0, Keras 3.13.2, NumPy 2.1.3, pandas 2.3.3, scikit-learn 1.6.1, CUDA 12.5.1, cuDNN 9. *Seed* 42 digunakan untuk pembagian data, pengacakan data latih, augmentasi, dan inisialisasi. Pembagian dikunci dalam split.csv dan diverifikasi dengan *hash* data uji (ae724ae9ba0c). Penelitian ini memakai satu *seed* sehingga tidak ada simpang baku antar-pelatihan.

2.6 Evaluasi dan Analisis Performa

Evaluasi memakai data uji yang tidak dilibatkan selama pelatihan (1.535 citra dari 20 grup). Penulis memantau stabilitas pelatihan melalui learning curve accuracy dan loss pada data latih dan validasi di setiap epoch. Kinerja pada data uji dihitung dari Confusion Matrix dengan Parasitized sebagai kelas positif, berupa True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Metrik yang dilaporkan adalah:

Vol. 7, No. 2, Desember 2026, 690 - 702

Accuracy = (TP + TN) / (TP + TN + FP + FN) (1)

Recall (Sensitivity) = TP / (TP + FN) (2)

Specificity = TN / (TN + FP) (3)

Balanced Accuracy = (Recall + Specificity) / 2 (4)

Precision = TP / (TP + FP) (5)

NPV = TN / (TN + FN) (6)

F1-score = 2 × (Precision × Recall) / (Precision + Recall) (7)

AUC adalah luas di bawah kurva ROC, yang memplot Recall terhadap False Positive Rate (FPR = 1 − Specificity) pada seluruh ambang probabilitas Sigmoid. Untuk metrik berbentuk proporsi (Accuracy, Recall, Specificity, Precision, NPV), penulis melaporkan Wilson confidence interval 95%, dengan x sebagai jumlah kejadian benar, n sebagai jumlah kasus yang relevan, $\hat{p} = x/n$, dan $z = 1{,}96$:

$$CI = [\ \hat{p} + z^2/(2n) \pm z \cdot \sqrt{(\ \hat{p}(1 - \hat{p})/n + z^2/(4n^2)\ )}\ ] / (1 + z^2/n) \quad (8)$$

Dua varian model dibandingkan pada data uji yang sama dengan uji McNemar eksak. Dengan b sebagai jumlah citra yang benar diklasifikasikan varian A tetapi salah oleh varian B, c sebagai kebalikannya, dan hipotesis nol bahwa pasangan diskordan sama mungkin berasal dari kedua arah, nilai p dua sisi adalah:

$$p = \min(\ 1,\ 2 \times \Sigma\ (i = 0 \text{ sampai } \min(b, c))\ C(b + c, i) \times 0{,}5\text{^}(b + c)\ ) \quad (9)$$

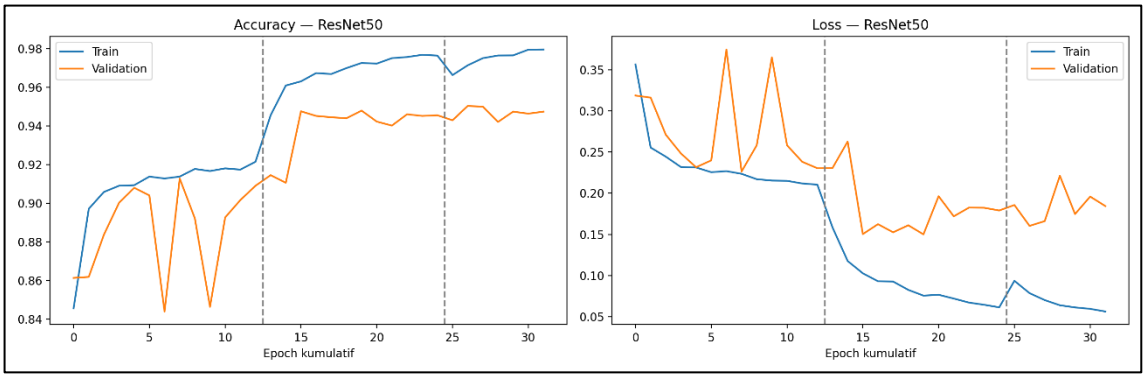
Grad-CAM memvisualisasikan area citra yang mendasari keputusan model. Skor yang diturunkan adalah logit kelas prediksi, sebelum aktivasi Sigmoid. Penulis menghitung gradien skor terhadap peta fitur pada lapisan sebelum Global Average Pooling, lalu merata-ratakannya pada seluruh posisi spasial untuk memperoleh bobot setiap kanal. Peta fitur dijumlahkan dengan bobot tersebut, diteruskan melalui ReLU, dan dinormalisasi min-max per citra ke rentang 0 sampai 1.

3. Hasil dan Pembahasan

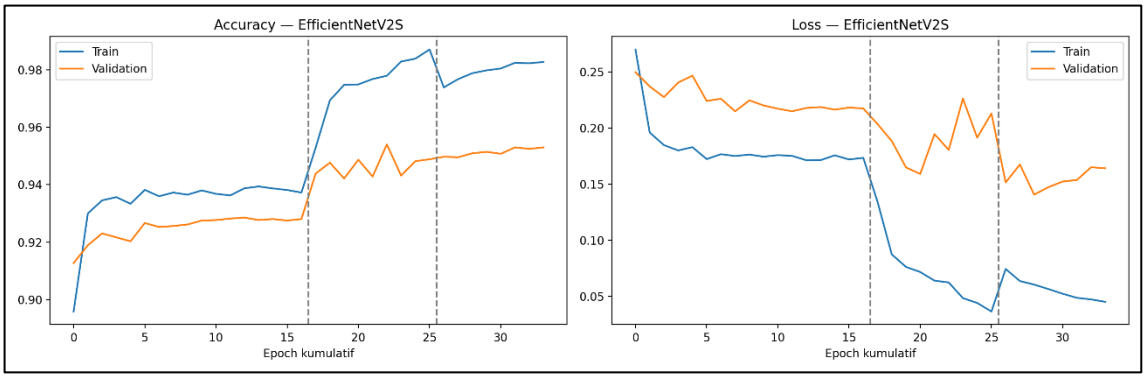
3.1 Dinamika Pelatihan dan *Progressive Fine-Tuning*

Seluruh model dilatih dengan *seed* 42 pada 20.214 citra latih dan dipantau pada 5.809 citra validasi. Gambar 2–4 menampilkan kurva akurasi dan *loss* pada ketiga tahap PFT; garis putus-putus menandai pergantian tahap, dan bobot yang dievaluasi pada akhir tiap tahap adalah bobot dengan *loss* validasi terendah pada tahap tersebut (restore_best_weights).

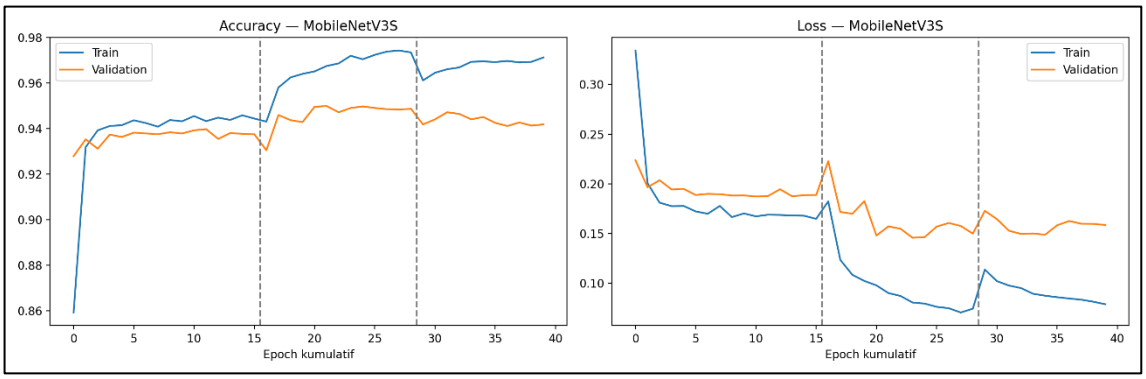
Pada Tahap 0 (*backbone* beku), EfficientNetV2-S dan MobileNetV3-S beradaptasi dengan mulus: akurasi validasi pada *epoch* pertama sekitar 91% dan 93%, dan sekitar 93% dan 94% pada akhir tahap. ResNet50 berbeda. Akurasi validasi *epoch* pertamanya sekitar 86%, dan kurva validasinya berfluktuasi besar walau seluruh *backbone* dibekukan: *loss* validasi melonjak hingga sekitar 0,37 dan akurasi turun ke sekitar 84–87% pada *epoch* ke-7 dan ke-10, sebelum mencapai sekitar 91% pada akhir tahap.



Gambar 2 Kurva akurasi dan loss ResNet50 selama PFT.



Gambar 3 Kurva akurasi dan loss EfficientNetV2-S selama PFT.



Gambar 4 Kurva akurasi dan loss MobileNetV3s selama PFT.

Pembukaan lapisan pada Tahap 1 menaikkan akurasi validasi sekitar 1–4 poin persentase, yaitu ResNet50 dari sekitar 91% menjadi 95%, EfficientNetV2-S dari 93% menjadi 95%, dan MobileNetV3-S dari 94% menjadi 95%. Tahap 2 hampir tidak menambahnya: kurva ResNet50 dan EfficientNetV2-S relatif datar, sedangkan MobileNetV3-S turun sekitar 0,6 poin. Setiap pembukaan lapisan juga diikuti kenaikan sesaat pada *loss* latih di awal tahap berikutnya, yang terlihat jelas pada awal Tahap 2.

Setelah lapisan dibuka, *loss* latih turun jauh lebih cepat daripada *loss* validasi. Pada akhir pelatihan, akurasi latih sekitar 98% (ResNet50), 98% (EfficientNetV2-S), dan 97% (MobileNetV3-S), berbanding akurasi validasi sekitar 95%, 95%, dan 94%. *Loss* validasi tertahan pada kisaran 0,15–0,20, sementara *loss* latih turun ke 0,05–0,08. Pola ini menunjukkan adanya celah generalisasi, sehingga kurva pelatihan tidak dapat dikatakan bebas dari *overfitting*.

Akurasi validasi (94–95%) lebih rendah daripada akurasi uji (97,3–98,2%) karena data validasi hampir seimbang (51,4% Parasitized) sedangkan data uji didominasi sel Uninfected (88,9%). Oleh karena itu, kedua akurasi tersebut tidak setara untuk dibandingkan secara langsung.

Tabel 3 Jumlah epoch (waktu dalam menit) per tahap PFT

Model	T0	T1	T2	Total PFT	*Full FT* langsung
ResNet50	13 (31,9)	12 (25,0)	7 (27,0)	32 (83,9)	9 (33,9)
EfficientNetV2-S	17 (35,0)	9 (27,5)	8 (33,6)	34 (96,1)	10 (40,4)
MobileNetV3-S	16 (27,6)	13 (22,9)	11 (20,5)	40 (71,0)	12 (21,2)

3.2 Evaluasi Kinerja Klasifikasi

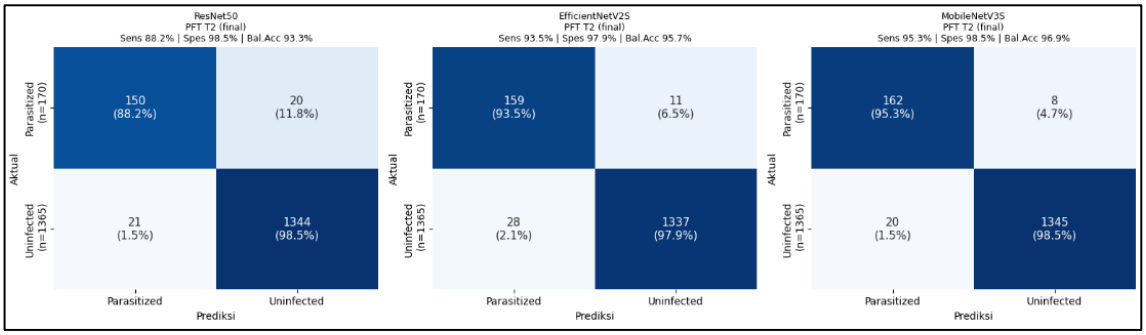
Evaluasi dilakukan pada 1.535 citra uji dari 20 slide/pasien yang terpisah sepenuhnya dari data latih dan validasi, terdiri atas 170 Parasitized dan 1.365 Uninfected (prevalensi Parasitized 11,1%). Karena prevalensi yang rendah, akurasi dilengkapi dengan sensitivitas (*recall* kelas Parasitized), spesifisitas, *balanced accuracy*, AUC, dan jumlah *False Negative* (FN) dan *False Positive* (FP). Tabel 4 merangkum empat varian per model: *feature extraction* (hasil Tahap 0), PFT Tahap 1, PFT Tahap 2 (hasil akhir PFT), dan *full fine-tuning* langsung.

Tabel 4. Kinerja pada data uji (n = 1.535; 170 Parasitized)

Model	Varian	Akurasi	Sensitivitas (CI 95%)	Spesifisitas	Bal. Acc.	AUC	FN	FP
ResNet50	Frozen	95,50	88,82 (83,2–92,7)	96,34	92,58	0,9712	19	50
	PFT T1	96,94	88,82 (83,2–92,7)	97,95	93,39	0,9839	19	28
	PFT T2	97,33	88,24 (82,5–92,3)	98,46	93,35	0,9865	20	21
	Full FT	97,00	94,71 (90,2–97,2)	97,29	96,00	0,9852	9	37
EfficientNetV2-S	Frozen	95,96	92,35 (87,4–95,5)	96,41	94,38	0,9803	13	49
	PFT T1	97,13	90,00 (84,6–93,7)	98,02	94,01	0,9835	17	27
	PFT T2	97,46	93,53 (88,8–96,3)	97,95	95,74	0,9917	11	28
	Full FT	97,72	92,35 (87,4–95,5)	98,39	95,37	0,9885	13	22
MobileNetV3-S	Frozen	96,42	96,47 (92,5–98,4)	96,41	96,44	0,9810	6	49
	PFT T1	97,72	94,71 (90,2–97,2)	98,10	96,40	0,9881	9	26
	PFT T2	98,18	95,29 (91,0–97,6)	98,53	96,91	0,9913	8	20
	Full FT	98,11	92,94 (88,1–95,9)	98,75	95,85	0,9909	12	17

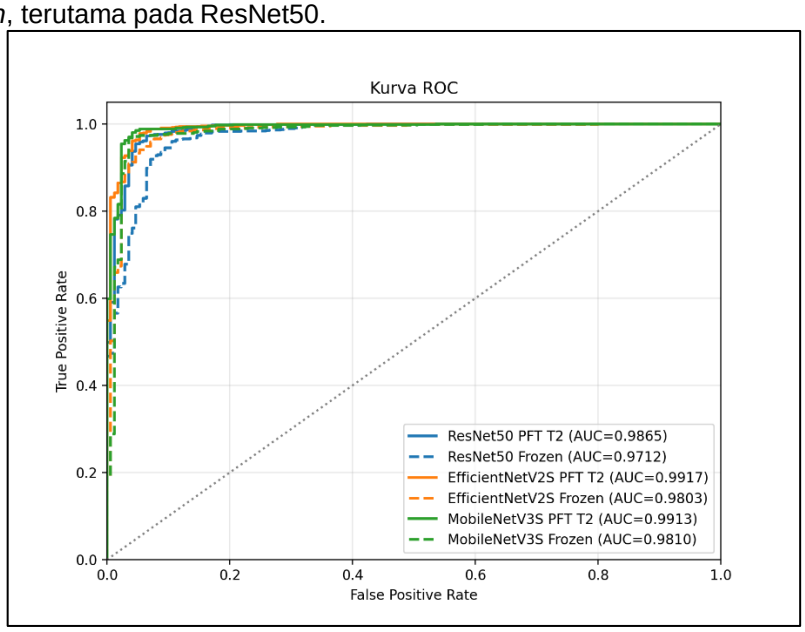
Setelah PFT (T2), MobileNetV3-S mencapai akurasi 98,18% dengan sensitivitas 95,29% (8 FN, 4,7% dari 170 sel terinfeksi) dan spesifisitas 98,53%. EfficientNetV2-S mencapai 97,46% dengan sensitivitas 93,53% (11 FN) dan spesifisitas 97,95%, sedangkan ResNet50 mencapai 97,33% dengan sensitivitas 88,24% (20 FN, 11,8%) dan spesifisitas 98,46%. Presisi kelas Parasitized pada T2 adalah 89,0% (MobileNetV3-S), 85,0% (EfficientNetV2-S), dan 87,7% (ResNet50), dengan F1-score kelas Parasitized 92,1%, 89,1%, dan 88,0%. Interval kepercayaan sensitivitas ketiga model saling tumpang tindih (MobileNetV3-S 91,0–97,6%, EfficientNetV2-S 88,8–96,3%, ResNet50 82,5–92,3%), sehingga peringkat antar-arsitektur tidak dapat dinyatakan signifikan secara statistik; uji antar-arsitektur tidak dilakukan.

Gambar 5 menampilkan *confusion matrix* PFT T2 untuk ketiga model; jumlah FN dan FP varian lainnya dirangkum pada Tabel 5. Warna dinormalisasi per baris agar kelas minoritas (Parasitized) tetap terlihat, sedangkan angka di dalam sel menunjukkan jumlah dan persentase.



Gambar 5 Confusion Matrix pada data uji (n = 1.535; 170 Parasitized, 1.365 Uninfected).

Kurva ROC pada Gambar 6 menunjukkan AUC setelah PFT sebesar 0,9917 (EfficientNetV2-S), 0,9913 (MobileNetV3-S), dan 0,9865 (ResNet50), dibandingkan 0,9803, 0,9810, dan 0,9712 pada *feature extraction*. Kurva ketiga model berhimpit di sudut kiri atas sehingga perbedaan antar-model sangat kecil; selisih yang terlihat jelas hanya antara PFT dan *feature extraction*, terutama pada ResNet50.



Gambar 6 Kurva ROC PFT T2 (garis tegas) dan feature extraction (garis putus-putus) pada data uji.

Perbandingan PFT dengan strategi pembanding. Tabel 5 menyajikan uji McNemar eksak antara PFT T2 dan setiap pembanding pada data uji yang sama. Karena sembilan perbandingan dilakukan, ambang signifikansi dikoreksi menurut Bonferroni ($\alpha = 0{,}05/9 \approx 0{,}0056$).

Tabel 5 Uji McNemar eksak, PFT T2 terhadap pembanding

Model	Pembanding	Benar hanya pembanding	Benar hanya PFT T2	p
ResNet50	Frozen	21	49	0,0011
	PFT T1	8	14	0,2863
	Full FT	18	23	0,5327
EfficientNetV2-S	Frozen	15	38	0,0022
	PFT T1	11	16	0,4421
	Full FT	12	8	0,5034
MobileNetV3-S	Frozen	4	31	<0,0001
	PFT T1	3	10	0,0923
	Full FT	10	11	1,0000

PFT T2 secara signifikan lebih baik daripada *feature extraction* pada ketiga model ($p \leq 0{,}0022$, melewati koreksi Bonferroni). Peningkatan tersebut terutama berasal dari berkurangnya FP (50 menjadi 21 pada ResNet50, 49 menjadi 28 pada EfficientNetV2-S, dan 49 menjadi 20 pada MobileNetV3-S), bukan dari peningkatan sensitivitas: sensitivitas berubah dari 88,82% menjadi 88,24% (ResNet50), 92,35% menjadi 93,53% (EfficientNetV2-S), dan 96,47% menjadi 95,29% (MobileNetV3-S). Model MobileNetV3-S dengan *feature extraction* bahkan mencatat FN terendah (6), dengan harga 49 FP. PFT T2 tidak berbeda signifikan dari PFT T1 maupun dari *full fine-tuning* langsung pada ketiga model.

Full fine-tuning langsung mencapai akurasi 97,00% (ResNet50), 97,72% (EfficientNetV2-S), dan 98,11% (MobileNetV3-S). Pada ResNet50 *full fine-tuning* memberikan sensitivitas yang lebih tinggi (94,71%; 9 FN) daripada PFT T2 (88,24%; 20 FN), dengan FP yang lebih banyak (37 vs 21). Uji McNemar menilai ketepatan keseluruhan, yang didominasi 1.365 sel Uninfected, sehingga perbedaan sensitivitas seperti pada ResNet50 (161 vs 150 dari 170 sel terinfeksi) belum diuji secara terpisah.

3.3 Kompleksitas Komputasi, Waktu Inferensi, dan Biaya Pelatihan

Waktu inferensi diukur pada GPU NVIDIA Tesla T4 (Kaggle Notebook) dengan predict_on_batch, sehingga mencakup transfer data CPU–GPU. Pada *batch* 1, pengukuran dilakukan setelah 20 iterasi pemanasan dan dirata-rata dari 200 pengukuran; pada *batch* 32, waktu dinyatakan per citra. Karena arsitekturnya sama, latensi *full fine-tuning* hampir identik dengan PFT (9,73; 9,59; dan 3,75 ms pada *batch* 1).

Tabel 6 Parameter, latensi, dan waktu pelatihan

Model	Parameter total	Latensi *batch* 1 (ms)	ms/citra (*batch* 32)	Rasio waktu latih PFT/*Full FT*
ResNet50	23.589.761	9,58 ± 0,23	2,922	2,47×
EfficientNetV2-S	20.332.641	9,80 ± 0,67	2,911	2,38×
MobileNetV3-S	939.697	3,79 ± 0,15	0,635	3,35×

MobileNetV3-S memiliki parameter sekitar 25 kali lebih sedikit daripada ResNet50 dan 22 kali lebih sedikit daripada EfficientNetV2-S. Latensinya 2,5–2,6 kali lebih rendah pada *batch* 1 dan 4,6 kali lebih rendah pada *batch* 32. Dari sisi biaya pelatihan, PFT memerlukan 71,0–96,1 menit, sedangkan *full fine-tuning* langsung hanya 21,2–40,4 menit, atau 2,4–3,4 kali lebih singkat.

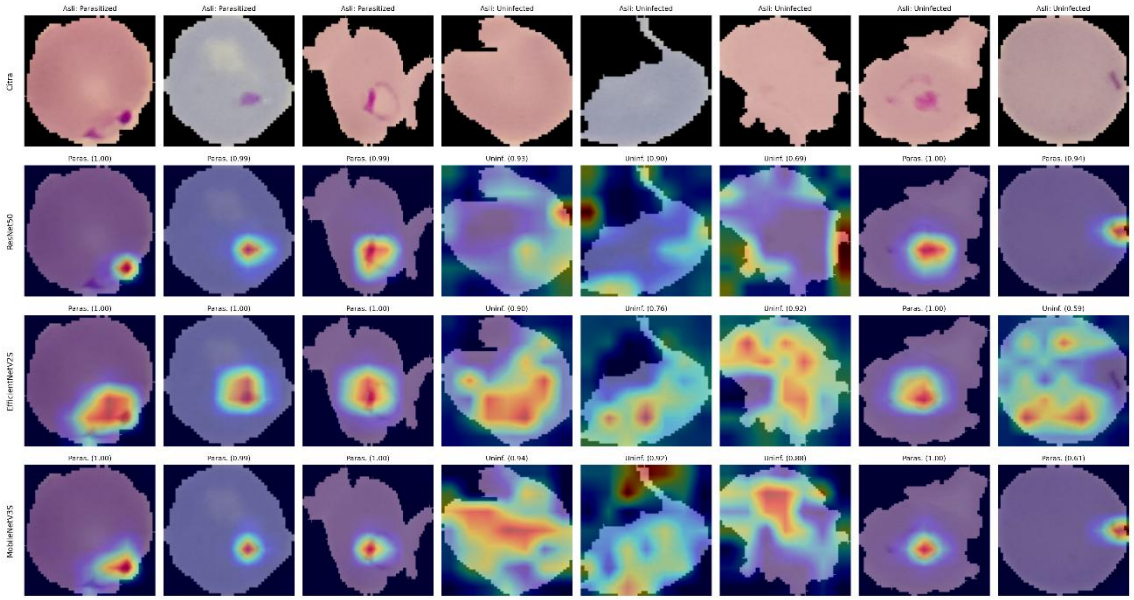
Tabel 7 membandingkan perilaku kurva *loss* validasi pada PFT dan *full fine-tuning*. Pola yang muncul tidak seragam antar-model: pada MobileNetV3-S, PFT lebih rendah pada kedua metrik; pada EfficientNetV2-S hanya simpang baku yang lebih rendah (lonjakan justru lebih tinggi); dan pada ResNet50 PFT lebih buruk pada keduanya. Selain itu, lonjakan pada PFT mencakup transisi antar-tahap saat lapisan dibuka, sehingga tidak setara dengan satu tahap *full fine-tuning*. Data ini tidak mendukung kesimpulan bahwa PFT menghasilkan pelatihan yang lebih stabil.

Tabel 7 Metrik kurva loss validasi: PFT vs full fine-tuning langsung.

Model	Strategi	Lonjakan maks. *loss* validasi	Simpang baku *loss* validasi, 5 *epoch* terakhir
ResNet50	PFT	0,1347	0,0192
	Full FT	0,0282	0,0081
EfficientNetV2-S	PFT	0,0457	0,0069
	Full FT	0,0364	0,0191
MobileNetV3-S	PFT	0,0340	0,0015
	Full FT	0,0691	0,0251

3.4 Interpretabilitas Model dengan Grad-CAM

Grad-CAM dihitung dari gradien logit kelas prediksi (sebelum fungsi Sigmoid) terhadap peta fitur terakhir sebelum *Global Average Pooling*, kemudian diproses dengan ReLU, dinormalisasi min–maks per citra, dan ditampilkan sebagai *overlay*. Gambar 7 memuat delapan citra uji yang sama untuk ketiga model PFT T2: tiga citra Parasitized dan tiga citra Uninfected yang diklasifikasikan benar oleh ketiga model, serta dua kasus sulit yang salah diklasifikasikan oleh sedikitnya satu model.



Gambar 7. Grad-CAM PFT T2 pada delapan citra uji. Baris: citra asli, ResNet50, EfficientNetV2-S, MobileNetV3-S. Label di atas peta menunjukkan kelas prediksi (keyakinan).

Pada citra Parasitized yang benar, ketiga model menghasilkan satu area aktivasi terfokus yang berimpit dengan struktur ungu parasit (keyakinan 0,99–1,00). Fokus ResNet50 dan MobileNetV3-S tampak lebih sempit, sedangkan EfficientNetV2-S menghasilkan area yang lebih lebar namun tetap berpusat pada struktur parasit. Pada citra Uninfected yang benar, aktivasi tampak menyebar di seluruh sel dan tepi tanpa satu pusat fokus; pada ResNet50, sebagian aktivasi terkonsentrasi di tepi sel.

Pada dua kasus sulit (keduanya berlabel Uninfected), pada kasus pertama ketiga model memprediksi Parasitized dengan keyakinan 1,00 dan aktivasi terpusat pada struktur ungu berbentuk cincin yang menyerupai parasit. Pada kasus kedua, ResNet50 (0,94) dan MobileNetV3-S (0,61) memprediksi Parasitized dengan aktivasi pada bintik ungu kecil di tepi sel, sedangkan EfficientNetV2-S memprediksi Uninfected (0,59) dengan aktivasi menyebar. Kesalahan pada kasus-kasus ini tampak terkait dengan struktur ungu yang menyerupai parasit, bukan dengan fokus pada latar belakang. Apakah citra tersebut merupakan galat label atau artefak pewarnaan memerlukan telaah ahli. Pengamatan ini bersifat kualitatif pada delapan citra tanpa metrik lokalisasi kuantitatif.

3. Kesimpulan

PFT meningkatkan akurasi dibandingkan *feature extraction* pada ResNet50, EfficientNetV2-S, dan MobileNetV3-S (dari 95,5–96,4% menjadi 97,3–98,2%; $p \leq 0{,}0022$), terutama dengan menurunkan *False Positive*, sedangkan sensitivitas tidak meningkat konsisten. PFT tidak berbeda signifikan dari *full fine-tuning* langsung ($p = 0{,}50$–$1{,}00$) dengan waktu pelatihan 2,4–3,4 kali lebih lama, dan klaim stabilitas pelatihan tidak didukung data. MobileNetV3-S memberikan hasil numerik tertinggi (akurasi 98,18%, sensitivitas 95,29%, 8 *False Negative*) dengan 0,94 juta parameter dan latensi 3,79 ms, meskipun perbedaan antar-arsitektur tidak diuji secara statistik. Hasil ini berasal dari satu dataset, satu *seed*, dan 170 sampel positif pada data uji, sehingga belum cukup untuk kesimpulan klinis.

Penelitian selanjutnya disarankan memakai beberapa *seed*, *fold* uji yang lebih besar dan seimbang, serta validasi eksternal pada dataset lain. Uji statistik khusus sensitivitas dan antar-arsitektur perlu ditambahkan, begitu juga pemilihan ambang keputusan pada data validasi dan agregasi prediksi pada tingkat pasien. Interpretabilitas dapat diperkuat dengan metrik lokalisasi kuantitatif dan telaah ahli, dan kelayakan MobileNetV3-S pada perangkat tepi perlu diuji.

Referensi

[1] "World malaria report 2025: Addressing the threat of antimalarial drug resistance".
[2] D. N. Aisyah *dkk.*, "The Changing Incidence of Malaria in Indonesia: A 9-Year Analysis of Surveillance Data," *Adv. Public Health*, vol. 2024, no. 1, hlm. 2703477, Jan 2024, doi: 10.1155/adph/2703477.
[3] S. Wantini dan M. Huda, "Pengaruh Konsentrasi dan Waktu Pengecatan Giemsa Pada Pemeriksaan Mikroskopik Malaria," *J. Anal. Kesehat.*, vol. 10, no. 1, hlm. 8, Jun 2021, doi: 10.26630/jak.v10i1.2715.
[4] L. Fransisca *dkk.*, "Enhanced data quality to improve malaria surveillance in Papua, Indonesia," *Malar. J.*, vol. 24, no. 1, hlm. 177, Jun 2025, doi: 10.1186/s12936-025-05358-x.
[5] F. D. Ojeniyi *dkk.*, "Performance and challenges of malaria rapid diagnostic tests in endemic regions of Africa," *Sci. Rep.*, vol. 15, no. 1, hlm. 43663, Des 2025, doi: 10.1038/s41598-025-97259-x.
[6] A. Koirala *dkk.*, "Deep Learning for Real-Time Malaria Parasite Detection and Counting Using YOLO-mp," *IEEE Access*, vol. 10, hlm. 102157–102172, 2022, doi: 10.1109/ACCESS.2022.3208270.
[7] W. Siłka, M. Wieczorek, J. Siłka, dan M. Woźniak, "Malaria Detection Using Advanced Deep Learning Architecture," *Sensors*, vol. 23, no. 3, hlm. 1501, Jan 2023, doi: 10.3390/s23031501.
[8] K. Hemachandran *dkk.*, "Performance Analysis of Deep Learning Algorithms in Diagnosis of Malaria Disease," *Diagnostics*, vol. 13, no. 3, hlm. 534, Feb 2023, doi: 10.3390/diagnostics13030534.
[9] D. Sukumarran *dkk.*, "Automated Identification of Malaria-Infected Cells and Classification of Human Malaria Parasites Using a Two-Stage Deep Learning Technique," *IEEE Access*, vol. 12, hlm. 135746–135763, 2024, doi: 10.1109/ACCESS.2024.3459411.
[10] N. Huda, S. Y. Prayogi, M. A. Ahmad, dan A. Y. Dewi, "Klasifikasi Malaria Menggunakan Metode Image Processing Dari Sel Darah Merah Dengan Algoritma Convolutional Neural Network," *JOINS J. Inf. Syst.*, vol. 7, no. 2, hlm. 166–177, Nov 2022, doi: 10.33633/joins.v7i2.7068.
[11] A. F. Saksenata, A. E. Minarno, dan Y. Azhar, "Klasifikasi Citra Sel Darah Untuk Penyakit Malaria Dengan Metode CNN," *J. Repos.*, vol. 4, no. 2, Feb 2024, doi: 10.22219/repositor.v4i2.31053.
[12] M. Tan dan Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," 2021, doi: 10.48550/ARXIV.2104.00298.
[13] J. G. Purnomo, S. Birowo, dan M. A. Maulana, "Identifikasi Malaria Pada Citra Darah Dengan Convolutional Neural Network," *Bit-Tech*, vol. 7, no. 2, hlm. 406–412, Des 2024, doi: 10.32877/bt.v7i2.1828.
[14] M. Mujahid *dkk.*, "Efficient deep learning-based approach for malaria detection using red blood cell smears," *Sci. Rep.*, vol. 14, no. 1, hlm. 13249, Jun 2024, doi: 10.1038/s41598-024-63831-0.
[15] A. W. Setiawan, "Perbandingan Kinerja Arsitektur Convolutional Neural Network Pada Deteksi Malaria Menggunakan Citra Sel Darah," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 12, no. 3, hlm. 643–652, Jun 2025, doi: 10.25126/jtiik.20258085.
[16] Y. Kumar *dkk.*, "Enhancing parasitic organism detection in microscopy images through deep learning and fine-tuned optimizer," *Sci. Rep.*, vol. 14, no. 1, hlm. 5753, Mar 2024, doi: 10.1038/s41598-024-56323-8.
[17] S. A. Ali, P. S. Abdulqadir, S. A. Abdullah, dan H. Yunusa, "M2ANET: Mobile Malaria Attention Network for efficient classification of plasmodium parasites in blood cells," 2024, *arXiv*. doi: 10.48550/ARXIV.2405.14242.
[18] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, dan T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Med. Imaging*, vol. 22, no. 1, hlm. 69, Des 2022, doi: 10.1186/s12880-022-00793-7.
[19] A. Davila, J. Colan, dan Y. Hasegawa, "Comparison of fine-tuning strategies for transfer learning in medical image classification," *Image Vis. Comput.*, vol. 146, hlm. 105012, Jun 2024, doi: 10.1016/j.imavis.2024.105012.
[20] I. A. M. Dioses, J. T. Dellosa, A. Hernandez, J. L. Dioses, J. La Madrid, dan A. S. P. Dulnuan, "An MSTL with a Progressive Fine-Tuning Framework for Brain Disease Classification using

3D-Rendered MRI Images," dalam *2026 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, Bangkok, Thailand: IEEE, Mar 2026, hlm. 1327–1332. doi: 10.1109/ICMLAS67792.2026.11483866.

[21] R. Parveen *dkk.*, "Trustworthy deep learning for malaria diagnosis using explainable artificial intelligence," *Sci. Rep.*, vol. 15, no. 1, hlm. 45037, Des 2025, doi: 10.1038/s41598-025-28387-7.

[22] A. G. Taye, S. Yemane, E. Negash, Y. Minwuyelet, M. Abebe, dan M. H. Asmare, "Automated Web-Based Malaria Detection System with Machine Learning and Deep Learning Techniques," 2024, *arXiv*. doi: 10.48550/ARXIV.2407.00120.

[23] M. Yebasse, K. J. Cheoi, dan J. Ko, "Malaria Disease Cell Classification With Highlighting Small Infected Regions," *IEEE Access*, vol. 11, hlm. 15945–15953, 2023, doi: 10.1109/ACCESS.2023.3245025.

[24] Y. Pamungkas dan D. S. Eljatin, "Hyperparameter Tuning of EfficientNet Method for Optimization of Malaria Detection System Based on Red Blood Cell Image," *J. Sisfokom Sist. Inf. Dan Komput.*, vol. 13, no. 3, hlm. 360–368, Nov 2024, doi: 10.32736/sisfokom.v13i3.2257.

[25] R. Mohit, E. Milani, A. Amini, M. Ahani, E. Nazari, dan T. Aldaghi, "Comparative Evaluation of Custom Convolutional Neural Networks and EfficientNet-B3 for Malaria Cell Image Classification: Impact of Targeted Data Augmentation on Model Performance," *IJBC*, vol. 17, no. 3, hlm. 62 EP – 72, Okt 2025, doi: 10.61882/ijbc.17.3.62.

[26] A. A. Laghari, W. Muhammad, M. L. Memon, A. Hussain, dan A. Kumar, "Malaria Parasite Cell Classification Using Transfer Learning with State-of-the-Art CNN Architectures," *Biology*, vol. 14, no. 12, hlm. 1792, Des 2025, doi: 10.3390/biology14121792.