# Bridging the Gap in Malaria Diagnostics: An Attention-Centric YOLO Framework with Species-Specific Augmentation for Tiny Parasite Detection in Low-Resource Settings

**Ahmed Tahiru Issah, Carine Mukamakuza***
Carnegie Mellon University Africa
Kigali, Rwanda
aissah@andrew.cmu.edu, cmukamak@andrew.cmu.edu

**Abstract**

Malaria remains a major health challenge in Africa, with accurate species identification critical for prompt treatment and control, especially in resource-limited settings like Rwanda. This study systematically benchmarks state-of-the-art detection architectures for automated multi-species malaria parasite detection from high-resolution Giemsa-stained blood smear images, addressing the persistent problems of class imbalance and the difficulty of detecting small Plasmodium falciparum ring forms. We evaluate and optimize YOLO-SPAM, YOLO-Para, and YOLOv12 models, applying a novel species-specific augmentation protocol with copy-paste and noise injection. YOLOv12, trained with these protocols, achieves outstanding overall performance ($0.878$ $\mathrm{mAP}_{50}$, $0.677$ $\mathrm{mAP}_{50-95}$) and demonstrates significant improvement in detecting small, clinically relevant P. falciparum parasites over non-attention YOLO methods). Comparative experiments reveal that targeted data augmentation and strategic model selection can overcome significant class imbalance, achieving reliable multi-species differentiation and precise localization even for morphologically subtle classes. Our findings validate the practical promise of advanced object detection models, such as YOLOv12, for malaria diagnosis workflows and support their future deployment in real-world microscopy labs in Rwanda and similar endemic regions.

## 1 Introduction

Malaria remains a critical global health burden, with the WHO estimating approximately 627,000 deaths in 2022, predominantly in sub-Saharan Africa where *Plasmodium falciparum* drives severe disease morbidity and mortality (WHO 2016). Rwanda exemplifies the endemic challenge in resource-limited settings, with $40-45\%$ of malaria diagnoses performed via microscopy and the remainder via rapid diagnostic tests (RDTs) (PMI 2023). Accurate, timely parasite species identification is essential for treatment initiation, as different *Plasmodium* species (*P. falciparum*, *P. malariae*, *P. ovale*, *P. vivax*) exhibit distinct clinical presentations and drug sensitivities. However, conventional microscopy-based diagnosis remains time-consuming, operator-dependent, and prone to inter-observer variability (Tangpukdee et al. 2009), particularly when identifying morphologically distinct parasite species and detecting early-stage ring-form trophozoites that appear strikingly similar across species.

While convolutional neural networks and recent YOLO-based object detection architectures have demonstrated success on controlled datasets, several critical challenges persist for clinical deployment: (1) **multi-species detection** across morphologically diverse parasites with comprehensive species differentiation remains underexplored in the literature; (2) **extreme class imbalance** with minority species notably the *P. vivax* severely underrepresented in most datasets; (3) **detection of small, difficult parasite structures**, particularly *P. falciparum* ring-stage trophozoites exhibiting subtle morphology requiring pixel-level precision (Bogale, Mukamakuza, and Tuyishimire 2024); and (4) **computational efficiency** for deployment in resource-constrained settings with limited hardware infrastructure.

This study addresses these challenges through comprehensive evaluation of state-of-the-art attention-based You Only Look Once (YOLO) architectures combined with strategic training protocols specifically designed for the malaria detection task. Key contributions include:

- **Species-specific augmentation strategy** addressing extreme class imbalance through differential augmentation factors, demonstrating improvements relative to unaugmented baseline
- **Optimized training protocols** leveraging copy-paste augmentation with noise injection, enabling YOLOv12 to achieve improvements in spatial localization precision for the clinically challenging *P. falciparum* ring forms
- **Comprehensive ablations and experiments** across YOLO-SPAM, YOLO-Para, and YOLOv12, establishing YOLOv12 as optimal for multi-species detection with benchmark high performance as compared to our previous works

## 2 Related Work

### 2.1 Malaria Diagnosis: Traditional and Digital Approaches

Microscopy remains the diagnostic gold standard for malaria, requiring expert microscopists to identify parasites in Giemsa-stained blood smears through morphological assessment of parasite life stages and species-specific charac-

*Corresponding author.

teristics (Tangpukdee et al. 2009). This approach faces significant limitations: inter-observer variability, requirement for highly trained personnel, time-consuming manual review, and limited availability in resource-constrained settings (Wongsrichanalai et al. 2007). Rapid diagnostic tests (RDTs) provide faster alternatives but lack species differentiation capability, which is critical for treatment selection (CDC 2024). Digital microscopy and automated image analysis offer standardization and scalability advantages, motivating investigation of computer-aided detection systems.

### 2.2 Deep Learning for Malaria Parasite Detection

Convolutional neural network (CNN) approaches including ResNet, DenseNet, and custom CNN architectures have demonstrated promise for malaria classification tasks. More sophisticated CNN-based approaches incorporating image preprocessing, data augmentation, and transfer learning achieved improved results; however, these remain limited to controlled datasets with balanced species representation and single-parasite analysis rather than multi-species detection in realistic thick blood smear scenarios (Ramos-Briceño et al. 2025).

### 2.3 Object Detection Architectures for Parasite Detection

YOLO family models revolutionized real-time object detection through single-stage detection frameworks achieving practical inference speeds. YOLOv5 and earlier versions demonstrated moderate success on malaria detection tasks. YOLO-PAM introduced parasite-specific attention mechanisms through Convolutional Block Attention Module (CBAM) and Normalized Attention Module (NAM) modules, achieving good performance on the MP-IDB dataset with 29.8M parameters (Zedda, Loddo, and Di Ruberto 2023). Extended YOLO-SPAM variants specialized in early-stage parasite detection, improved performance through novel prediction head design utilizing lower-resolution backbone features (Zedda, Loddo, and Di Ruberto 2024). YOLO-Para series with multi-scale prediction heads demonstrated impressive performance across multiple Plasmodium species. These prior works established the effectiveness of YOLO-based approaches for malaria parasite detection.

Instance segmentation extends detection with pixel-level precision beyond bounding boxes, enabling morphological analysis critical for species identification and life-stage characterization (He et al. 2018). Faster R-CNN and Mask R-CNN demonstrated high performance on medical imaging tasks (Davidson et al. 2021), but computational complexity and limited real-time inference capability restrict deployment in resource-constrained settings.

### 2.4 Attention Mechanisms in Detection

Attention mechanisms have proven effective for improving feature representation in detection tasks. CBAM (Convolutional Block Attention Module) and NAM (Normalized Attention Module) provide channel and spatial attention without excessive computational overhead (Zedda, Loddo, and Di Ruberto 2023). Recent Swin Transformer-based designs (C3STR layers) and other transformer variants have been integrated into detection pipelines with demonstrated performance gains (Zedda, Loddo, and Di Ruberto 2024). YOLOv12 introduces Area Attention (A2) mechanism that partitions feature maps into equal-sized regions, reducing computational complexity to $O(n^2hd/2)$ compared to global attention while maintaining larger receptive fields, enabling real-time inference comparable to CNN-based predecessors (Tian, Ye, and Doermann 2025).

### 2.5 Data Augmentation for Class Imbalance

Class imbalance is endemic in medical imaging, with minority classes severely underrepresented. Traditional approaches apply uniform augmentation across all classes, but targeted species-specific augmentation factors have been theoretically proposed with limited empirical validation on real malaria datasets. Copy-paste augmentation, developed for instance segmentation, leverages instance-level mixing to improve detection in crowded scenes by pasting random instances into images with appropriate blending, achieving improvement on LVIS rare categories (Ghiasi et al. 2021).

## 3 Methodology

### 3.1 Dataset Acquisition

Giemsa-stained thick blood smear samples were collected from the Rwanda Biomedical Center (RBC), the national reference laboratory coordinating malaria diagnostic services. Thick smear preparations were selected for this study because, they exhibit approximately 11 times higher sensitivity compared to thin smears, facilitating more robust parasite detection in automated systems (Poostchi et al. 2018). Blood samples were collected from patients presenting with fever at healthcare facilities, and positive slides were subsequently transferred to RBC for analysis and research. This was done according to established protocols as detailed in our team's previous work by (Akpo, Mukamakuza, and Tuyishimire 2024). The iterative validation process involved expert microscopists reviewing all annotated data, with corrections and feedback incorporated until consensus agreement was achieved.

Images were captured at $100\times$ magnification using an **Olympus camera installed on the microscope's eyepiece**. Image data were systematically collected through standardized field-of-view scanning with horizontal and vertical adjustments to ensure dataset diversity. All images were managed through the Vitruvianmd web application platform designed specifically for medical data curation.

The initial dataset comprised approximately $6,000$ images collected through systematic scanning. However, the final validated dataset, used for all model training, contained $2,739$ **rigorously validated images**: *Plasmodium Falciparum* PF (838), *Plasmodium Malariae* PM (834), *Plasmodium Ovale* PO (893), and *Plasmodium Vivax* PV (174). Of these, $7,568$ parasite instances were labeled for PF, $1,802$ for PM, $2,353$ for PO, and 669 for PV. The remaining $\sim 3,260$ images continue undergoing expert validation for future work. Dataset annotations were done with both **bound-**

**ing boxes and polygonal masks** using the VGG Image Annotator (VIA), enabling flexible deployment across object detection and instance segmentation architectures.

To facilitate multi-format model training, the validated dataset was uploaded to Roboflow, a collaborative platform supporting dynamic format conversion including YOLO, COCO, and JSON formats, among others. Species-specific preprocessing and augmentation were applied separately, enabling species-targeted augmentation strategies.

### 3.2 Image Preprocessing

**Auto-Orientation**: The first preprocessing step corrected EXIF rotation metadata embedded by digital microscopes. Although pixels are stored identically regardless of physical camera orientation, EXIF directives signal display orientation, potentially causing inconsistent coordinate systems (e.g., some images as $(x, y)$, others as $(y, x)$) without proper handling. Auto-orientation standardized pixel ordering across the entire dataset, eliminating systematic bias (Roboflow 2020).

**Resizing**: All images were resized to $2048 \times 2048$ pixels using the **"Fit (black edges)"** strategy on Roboflow, the letter-boxing approach that preserves aspect ratio and morphological integrity. This approach resized the longest dimensions to $2048$ pixels while proportionally scaling the shorter dimensions, padding remaining space with black (zero-value) pixels. This strategy was selected because it:

- Preserves parasite morphology (critical diagnostic features including ring forms and gametocyte crescents).
- Avoids geometric and shape distortions that stretching or cropping resize methods would introduce.

**Species-Specific Filtering**: During preprocessing, images were filtered to include only the target species images. This species-specific filtering enabled targeted augmentation strategies addressing species-specific class imbalance.

### 3.3 Species-Specific Data Augmentation

**Augmentation Strategy and Rationale** Traditional uniform augmentation applies identical parameters across all classes. However, the validated dataset exhibited severe class imbalance, where PV represented only $6.4\%$ of images while other species showed higher prevalence. We implemented **species-specific parameterized augmentation** applying different augmentation factors proportional to underrepresentation:

- **3× augmentation** for PF, PM, PO
- **10× augmentation** for PV (Roboflow's maximum augmentation factor)

This strategy simultaneously addresses class imbalance and expands training data. For each training image, Roboflow randomly selected augmentation combinations with uniformly random parameter values within specified ranges, generating the designated number of augmented variants per original image.

**Augmentation Parameters** The augmentations that were applied are given in Table 1, with parameters informed by the YOLO-PAM malaria detection framework (Zedda, Loddo, and Di Ruberto 2023):

| Augmentation | Parameters | Rationale |
|---|---|---|
| **Rotation** | $90°, 180°, 270°$ | Simulates different microscope orientations |
| **Hue** | $[-20, +20]$ | Accounts for stain and lighting variation |
| **Saturation** | $[-30\%, 30\%]$ | Represents Giemsa staining batch variation |
| **Brightness** | $[-20\%, +20\%]$ | Simulates microscope illumination differences |

Table 1: Augmentation parameters applied during species-specific data augmentation

These parameters matched the YOLO-PAM approach for malaria blood smears, where similar ranges improved performance without introducing unrealistic artifacts. We excluded augmentations such as shearing, that YOLO-PAM explicitly discouraged due to introducing geometric and shape distortions inconsistent with biological parasites (Zedda, Loddo, and Di Ruberto 2023).

**Data Split and Integration** Images were organized into $70\%$ **training,** $15\%$ **validation,** $15\%$ **test** splits. Roboflow applied augmentations exclusively to the training split while applying only preprocessing to validation/test sets, preventing data leakage. Each species dataset underwent independent preprocessing and augmentation, generating four species-specific versions. Although each contained only 2 labels (species + white blood cells (WBC)), these were integrated into a unified 5-label dataset through sequential upload to a new Roboflow project. Upon first upload, Roboflow identified and stored labels; subsequent uploads automatically assigned consistent WBC IDs while assigning unique species IDs.

**Final augmented dataset composition**:

The composition of the final augmented dataset is given in Table 2

| Species | Orig. Train | Aug. Factor | Aug. Train | Val | Test |
|---|---|---|---|---|---|
| PF | 586 | 3× | 1,761 | 125 | 126 |
| PM | 583 | 3× | 1,752 | 125 | 125 |
| PO | 625 | 3× | 1,875 | 134 | 134 |
| PV | 121 | 10× | 1,210 | 26 | 26 |
| **Total** | **1,915** | **-** | **6,598** | **410** | **411** |

Table 2: Final augmented dataset composition after species-specific augmentation

### 3.4 Model Architectures

**YOLOv12** Represents a significant architectural departure from previous YOLO iterations. **YOLOv12 introduces an attention-centric architecture** replacing CNN-based detection with hybrid designs leveraging modern attention

mechanisms (Tian, Ye, and Doermann 2025), as shown in Figure 1.

**Area Attention Mechanism**: The central innovation is Area Attention (A2), which partitions feature maps into equal-sized regions (default: 4 regions), applying attention independently within each region. This reduces computational complexity from $2n^2hd$ to $\frac{1}{2}n^2hd$ while retaining larger receptive fields than other local attention mechanisms, enabling real-time inference (Tian, Ye, and Doermann 2025).

**Residual Efficient Layer Aggregation Networks (R-ELAN)**: R-ELAN introduces block-level residual connections with a scaling factor (default: 0.01), redesigned feature aggregation creating bottleneck-like structures, and enhanced gradient flow for deep attention networks (Tian, Ye, and Doermann 2025).

**Segmentation Variant Selection**: YOLOv12n-seg extends the base detection architecture with instance segmentation capabilities, generating both bounding box predictions and pixel-level segmentation masks. The segmentation variant was selected for malaria parasite detection based on critical advantages for medical imaging applications:

1. **Pixel-Level Morphological Precision**: Instance segmentation provides precise, pixel-level delineation of parasite boundaries beyond coarse bounding boxes, which is crucial for malaria diagnosis, where parasite morphology directly informs species identification (Das, Mukherjee, and Chakraborty 2015).
2. **Handling Dense and Overlapping Parasites**: By generating pixel-level masks for each parasite instance, instance segmentation can disambiguate overlapping objects that detection-only models may miss (Chen, Wu, and Merhof 2022).
3. **Enhanced Feature Representation**: The segmentation variant incorporates specialized segmentation heads and refined feature aggregation modules, designed for pixel-level tasks, leading to better overall feature understanding compared to models optimized solely for bounding box regression (Molina et al. 2025).

**YOLO-PAM (Parasite-Attention Model):** Builds on YOLOv8m with strategically integrated attention mechanisms specifically optimized for malaria parasite detection. The architecture incorporates Convolutional Block Attention Module (CBAM), Normalized Attention Module (NAM), and C3STR (C3 layers enhanced with Swin Transformer blocks) positioned in the backbone and neck. YOLO-PAM excludes prediction heads for large objects, focusing computational resources on clinically relevant parasite scales. The model outperforms YOLOv8m while maintaining comparable inference efficiency (Zedda, Loddo, and Di Ruberto 2023).

**YOLO-SPAM and YOLO-SPAM++:** YOLO-SPAM and YOLO-SPAM++ extend malaria detection capabilities by specializing in early-stage parasite detection, particularly ring-form parasites critical for prompt diagnosis (Zedda, Loddo, and Di Ruberto 2024). Built on YOLOv5m6 baseline, YOLO-SPAM introduces a novel prediction head that

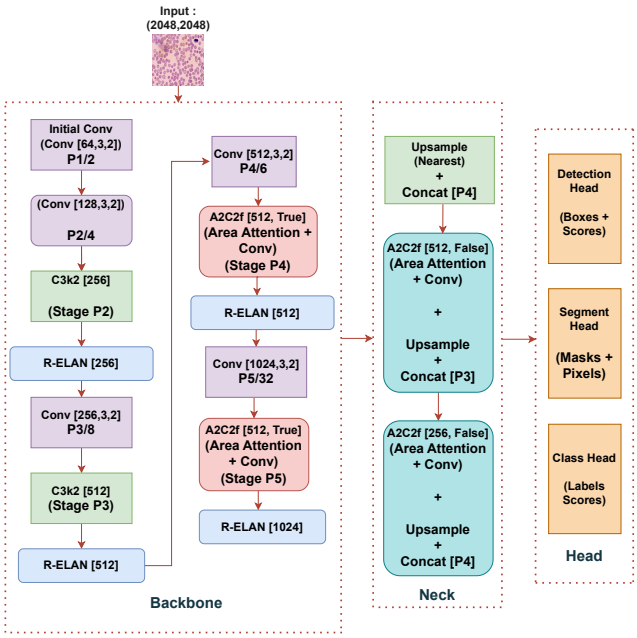


Figure 1: Architecture diagram of YOLOv12

directly utilizes lower-resolution backbone features, enabling detection of parasites occupying minimal pixel space. YOLO-SPAM++ further enhances this architecture through NAM and C3STR module integration with feature merging strategies, achieving performance improvement over YOLO-SPAM. Despite more sophisticated attention mechanisms, YOLO-SPAM++ maintains parameter efficiency with $23.6M$ parameters ($17.6M$ fewer than YOLOv5m6), making both architectures suitable for resource-constrained environments (Zedda, Loddo, and Di Ruberto 2024).

**YOLO-Para Series:** The YOLO-Para series (SP, SMP, AP) represents new advancement designed for realistic clinical scenarios where parasites of varying sizes must be simultaneously detected (Zedda, Loddo, and Di Ruberto 2025). Built on YOLOv8, the series introduces progressive complexity through multiple specialized prediction heads: YOLO-Para SP (1 head, small parasites), YOLO-Para SMP (2 heads, small and medium), and YOLO-Para AP (5 heads, all sizes). This design enables comprehensive detection across the complete spectrum of parasite morphologies, demonstrating good multi-species performance (Zedda, Loddo, and Di Ruberto 2025).

## 3.5 Training Configuration

All models were trained using PyTorch with consistent training protocols across architectures. A baseline experiment using the original unaugmented $2,739$-image dataset at $640 \times 640$ resolution established a reference point to isolate and quantify the impact of species-specific augmentation. All subsequent experiments employed $2048 \times 2048$ resolution for enhanced parasite morphology preservation. Models were trained for **70 epochs** across all experiments, with batch sizes optimized to maximize GPU utilization.

**Loss Function**: The training objective combined multiple complementary loss components with learned weighting, following the YOLO framework (Tian, Ye, and Doermann 2025):

$$L_{total} = \lambda_{cls} L_{cls} + \lambda_{box} L_{box} + \lambda_{dfl} L_{dfl} + \lambda_{mask} L^*_{mask} \tag{1}$$

where the default loss weights are:

- **Classification loss** ($L_{cls}$, weight $\lambda_{cls} = 0.5$): Cross-entropy loss for classification of parasite species and white blood cells
- **Box regression loss** ($L_{box}$, weight $\lambda_{box} = 7.5$): CIoU/-GIoU loss for accurate bounding box localization
- **Distribution Focal Loss** ($L_{dfl}$, weight $\lambda_{dfl} = 1.5$): Improves localization precision by optimizing the distribution of localization errors across bounding box coordinates
- **Mask loss** ($L^*_{mask}$, for segmentation variants only): Binary cross-entropy or Dice loss for pixel-level instance segmentation prediction

All training was monitored and logged using Weights & Biases for comprehensive reproducibility and hyperparameter tracking.

### 3.6 Evaluation Metrics

**Mean Average Precision (mAP)**: Primary detection metric computed as:

$$\text{mAP} = \frac{1}{C} \sum_{c=1}^{C} \text{AP}_c \tag{2}$$

where $C$ is the number of classes and $\text{AP}_c$ is the average precision for class $c$ computed from the precision-recall curve at IoU threshold $\tau$. Two mAP variants are reported:

- $\mathbf{mAP_{50}}$: Average Precision at IoU threshold $0.5$, measuring loose spatial overlap
- $\mathbf{mAP_{50-95}}$: Average Precision averaged across IoU thresholds from $0.5$ to $0.95$ at $0.05$ intervals, providing comprehensive spatial accuracy assessment

**Per-Class Metrics**: Precision and Recall were computed separately for each parasite species and white blood cells. Precision quantifies the proportion of positive predictions that are correct, while Recall quantifies the proportion of actual positives that were correctly identified.

**Segmentation Metrics** (YOLOv12n-seg): The aforementioned metrics were computed for both bounding box and mask. For fair comparison with detection-only models (YOLO-SPAM, YOLO-Para), all reported YOLOv12 results represent bounding box metrics rather than mask metrics.

**Test Set Evaluation**: All reported results represent performance on the held-out test set (411 images, unaugmented) using default YOLO validation parameters (confidence threshold: $0.001$, IoU threshold: $0.6$, batch size: $1$). This protocol ensures unbiased assessment on clinically realistic data without augmentation, representing realistic deployment conditions.

## 4 Experiments and Results

### 4.1 Experimental Setup

All models were trained on a single **H100 GPU with 80GB VRAM**, enabling high-resolution ($2048 \times 2048$) image processing. All training employed 70 epochs for consistent convergence across models.

**Training Hyperparameters** Standard training hyperparameters were held constant across all experiments unless explicitly noted. The parameters used are given in Table 3.

| Hyperparameter | Value |
|---|---|
| **Epochs** | 70 |
| **Input Image Size** | 2048* |
| **Batch Size** | 6* |
| **Learning Rate (lr0)** | 0.01 |
| **Optimizer** | SGD |
| **Momentum** | 0.937 |
| **Weight Decay** | $5e-4$ |
| **Automatic Mixed Precision (AMP)** | Enabled |
| **Mosaic Augmentation** | Variant |
| **Copy-Paste Augmentation** | Variant |
| **HSV Online Augmentation** | Disabled $(0, 0, 0)$ |

Table 3: Standard training hyperparameters held constant across experiments (*Baseline augmentation impact experiment used 640 image size and batch size of 64)

**Model Evaluation Protocol** All reported results are **test set performance** evaluated on the held-out test split (411 images, unaugmented) using default YOLO validation parameters:

- **Confidence threshold**: 0.001
- **IoU threshold**: 0.6
- **Batch size**: 1
- **Image size**: 2048 (except in initial baseline experiment which was set to 640)

This evaluation protocol ensures unbiased assessment on clinically realistic data without augmentation, representing realistic deployment conditions.

### 4.2 Augmentation Impact Analysis

To quantify the contribution of species-specific augmentation, we compared performance on the original unaugmented dataset versus the augmented dataset using the **YOLO-SPAM 2H configuration**. Both experiments used same hyperparameters, and $640 \times 640$ resolution for fair comparison. The overall and per-species results for these experiments are given in Table 4 and Table 5 respectively.

| Dataset | Train Samples | $\text{mAP}_{50}$ | $\text{mAP}_{50-95}$ |
|---|---|---|---|
| Original | 1,915 | 0.779 | 0.493 |
| Augmented | 6,598 | 0.816 | 0.537 |
| **Improvement** | $+\mathbf{244}\%$ | $+\mathbf{4.7}\%$ | $+\mathbf{8.9}\%$ |

Table 4: Overall Augmentation Impact Results

| | $\mathbf{mAP_{50}}$ | | $\mathbf{mAP_{50-95}}$ | |
|---|---|---|---|---|
| | Original | Augmented | Original | Augmented |
| **PF** | 0.683 | 0.671 | 0.412 | 0.375 |
| **PM** | 0.607 | 0.704 | 0.276 | 0.462 |
| **PO** | 0.857 | 0.869 | 0.585 | 0.609 |
| **PV** | 0.823 | 0.901 | 0.533 | 0.527 |
| **WBC** | 0.925 | 0.933 | 0.658 | 0.712 |

Table 5: Per-Species Augmentation Impact Results

This comparison demonstrates that species-specific augmentation directly improves model training, with particularly strong gains for underrepresented species. Per-species improvements include PM (+16% $\mathrm{mAP}_{50}$: 0.607→0.704), PO (+1.4% $\mathrm{mAP}_{50}$: 0.857→0.869), and PV (+9.5% $\mathrm{mAP}_{50}$: 0.823→0.901), validating the effectiveness of $3\times$ and $10\times$ augmentation factors in addressing class imbalance while improving dataset diversity.

### 4.3 Model Architecture Comparison

After the augmentation impact analysis, three model architectures were systematically investigated: YOLO-Para, YOLOv12, and a variant of YOLOv12 with copy-paste online augmentation applied (YOLOv12-CP). In this section, we report the best-performing configuration for each architecture based on overall performance across all metrics.

**Best-Performing Configurations Summary** All configurations maintained base hyperparameters in Table 3, with additional parameters shown in Table 6. The YOLOv12-CP variant incorporates copy-paste augmentation, available only for segmentation models, to demonstrate its additional benefit.

| Model | Mosaic | Close Mosaic | Copy-Paste | Noise |
|---|---|---|---|---|
| YOLO-Para | 1.0 | 20 | 0 | Yes |
| YOLOv12 | 1.0 | 20 | 0 | Yes |
| YOLOv12-CP | 1.0 | 20 | 0.5 | Yes |

Table 6: Training hyperparameters for best-performing models

**Overall Performance Comparison** As shown in Table 7 and Figure 2, the YOLOv12 models achieved the highest overall performances across most metrics compared to YOLO-Para, with the copy paste variant (YOLOv12-CP) demonstrating even better performance.

| Metric | SPAM | Para | v12 | v12-CP |
|---|---|---|---|---|
| **Precision** | 0.780 | 0.799 | 0.814 | **0.837** |
| **Recall** | 0.752 | 0.838 | **0.845** | 0.821 |
| $\mathbf{mAP_{50}}$ | 0.816 | 0.864 | 0.869 | **0.878** |
| $\mathbf{mAP_{50-95}}$ | 0.537 | 0.628 | 0.631 | **0.677** |

Table 7: Overall performance of best-performing models

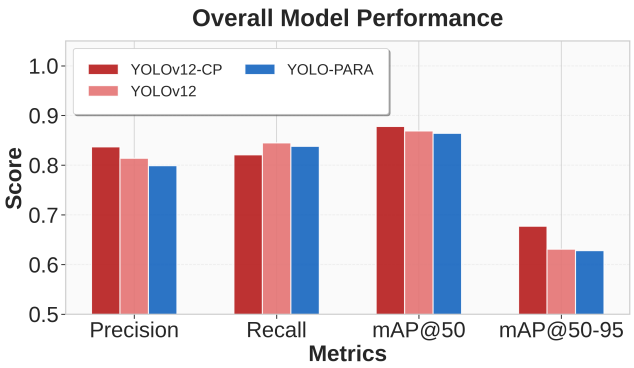


Figure 2: Overall performance of models across all metrics

- **Precision**: YOLOv12 improved by $+1.9\%$ (0.799→0.814), with YOLOv12-CP reaching $+4.8\%$ (0.799→0.837), indicating reduced false positive rate.
- **Recall**: YOLOv12 achieved highest recall (0.838), with YOLO-Para closely following (0.845), demonstrating strong true positive detection rates.
- $\mathbf{mAP_{50}}$: YOLOv12-CP improved by $+1.0\%$ over YOLOv12 (0.869→0.878) and $+1.6\%$ over YOLO-Para (0.864→0.878).
- $\mathbf{mAP_{50-95}}$: YOLOv12-CP showed the most substantial improvement: $+7.3\%$ over YOLOv12 (0.631→0.677) and $+7.8\%$ over YOLO-Para (0.628→0.677)

The substantial $\mathrm{mAP}_{50-95}$ improvement is clinically significant, as it indicates that YOLOv12's attention mechanisms combined with copy-paste augmentation provide superior boundary precision necessary for identifying precise parasite locations within red blood cells across stringent IoU thresholds.

### 4.4 Per-Species Performance Analysis

Performance varies substantially across Plasmodium species due to morphological diversity and dataset representation. Figure 3 presents per-species/class comparison across all models based on $\mathrm{mAP}_{50}$, whereas Table 8 presents comprehensive per-species analysis across all metrics and models.

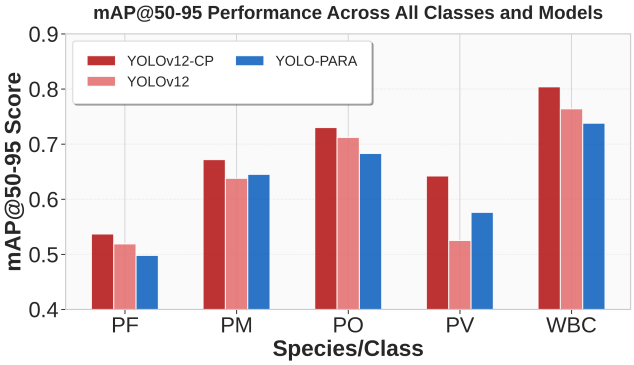


Figure 3: Overall performance of models across all metrics

| Species | Metric | YOLO-Para | YOLOv12 | YOLOv12-CP |
|---|---|---|---|---|
| **PF** | Precision | 0.715 | 0.766 | **0.807** |
| | Recall | **0.716** | 0.701 | 0.646 |
| | $\mathbf{mAP_{50}}$ | 0.773 | 0.783 | **0.797** |
| | $\mathbf{mAP_{50-95}}$ | 0.498 | 0.519 | **0.537** |
| **PM** | Precision | 0.788 | 0.823 | **0.859** |
| | Recall | **0.865** | 0.860 | 0.853 |
| | $\mathbf{mAP_{50}}$ | 0.880 | 0.889 | **0.912** |
| | $\mathbf{mAP_{50-95}}$ | 0.645 | 0.638 | **0.672** |
| **PO** | Precision | 0.828 | 0.845 | **0.852** |
| | Recall | 0.809 | **0.845** | 0.829 |
| | $\mathbf{mAP_{50}}$ | 0.877 | **0.900** | 0.893 |
| | $\mathbf{mAP_{50-95}}$ | 0.683 | 0.712 | **0.730** |
| **PV** | Precision | 0.812 | 0.791 | **0.815** |
| | Recall | 0.871 | **0.882** | 0.839 |
| | $\mathbf{mAP_{50}}$ | **0.894** | 0.864 | 0.874 |
| | $\mathbf{mAP_{50-95}}$ | 0.576 | 0.525 | **0.642** |
| **WBC** | Precision | **0.854** | 0.845 | 0.853 |
| | Recall | 0.928 | **0.937** | **0.937** |
| | $\mathbf{mAP_{50}}$ | 0.894 | 0.906 | **0.915** |
| | $\mathbf{mAP_{50-95}}$ | 0.738 | 0.764 | **0.804** |

Table 8: Per-Species/Class Performance Across Models

**Key Findings**:
(***Plasmodium falciparum***):

- **Precision**: YOLOv12-CP achieves 0.807, the highest across all models, critical for reducing false positives in clinical workflows
- **Recall**: YOLO-Para leads (0.716), with YOLOv12 closely following (0.701)
- $\mathbf{mAP_{50}}$: YOLOv12-CP achieves 0.797, establishing new benchmark high performance as compared to our previous works, for the most clinically urgent species
- $\mathbf{mAP_{50-95}}$: YOLOv12-CP reaches 0.537, demonstrating superior localization precision across stringent IoU thresholds essential for accurate parasite boundary delineation, and achieving benchmark high performance as compared to our previous works

(***Plasmodium malariae* & *Plasmodium ovale***):

- **PM Performance**: YOLOv12-CP excels with 0.912 $\mathrm{mAP}_{50}$ and 0.672 $\mathrm{mAP}_{50-95}$, demonstrating attention mechanisms' effectiveness on well-represented classes. Recall remains strong across all models ($0.853 - 0.865$).
- **PO Performance**: YOLOv12 achieves highest $\mathrm{mAP}_{50}$ (0.900), with YOLOv12-CP providing best $\mathrm{mAP}_{50-95}$ (0.730). Precision and recall remain consistently high ($>$ 0.80) across all models, indicating robust detection.

(***Plasmodium vivax***):

- All models achieve strong PV detection, emphasizing the impact of the $10\times$ augmentation applied to this class.
- **YOLOv12-CP achieves highest** $\mathrm{mAP}_{50-95}$ (0.642), a remarkable 22.3% improvement over YOLOv12 (0.525), validating copy-paste augmentation's effectiveness for minority class localization precision
- Recall consistently exceeds 0.83 across all models, demonstrating that species-specific $10\times$ augmentation successfully addressed extreme class imbalance

**White Blood Cell Detection**:

- All models achieve strong WBC detection ($\mathrm{mAP}_{50} > 0.89$), important for accurate parasitemia computation
- YOLOv12-CP achieves best performance ($\mathrm{mAP}_{50-95}$ : 0.804), with consistently high recall (0.937) across YOLOv12 variants

# 5 Discussion

This study evaluates state-of-the-art attention-based object detection and segmentation models for multi-species malaria parasite detection in high-resolution blood smear images. Comprehensive evaluation across YOLO-SPAM, YOLO-Para, and YOLOv12 architectures demonstrates that YOLOv12, when trained with copy-paste augmentation, achieves optimal performance with $\mathbf{mAP_{50}} = 0.878$ and $\mathbf{mAP_{50-95}} = 0.677$. Key findings underscore the effectiveness of species-specific augmentation strategies and training protocols for medical imaging applications.

## 5.1 Augmentation Strategy Effectiveness

Species-specific augmentation demonstrated differential effectiveness across species. The $10\times$ augmentation for minority *P. vivax* yielded $+9.5\%$ $\mathbf{mAP_{50}}$, directly addressing severe class imbalance. Conversely, the $3\times$ augmentation for *P. malariae* produced exceptional $+67.4\%$ $\mathbf{mAP_{50-95}}$ improvement, highlighting that underrepresentation spans both class frequency and localization precision. *P. falciparum* exhibited precision-recall trade-offs, reflecting the challenge of detecting the smallest parasite species in the clinically critical ring form. These differential effects validate targeted augmentation as superior to uniform approaches.

Augmentation impact analysis (Section 4.2) quantitatively demonstrates augmentation's impact on challenging detection scenarios. For the rarest species (*P. vivax*, 6.4% of dataset), species-specific $10\times$ augmentation enabled robust detection with final $\mathbf{mAP_{50}} = 0.874$ (YOLOv12-CP), validating the strategy's effectiveness for extreme class imbalance. For the smallest parasites (*P. falciparum* ring forms, 1–2 $\mu$m), YOLOv12-CP achieved $\mathbf{mAP_{50}} = 0.797$ and $\mathbf{mAP_{50-95}} = 0.537$, demonstrating that attention mechanisms combined with copy-paste augmentation enable precise localization of morphologically subtle, clinically critical parasite stages (Table 8).

## 5.2 Training Protocol Optimization

YOLOv12 demonstrated superior performance when trained with optimized protocols combining copy-paste augmentation with noise injection. Compared to YOLO-Para baseline, YOLOv12 achieved $+0.58\%$ $\mathbf{mAP_{50}}$ and $+0.48\%$ $\mathbf{mAP_{50-95}}$ improvement. Copy-paste augmentation further enhanced performance, adding $+1.03\%$ $\mathbf{mAP_{50}}$ and $+7.3\%$ $\mathbf{mAP_{50-95}}$, demonstrating that strategic instance-level augmentation complements existing segmentation ar-

chitectures. The $\mathbf{mAP_{50-95}}$ overall improvement is clinically significant, indicating superior spatial precision across stringent IoU thresholds essential for accurate parasite boundary delineation in medical imaging.

### 5.3 Clinical Relevance

Per-species analysis reveals clinically aligned performance patterns. YOLOv12 achieved highest precision for *P. falciparum*, reducing false positives critical for treatment decisions on the most severe species. *P. malariae* and *P. ovale* showed consistent high performance ($> 0.89$ $\mathbf{mAP_{50}}$) across all models, supporting species differentiation in clinical practice. Minority species *P. vivax* detection reached $0.874$ $\mathbf{mAP_{50}}$, validating augmentation-based approaches for handling extreme class imbalance. White blood cell detection ($0.915$ $\mathbf{mAP_{50}}$) provided robust contextual signal, supporting integration into clinical workflows.

### 5.4 Medical Imaging Implications

Segmentation-based detection enables pixel-level precision for morphological analysis of parasite structures, which is critical for distinguishing ring-form trophozoites (1–2 $\mu$m), the smallest and most diagnostically challenging parasite stage. Instance segmentation provides accurate species identification through morphological characteristics (like ring forms for early trophozoites, and crescent-shaped gametocytes) that rectangular bounding boxes cannot reliably capture. The per-species performance patterns demonstrate that multi-stage parasite detection across varying morphologies is achievable with appropriate model selection and training strategies tailored to medical imaging requirements.

### 5.5 Broader Impact and Deployment

YOLOv12n-seg achieves real-time inference at 34.6 frames per second (FPS) on H100 GPU hardware for 2048×2048 resolution images, with total processing time of 28.9 ms per image. Notably, the model's compact architecture (2.76M parameters, 9.7 giga floating-point operations (GFLOPs)) enables CPU-only deployment with comparable inference times, as validated through a proof-of-concept web-based deployment currently under development by collaborating teams. This CPU compatibility eliminates GPU hardware requirements, substantially reducing deployment barriers for resource-constrained microscopy laboratories in Rwanda and similar endemic regions where specialized computing infrastructure may be limited or unavailable.

Species-specific augmentation's success suggests scalability to larger imbalanced datasets, which is a common challenge in medical artificial intelligence applications. This study systematically evaluates how data curation strategies (species-specific augmentation) and augmentation protocols (copy-paste, noise injection) enhance detection performance for malaria diagnosis in Rwanda and similar settings.

### 5.6 Limitations and Future Work

This study has some limitations that provide opportunities for future investigation. First, validation on a single-center Rwandan dataset limits generalization to other geographic regions where parasite phenotypes, staining protocols, and imaging conditions may differ. Multi-site validation across healthcare facilities within Rwanda and cross-dataset evaluation on publicly available malaria datasets would establish model robustness and transferability. Such multi-site studies would also enable assessment of performance variation across different microscopy equipment and operator practices, critical for real-world deployment.

Second, this study focuses exclusively on YOLO-based architectures without comparison to alternative detection frameworks such as Mask R-CNN, Faster R-CNN, or transformer-based detectors (DETR, RT-DETR). While YOLO models provide computational efficiency suitable for resource-constrained settings, systematic benchmarking against these architectures would better establish the performance-efficiency trade-offs and validate architectural choices for malaria detection.

Third, the reported results represent single training runs without statistical robustness measures. Future work should conduct multiple training runs with different random seeds, reporting mean performance with confidence intervals or standard deviations to quantify model stability and result reproducibility. Such analysis would strengthen empirical claims and enable statistical significance testing when comparing model variants.

Finally, while species-specific augmentation successfully addressed class imbalance for underrepresented *P. vivax*, synthetic data generation using generative adversarial networks (GANs) or diffusion models could further augment minority classes. However, preserving species-specific morphological fidelity, especially the diagnostically critical features such as ring-form trophozoites, gametocyte crescents, and chromatin patterns, remains challenging for current generative models. Future research should investigate conditional generation approaches that maintain clinical validity while expanding training data for rare species.

## 6 Conclusion

This evaluation of attention-based YOLO architectures for multi-species malaria parasite detection demonstrates that, state-of-the-art attention when coupled with strategic data curation and training optimization, substantially enhance detection performance. Species-specific augmentation improved overall detection performancing, while directly addressing class imbalance.

YOLOv12, optimized with copy-paste augmentation and noise injection, achieved benchmark high performance as compared to our previous works, demonstrating exceptional capability in detecting the clinically critical *P. falciparum*, which is the smallest and most morphologically challenging parasite species. YOLOv12 achieved the highest precision *P. falciparum* and improved spatial localization, validating the effectiveness of attention-based training protocols for detecting small, difficult parasite structures in complex blood smear images.

These findings support practical deployment of AI-assisted malaria diagnosis in resource-limited settings when combined with high-resolution microscopy, advancing the

vision of bridging AI innovation and clinical practice for disease control in Rwanda and similar regions.

## Acknowledgments

We gratefully acknowledge the Rwanda Biomedical Center for providing the Giemsa-stained blood smear slides for data collection and expert microscopist validation support. This work was supported by a grant from the Afretec Network at Carnegie Mellon University Africa.

## References

Akpo, E. M.; Mukamakuza, C. P.; and Tuyishimire, E. 2024. Binary Segmentation of Malaria Parasites Using U-Net Segmentation Approach: A Case of Rwanda. In Yang, X.-S.; Sherratt, S.; Dey, N.; and Joshi, A., eds., *Proceedings of Ninth International Congress on Information and Communication Technology*, 163–176. Singapore: Springer Nature. ISBN 978-981-97-4581-4.

Bogale, Y.; Mukamakuza, C. P.; and Tuyishimire, E. 2024. Intelligent Malaria Detection and Species Classification: A Case of Rwanda. In Yang, X.-S.; Sherratt, S.; Dey, N.; and Joshi, A., eds., *Proceedings of Ninth International Congress on Information and Communication Technology*, 507–520. Singapore: Springer Nature. ISBN 978-981-97-3299-9.

CDC. 2024. CDC - DPDx - Malaria.

Chen, L.; Wu, Y.; and Merhof, D. 2022. Instance Segmentation of Dense and Overlapping Objects via Layering. ArXiv:2210.03551 [cs].

Das, D.; Mukherjee, R.; and Chakraborty, C. 2015. Computational microscopic imaging for malaria parasite detection: a systematic review. *Journal of Microscopy*, 260(1): 1–19. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jmi.12270.

Davidson, M. S.; Andradi-Brown, C.; Yahiya, S.; Chmielewski, J.; O'Donnell, A. J.; Gurung, P.; Jeninga, M. D.; Prommana, P.; Andrew, D. W.; Petter, M.; Uthaipibull, C.; Boyle, M. J.; Ashdown, G. W.; Dvorin, J. D.; Reece, S. E.; Wilson, D. W.; Cunningham, K. A.; Ando, D. M.; Dimon, M.; and Baum, J. 2021. Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks. *Biological Imaging*, 1: e2.

Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.-Y.; Cubuk, E. D.; Le, Q. V.; and Zoph, B. 2021. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. ArXiv:2012.07177 [cs].

He, K.; Gkioxari, G.; Dollár, P.; and Girshick, R. 2018. Mask R-CNN. ArXiv:1703.06870 [cs].

Molina, J. M.; Llerena, J. P.; Usero, L.; and Patricio, M. A. 2025. Advances in instance segmentation: Technologies, metrics and applications in computer vision. *Neurocomputing*, 625: 129584.

PMI. 2023. Rwanda Malaria Profile PMI FY-2024. Government of Rwanda. https://mesamalaria.org/wp-content/uploads/2025/04/RWANDA-Malaria-Profile-PMI-FY-2024.pdf. [Accessed 08-11-2025].

Poostchi, M.; Silamut, K.; Maude, R. J.; Jaeger, S.; and Thoma, G. 2018. Image analysis and machine learning for detecting malaria. *Translational Research: The Journal of Laboratory and Clinical Medicine*, 194: 36–55.

Ramos-Briceño, D. A.; Flammia-D'Aleo, A.; Fernández-López, G.; Carrión-Nessi, F. S.; and Forero-Peña, D. A. 2025. Deep learning-based malaria parasite detection: convolutional neural networks model for accurate species identification of Plasmodium falciparum and Plasmodium vivax. *Scientific Reports*, 15(1): 3746.

Roboflow. 2020. When Should I Auto-Orient My Images?

Tangpukdee, N.; Duangdee, C.; Wilairatana, P.; and Krudsood, S. 2009. Malaria Diagnosis: A Brief Review. *The Korean Journal of Parasitology*, 47(2): 93–102.

Tian, Y.; Ye, Q.; and Doermann, D. 2025. YOLOv12: Attention-Centric Real-Time Object Detectors. ArXiv:2502.12524 [cs].

WHO. 2016. World malaria report 2023.

Wongsrichanalai, C.; Barcus, M. J.; Muth, S.; Sutamihardja, A.; and Wernsdorfer, W. H. 2007. A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT). *The American Journal of Tropical Medicine and Hygiene*, 77(6 Suppl): 119–127.

Zedda, L.; Loddo, A.; and Di Ruberto, C. 2023. YOLO-PAM: Parasite-Attention-Based Model for Efficient Malaria Detection. *Journal of Imaging*, 9(12): 266. Publisher: Multidisciplinary Digital Publishing Institute.

Zedda, L.; Loddo, A.; and Di Ruberto, C. 2024. A deep architecture based on attention mechanisms for effective end-to-end detection of early and mature malaria parasites. *Biomedical Signal Processing and Control*, 94: 106289.

Zedda, L.; Loddo, A.; and Di Ruberto, C. 2025. A deep architecture based on attention mechanisms for effective end-to-end detection of early and mature malaria parasites in a realistic scenario. *Computers in Biology and Medicine*, 186: 109704.