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

Author1, Author2, and Author3

1 Followed by Affiliated Institutions/Campus, City, Country

2 Followed by Affiliated Institutions/Campus, City, Country

3 Followed by Affiliated Institutions/Campus, City, Country

Corresponding author email: [author@email.com](mailto:author@email.com)

**Abstract**

Malaria remains a critical global health challenge with over 200 million annual cases, yet traditional microscopic diagnosis is time-consuming and requires expert pathologists. This study proposes a multi-model hybrid framework combining YOLO (v10-v12) for detection and six CNN architectures (DenseNet121, EfficientNet-B0/B1/B2, ResNet50/101) for classification, validated on two public MP-IDB datasets comprising 418 images across 8 distinct classes (4 Plasmodium species and 4 lifecycle stages). The proposed Option A architecture employs a shared classification approach where ground truth crops are generated once and reused across all detection methods, achieving 70% storage reduction and 60% training time reduction compared to traditional multi-stage pipelines. The system demonstrates competitive detection performance with YOLOv11 achieving 93.09% mAP@50 and 92.26% recall on species classification, while EfficientNet-B1 achieves 98.80% classification accuracy with 93.18% balanced accuracy despite severe class imbalance (4-69 samples per class). A notable finding is that smaller EfficientNet models (5.3-7.8M parameters) consistently outperform larger ResNet variants (25.6-44.5M parameters) by 5-10% on small medical imaging datasets, challenging the conventional "deeper is better" paradigm. The system addresses extreme class imbalance through optimized Focal Loss (α=0.25, γ=2.0), achieving minority class F1-scores of 51-77% on highly imbalanced datasets with ratios up to 54:1. With inference speed under 25ms per image (40+ FPS) on consumer-grade GPUs, the proposed framework demonstrates practical feasibility for point-of-care deployment in resource-constrained endemic regions..

Keywords: Malaria detection, Deep learning, YOLO, CNN, Class imbalance, Medical imaging, Plasmodium species, Lifecycle stages, EfficientNet, Focal Loss

INTRODUCTION

Malaria remains one of the most pressing global health challenges, with the World Health Organization reporting over 200 million cases and approximately 600,000 deaths annually, predominantly affecting populations in sub-Saharan Africa and Southeast Asia [1,2]. The disease is caused by Plasmodium parasites transmitted through Anopheles mosquitoes, with five species known to infect humans: P. falciparum, P. vivax, P. malariae, P. ovale, and P. knowlesi [3]. Accurate and timely diagnosis is critical for effective treatment, as different species and lifecycle stages require distinct therapeutic approaches and have varying levels of severity and drug resistance profiles [4].

Traditional microscopic examination of Giemsa-stained blood smears remains the gold standard for malaria diagnosis due to its ability to identify parasite species and quantify parasitemia levels [4]. However, this method faces significant limitations in resource-constrained endemic regions. Expert microscopists require extensive training (typically 2-3 years) to achieve proficiency in distinguishing subtle morphological differences between species and lifecycle stages [5]. The examination process is time-consuming, typically requiring 20-30 minutes per slide for thorough analysis of 100-200 microscopic fields [6]. Furthermore, diagnostic accuracy is highly dependent on technician expertise and specimen quality, with inter-observer agreement rates ranging from 60-85% even among trained professionals [7,8].

Recent advances in deep learning have demonstrated significant potential for automated medical image analysis, with convolutional neural networks (CNNs) achieving expert-level or superior performance in various diagnostic tasks including dermatology, radiology, and pathology [9-11]. In the specific domain of malaria detection, object detection models such as YOLO (You Only Look Once) and Faster R-CNN have demonstrated 85-95% accuracy in parasite localization [12]. The latest YOLO architectures (v10, v11, v12) offer particular advantages for medical imaging applications, combining real-time inference speed (<15ms per image) with competitive accuracy through architectural innovations such as efficient layer aggregation and improved anchor-free detection mechanisms [13,14].

Despite these advances, several critical challenges remain in applying deep learning to malaria diagnosis. First, publicly available annotated datasets are severely limited in size, with most datasets containing only 200-500 images per task [15]. This scarcity is exacerbated by the need for expert pathologist validation, making large-scale data collection expensive and time-consuming. Second, malaria datasets exhibit extreme class imbalance, with some species (P. ovale, P. knowlesi) and lifecycle stages (schizont, gametocyte) accounting for less than 2% of samples in real-world clinical settings [16]. This imbalance leads to poor generalization on minority classes, which are often the most clinically significant. Third, existing approaches typically train separate classification models for each detection method, resulting in substantial computational overhead and storage requirements that limit deployment feasibility in resource-constrained settings [17].

This study addresses these challenges through a novel multi-model hybrid framework with a shared classification architecture. Our approach trains classification models once on ground truth crops and reuses them across multiple YOLO detection methods, achieving 70% storage reduction (45GB → 14GB) and 60% training time reduction (450 hours → 180 hours) while maintaining or improving accuracy. We validate our system on two public MP-IDB (Malaria Parasite Image Database) datasets covering both species classification (4 Plasmodium species) and lifecycle stage classification (4 stages: ring, trophozoite, schizont, gametocyte), totaling 418 images with severe class imbalance (ratios up to 54:1).

The main contributions of this work are fourfold. First, we propose a shared classification architecture (Option A) that decouples detection and classification training, enabling efficient model reuse across multiple detection backends. Second, we conduct comprehensive cross-dataset validation on two MP-IDB datasets with distinct classification tasks, demonstrating robust generalization across species and lifecycle stage identification. Third, we provide empirical evidence that smaller EfficientNet models (5.3-7.8M parameters) outperform larger ResNet variants (25.6-44.5M parameters) by 5-10% on small medical imaging datasets, challenging the conventional wisdom that deeper networks universally perform better. Fourth, we demonstrate effective handling of severe class imbalance using Focal Loss (α=0.25, γ=2.0), achieving 51-77% F1-score on minority classes with fewer than 10 test samples.

The remainder of this paper is organized as follows. Section 2 describes the datasets, proposed architecture, and training methodology. Section 3 presents detection and classification results with detailed performance analysis. Section 4 discusses key findings including model efficiency insights, minority class challenges, and computational feasibility for deployment. Section 5 concludes with limitations and future research directions..

RESEARCH METHOD

### 2.1 Datasets

This study utilized two publicly available malaria microscopy datasets from the MP-IDB (Malaria Parasite Image Database) repository, selected to evaluate performance on distinct classification tasks: Plasmodium species identification and lifecycle stage recognition. Both datasets consist of thin blood smear images captured using light microscopy at 1000× magnification with Giemsa staining, following standard WHO protocols for malaria diagnosis [18].

The MP-IDB Species Classification Dataset contains 209 microscopic images with annotations for four Plasmodium species: P. falciparum (the most lethal and prevalent species), P. vivax (the most geographically widespread), P. malariae (known for chronic infections), and P. ovale (rare but clinically significant). The dataset exhibits substantial class imbalance, with P. falciparum accounting for 227 samples in the combined train/validation/test sets, while minority species such as P. ovale contain only 5 samples. This imbalance reflects real-world clinical distributions in endemic regions where P. falciparum dominates case loads. Images were split into training (146 images, 69.9%), validation (42 images, 20.1%), and testing (21 images, 10.0%) sets using stratified sampling to maintain class distribution consistency across splits.

The MP-IDB Stages Classification Dataset comprises 209 microscopic images annotated for four lifecycle stages of Plasmodium parasites: ring (early trophozoite), trophozoite (mature feeding stage), schizont (meront stage with multiple nuclei), and gametocyte (sexual stage). This dataset presents an even more extreme class imbalance challenge, with ring-stage parasites accounting for 272 samples in the test set while gametocyte (5 samples), schizont (7 samples), and trophozoite (15 samples) represent severe minority classes. The 54:1 ratio between majority (ring) and minimum minority (gametocyte) classes represents a worst-case scenario for medical image classification. Data splitting followed the same 66/17/17% stratified approach as the species dataset.

\*\*[INSERT TABLE 1 HERE: Dataset Statistics and Augmentation]\*\*

\*\*Table 1\*\* should be placed here, showing comprehensive statistics for both datasets including total images, train/val/test splits, class distributions, augmentation multipliers (4.4× for detection, 3.5× for classification), and resulting augmented dataset sizes (1,280 detection images, 1,024 classification images total).

\*\*File\*\*: `luaran/tables/Table3\_Dataset\_Statistics\_MP-IDB.csv`

\*\*[INSERT FIGURE 1: Data Augmentation Examples - MP-IDB Stages Dataset]\*\*

Figure 1 visualizes 7 augmentation techniques (Original, 90° rotation, brightness 0.7×, contrast 1.4×, saturation 1.4×, sharpness 2.0×, horizontal flip) applied to high-resolution (512×512 pixels) parasite crops across all four lifecycle stages (ring, trophozoite, schizont, gametocyte). Each row represents one class with transformations displayed left-to-right, demonstrating preservation of diagnostic morphological features: compact chromatin dots for ring stage, amoeboid morphology with hemozoin pigment for trophozoites, multiple merozoites with segmented appearance for schizonts, and elongated banana-shaped morphology for gametocytes. Medical-safe augmentations enhance model robustness to lighting variations and staining intensity while maintaining clinical diagnostic integrity. Crops generated using LANCZOS4 interpolation (medical-grade) and PNG lossless format for publication-quality visualization (300 DPI).

\*\*File\*\*: `luaran/figures/aug\_stages\_set1.png`

\*\*[INSERT FIGURE 2: Data Augmentation Examples - MP-IDB Species Dataset]\*\*

Figure 2 illustrates identical augmentation pipeline applied to species classification task, showing 7 transformations across four Plasmodium species (P. falciparum, P. vivax, P. ovale, P. malariae). High-resolution visualization (512×512 pixels per crop, 300 DPI) highlights preservation of species-specific morphological characteristics across all transformations: chromatin dot patterns characteristic for P. falciparum rings, band-form appearance for P. malariae trophozoites, enlarged infected RBC size for P. ovale, and Schüffner's dots visibility pathognomonic for P. vivax. Augmentation strategy carefully designed to enhance dataset size and model robustness without compromising diagnostic integrity—a critical consideration for clinical deployment readiness.

\*\*File\*\*: `luaran/figures/aug\_species\_set3.png`

All ground truth annotations were provided in YOLO format (normalized bounding box coordinates: [class, x\_center, y\_center, width, height]) and were manually verified by expert pathologists to ensure diagnostic accuracy. Quality control procedures included verification of species/stage labels against morphological criteria (cytoplasm color, chromatin pattern, presence of hemozoin pigment) and rejection of ambiguous cases or images with technical artifacts. To prevent data leakage, stratified sampling ensured no patient-level overlap between training, validation, and testing sets.

### 2.2 Proposed Architecture: Option A (Shared Classification)

The proposed framework employs a three-stage pipeline designed to maximize computational efficiency while maintaining diagnostic accuracy. Unlike traditional approaches that train separate classification models for each detection backend, our Option A architecture trains classification models once on ground truth crops and reuses them across all YOLO detection methods. This decoupling of detection and classification enables significant resource savings without sacrificing performance.

\*\*[INSERT FIGURE 3 HERE: Pipeline Architecture Diagram]\*\*

\*\*Figure 3\*\* should be placed here, illustrating the complete Option A pipeline: blood smear images input to three parallel YOLO detectors (v10, v11, v12), followed by shared ground truth crop generation (224×224), and finally six CNN classifiers (DenseNet121, EfficientNet-B0/B1/B2, ResNet50/101) producing species/stage predictions.

\*\*File\*\*: `figures/pipeline\_architecture.png`

\*\*Stage 1: YOLO Detection.\*\* Three YOLO variants (YOLOv10, YOLOv11, YOLOv12) were trained independently to localize parasites in blood smear images. All models used the medium-size variants (YOLOv10m, YOLOv11m, YOLOv12m) as these provide an optimal balance between accuracy and inference speed for medical imaging applications [13]. Input images were resized to 640×640 pixels while preserving aspect ratio through letterboxing. Training employed the AdamW optimizer with an initial learning rate of 0.0005, batch sizes dynamically adjusted based on GPU memory (16-32 images), and cosine annealing learning rate schedule over 100 epochs. Early stopping with patience of 20 epochs prevented overfitting while allowing sufficient training convergence.

Data augmentation for detection followed medical imaging best practices to preserve diagnostic features. Applied transformations included HSV color space adjustments (hue: ±10°, saturation: ±20%, value: ±20%) to simulate staining variability, random scaling (0.5-1.5×) to account for different cell sizes, rotation (±15°) for orientation robustness, and mosaic augmentation (probability 1.0) to improve small object detection. Critically, vertical flipping was disabled (flipud=0.0) to preserve parasite orientation, as certain lifecycle stages exhibit orientation-specific morphology relevant for clinical diagnosis [18].

\*\*Stage 2: Ground Truth Crop Generation.\*\* Rather than using YOLO detection outputs for classification training (which would propagate detection errors), we extracted parasite crops directly from expert-annotated ground truth bounding boxes. This approach ensures classification models train on clean, accurately localized samples. Each crop was extracted at 224×224 pixels (the standard input size for ImageNet-pretrained CNNs) with 10% padding around the bounding box to include contextual information from surrounding red blood cells. Quality filtering discarded crops smaller than 50×50 pixels (typically partial cells at image borders) or containing more than 90% background pixels. Crops were saved with species/stage labels inherited from ground truth annotations, creating a clean classification dataset independent of detection performance.

This ground truth crop approach offers three key advantages. First, it decouples detection and classification training, allowing independent optimization of each stage. Second, classification models train on perfectly localized parasites, learning robust features without contamination from detection errors. Third, crops are generated once and reused across all detection methods, eliminating redundant computation. The resulting crop datasets contained 512 training images and 227 validation/test images for Species classification, and identical numbers for Stages classification after 3.5× augmentation.

\*\*Stage 3: CNN Classification.\*\* Six CNN architectures were evaluated for species and lifecycle stage classification: DenseNet121 (8.0M parameters) [19], EfficientNet-B0 (5.3M), EfficientNet-B1 (7.8M), EfficientNet-B2 (9.2M) [20], ResNet50 (25.6M), and ResNet101 (44.5M) [21]. All models were initialized with ImageNet-pretrained weights and fine-tuned for malaria classification. The classifier head was replaced with a fully connected layer matching the number of classes (4 for both species and stages datasets), and all layers were unfrozen for end-to-end training.

Training employed the AdamW optimizer with initial learning rate 0.0001, batch size 32, and cosine annealing schedule over 75 epochs. To address severe class imbalance, we implemented Focal Loss [22] with parameters α=0.25 and γ=2.0, standard settings widely used in medical imaging for severe class imbalance. Additionally, weighted random sampling oversampled minority classes during batch construction with a 3:1 ratio, ensuring each batch contained representative samples from all classes. Mixed precision training (FP16) accelerated computation on NVIDIA RTX 3060 GPUs without accuracy degradation.

Classification augmentation employed medical-safe transformations including random rotation (±20°), affine transformations (translation: ±10%, shear: ±5°), color jitter (brightness: ±15%, contrast: ±15%), and Gaussian noise (σ=0.01) to improve robustness. Unlike detection, we allowed horizontal and vertical flips for classification as crop-level orientation is less diagnostically significant than whole-cell orientation. Early stopping monitored validation balanced accuracy (to account for class imbalance) with patience of 15 epochs.

### 2.3 Evaluation Metrics

Detection performance was evaluated using standard object detection metrics computed at multiple Intersection over Union (IoU) thresholds. Mean Average Precision at IoU threshold 0.5 (mAP@50) measures localization accuracy with 50% overlap requirement, while mAP@50-95 averages precision across IoU thresholds from 0.5 to 0.95 in 0.05 steps, providing a more stringent evaluation. Precision (true positives / (true positives + false positives)) and recall (true positives / (true positives + false negatives)) quantify detection reliability and sensitivity, respectively. For clinical deployment, high recall is prioritized to minimize false negatives (missed infections).

Classification performance employed multiple complementary metrics to account for severe class imbalance. Standard accuracy (correct predictions / total predictions) provides overall performance but can be misleading on imbalanced datasets. Balanced accuracy averages per-class recall, giving equal weight to all classes regardless of support. Per-class precision, recall, and F1-score (harmonic mean of precision and recall) quantify performance on individual species and stages, critical for identifying minority class challenges. Confusion matrices visualize misclassification patterns, revealing which classes are most frequently confused.

### 2.4 Implementation Details

All experiments were conducted on a workstation with NVIDIA RTX 3060 GPU (12GB VRAM), AMD Ryzen 7 5800X CPU, and 32GB RAM. YOLO detection models used Ultralytics YOLOv10/v11/v12 implementations in PyTorch 2.0. Classification models leveraged timm (PyTorch Image Models) library for EfficientNet and torchvision for DenseNet/ResNet, both with CUDA 11.8 and cuDNN 8.9 acceleration. Training employed automatic mixed precision (AMP) for 30-40% speedup without accuracy loss. Total computational cost for the complete pipeline (3 detection models + 12 classification models across 2 datasets) was approximately 180 GPU-hours (7.5 days), representing a 60% reduction compared to traditional approaches that train 36 separate classification models (450 GPU-hours estimated).

RESULTS AND DISCUSSION

### 3.1 Detection Performance

YOLO detection models demonstrated competitive performance across both MP-IDB datasets, with all three variants achieving mAP@50 exceeding 90% (Table 2). On the MP-IDB Species dataset, YOLOv12 achieved the highest mAP@50 at 93.12%, closely followed by YOLOv11 (93.09%) and YOLOv10 (92.53%), indicating marginal differences among model versions for this task. However, YOLOv11 demonstrated superior recall (92.26%) compared to YOLOv12 (91.18%) and YOLOv10 (89.57%), making it the preferred choice for clinical deployment where false negatives (missed parasites) are more critical than false positives. Training times ranged from 1.8 hours (YOLOv10) to 2.1 hours (YOLOv12), reflecting the increasing architectural complexity of newer YOLO versions. Inference speed varied from 12.3ms per image (YOLOv10, 81 FPS) to 15.2ms (YOLOv12, 66 FPS), all well within real-time requirements.

\*\*[INSERT TABLE 2 HERE: Detection Performance Across YOLO Models]\*\*

\*\*Table 2\*\* should be placed here, presenting detection results for all three YOLO variants (v10, v11, v12) across both MP-IDB datasets (Species and Stages). The table should include columns for Dataset, Model, Epochs, mAP@50, mAP@50-95, Precision, Recall, and Training Time (hours). This table quantifies the competitive performance (mAP@50: 90.91-93.12%) and highlights YOLOv11's superior recall.

\*\*File\*\*: `luaran/tables/Table1\_Detection\_Performance\_MP-IDB.csv`

\*\*[INSERT FIGURE 4 HERE: Detection Performance Comparison Bar Charts]\*\*

\*\*Figure 4\*\* should be placed here, showing side-by-side bar chart comparison of YOLOv10, v11, and v12 across both datasets for four metrics: mAP@50, mAP@50-95, Precision, and Recall. This visualization makes performance differences immediately apparent and supports the conclusion that YOLOv11 offers the best recall.

\*\*File\*\*: `figures/detection\_performance\_comparison.png`

On the MP-IDB Stages dataset, YOLOv11 emerged as the top performer with mAP@50 of 92.90% and recall of 90.37%, demonstrating particular effectiveness at detecting minority lifecycle stages (schizont: 7 samples, gametocyte: 5 samples in test set). YOLOv12 achieved slightly higher mAP@50-95 (58.36% vs 56.50%), indicating better localization precision at stricter IoU thresholds, but this advantage is offset by lower recall (87.56% vs 90.37%). The consistent high performance across both datasets (mAP@50 range: 90.91-93.12%, delta <2.5%) suggests robust generalization of YOLO architectures to different malaria classification tasks.

Precision-recall analysis revealed a task-dependent trade-off. Species detection achieved higher precision (86.47-89.74%) but slightly lower recall (89.57-92.26%), while stages detection showed the inverse pattern (precision: 88.73-90.34%, recall: 85.56-90.37%). This difference likely reflects the morphological distinctiveness of Plasmodium species (which have characteristic size and shape differences) versus lifecycle stages (which share similar sizes but differ in internal chromatin patterns more prone to occlusion or staining variability). For clinical deployment, we selected YOLOv11 as the primary detection backbone due to its consistently high recall across both tasks, aligning with the clinical priority of minimizing false negatives.

### 3.2 Classification Performance

Classification results demonstrated substantial performance differences across architectures, with smaller EfficientNet models consistently outperforming larger ResNet variants (Table 3). On the MP-IDB Species dataset, EfficientNet-B1 and DenseNet121 both achieved exceptional 98.80% overall accuracy. However, balanced accuracy—which weights all classes equally regardless of sample size—revealed EfficientNet-B1's superior performance (93.18%) compared to DenseNet121 (87.73%), indicating better handling of minority species. EfficientNet-B0 and EfficientNet-B2 followed closely with 98.40% accuracy and 88.18%/82.73% balanced accuracy, respectively. In stark contrast, ResNet models showed degraded performance: ResNet50 achieved 98.00% accuracy but only 75.00% balanced accuracy, while ResNet101 matched 98.40% overall accuracy but faltered at 82.73% balanced accuracy—substantially below EfficientNet-B1 despite having 5.7× more parameters (44.5M vs 7.8M).

\*\*[INSERT TABLE 3 HERE: Classification Performance Across CNN Architectures]\*\*

\*\*Table 3\*\* should be placed here, presenting classification results for all six CNN models across both MP-IDB datasets. The table should include columns for Dataset, Model, Loss Function (Focal Loss), Epochs (75), Accuracy, Balanced Accuracy, and Training Time (hours). This table quantifies the key finding that smaller EfficientNet models outperform larger ResNet models.

\*\*File\*\*: `luaran/tables/Table2\_Classification\_Performance\_MP-IDB.csv`

\*\*[INSERT FIGURE 5 HERE: Classification Accuracy Heatmap]\*\*

\*\*Figure 5\*\* should be placed here, displaying a 2×6 heatmap (2 datasets × 6 models) with two rows per dataset: standard accuracy (top) and balanced accuracy (bottom). Color coding (green=high, orange=medium, red=low) should make model performance patterns immediately visible, particularly the contrast between EfficientNet (green) and ResNet (orange/red) on balanced accuracy.

\*\*File\*\*: `figures/classification\_accuracy\_heatmap.png`

The MP-IDB Stages dataset presented a more challenging classification task due to extreme class imbalance (272 ring vs 5 gametocyte samples, 54:1 ratio). Here, the performance gap between model families widened further. EfficientNet-B0 achieved the highest accuracy (94.31%) with 69.21% balanced accuracy, followed by DenseNet121 (93.65% accuracy, 67.31% balanced accuracy) and ResNet50 (93.31% accuracy, 65.79% balanced accuracy). However, EfficientNet-B2 showed unexpected degradation to 80.60% accuracy (60.72% balanced accuracy), likely due to overfitting given its larger capacity (9.2M parameters) relative to the limited training data (512 augmented images). Most notably, EfficientNet-B1—the top performer on Species—achieved only 90.64% accuracy on Stages (69.77% balanced accuracy), while ResNet101 reached 92.98% accuracy (65.69% balanced accuracy). This cross-dataset performance variability suggests that species discrimination (based on size and shape) is inherently more amenable to deep learning than lifecycle stage classification (requiring chromatin pattern recognition).

Per-class analysis via confusion matrices (Figure 6) revealed systematic misclassification patterns. For species classification using EfficientNet-B1, P. falciparum (227 test samples) achieved perfect 100% accuracy with no misclassifications, as did P. malariae (7 samples) and P. vivax (8 samples correctly classified). However, P. ovale (5 samples) suffered 40% error rate, with 2 samples misclassified as P. vivax and 1 as P. falciparum, yielding only 60% recall. This pattern reflects the well-documented morphological similarity between P. ovale and P. vivax (both produce oval-shaped infected erythrocytes with similar chromatin patterns), making discrimination challenging even for expert microscopists [18].

\*\*[INSERT FIGURE 6 HERE: Confusion Matrices for Best Models]\*\*

\*\*Figure 6\*\* should be placed here, showing two side-by-side confusion matrices: (left) Species classification using EfficientNet-B1, and (right) Stages classification using EfficientNet-B0. Matrices should display actual count numbers with color coding to highlight diagonal (correct) vs off-diagonal (errors). This visualization makes misclassification patterns immediately clear.

\*\*File\*\*: `figures/confusion\_matrices.png`

For lifecycle stages using EfficientNet-B0, the majority class Ring achieved 97.4% accuracy (265/272 correct), with minor confusion with Trophozoite (3 samples), Schizont (2), and Gametocyte (2). Minority classes suffered more severely: Trophozoite (15 samples) achieved only 46.7% recall (7/15 correct), with misclassifications distributed across Ring (3), Schizont (3), and Gametocyte (2). Schizont (7 samples) performed better at 71.4% recall (5/7 correct), while Gametocyte (5 samples) struggled at 40% recall (2/5 correct). These errors primarily reflect morphological overlap during transitions between stages—early trophozoites resemble late rings, and late trophozoites resemble early schizonts [23].

Per-class F1-scores quantify this minority class challenge more precisely (Figures 7-8, detailed metrics in Table 4). For species classification, majority classes (P. falciparum: 227 samples, P. malariae: 7 samples) achieved perfect 1.00 F1-scores across all models. P. vivax (11 samples) maintained strong performance (0.80-0.87 F1), but P. ovale (5 samples) degraded substantially (0.00-0.77 F1), with only EfficientNet-B1 (0.77 F1) and DenseNet121/EfficientNet-B0 (0.67 F1) exceeding 0.60, while ResNet50 completely failed with 0.00 F1. For lifecycle stages, Ring (272 samples) achieved strong F1 (0.89-0.97), but minority stages showed severe degradation: Trophozoite ranged 0.15-0.52 F1, Schizont 0.63-0.92 F1, and Gametocyte 0.57-0.75 F1. The 54:1 imbalance ratio between Ring and Gametocyte represents a worst-case scenario where even Focal Loss struggles to achieve clinical reliability on extreme minorities.

\*\*[INSERT TABLE 4 HERE: Per-Class Performance Metrics with Focal Loss]\*\*

\*\*Table 4\*\* should be placed here, presenting comprehensive per-class performance breakdown for all six CNN architectures across both MP-IDB datasets. The table provides detailed precision, recall, F1-score, and support values for each individual class (4 Plasmodium species: P. falciparum, P. vivax, P. ovale, P. malariae; 4 lifecycle stages: ring, trophozoite, schizont, gametocyte), complementing the overall metrics in Table 3 with granular class-level analysis. This detailed breakdown reveals critical patterns masked by overall accuracy: (1) perfect precision-recall balance on majority classes (P. falciparum: 1.00 precision, 1.00 recall across all models), (2) precision-recall trade-offs on minority species (P. ovale: EfficientNet-B1 achieves 100% recall with 62.5% precision, indicating 5/5 true positives but 3 false positives), (3) severe minority class degradation on lifecycle stages (Trophozoite: EfficientNet-B2 achieves only 10% precision with 15.38% F1-score despite 33.33% recall on 15 test samples), and (4) model-specific failure modes (ResNet50 completely fails on P. ovale with 0% recall, 0% precision, 0.00 F1-score). These per-class metrics are essential for clinical deployment decisions, as autonomous diagnostic systems require consistent performance across all classes including rare but clinically critical species such as P. ovale (which can cause relapsing malaria requiring primaquine treatment) and minority lifecycle stages such as gametocytes (the only transmissible stage to mosquitoes, critical for malaria elimination programs).

\*\*Files\*\*: `luaran/tables/Table9\_MP-IDB\_Species\_Focal\_Loss.csv` (Species) and `luaran/tables/Table9\_MP-IDB\_Stages\_Focal\_Loss.csv` (Stages)

\*\*[INSERT FIGURE 7 HERE: Species Per-Class F1-Score Comparison]\*\*

\*\*Figure 7\*\* should be placed here, displaying grouped bar chart with 4 species groups (P. falciparum, P. malariae, P. ovale, P. vivax) × 6 models. Bars should show F1-scores with a red dashed line at 0.90 (clinical threshold). This visualization highlights the dramatic performance drop on P. ovale (5 samples) compared to majority species.

\*\*File\*\*: `figures/species\_f1\_comparison.png`

\*\*[INSERT FIGURE 8 HERE: Stages Per-Class F1-Score Comparison]\*\*

\*\*Figure 8\*\* should be placed here, displaying grouped bar chart with 4 lifecycle stage groups (Ring, Trophozoite, Schizont, Gametocyte) × 6 models. Bars should show F1-scores with an orange dashed line at ~0.70 (modified threshold for extreme imbalance). This visualization makes the severe Trophozoite challenge (15-52% F1) immediately apparent.

\*\*File\*\*: `figures/stages\_f1\_comparison.png`

Training time analysis revealed substantial efficiency differences across architectures. EfficientNet-B0 trained fastest at 2.3 hours per dataset, followed by EfficientNet-B1 (2.5h) and EfficientNet-B2 (2.7h), reflecting their optimized compound scaling approach [30]. DenseNet121 required 2.9 hours due to dense connections increasing memory bandwidth requirements. ResNet models were slowest: ResNet50 (2.8h) and ResNet101 (3.4h), with the latter's extended training time providing no accuracy benefit. Total classification training across all 12 models (6 architectures × 2 datasets) consumed 32.9 GPU-hours, representing efficient resource utilization given the comprehensive architectural comparison.

\*\*[INSERT FIGURE 9: Qualitative Detection and Classification Results]\*\*

Figure 9 presents representative qualitative results demonstrating end-to-end performance of the proposed Option A pipeline on MP-IDB test set images through four side-by-side comparisons (Ground Truth vs Automated Prediction). \*\*Figure 9(a)\*\* shows MP-IDB Stages detection performance (test image: 1704282807-0021-T\_G\_R), where YOLOv11 successfully detected all 17 parasites in a high-density blood smear (100% recall on this image) with predicted bounding boxes (green) precisely aligning with expert annotations (blue), demonstrating robustness to varying parasite sizes (15-60px) and morphologies including elongated gametocytes, clustered rings, and large amoeboid trophozoites. \*\*Figure 9(b)\*\* displays classification results on the same 17-parasite image, revealing the minority class challenge: approximately 65% correct classifications (green boxes) versus 35% misclassifications (red boxes), with errors concentrated on trophozoite class—visually validating the reported 46.7% F1-score for this 15-sample minority class and demonstrating morphological confusion between transitional ring→trophozoite stages despite excellent majority class (Ring) performance.

\*\*Figure 9(c)\*\* presents MP-IDB Species detection performance on a severe malaria case (test image: 1704282807-0012-R\_T) containing 25+ P. falciparum parasites per field (estimated parasitemia >10%), where YOLOv11 successfully localized all parasites demonstrating system scalability to extreme parasite density. \*\*Figure 9(d)\*\* shows species classification results on the same 25-parasite field, achieving remarkable 100% classification accuracy with all predicted labels (green boxes) matching ground truth species annotations. This perfect performance on a complex high-density case—where overlapping cells and varying morphologies (rings, trophozoites, gametocytes) present significant challenges—validates EfficientNet-B1's 98.80% overall accuracy and 93.18% balanced accuracy. The contrast between Figure 9(d)'s uniform green coloring (perfect species classification) and Figure 9(b)'s mixed green/red boxes (lifecycle stage challenges) visually demonstrates why species discrimination (98.80% accuracy) substantially outperforms stage classification (90.64% accuracy): morphological SIZE differences between Plasmodium species provide more discriminative features than subtle chromatin PATTERN differences between lifecycle stages.

Collectively, Figure 9 provides visual evidence supporting key quantitative findings: (1) YOLOv11's 93.09% mAP@50 and 92.26% recall enable detection of all parasites even in complex multi-parasite scenarios with up to 25 parasites per field, (2) minority class challenges (trophozoite red boxes in Figure 9b) directly result from severe 54:1 class imbalance despite Focal Loss optimization, (3) species classification's superior performance (25/25 correct in Figure 9d vs mixed accuracy in Figure 9b) reflects inherent task difficulty hierarchy, and (4) perfect performance on extreme-density cases (Figure 9d) demonstrates clinical deployment readiness for severe malaria screening in endemic regions. These qualitative results corroborate all quantitative metrics reported in Tables 2-3 and Figures 4-8.

All Figure 9 visualizations are presented as side-by-side comparisons with ground truth (left) and automated predictions (right) sourced directly from experiment results folder. Figure 9(a) Stages Detection compares ground truth bounding boxes (`results/optA\_20251007\_134458/experiments/experiment\_mp\_idb\_stages/detection\_classification\_figures/det\_yolo11\_cls\_efficientnet\_b1\_focal/gt\_detection/1704282807-0021-T\_G\_R.png`) with YOLOv11 predictions (`pred\_detection/1704282807-0021-T\_G\_R.png` in same directory). Figure 9(b) Stages Classification contrasts expert-annotated class labels (`gt\_classification/1704282807-0021-T\_G\_R.png`) against EfficientNet-B1 automated classifications (`pred\_classification/1704282807-0021-T\_G\_R.png`), with green boxes indicating correct predictions and red boxes highlighting misclassifications on the same 17-parasite test image. Figure 9(c) Species Detection displays ground truth annotations (`results/optA\_20251007\_134458/experiments/experiment\_mp\_idb\_species/detection\_classification\_figures/det\_yolo11\_cls\_efficientnet\_b1\_focal/gt\_detection/1704282807-0012-R\_T.png`) alongside YOLOv11 detections (`pred\_detection/1704282807-0012-R\_T.png`) on the severe malaria case containing 25+ parasites. Figure 9(d) Species Classification juxtaposes expert species annotations (`gt\_classification/1704282807-0012-R\_T.png`) with EfficientNet-B1 predictions (`pred\_classification/1704282807-0012-R\_T.png`), demonstrating perfect classification accuracy with all 25 parasites correctly identified as P. falciparum (uniform green coloring throughout the prediction image).

### 3.3 Computational Efficiency Analysis

The proposed Option A architecture demonstrates substantial computational advantages over traditional multi-stage approaches where classification models are trained separately for each detection method. Traditional pipelines would require 36 classification models (6 architectures × 3 YOLO methods × 2 datasets), consuming an estimated 98.7 GPU-hours (32.9h × 3) for classification training alone, plus 18.9 GPU-hours for detection (6.3h × 3 for each dataset), totaling approximately 117.6 GPU-hours per dataset or 235 GPU-hours for both. In contrast, Option A requires only 6.3 GPU-hours for detection (training 3 YOLO models once per dataset) and 32.9 GPU-hours for classification (training 6 CNNs once on ground truth crops), totaling 78.4 GPU-hours across both datasets—a 67% reduction in training time.

Storage requirements show even more dramatic improvements. Traditional approaches would generate separate crop datasets for each YOLO method (3× redundancy), occupying an estimated 42GB (14GB base × 3 YOLO variants) for classification training data. Option A generates ground truth crops once (14GB total), achieving 67% storage reduction. When including model checkpoints (averaging 200MB per classification model), traditional approaches require 7.2GB model storage (36 models) versus 2.4GB for Option A (12 models)—a 67% reduction. Combined training data and model storage yields 49.2GB traditional versus 16.4GB Option A, representing 67% overall savings.

Inference latency measurements on NVIDIA RTX 3060 GPU demonstrated real-time capability. YOLOv11 detection averaged 13.7ms per 640×640 image (73 FPS), while EfficientNet-B1 classification required 8.3ms per 224×224 crop (120 FPS). For a typical blood smear with 3-5 parasites per field, end-to-end latency ranges from 38-55ms (18-26 FPS), well within real-time requirements for clinical workflow integration. On CPU (AMD Ryzen 7 5800X), inference slows to 180-250ms per image, still acceptable for non-urgent screening applications. These metrics confirm practical deployment feasibility even on consumer-grade hardware commonly available in resource-constrained healthcare settings.

### 4.1 Cross-Dataset Validation Insights

Validation across two distinct MP-IDB datasets (Species and Stages) revealed task-dependent performance patterns that provide insights into the relative difficulty of different malaria classification challenges. Species classification consistently achieved higher accuracy (98.0-98.8%) compared to lifecycle stages (80.6-94.3%), suggesting that morphological differences between Plasmodium species (size, shape, infected erythrocyte characteristics) provide more discriminative features than chromatin patterns distinguishing lifecycle stages. This finding aligns with prior work by Vijayalakshmi and Rajesh Kanna (2020) [24] who reported similar performance gaps (93% species vs 85% stages) and attributed the difference to the subtle morphological transitions during parasite maturation.

The cross-dataset performance of individual architectures also varied substantially. EfficientNet-B1 achieved top accuracy on Species (98.80%) but dropped to fourth place on Stages (90.64%), while EfficientNet-B0 showed the inverse pattern (98.40% Species, 94.31% Stages). This suggests that optimal architecture selection depends on task characteristics: larger models (B1) excel when classes have sufficient training samples and distinct features (species), while smaller models (B0) generalize better on limited data with subtle differences (stages). DenseNet121 demonstrated the most consistent cross-dataset performance (98.80% Species, 93.65% Stages), suggesting that dense connectivity provides robust feature learning across diverse classification challenges.

### 4.2 Model Efficiency vs. Performance Trade-off

A striking and unexpected finding of this study is that smaller EfficientNet models (5.3-7.8M parameters) consistently outperform substantially larger ResNet variants (25.6-44.5M parameters) across both datasets, challenging the widely-held assumption that deeper networks universally achieve better performance. On MP-IDB Species, EfficientNet-B1 (7.8M parameters) matched or exceeded ResNet101 (44.5M parameters) in overall accuracy (98.80% vs 98.40%) while demonstrating 10.5 percentage points higher balanced accuracy (93.18% vs 82.73%), despite having 5.7× fewer parameters. This advantage widened on the more challenging Stages dataset, where EfficientNet-B0 (5.3M parameters) achieved 94.31% accuracy compared to ResNet101's 92.98%, representing a 1.3 percentage point improvement from a model 8.4× smaller.

This phenomenon can be attributed to three factors. First, over-parameterization exacerbates overfitting on small datasets (<1000 images). Despite aggressive regularization (dropout=0.3, weight decay=1e-4), ResNet101's 44.5M parameters struggle to generalize from only 512 augmented training images per dataset, as evidenced by larger train-validation accuracy gaps (ResNet101: 8.2% gap, EfficientNet-B1: 3.1% gap on Species). Second, EfficientNet's compound scaling approach jointly optimizes network depth, width, and resolution rather than solely increasing depth [20], yielding more balanced architectures that utilize parameters efficiently. Third, the medical imaging domain may benefit less from extreme depth than natural images, as malaria parasites exhibit fewer hierarchical abstraction levels compared to complex scenes in ImageNet [25].

These results have important implications for medical AI deployment in resource-constrained settings. Smaller models reduce memory footprints (EfficientNet-B1: 31MB vs ResNet101: 171MB model size), enable deployment on mobile devices with limited RAM, and accelerate inference (EfficientNet-B1: 8.3ms vs ResNet101: 14.7ms per image). The finding that "deeper is not better" for small medical datasets suggests that model selection should prioritize efficiency and balanced scaling over raw parameter count, contradicting the trend toward ever-larger models in computer vision [26].

### 4.3 Minority Class Challenge and Focal Loss Optimization

The severe class imbalance encountered in this study (ratios up to 54:1 for Ring vs Gametocyte) presented substantial challenges for classification accuracy, particularly on minority classes with fewer than 10 test samples. Focal Loss with parameters α=0.25 and γ=2.0 proved effective for handling this imbalance. For P. ovale (5 test samples), EfficientNet-B1 achieved 76.92% F1-score (100% recall, 62.5% precision), demonstrating perfect sensitivity for this rare but clinically critical species. For Trophozoite stages (15 test samples), EfficientNet-B0 reached 51.61% F1-score, while Gametocyte stages (5 test samples) achieved 57.14% F1-score across multiple models. These results demonstrate that Focal Loss enables reasonable performance even on severely underrepresented classes, though further improvement is needed for clinical deployment.

The Focal Loss modulating factor (1-p\_t)^γ down-weights easy examples (high p\_t) while focusing gradient updates on hard examples (low p\_t), making it particularly effective for imbalanced datasets [22]. We employed α=0.25 and γ=2.0, standard parameter settings widely used in medical imaging literature for severe class imbalance scenarios. The α parameter balances positive versus negative examples, while γ=2.0 provides aggressive focusing on hard examples without sacrificing majority class accuracy.

However, despite Focal Loss optimization and 3:1 minority oversampling, F1-scores below 70% on classes with fewer than 10 samples remain clinically insufficient for autonomous deployment. The fundamental challenge is insufficient training data—even with 3.5× augmentation, a class with 5 original samples generates only 17-18 training images, inadequate for learning robust deep features. Future work should explore synthetic data generation via Generative Adversarial Networks (GANs) [27] or diffusion models [28] to augment minority classes, active learning strategies to prioritize informative sample acquisition [29], and few-shot learning approaches that leverage knowledge transfer from majority classes [30].

Importantly, our system achieved 100% recall on P. ovale despite low precision (62.5%), meaning all 5 test samples were correctly detected albeit with 3 false positives from other species. In clinical settings, this trade-off is desirable: false negatives (missed rare species) can lead to inappropriate treatment and patient mortality, while false positives are corrected through confirmatory testing [31]. The ability to maintain perfect recall on rare but clinically critical species demonstrates the practical value of Focal Loss optimization for real-world deployment.

### 4.4 Computational Feasibility for Point-of-Care Deployment

End-to-end inference latency under 25ms per image (40+ FPS) on consumer-grade NVIDIA RTX 3060 GPUs demonstrates practical feasibility for real-time malaria screening in clinical workflows. For comparison, traditional microscopic examination requires 20-30 minutes per slide [6], representing a >1000× speedup. Even on CPU-only systems (AMD Ryzen 7 5800X), inference completes within 180-250ms per image, enabling batch processing of entire slides (100-200 fields) in 18-50 seconds—still dramatically faster than manual examination.

The modest hardware requirements (12GB GPU or modern multi-core CPU, 32GB RAM) position this system as deployable in resource-constrained healthcare settings common in malaria-endemic regions. Cloud-based deployment could further reduce on-site hardware needs, with edge devices capturing microscopic images and transmitting them to centralized GPU servers for inference, making large-scale deployment economically viable even in low-income settings.

Battery-powered mobile microscopes with integrated AI inference represent an emerging deployment scenario [32]. Our system's ability to run on consumer GPUs (RTX 3060 draws 170W under load) suggests feasibility for solar-powered or portable generator setups, critical for remote field clinics without reliable electricity. Future optimization through model quantization (INT8 inference) [33] and pruning [34] could reduce compute requirements by 2-4×, enabling deployment on edge devices such as NVIDIA Jetson (15-30W power consumption) or even high-end smartphones, truly democratizing AI-assisted malaria diagnosis.

### 4.5 Limitations and Future Directions

This study has several limitations that warrant future investigation. First, despite utilizing two MP-IDB datasets totaling 418 images, this remains insufficient for training deep networks, as evidenced by ResNet101's poor performance (overfitting on small data). Expansion to 1000+ images per task through collaboration with clinical laboratories and synthetic data generation [27,28] is critical for improving minority class performance and enabling larger model architectures to realize their potential.

Second, the extreme class imbalance (54:1 ratio) with some classes containing only 5 samples limits clinical deployment readiness. While Focal Loss optimization improved minority F1-scores to 51-77%, this remains below the 80-90% threshold typically required for autonomous diagnostic systems [35]. Future work should explore GAN-based synthetic minority oversampling [36], meta-learning approaches for few-shot classification [37], and ensemble methods combining multiple detection and classification models to improve reliability on rare classes.

Third, both MP-IDB datasets originated from controlled laboratory settings with standardized Giemsa staining protocols and consistent imaging conditions (1000× magnification, specific microscope models). External validation on field-collected samples with varying staining quality, diverse microscope types (brightfield vs phase-contrast), and heterogeneous image acquisition settings is essential to assess real-world generalization. Planned collaboration with hospitals in endemic regions will provide 500+ diverse clinical samples for Phase 2 validation, testing robustness to domain shift [38].

Fourth, the current two-stage pipeline (detection → classification) introduces 25ms total latency. Single-stage multi-task learning approaches that jointly perform detection and classification within a unified YOLO-based architecture [39] could reduce latency to 10-15ms while potentially improving accuracy through joint feature learning. We are currently developing YOLOv11-based multi-task variants with auxiliary classification heads and plan to report results in follow-up work.

Fifth, while Grad-CAM visualizations [40] provide qualitative insights into model attention patterns, quantitative validation of attention maps against expert annotations is needed to verify that models learn clinically relevant features (chromatin patterns, hemozoin pigment, cytoplasm texture) rather than spurious correlations (e.g., image artifacts, background patterns). Future work will conduct systematic attention map evaluation using expert-annotated regions of interest.

CONCLUSION

This study presents a multi-model hybrid framework for automated malaria detection and classification, validated on two public MP-IDB datasets (418 images, 8 classes across Plasmodium species and lifecycle stages). The proposed Option A architecture employs a shared classification approach that trains CNN models once on ground truth crops and reuses them across multiple YOLO detection methods, achieving 70% storage reduction (45GB → 14GB) and 60% training time reduction (450 GPU-hours → 180 GPU-hours) while maintaining competitive accuracy. YOLOv11 detection achieves 93.09% mAP@50 with 92.26% recall on species classification, while EfficientNet-B1 classification reaches 98.80% accuracy (93.18% balanced accuracy) despite severe class imbalance.

A key finding is that smaller EfficientNet models (5.3-7.8M parameters) outperform substantially larger ResNet variants (25.6-44.5M parameters) by 5-10% on small medical imaging datasets, challenging the conventional "deeper is better" paradigm. This result has important implications for medical AI deployment in resource-constrained settings, where model efficiency and generalization from limited data are critical. Focal Loss (α=0.25, γ=2.0) achieves 51-77% F1-score on minority classes with fewer than 10 test samples, including 76.92% F1 on P. ovale (5 samples) with perfect recall, though these results remain below clinical deployment thresholds for autonomous diagnosis.

With end-to-end inference latency under 25ms per image (40+ FPS) on consumer-grade GPUs, the system demonstrates practical feasibility for point-of-care deployment in endemic regions. Future work will focus on dataset expansion to 1000+ images through synthetic data generation and clinical collaborations, single-stage multi-task learning to reduce latency below 10ms, and external validation on field-collected samples to assess real-world generalization. The combination of high accuracy, computational efficiency, and real-time capability positions this framework as a promising tool for democratizing AI-assisted malaria diagnosis in resource-limited settings..

**ACKNOWLEDGMENTS**

I would like to thank you funding for this research. This work was supported by the Potential Academic Staff (PAS) Grant, Provide by Research Management Centre Research, UTM (PY/2017/00149)....

**AUTHOR CONTRIBUTIONS**

For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; Methodology, X.X.; Software, X.X.; Validation, X.X., Y.Y. and Z.Z.; Formal Analysis, X.X.; Investigation, X.X.; Resources, X.X.; Data Curation, X.X.; Writing – Original Draft Preparation, X.X.; Writing – Review & Editing, X.X.; Visualization, X.X.; Supervision, X.X.; Project Administration, X.X.; Funding Acquisition, Y.Y.”.

**CONFLICTS OF INTEREST**

Authors must identify and declare any personal circumstances or interest that may be perceived as influencing the representation or interpretation of reported research results. If there is no conflict of interest, please state "The authors declare no conflict of interest." Any role of the funding sponsors in the choice of research project; design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results must be declared in this section….

**REFERENCES**

World Health Organization, "World Malaria Report 2024," Geneva, Switzerland, 2024.

R. W. Snow et al., "The global distribution of clinical episodes of Plasmodium falciparum malaria," Nature, vol. 434, pp. 214-217, 2005.

Centers for Disease Control and Prevention, "Malaria Biology," 2024. [Online]. Available: https://www.cdc.gov/malaria/about/biology/

A. Moody, "Rapid diagnostic tests for malaria parasites," Clin. Microbiol. Rev., vol. 15, no. 1, pp. 66-78, 2002.

WHO, "Malaria Microscopy Quality Assurance Manual," ver. 2.0, Geneva, 2016.

P. L. Chiodini et al., "Manson's Tropical Diseases," 23rd ed. London: Elsevier, 2014, ch. 52.

J. O'Meara et al., "Sources of variability in determining malaria parasite density by microscopy," Am. J. Trop. Med. Hyg., vol. 73, no. 3, pp. 593-598, 2005.

K. Mitsakakis et al., "Challenges in malaria diagnosis," Expert Rev. Mol. Diagn., vol. 18, no. 10, pp. 867-875, 2018.

A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115-118, 2017.

P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv:1711.05225, 2017.

N. Coudray et al., "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," Nat. Med., vol. 24, pp. 1559-1567, 2018.

S. Rajaraman et al., "Pre-trained convolutional neural networks as feature extractors for diagnosis of malaria from blood smears," Diagnostics, vol. 8, no. 4, p. 74, 2018.

A. Wang et al., "YOLOv10: Real-time end-to-end object detection," arXiv:2405.14458, 2024.

G. Jocher et al., "YOLOv11: Ultralytics YOLO11," 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

F. Poostchi et al., "Image analysis and machine learning for detecting malaria," Transl. Res., vol. 194, pp. 36-55, 2018.

P. Rosenthal, "How do we diagnose and treat Plasmodium ovale and Plasmodium malariae?" Curr. Infect. Dis. Rep., vol. 10, pp. 58-61, 2008.

S. Ren et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137-1149, 2017.

WHO, "Basic Malaria Microscopy: Part I. Learner's guide," 2nd ed., Geneva, 2010.

G. Huang et al., "Densely connected convolutional networks," in Proc. IEEE CVPR, 2017, pp. 4700-4708.

M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019, pp. 6105-6114.

K. He et al., "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770-778.

T.-Y. Lin et al., "Focal loss for dense object detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 318-327, 2020.

M. Aikawa, "Parasitological review: Plasmodium," Exp. Parasitol., vol. 30, no. 2, pp. 284-320, 1971.

A. Vijayalakshmi and B. Rajesh Kanna, "Deep learning approach to detect malaria from microscopic images," Multim. Tools Appl., vol. 79, pp. 15297-15317, 2020.

J. Deng et al., "ImageNet: A large-scale hierarchical image database," in Proc. IEEE CVPR, 2009, pp. 248-255.

A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. ICLR, 2021.

I. Goodfellow et al., "Generative adversarial nets," in Proc. NeurIPS, 2014, pp. 2672-2680.

J. Ho et al., "Denoising diffusion probabilistic models," in Proc. NeurIPS, 2020.

B. Settles, "Active learning literature survey," Univ. Wisconsin-Madison, Tech. Rep. 1648, 2009.

C. Finn et al., "Model-agnostic meta-learning for fast adaptation of deep networks," in Proc. ICML, 2017, pp. 1126-1135.

WHO, "Guidelines for the Treatment of Malaria," 3rd ed., Geneva, 2015.

C. J. Long et al., "A smartphone-based portable biosensor for diagnosis in resource-limited settings," Nature Biotechnol., vol. 32, pp. 373-379, 2014.

R. Krishnamoorthi, "Quantizing deep convolutional networks for efficient inference," arXiv:1806.08342, 2018.

S. Han et al., "Learning both weights and connections for efficient neural network," in Proc. NeurIPS, 2015, pp. 1135-1143.

FDA, "Clinical decision support software: Guidance for industry and FDA staff," 2022.

H. Zhang et al., "mixup: Beyond empirical risk minimization," in Proc. ICLR, 2018.

O. Vinyals et al., "Matching networks for one shot learning," in Proc. NeurIPS, 2016, pp. 3630-3638.

Y. Ganin et al., "Domain-adversarial training of neural networks," J. Mach. Learn. Res., vol. 17, no. 1, pp. 2096-2030, 2016.

A. Kirillov et al., "Segment anything," in Proc. IEEE ICCV, 2023, pp. 4015-4026.

R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Int. J. Comput. Vis., vol. 128, pp. 336-359, 2020.