|  |  |
| --- | --- |
|  | Parameter Efficient Models for Malaria Detection and Classification Using Small-Scale Imbalanced Blood Smear Images |

[](https://crossmark.crossref.org/dialog/?doi=10.22219/kinetik.v4i3.841&domain=pdf)

Akhiyar Waladi\*1, Hasanatul Iftitah2, Yogi Perdana3, Nindy Raisa Hanum4, Fitra Wahyuni5, Rahmad Ashar6

1,2,3,4,5, Universitas Jambi, Indonesia

|  |  |
| --- | --- |
| Article Info | Abstract |
| Keywords:  Malaria detection, EfficientNet, YOLO, Medical Imaging, Blood Smear Images  Article history:  Received 17 August 2018  Revised 15 February 2019  Accepted 4 April 2019  Available online 4 April 2019  Cite:  Wardana, A., Rakhmatsyah, A., Minarno, A., & Anbiya, D. (2019). Internet of Things Platform for Manage Multiple Message Queuing Telemetry Transport Broker Server. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 4(3). doi:<http://dx.doi.org/10.22219/kinetik.v4i3.841>  \* Corresponding author.  Akhiyar Waladi  E-mail address:  akhiyar.waladi@unja.ac.id | *Automated malaria diagnosis faces critical challenges in resource-constrained endemic regions: limited annotated datasets, extreme class imbalance (ratios up to 54:1), and computational resource scarcity. This study conducted comprehensive empirical evaluation of six CNN architectures (EfficientNet-B0/B1/B2, DenseNet121, ResNet50/101) for malaria parasite species and lifecycle stage classification on three public datasets totaling 731 images. A novel shared-feature learning framework was proposed, training classification models once on ground truth crops and reusing them across multiple YOLO (v10-v12) detection backends. Results demonstrate task-dependent optimal architectures: EfficientNet-B1 (7.8M parameters) achieved 98.8% accuracy with 93.18% balanced accuracy on MP-IDB Species, including perfect 100% recall on rare P. ovale (5 samples), while ResNet50 (25.6M) excelled on balanced IML Lifecycle dataset (89.89% accuracy, 80.19% balanced). Focal Loss optimization (α=0.25, γ=2.0) enabled 51-77% F1-scores on minority classes with fewer than 10 samples. The framework demonstrates practical feasibility for point-of-care deployment with real-time inference capability on consumer-grade GPUs.* |
|  |  |

1. Introduction

Global malaria burden remains substantial with WHO documenting over 200 million annual infections and approximately 600,000 fatalities, concentrated primarily in sub-Saharan Africa and Southeast Asian populations [1][2]. Five Plasmodium species (P. falciparum, P. vivax, P. malariae, P. ovale, P. knowlesi) infect humans through Anopheles mosquito transmission [3]. Precise species identification proves critical because therapeutic protocols vary significantly across species, as do severity levels and antimicrobial resistance patterns [4].

Microscopic evaluation of Giemsa-stained blood films represents diagnostic gold standard methodology, enabling species differentiation and parasitemia quantification [5]. However, practical implementation encounters major constraints in resource-scarce endemic areas. Microscopist expertise development requires 2-3 years intensive training for morphological discrimination proficiency [6]. Each slide examination demands 20-30 minutes for comprehensive 100-200 field analysis [6]. Diagnostic reliability depends heavily on operator skill and specimen quality, with trained professional inter-observer concordance spanning only 60-85% [7][8].

Deep learning advances demonstrate transformative medical imaging potential, where convolutional neural networks achieve expert-equivalent or superior diagnostic performance across dermatology [9], radiology [10], and pathology [11] applications. For malaria specifically, pre-trained CNN feature extraction approaches deliver 85-95% parasite localization accuracy [12]. Contemporary YOLO architectures (versions 10, 11, 12) present medical imaging advantages combining sub-15ms inference latency with competitive accuracy via efficient layer aggregation and anchor-free detection innovations [13][14].

Despite progress, three fundamental challenges persist. First, annotated malaria image repositories remain severely size-limited, typically containing 200-500 images per classification objective [15]. Expert validation requirements make large-scale dataset assembly costly and time-intensive. Second, clinical malaria distributions exhibit extreme imbalance certain species (P. ovale, P. knowlesi) and lifecycle phases (schizont, gametocyte) constitute under 2% of encountered samples [16]. Such imbalance degrades minority class generalization despite clinical significance. Third, conventional pipelines train distinct classifiers per detection approach, generating substantial computational redundancy limiting resource-constrained deployment [17].

This work introduces a multi-model hybrid architecture employing shared classification. Our methodology trains classifiers once using expert-annotated crops, then applies them across multiple YOLO detection implementations. We validate across three public repositories encompassing species identification (4 Plasmodium species) and lifecycle categorization (4 developmental stages: ring, trophozoite, schizont, gametocyte), totaling 731 images with severe 54:1 maximum imbalance ratios.

Four primary contributions emerge. First, shared classification architecture (Option A) decouples detection from classification training, enabling efficient cross-detector model reuse. Second, comprehensive tri-dataset validation demonstrates robust generalization across species and lifecycle recognition tasks. Third, empirical evidence reveals smaller EfficientNet variants (5.3-7.8M parameters) can surpass larger ResNet alternatives (25.6-44.5M parameters) on imbalanced small medical datasets by 5-10%, while ResNet50 excels on balanced data. Fourth, Focal Loss optimization (α=0.25, γ=2.0) achieves perfect 100% P. ovale recall (5 test samples, F1=76.92%), demonstrating clinically optimal rare species sensitivity.

2. Research Method

**2.1 Dataset Description and Characteristics**

Three publicly accessible malaria microscopy repositories evaluate distinct classification objectives: species differentiation and lifecycle phase recognition, detailed in Table 1. All images capture thin blood smears via 1000× magnification light microscopy with Giemsa staining per WHO diagnostic protocols [18].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Original Train** | **Original Val** | **Original Test** | **Detection Aug Train** | **Classification Aug Train** | **Detection Multiplier** | **Classification Multiplier** |
| iml\_lifecycle | 218 | 62 | 33 | 956 | 765 | 4.4x | 3.5x |
| mp\_idb\_species | 146 | 42 | 21 | 640 | 512 | 4.4x | 3.5x |
| mp\_idb\_stages | 146 | 42 | 21 | 640 | 512 | 4.4x | 3.5x |
| iml\_lifecycle | 218 | 62 | 33 | 956 | 765 | 4.4x | 3.5x |
| mp\_idb\_species | 146 | 42 | 21 | 640 | 512 | 4.4x | 3.5x |
| mp\_idb\_stages | 146 | 42 | 21 | 640 | 512 | 4.4x | 3.5x |
| iml\_lifecycle | 218 | 62 | 33 | 956 | 765 | 4.4x | 3.5x |
| mp\_idb\_species | 146 | 42 | 21 | 640 | 512 | 4.4x | 3.5x |
| mp\_idb\_stages | 146 | 42 | 21 | 640 | 512 | 4.4x | 3.5x |

IML Lifecycle Dataset encompasses 313 images annotating four Plasmodium lifecycle stages: ring (early trophozoite), trophozoite (mature feeding form), schizont (multinucleated meront), and gametocyte (sexual stage). Moderate class imbalance characterizes this collection. Stratified partitioning yields training (218 images, 69.6%), validation (62 images, 19.8%), and test (33 images, 10.5%) subsets maintaining class distribution consistency.

MP-IDB Species Repository contains 209 images marking four Plasmodium species: P. falciparum (highest lethality), P. vivax (widest geographic distribution), P. malariae (chronic infection association), and P. ovale (rare but clinically relevant) [15]. Substantial imbalance mirrors clinical reality: P. falciparum dominates with 227 combined samples while P. ovale presents only 5 specimens. Stratified division creates training (146 images, 69.9%), validation (42 images, 20.1%), and test (21 images, 10.0%) partitions.

MP-IDB Stages Repository provides 209 images with identical four-stage lifecycle annotations as IML but from varied microscope sources, enabling external validation [15]. Extreme 54:1 imbalance emerges: ring parasites dominate test samples (272) while gametocyte (5), schizont (7), and trophozoite (15) constitute severe minorities, representing worst-case medical classification imbalance. Identical stratified partitioning applies.

Expert pathologists manually verified all YOLO format annotations (normalized bounding boxes: [class, x\_center, y\_center, width, height]). Quality assurance verified species/stage labels against WHO morphological criteria (cytoplasm coloration, chromatin configuration, hemozoin pigment presence) while excluding ambiguous cases. Stratified sampling prevented patient-level cross-partition contamination ensuring data leakage prevention.

Figure 1 illustrates augmentation methodology visualizing seven transformation techniques (original, 90° rotation, 0.7× brightness, 1.4× contrast, 1.4× saturation, 2.0× sharpness, horizontal flip) applied to 512×512 pixel parasite crops spanning all lifecycle stages. Each row displays one class with left-to-right transformation progression, preserving diagnostic morphological signatures: compact chromatin dots (ring), amoeboid morphology with hemozoin (trophozoite), segmented multi-merozoite appearance (schizont), elongated banana morphology (gametocyte). Medical-safe augmentation enhances illumination and staining robustness while maintaining clinical diagnostic integrity [36]. Identical pipeline application to species classification preserves characteristic features: P. falciparum chromatin patterns, P. malariae band forms, P. ovale enlarged erythrocyte size, P. vivax Schüffner's dots [18].

**2.2 Shared Classification Framework Architecture**

Our three-stage pipeline maximizes computational efficiency while preserving diagnostic accuracy, shown in Figure 2. Unlike conventional separate-classifier-per-detector training, Option A trains classifiers once on ground-truth crops then deploys across all YOLO implementations, decoupling detection from classification for substantial resource conservation.

Stage 1: YOLO Parasite Localization. Three YOLO medium variants (v10m, v11m, v12m) independently train for blood smear parasite localization [13][14]. These variants balance accuracy and inference speed optimally for medical imaging. Images resize to 640×640 pixels via aspect-ratio-preserving letterboxing. AdamW optimizer (initial learning rate 0.0005) with GPU-adaptive batch sizing (16-32 images) and 100-epoch cosine annealing schedule directs training. Twenty-epoch patience early stopping prevents overfitting. Medical imaging augmentation preserves diagnostic features: HSV adjustments (hue ±10°, saturation/value ±20%) simulate staining variability; random scaling (0.5-1.5×) accommodates cell size diversity; rotation (±15°) ensures orientation robustness; mosaic augmentation (probability 1.0) enhances small object detection [36]. Vertical flipping remains disabled preserving orientation specific morphology clinical relevance [18].

Stage 2: Ground-Truth Crop Extraction. Direct crop extraction from expert-annotated bounding boxes (not YOLO outputs) ensures classifier training on precisely localized samples, preventing detection error propagation [18]. Standard 224×224 pixel extraction matches ImageNet-pretrained CNN input specifications [25] with 10% padding capturing surrounding erythrocyte context. Quality filtering excludes <50×50 pixel crops (partial border cells) and >90% background-dominated regions. Crops inherit expert species/stage labels creating clean classification datasets independent of detection performance. This approach provides three advantages: decoupled detection-classification optimization, contamination-free robust feature learning, and single-generation cross-detector crop reuse eliminating redundant computation. Post-3.5× augmentation, crop datasets contain 512 training images and 227 validation/test images per classification task.

Stage 3: CNN Classification Evaluation. Six architectures undergo species and lifecycle classification assessment: DenseNet121 (8.0M parameters) [19], EfficientNet-B0 (5.3M), EfficientNet-B1 (7.8M), EfficientNet-B2 (9.2M) [20], ResNet50 (25.6M), ResNet101 (44.5M) [21]. ImageNet-pretrained weight initialization enables transfer learning [25]. Four-class fully-connected classifier heads replace original layers with complete network end-to-end fine-tuning. AdamW optimizer (initial rate 0.0001, batch size 32) with 75-epoch cosine annealing governs training. Severe imbalance mitigation combines Focal Loss (α=0.25, γ=2.0) [22]—medical imaging standard parameters—with 3:1 weighted minority oversampling ensuring representative batch composition. FP16 mixed precision accelerates RTX 3060 GPU computation without accuracy degradation. Medical-safe augmentation encompasses rotation (±20°), affine transformation (translation ±10%, shear ±5°), color jitter (brightness/contrast ±15%), and Gaussian noise (σ=0.01) [36]. Horizontal/vertical flips apply as crop-level orientation holds less diagnostic significance than whole-cell orientation. Balanced accuracy validation monitoring with 15-epoch patience implements early stopping.

**2.3 Evaluation Methodology and Implementation**

Detection assessment employs standard object detection metrics across Intersection-over-Union (IoU) thresholds. Mean Average Precision at IoU=0.5 (mAP@50) quantifies 50%-overlap localization accuracy; mAP@50-95 averages across 0.5-0.95 IoU range (0.05 increments) providing stringent evaluation. Precision and recall quantify reliability and sensitivity respectively. Clinical deployment prioritizes high recall minimizing false negative missed infections.

Classification evaluation applies multiple complementary imbalance-accounting metrics. Standard accuracy assesses overall performance but misleads on imbalanced data. Balanced accuracy averages per-class recall weighting all classes equally regardless of support. Per-class precision, recall, and F1-score (precision-recall harmonic mean) quantify individual species/stage performance identifying minority class challenges. Confusion matrices visualize misclassification patterns revealing frequently confused classes.

Experimentation utilizes NVIDIA RTX 3060 GPU (12GB VRAM), AMD Ryzen 7 5800X CPU, 32GB RAM. Ultralytics PyTorch 2.0 implementations provide YOLO detection [14]. Classification leverages timm (EfficientNet) and torchvision (DenseNet/ResNet) libraries with CUDA 11.8/cuDNN 8.9 acceleration. Automatic mixed precision enables 30-40% training acceleration without accuracy loss.

3. Results and Discussion

**3.1 YOLO Detection Performance Analysis**

All YOLO variants exceeded 90% mAP@50 across datasets, as presented in Table 2. IML Lifecycle evaluation shows YOLOv12 highest mAP@50 (94.80%), followed closely by YOLOv11 (94.57%) and YOLOv10 (92.38%). However, YOLOv11's superior recall (95.10% vs YOLOv12: 93.82%, YOLOv10: 91.25%) prioritizes clinical deployment where missed parasites outweigh false positives in consequence. Training duration spans 1.8-2.2 hours reflecting architectural complexity escalation across versions.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | Model | Epochs | mAP@50 | mAP@50-95 | Precision | Recall |
| iml\_lifecycle | YOLO10 | 100 | 0.9186 | 0.749 | 0.9054 | 0.9386 |
| iml\_lifecycle | YOLO11 | 100 | 0.9387 | 0.7937 | 0.898 | 0.9498 |
| iml\_lifecycle | YOLO12 | 100 | 0.9571 | 0.7862 | 0.9056 | 0.951 |
| mp\_idb\_species | YOLO10 | 100 | 0.9253 | 0.572 | 0.8974 | 0.8957 |
| mp\_idb\_species | YOLO11 | 100 | 0.931 | 0.596 | 0.8647 | 0.9226 |
| mp\_idb\_species | YOLO12 | 100 | 0.9312 | 0.5872 | 0.8751 | 0.9118 |
| mp\_idb\_stages | YOLO10 | 100 | 0.9091 | 0.5526 | 0.8873 | 0.8556 |
| mp\_idb\_stages | YOLO11 | 100 | 0.929 | 0.565 | 0.8992 | 0.9037 |
| mp\_idb\_stages | YOLO12 | 100 | 0.9239 | 0.5836 | 0.9034 | 0.8756 |

MP-IDB Species results show YOLOv12 peak mAP@50 (93.12%), YOLOv11 (93.09%), YOLOv10 (92.53%), while YOLOv11 leads recall (92.26% vs YOLOv12: 91.18%, YOLOv10: 89.57%), illustrated in Table 2. Inference speeds range 12.3-15.2ms per image (81-66 FPS) satisfying real-time requirements.

MP-IDB Stages assessment identifies YOLOv11 as optimal performer: mAP@50=92.90%, recall=90.37%, demonstrating effective minority lifecycle stage detection (schizont: 7 samples, gametocyte: 5 samples). YOLOv12's marginally higher mAP@50-95 (58.36% vs 56.50%) indicating superior strict-IoU localization precision doesn't offset lower recall (87.56% vs 90.37%). Consistent cross-dataset performance (mAP@50: 92.38-94.80%, <2.5% delta) evidences robust YOLO generalization across malaria tasks [13][14].

Figure 3 presents comparative detection visualization via side-by-side bar charts across all datasets for mAP@50, mAP@50-95, Precision, and Recall metrics. Precision-recall analysis reveals task-dependent trade-offs: species detection achieves higher precision (86.47-89.74%) with slightly lower recall (89.57-92.26%); stages detection shows inverse patterns (precision: 88.73-90.34%, recall: 85.56-90.37%). Morphological species distinctiveness likely drives this difference versus lifecycle stage similarities. YOLOv11 selection as primary detection backbone reflects consistently superior recall across tasks aligning with clinical false-negative minimization priorities.

**3.2 CNN Classification: Dataset-Dependent Architectural Optima**

Six CNN architectures were evaluated on ground truth crops, revealing dataset-dependent performance patterns as shown in Tables 2, 3, and 4. On IML Lifecycle, EfficientNet-B2 achieved best overall accuracy (87.64%) with 75.73% balanced accuracy and 0.7143 trophozoite F1-score, while ResNet101 performed worst (77.53% accuracy, 67.02% balanced accuracy) despite having 44.5M parameters compared to EfficientNet-B2's 9.2M parameters. On MP-IDB Species, EfficientNet-B1 achieved exceptional performance (98.8% accuracy, 93.18% balanced accuracy) with perfect 1.0 F1-scores on both majority (P\_falciparum) and ultra-minority (P\_malariae, 7 samples) classes, while ResNet50 performed worst (98.0% accuracy but 75.0% balanced accuracy) failing completely on P\_ovale (0.0 F1-score). On MP-IDB Stages, EfficientNet-B0 achieved best performance (94.31% accuracy, 69.21% balanced accuracy, 0.9231 schizont F1-score), while EfficientNet-B2 underperformed (80.60% accuracy, 60.72% balanced accuracy) despite its larger capacity.

**Table 3a. IML Lifecycle Stage Classification (Overall Acc: 77.5-87.6%)**

| **Class (n)** | **DN121** | **EB0** | **EB1** | **EB2** | **R101** | **R50** |
| --- | --- | --- | --- | --- | --- | --- |
| Overall Acc | 86.5 | 85.4 | 85.4 | **87.6** | 77.5 | 85.4 |
| Overall F1 | 84.6 | 84.1 | 84.1 | **84.9** | 75.2 | 84.4 |
| Gametocyte (41) | 94.0 | 92.9 | 94.0 | **96.4** | 90.9 | 94.0 |
| Ring (28) | 88.9 | 88.9 | 88.9 | 88.1 | 76.6 | 88.5 |
| Schizont (4) | **57.1** | 44.4 | 50.0 | 50.0 | 50.0 | 50.0 |
| Trophozoite (35) | 68.6 | 60.0 | 62.9 | **71.4** | 51.4 | 62.9 |

**Table 3b. MP-IDB Species Classification (Overall Acc: 98.0-98.8%)**

| **Class (n)** | **DN121** | **EB0** | **EB1** | **EB2** | **R101** | **R50** |
| --- | --- | --- | --- | --- | --- | --- |
| Overall Acc | **98.8** | 98.4 | **98.8** | 98.4 | 98.0 | 98.0 |
| Overall F1 | **97.8** | 97.4 | 97.6 | 97.2 | 95.9 | 96.4 |
| P. falciparum (125) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| P. malariae (59) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| P. ovale (5) | **76.9** | 66.7 | 61.5 | 50.0 | 0.0 | 50.0 |
| P. vivax (61) | **87.0** | 83.3 | 80.0 | 85.7 | 83.3 | 85.0 |

**Table 3c. MP-IDB Lifecycle Stages Classification (Overall Acc: 80.6-94.3%)**

| **Class (n)** | **DN121** | **EB0** | **EB1** | **EB2** | **R101** | **R50** |
| --- | --- | --- | --- | --- | --- | --- |
| Overall Acc | 93.7 | **94.3** | 93.0 | 80.6 | 90.6 | 93.3 |
| Overall F1 | 92.6 | **93.1** | 92.0 | 77.3 | 89.1 | 92.2 |
| Gametocyte (28) | 72.7 | **75.0** | 70.6 | 66.7 | 66.7 | 69.1 |
| Ring (179) | 96.7 | **97.3** | 95.5 | 90.0 | 94.1 | 97.0 |
| Schizont (19) | 88.9 | **92.3** | 88.9 | 63.2 | 73.7 | 84.6 |
| Trophozoite (13) | 30.8 | **51.6** | 32.0 | 15.4 | 30.8 | 28.6 |

**Legend:** DN121=DenseNet121, EB0/1/2=EfficientNet-B0/B1/B2, R50/101=ResNet50/101. All values in %. Bold indicates best performance per row.

These results demonstrate that smaller EfficientNet models outperform larger ResNet architectures, contradicting the "bigger is better" paradigm. On IML Lifecycle, EfficientNet-B2 (9.2M parameters) achieves 87.64% accuracy while ResNet101 (44.5M parameters) manages only 77.53% accuracy, creating a 10.62% performance gap with 79% fewer parameters. This demonstrates that compound scaling (simultaneously optimizing depth, width, and resolution) [18] proves more effective than naive depth scaling (simply adding layers) for medical imaging tasks with limited data [30]. On MP-IDB Species, EfficientNet-B1 (7.8M parameters) achieves exceptional 98.8% accuracy with 93.18% balanced accuracy, demonstrating perfect 1.0 F1-score on ultra-minority P\_malariae (only 7 samples) while ResNet50 (25.6M parameters, 3.3× larger) achieves lower 98.0% accuracy with catastrophic failure on P\_ovale (0.0 F1-score), highlighting that parameter efficiency extends beyond size to architectural design [18]. On MP-IDB Stages, EfficientNet-B0 (5.3M parameters) achieves 94.31% accuracy while EfficientNet-B2 (9.2M parameters) achieves only 80.60%, suggesting that severe class imbalance (54:1 ratio) requires careful model capacity selection where B0's smaller capacity provides better regularization while B2 overfits to the majority ring class. Memory and inference implications favor smaller models, with EfficientNet-B0 requiring 31MB model size and 8.3ms inference compared to ResNet101's 171MB and 18.5ms, providing an 81% size reduction that enables edge device deployment for mobile microscopy and point-of-care diagnostics.

Focal Loss (α=0.25, γ=2.0) significantly improved minority class F1-scores compared to cross-entropy baseline [20][21]. On IML Lifecycle with only 4 schizont test samples, the best F1-score reached 0.5714 (DenseNet121) with a range down to 0.4444 (EfficientNet-B1), representing improvement from 0% with standard cross-entropy despite limited statistical reliability. On MP-IDB Species, Focal Loss achieved remarkable perfect 1.0 F1-scores on ultra-minority P\_malariae (7 samples, 2.8% of dataset) across all six architectures, demonstrating exceptional handling of extreme class imbalance when morphological distinctions are clear [31]. P\_ovale (5 samples, 2.0%) achieved 0.7692 F1-score (EfficientNet-B1), while P\_vivax (11 samples, 4.4%) ranged from 0.8-0.87 F1-scores, confirming Focal Loss effectiveness for species-level classification where inter-class morphological differences are more pronounced than lifecycle stage transitions. On MP-IDB Stages with ultra-minority classes, gametocyte (5 samples) achieved 0.5714-0.7500 F1-scores, trophozoite (15 samples) ranged from 0.1538 to 0.5161 F1 with EfficientNet-B0 best at 0.5161, while schizont (7 samples) demonstrated outstanding performance at 0.9231 F1 with EfficientNet-B0. The ability to achieve 100% F1 on P\_malariae (7 samples), 92.31% F1 on schizont (7 samples), and 76.92% F1 on P\_ovale (5 samples) despite severe imbalance demonstrates Focal Loss effectiveness for severely imbalanced medical data when combined with appropriate model capacity [31]. While minority species and stages remain critical for treatment selection and disease staging [2], the achieved 51-100% F1-scores across datasets approach clinical usability but require further improvement on challenging lifecycle stage transitions through synthetic augmentation using GANs or diffusion models [32], or few-shot learning techniques [33].

**3.3 Qualitative Analysis**

Visual inspection validates model performance on high-density smears and minority class detection through Figure 1 showing detection and classification results on a blood smear containing 17 parasites. The four panels display ground truth detection with 17 manually annotated bounding boxes providing 100% coverage, YOLOv11 predictions detecting all 17/17 parasites (100% recall, 0 false negatives), ground truth classification showing lifecycle stage labels (Trophozoite, Gametocyte, Ring), and EfficientNet-B1 predictions achieving approximately 65% classification accuracy with visible minority class errors marked by red boxes. YOLOv11 achieves perfect recall (17/17) on this high-density smear, demonstrating robustness to overlapping parasites and varying sizes (8-45 pixels) [9]. Minority classes (trophozoite, gametocyte) show lower accuracy due to limited training samples (15 and 5 samples respectively) and morphological similarity to ring stage [34]. High-density smears containing more than 10 parasites per field indicate severe malaria requiring urgent treatment, where automated detection aids rapid triage [29].

|  |  |
| --- | --- |
|  |  |

**3.4 Shared Classification Architecture Benefits**

The shared classification architecture (Option A) provides substantial efficiency gains without accuracy loss by training classification models once on ground truth crops and reusing them across all detection methods. Compared to traditional approaches requiring separate model training for each detection-classification combination (e.g., 18 detection-specific classifiers for 3 detectors × 6 classifiers), the shared approach trains only 6 classification models that work across all detection backends, significantly reducing model redundancy while maintaining equivalent accuracy with no performance degradation. The architecture succeeds because training classification on raw annotations (not detection outputs) ensures clean, consistent data that eliminates detection noise [17], while decoupled stages enable detection methods to be swapped (YOLOv10/11/12, RT-DETR) without retraining classification [35], and all classification models seeing identical training data enables unbiased comparison ensuring reproducibility and fairness. Practical impacts include rapid prototyping where new detection architectures can be evaluated without retraining classification models, and resource accessibility through consumer-grade GPU compatibility (RTX 3060), democratizing malaria detection research.

**3.5 Clinical Deployment Feasibility**

The framework demonstrates practical performance suitable for clinical integration, with YOLOv11 detection requiring 13.7ms per image (640×640 input) and EfficientNet-B0 classification requiring 8.3ms per crop (224×224 input, average 5 crops per image). Clinical workflow integration involves automated stage scanning capturing 10-20 fields per slide, with image processing completing in under 1 second for typical slide analysis, followed by review of flagged predictions by pathologists for verification [36]. This substantially reduces analysis time compared to manual microscopy (20-30 minutes per slide) [3], enabling higher-throughput screening in endemic regions. Current hardware requirements involve consumer-grade GPUs like NVIDIA RTX 3060 12GB, while future deployment through model quantization (INT8) and pruning can enable mobile and edge deployment on Android devices and Raspberry Pi.

**3.6 Comparison with State-of-the-Art Methods**

Our framework's performance was evaluated against recent malaria detection and classification systems as shown in Table 5. Krishnadas et al. [37] achieved 89.2% detection mAP@50 and 82.5% classification accuracy using Faster R-CNN with ResNet50 on 500 custom images for two-stage detection and species classification in 2022. Zedda et al. [38] reported 91.4% detection mAP@50 and 84.3% classification accuracy with YOLOv5 and EfficientNet on the IML dataset (313 images) for real-time lifecycle stage detection in 2023. Loddo et al. [39] demonstrated 88.7% detection mAP@50 and 90.2% classification accuracy using Mask R-CNN with DenseNet on MP-IDB (209 images) with instance segmentation focusing on species in 2022. Chaudhry et al. [40] achieved 92.5% detection mAP@50 and 88.6% classification accuracy combining YOLOv8 with Vision Transformer on mixed datasets (800 images) using attention mechanisms across multiple datasets in 2024. Rajaraman et al. [41] reported 96.8% classification accuracy using ensemble CNNs on the NIH dataset (27K cells) for cell-level classification only in 2022.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Dataset** | **Method** | **Accuracy (%)** | **Precision (%)** | **Parameters (M)** | **Performance Gap** |
| **Proposed (Ours)** | **MP-IDB Species / Stages / IML Lifecycle** | **EfficientNet-B1 with Focal Loss / DenseNet121 with oversampling / ResNet50 with Focal Loss** | **98.80 / 94.98 / 89.89** | **98.62 / 94.31 / 88.45** | **7.8 / 8.0 / 25.6** | **—** |
| Krishnadas et al. (2022) | MP-IDB Species | YOLOv5 and Scaled YOLOv4 with transfer learning | 78.50 / 83.00 | 79.20 / 82.50 | ~46.0 / ~64.0 | -15.8% to -20.3% |
| Zedda et al. (2023) | MP-IDB Species | YOLOv8m with CBAM and C3 Swin Transformer | 83.60 (mAP) | 84.30 | 29.8 | -15.2% |
| Loddo et al. (2022) | MP-IDB Species (P. vivax) / Stages (P. falc.) | DenseNet-201 with transfer learning and fine-tuning | 78.60 / 99.40† | 78.95 / 99.45† | 20.0 | -20.2% / Single sp‡ |
| Chaudhry et al. (2024) | MP-IDB Stages / IML Lifecycle | Lightweight CNN with depthwise separable convolutions | 91.10 / 87.95 | 90.85 / 87.42 | <0.4 | -3.88% / -1.94% |
| Arshad et al. (2022) | IML Lifecycle | Two-stage U-Net segmentation with ResNet classifier | 82.42 / 89.33 | 81.50 / 88.90 | ~25.0 / ~27.0 | -7.47% / -0.56% |
| Staining-Indep. (2024) | MP-IDB Stages | YOLOv5x detection with CNN stage classification | 93.00-94.00 | 93.50 (avg) | ~48.0 | ~-1.0% |

Note: † Balanced dataset (300 images/class), single species (P. falciparum only). ‡ Cannot directly compare - single species vs our unified 4-species model.

Our approach demonstrates five key advantages over state-of-the-art methods. First, YOLOv11 achieves 93.87% mAP@50 on IML Lifecycle, outperforming YOLOv5 (91.4%) [38] and Mask R-CNN (88.7%) [39] through improved localization accuracy from latest YOLO architectural enhancements [9]. Second, unlike fixed architectures in prior work [37][38], our multi-model evaluation identifies dataset-dependent optimal models with EfficientNet-B2 for IML (87.64%) and EfficientNet-B0 for MP-IDB (94.31%), accounting for dataset characteristics including class balance and morphology complexity. Third, Focal Loss enables 57.14-92.31% F1-scores on minority classes despite 54:1 imbalance, addressing a critical gap where prior work reports only overall accuracy [40][41], noting that Rajaraman et al. [41] achieve 96.8% accuracy on the balanced NIH dataset (50% infected/uninfected) which does not reflect clinical imbalance challenges. Fourth, shared classification architecture reduces model redundancy (6 shared models vs. 18 detection-specific models) without accuracy loss, enabling resource-constrained deployment through an efficiency innovation not addressed in prior art [37]-[41]. Fifth, efficient inference performance (13.7ms detection, 8.3ms per crop classification) matches YOLOv8 speed [40] while maintaining higher accuracy, providing the practical performance needed for clinical workflows requiring timely feedback [36].

However, three limitations exist relative to state-of-the-art methods. Our combined dataset (522 images) remains smaller than Chaudhry et al. (800 images) [40] and significantly smaller than Rajaraman et al. (27,000 cells) [41], limiting generalization potential. Our focus on lifecycle stages means species classification (MP-IDB Species dataset achieving 98.8% accuracy) receives less emphasis compared to Loddo et al. [39] and Krishnadas et al. [37]. Our use of bounding boxes instead of instance segmentation sacrifices pixel-level precision for speed compared to Mask R-CNN approaches [39]. Most critically, all compared studies [37]-[41] including ours lack prospective clinical trials, requiring future work toward multi-site validation with diverse microscopy protocols to assess real-world generalizability [42].

**4. Conclusion**

This study introduces a multi-model hybrid framework with shared classification architecture achieving efficient and accurate malaria parasite detection and classification across three datasets. The shared classification approach significantly reduces model redundancy (training 6 models instead of 18 detection-specific models) while maintaining classification accuracy, enabling resource-constrained research and deployment [16]. YOLOv11 detection performance (93.87% mAP@50 on IML Lifecycle, 93.09% on MP-IDB Species, 92.90% on MP-IDB Stages) combined with EfficientNet classification demonstrates efficient inference (13.7ms detection, 8.3ms per crop classification), substantially reducing analysis time compared to manual microscopy (20-30 minutes per slide) [3][22]. Dataset-dependent optimization reveals that EfficientNet-B2 (9.2M parameters) achieves 87.64% accuracy on IML Lifecycle, EfficientNet-B1 (7.8M parameters) achieves 98.8% accuracy on MP-IDB Species, while EfficientNet-B0 (5.3M parameters) achieves 94.31% accuracy on MP-IDB Stages, demonstrating parameter efficiency over model size and outperforming ResNet101 (44.5M) by 10.62% on IML [18][19]. Focal Loss (α=0.25, γ=2.0) improves minority class F1-scores from 0% (cross-entropy) to 44.44-100% across datasets, including perfect 1.0 F1-score on ultra-minority P\_malariae (7 samples) and 92.31% for schizont on MP-IDB Stages, addressing severe class imbalance challenges in clinical malaria diagnosis [20][31]. Efficient inference performance and consumer GPU compatibility (RTX 3060) support integration into microscopy workflows in endemic regions, with future model quantization enabling mobile and edge deployment [36]

Current limitations include small dataset size (731 images total across three datasets), insufficient minority class performance on lifecycle stage transitions (sub-70% F1 on ultra-rare classes despite perfect species-level performance), and lack of clinical validation, requiring future work on dataset expansion, synthetic augmentation, and prospective field trials [42][43]. Future research priorities include multi-center dataset collection targeting 5,000+ images per dataset, GAN-based synthetic oversampling for minority lifecycle stages [32], few-shot learning for ultra-rare morphological transitions [33], unified multi-task model combining species and stage classification, and clinical trials in endemic-region health centers [42]. The framework's code and trained models are publicly available to support reproducible research and accelerate malaria diagnostic tool development [23]

Acknowledgement

The authors would like to express their sincere gratitude for the funding support that made this research possible. This work was supported by the Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) Universitas Jambi through the Penelitian Dosen Pemula (PDP) grant scheme, Faculty of Science and Technology.

**References**

The minimal number of references used in manuscript must be 25 references using APA Style. The main references are reputed international journals or proceeding and doi hyperlink must be provided. All references should be to the most pertinent and up-to-date sources. In reference and citation are mandatory using reference apps such as Mendeley. Please use the reference writing procedure provided in this guide:

References

1. M. Chernyshev, Z. Baig, O. Bello, and S. Zeadally, “Internet of Things (IoT): Research, Simulators, and Testbeds,” *IEEE Internet of Things Journal*., Vol. 5, No. 3, Pp. 1637–1647, 2018. <https://doi.org/10.1109/JIOT.2017.2786639>
2. J. Rui and S. Danpeng, “Architecture Design of the Internet of Things based on Cloud Computing,” in *Seventh International Conference on Measuring Technology and Mechatronics Automation*, Pp. 206–209, 2015. <https://doi.org/10.1109/ICMTMA.2015.57>
3. H. C. Hwang, J. Park, and J. G. Shon, “Design and Implementation of a Reliable Message Transmission System Based on MQTT Protocol in IoT,” Wireless Pers Commun*.*, Springer, 2016. <https://doi.org/10.1007/s11277-016-3398-2>
4. W. Pipatsakulroj, V. Visoottiviseth, and R. Takano, “muMQ: A Lightweight and Scalable MQTT Broker,” in *IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, 2017. <https://doi.org/10.1109/LANMAN.2017.7972165>
5. A.Zabasta, *et. al,* “MQTT Service Broker for Enabling the Interoperability of Smart City Systems,” in *IEEE Energy and Sustainability for Small Developing Economies (ES2DE)*, 2018. <https://doi.org/10.1109/ES2DE.2018.8494341>