Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automated Detection of Soil transmitted Helminthes and Schistosomiasis Using YOLO Based Deep Learning Model

Abrham Adamu

abrhamadamu05@gmail.com

Haramaya University

Etefa Belachew

Jimma University – Institute of Technology

Kris Calpotura

Jimma University – Institute of Technology

Aregawi Gebremedhin

Mekelle University, Ethiopian Institute of Technology

Research Article

Keywords: Soil transmitted helminths and schistosomiasis, Yolo, Object detection, parasite egg detection

Posted Date: February 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8887946/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Parasitology Research on October 24th, 2026. See the published version at https://doi.org/10.1007/s00436-026-08790-3.

Automated Detection of Soil transmitted Helminthes and Schistosomiasis Using YOLO Based Deep Learning Model

Abrham Adamu [1], Etefa Belachew[2], Kris Calpotura[2], Aregawi Gebremedhin[3]
[1] Haramaya University, School of Electrical and Computer Engineering, Dire Dawa, Ethiopia
[2]Faculty of Electrical and Computer Engineering, Jimma University – Institute of Technology, Jimma, Ethiopia
[3]Mekelle University, Ethiopian Institute of Technology, Mekelle, Ethiopia

Correspondence should be addressed to Abrham Adamu; abrhamadamu05@gmail.com

Abstract

Soil-transmitted helminths (STHs) and schistosomiasis remain prevalent public health concerns in tropical and resource constrained regions necessitating accurate and timely diagnostic approaches for effective treatment and control. However, conventional microscopic examination of parasite eggs is laborious, time-intensive and dependent on expert interpretation which can introduce variability and diagnostic errors. To address these challenges, this study proposes a real time deep learning-based detection and classification using YOLO-based object detection models for automated analysis of microscopy images. Specifically, YOLOv11 and YOLOv12 architectures were evaluated across three lightweights to medium model variants (Nano, Small and Medium) to assess tradeoffs between detection accuracy, inference speed and computational efficiency. A custom annotated dataset comprising 1713 images from Ethiopian health institute with four clinically relevant parasite classes (Ascaris, hookworm, Schistosomia and Trichuris) was used for independent training, validation, and testing. Model performance was evaluated using standard object detection metrics including mean Average Precision at IoU 0.5 (mAP@0.5, precision, recall and F1-score) and inference speed. A 5-fold cross validation and statistical significance analysis ($p < 0.001$) were conducted to ensure robustness and reproducibility. Experimental results indicate that YOLOv12m achieves the highest detection performance (mAP@0.5 = 94%) and mAP@0.5–0.9 of 67.4%, while yolo11m scored the highest recall of 88.6% indicating fewer false Negative. In contrast, YOLOv12n offers the lowest computational cost 6.3 GFLOPs by delivering the fast inference of 159FPS with reduced sensitivity. These findings demonstrate the effectiveness of YOLO-based models for scalable, real time parasite detection and provide practical guidance for deployment in resource-constrained healthcare settings.

Keywords: Soil transmitted helminths and schistosomiasis, Yolo, Object detection, parasite egg detection.

1. Introduction

Soil transmitted helminths (STHs) and schistosomiasis remain major public health concerns in tropical and resource limited regions, where timely and accurate diagnosis is essential for effective treatment and control. These infections contribute significantly to morbidity, particularly in children through malnutrition, anemia and impaired cognitive development [14]. Conventional microscopic examination of parasite eggs though widely used is labor intensive, time consuming and dependent on expert interpretation which can introduce variability and diagnostic errors [1]. Current diagnostic gold standard, manual microscopy using the Kato-Katz technique, is labor intensive and slow. It suffers from poor sensitivity in low intensity infections and is impacted by significant inter-observer variability, limiting its suitability for large-scale screening [2]. These limitations create a critical bottleneck, hindering the data-driven decisions needed for efficient public health response. Artificial intelligence (AI), particularly deep learning (DL), offers a paradigm-shifting solution. Our previous work demonstrated the high efficacy of a hybrid CNN Machine Learning (ML) approach for classifying parasite eggs from microscopy images, achieving over 99% accuracy [3]. However, such a classification-based model requires a step that itself demands significant manual labor or a separate detection system. YOLO models are highly relevant to this diagnostic gap because they unify detection and classification in a single stage, enabling real time localization and identification of parasite eggs without the need for prior region of interest (ROI) extraction [10]. This integrated approach directly addresses the limitations of CNN-based classification pipelines, which, although effective in enhancing diagnostic accuracy, typically require a separate detection or

segmentation step that introduces additional manual labor and complexity [4]. While prior studies have applied various YOLO versions (v3-v8) to parasitology [1, 4, 13, 26], the field lacks a systematic, comparative analysis of the very latest architectures like YOLOv11 and YOLOv12 across their efficiency spectrum (Nano, Small, Medium). Crucially, the trade-offs between detection accuracy, inference speed and computational cost for these newest models remain unquantified leaving a gap in practical guidance for deployment in resource constrained settings [7, 10]. Building directly on our foundational hybrid model research [3], this study addresses this gap by conducting a rigorous, comparative benchmark of YOLOv11 and YOLOv12. We evaluate these architectures across their scalable variants to map the performance efficiency frontier explicitly. Using a curated dataset of four key helminth egg types (Ascaris lumbricoides, hookworm, Schistosomia mansoni, Trichuris trichiura), rigorous k-fold cross validation and comprehensive metrics (mAP, FPS, GFLOPs), we aim to identify the optimal model variants for different deployment scenarios. This work provides the essential, empirical evidence needed to transition from a high accuracy classification tool to a scalable, real time detection system directly supporting the development of integrated AI solutions for global NTD control programs.

2. Related work

A significant contribution in this domain is our previous study, which developed a high-accuracy predictive model for STH and schistosomiasis. This work employed a hybrid CNN-ML methodology, using pretrained CNNs (VGG16, DenseNet121) as Feature extractors coupled with classical ML classifiers (SVM, XGBoost), achieving test accuracies exceeding 99%. This study conclusively demonstrated that hybrid models can outperform standalone deep learning classifiers on curated datasets, providing a powerful tool for the precise classification of pre-identified objects. However, it also highlighted a key dependency: the need for a prior step to locate eggs within a full microscopy slide, which remains a manual bottleneck [3]. A Scientific Reports study comparing AI-supported versus manual microscopy of Kato–Katz smears showed that AI can improve diagnostic consistency, though the system lacked detailed architectural benchmarking and efficiency analysis [2]. Study [8] introduced a deep learning-based platform for helminth egg recognition, demonstrating that modern CNN architectures can achieve robust performance across multiple parasite species. Additionally study [16] demonstrated the feasibility of CNN-based classification for intestinal parasites in reptile stool samples. This study validated the potential of AI to enhance diagnostic accuracy, while CNN-attention models have also been applied to parasitic egg recognition further highlighting the potential of classification-based systems [9] and comparative evaluation of multiple deep learning architectures for helminth diagnosis [10].

Although these approaches improved classification accuracy, they lacked real-time object localization capabilities limiting their applicability in routine diagnostics. To eliminate the need for manual cropping, research has increasingly shifted toward end-to-end object detection, particularly YOLO-based architectures, their computational cost was prohibitive for real time use. Study [24] demonstrated the feasibility of YOLO based detection for parasitic products in stool samples, achieving substantial improvements in speed and accuracy. One of the most comprehensive evaluations to date is the study by Mohammed et al. [1], which assessed multiple YOLOv7 variants for detecting STH and Schistosomia mansoni eggs in Kato–Katz smear images. Their results showed strong performance across accuracy, robustness and inference speed reinforcing YOLO's suitability for real-time parasitological diagnostics. Further studies have expanded YOLO's application scope. Study [4] developed a YOLO-based multiclass detection system for STH and Schistosomia mansoni eggs, achieving high accuracy on field acquired fecal smear images. These results reinforce the growing consensus that YOLO architectures offer a practical balance between speed and accuracy for automated microscopy. YOLOv5 to detect Schistosomia mansoni in drug-screening assays [13], while study [9] proposed a lightweight deep learning model (yolov5n) optimized for low resource environments, demonstrating that compact architectures can still achieve competitive detection performance. Study [23] reviewed optical devices for STH and schistosomiasis diagnosis including portable and smartphone-based microscopy systems.

The evolution of YOLO architectures has been extensively documented with major improvements in speed, accuracy and small object detection from YOLOv1 to YOLOv8 [10]. Recent releases such as YOLOv11 and YOLOv12 [26, 27] offer enhanced

multi scale feature extraction improved detection of small objects and reduced computational cost making them strong candidates for real time parasitological diagnostics. Most existing studies rely on earlier YOLO versions (v3–v7) with limited evaluation of the newest architectures. This study addresses these gaps by evaluating YOLOv11 and YOLOv12 across three model sizes using a custom annotated dataset of clinically relevant parasite classes. Through rigorous evaluation including 5-fold cross validation, statistical significance testing and efficiency tradeoffs, this work advances the state of automated parasitic egg detection and provides practical insights for deployment in resource constrained healthcare settings.

3. Experimental Design, Materials and Methods

3.1. System Architecture

The Figure 1 outlines the workflow used to develop fine-tuned deep learning models for automatic detection of soil-transmitted helminths (STH) and schistosomiasis from microscopic images. The workflow starts with data collection, followed by preprocessing steps such as removal of duplicate images and image resizing to maintain uniformity. The images are then annotated to identify STH and Schistosomia species for supervised training. Data augmentation methods, including image rotations and flips, are applied to improve model robustness. The dataset is subsequently partitioned into training, validation, and test sets. Variants of YOLOv11 and YOLOv12 ranging from nano to medium are trained and evaluated separately using precision recall and mAP@50 metrics. Model performance is further refined through hyperparameter tuning until convergence. The final model selection is based on the average mAP@50 scores obtained through five-fold cross-validation, and statistical significance among models is assessed using one-way ANOVA followed by Tukey's post hoc analysis.

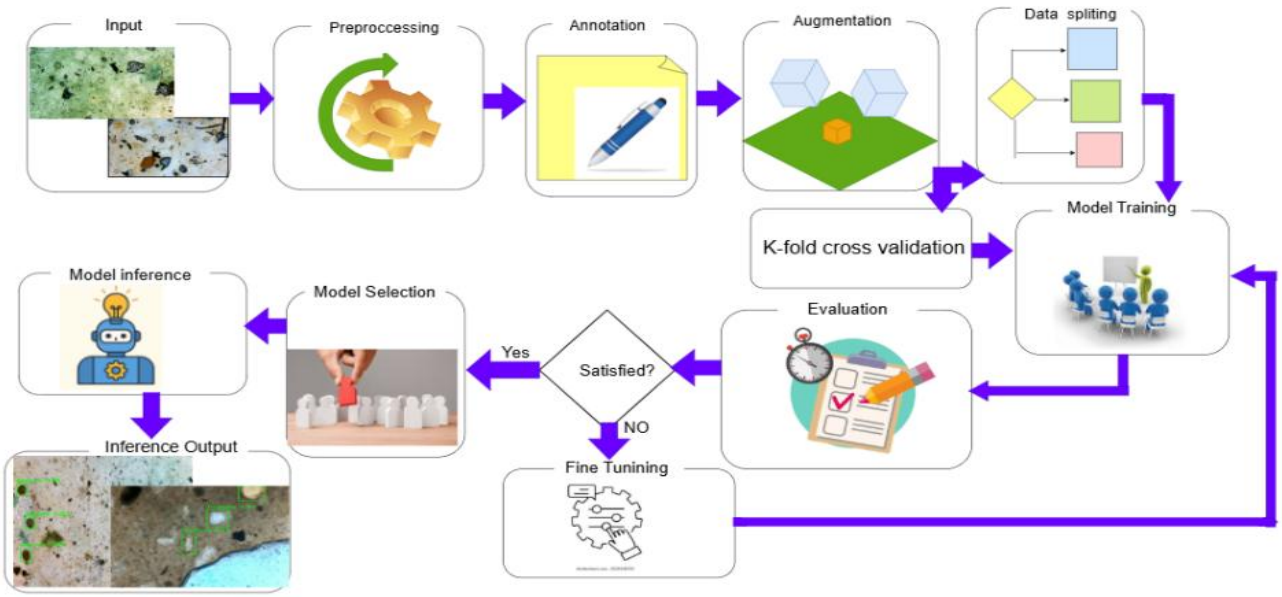


Figure 1: Workflow of Proposed Method. Overview of the machine learning lifecycle, from data preparation and model training to validation and iterative fine-tuning.

3.2. Data Collection and Sample Preparation

3.2.1. Source and Location

The dataset was systematically collected through a collaborative effort between researchers and the Ethiopian Public Health Institute (EPHI), ensuring adherence to standardized diagnostic protocols. Previously published in our earlier work [3]. Microscopy slides were obtained from two well-equipped laboratory facilities: the EPHI National Laboratory Center in Addis Ababa, which serves as

Ethiopia′s primary reference lab for infectious diseases, and the pathology laboratory at Jimma University, a leading institution in medical research and regional healthcare. A total of 1713 high quality images were captured at 40×magnification from two laboratory sites with 645 (Ascaris), 378 (Schistosomia), 423 (Trichuris) and 267 (Hookworm) images per class creating a robust and representative dataset for model training. The details are summarized in Table 1.

Table 1: Datasets collated from the Ethiopian Public Health Institute. As shown in the table, the dataset consists of 1713 microscopic images collected from 373 slides distributed among four parasite categories and negative samples.

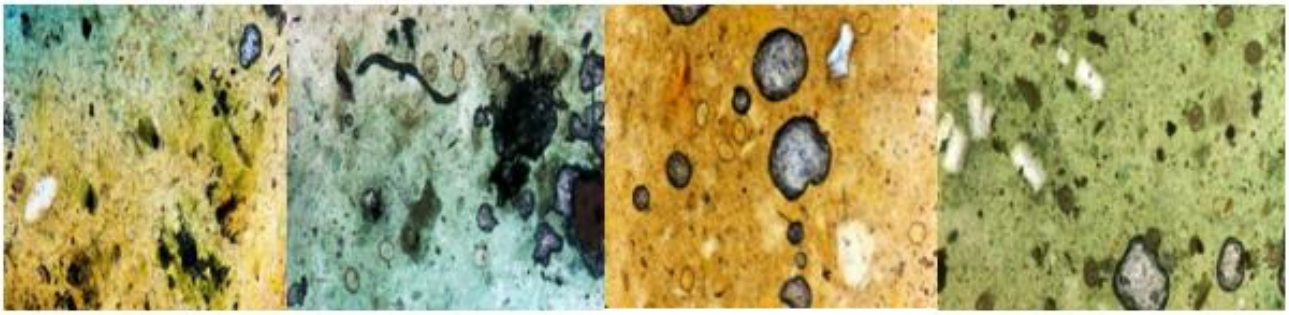
Classes	Images taken	Number of slides	Format
Trichuris trichiura	423	94	.Jpeg.bmp. jpg
Ascaris	645	110	.Jpeg.bmp. jpg
Schistosomiasis	378	83	.Jpeg.bmp. jpg
Hookworm	267	42	.Jpeg.bmp. jpg

Using JPEG and BMP formats offers distinct advantages for image preprocessing and machine learning. The JPEG (Joint Photographic Experts Group) format employs lossy compression which significantly reduces file size often by 80-90% compared to uncompressed formats while maintaining perceptual image quality. Conversely, the BMP (Bitmap) format is uncompressed and lossless preserving every pixel's original data as captured by the microscope sensor. This ensures the highest fidelity for detailed morphological analysis which is paramount for distinguishing subtle differences between parasite egg species [3, 6].

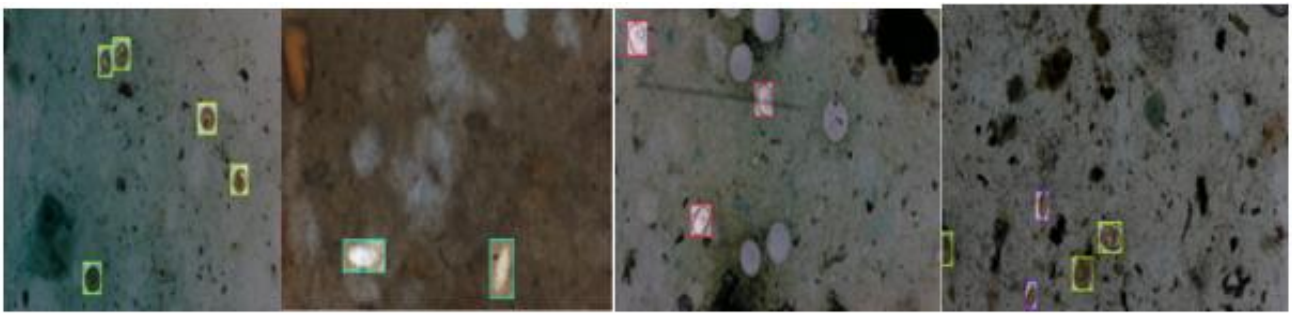
3.2.2. Sample Collection and Labeling

Microscopic images of STH and schistosomiasis eggs were collected from primary sources, including the EPHI, clinics and research laboratories. These images were captured using high resolution microscopes and standardized imaging protocols. They were then preprocessed through resizing, normalization and augmentation. Experts from EPHI and Jimma University manually labeled STH and schistosomiasis image datasets, enhancing them with metadata. The data was collected from high prevalence Ethiopian regions using the Kato–Katz technique. The Kato–Katz technique, a WHO recommended quantitative method for diagnosing STH, involves preparing standardized fecal smears (41.7 mg) cleared with glycerin-coated cellophane. Eggs are microscopically counted to determine infection intensity (eggs per gram). While it is cost effective and field friendly, its sensitivity depends on operator skill and egg stability. To ensure accuracy, duplicate smears were analyzed, and two microscopists independently reviewed each slide with a third expert resolving discrepancies. This rigorous protocol enhanced dataset reliability for predictive modeling [3]. The figure 2 illustrate the dataset sample images of the microscopic inputs used to train the object detection models and their corresponding annotated versions, which together form the foundation for supervised learning in this study. The raw images (2A) consist of microscope-captured slides containing soil transmitted helminth (STH) and schistosomiasis eggs, acquired under varying illumination conditions, magnifications and slide preparations. These variations reflect real world diagnostic settings and introduce challenges such as background noise, overlapping structures, and differences in egg size, shape and contrast.

Alongside each raw image, an annotated counterpart (2B) is provided, where parasite eggs are enclosed within bounding boxes and labeled according to their respective classes (Ascaris with yellow bounding box, Hookworm with blue, Trichuris with green and Schistosomia with red bounding box). The annotations follow the YOLO format, in which each object is represented by its class label and normalized bounding box coordinates. These labeled images serve as ground truth, enabling the models to learn both where the parasite eggs are located (localization) and what type of parasite each egg represents (classification). The annotated samples demonstrate careful delineation of egg boundaries, even in cases where eggs are partially occluded, clustered or exhibit low contrast against the background. This precise annotation is crucial for training robust detection models, as it guides the learning process to focus on biologically relevant features while ignoring irrelevant background structures. Together, the raw and annotated dataset samples highlight the diversity and quality of the training data and illustrate how supervised annotations enable the YOLOv11 and YOLOv12 models to learn accurate and reliable detection of STH and schistosomiasis eggs.



A



B

Figure 2: Sample images of STH and Schistosomia A) Unannotated images of STH and Schistosomia B) Annotated images of STH and Schistosomia displaying boxes around parasite egg.

3.2.3. Data Augmentation

In this study, data augmentation was applied to enhance the robustness and generalization ability of the YOLOv11 and YOLOv12 models for STH and schistosomiasis egg detection. Because microscopic images often suffer from limited data availability, class imbalance and variability in illumination, orientation and egg appearance augmentation was essential to simulate real world variations encountered during routine parasite diagnosis. Geometric augmentations such as random rotation, horizontal and vertical flipping, scaling and translation were used to make the models invariant to egg orientation and position within the microscopic field of view [7]. This is particularly important for STH and schistosomiasis eggs, which can appear at arbitrary angles and locations depending on slide preparation and microscope alignment. Advanced YOLO specific augmentations (Mosaic augmentation) were also utilized. Mosaic combines multiple images into a single training sample allowing the model to learn from eggs at different scales and contexts simultaneously [10, 11]. This significantly improved small object detection performance which is critical since parasite eggs occupy only a small portion of the image. In addition, Mix-up augmentation was applied to reduce overfitting and encourage smoother decision boundaries by blending image label pairs. The augmentation process was carried out using the Roboflow platform [24].

3.3. Model descriptions

3.3.1. Yolo

The YOLO series represents a collection of object detectors based on DCNNs, known for their unified approach. Unlike two stage detectors such as Faster R-CNN which require separate components for proposing bounding box regions and predicting class probabilities, YOLO models are trained end to end jointly optimizing both bounding box regression and object classification in a single

network [1, 10]. This allows them to predict multiple bounding boxes and class probabilities concurrently. Owing to their continuous advancements and success they have emerged as the preferred method for object detection tasks in medical applications. YOLOv11 and YOLOv12 represent the most recent advancements in the YOLO family [26, 27]. Accordingly, our experiments focus on these two versions as they incorporate the latest architectural refinements and optimization strategies available at the time of writing. This work provides a robust and representative foundation for understanding the performance characteristics and tradeoffs of current and forthcoming YOLO based object detection models.

3.3.2. YOLOv11 Architecture Overview

YOLOv11 is a single stage real time object detection architecture designed to achieve a favorable balance between detection accuracy and computational efficiency. Similar to earlier YOLO variants YOLOv11 follows an end-to-end detection paradigm in which object localization and classification are performed simultaneously within a unified neural network. The architecture is composed of three main components: the backbone, neck, and detection head [26].

The backbone is responsible for extracting hierarchical feature representations from the input image. It consists of a series of convolutional layers and optimized feature extraction blocks that progressively reduce spatial resolution while increasing semantic richness. These deep features enable the model to capture both low level visual details and high-level contextual information, which are crucial for detecting objects of varying scales. The neck serves as a feature aggregation module that enhances multi scale representation. By fusing features from different stages of the backbone the neck improves the model's ability to detect small, medium and large objects [26]. This component typically employs feature pyramid-based mechanisms to combine fine-grained spatial details with high-level semantic features, thereby improving detection robustness across scales. The detection head performs the final object prediction. It outputs bounding box coordinates, object confidence scores and class probabilities for each detected object. YOLOv11 utilizes advanced loss formulations, including bounding box regression loss, classification loss and Distribution Focal Loss (DFL) to improve localization precision and classification accuracy [10, 26]. These losses are jointly optimized during training in an end-to-end manner.

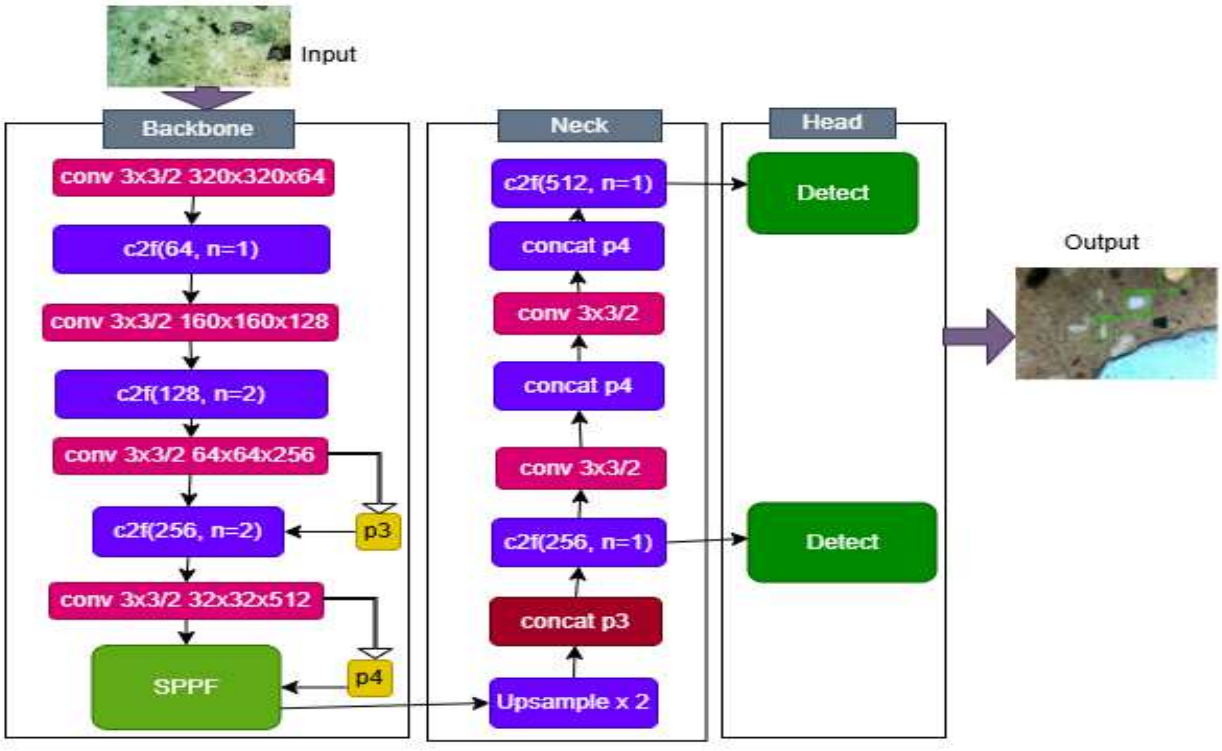


Figure 3: Schematic of the YOLOv11 object detection architecture, illustrating the backbone's convolutional and C2f modules, the neck's multi-scale feature fusion, and the detection head for final prediction output.

3.3.3. YOLOv12 Architecture Overview

YOLOv12 is a single stage, real time object detection architecture designed to achieve an improved balance between detection accuracy computational efficiency and scalability across different model sizes. It follows the core YOLO principle of end-to-end detection, where object localization and classification are performed simultaneously.

As shown in the figure 4, YOLOv12 is composed of three main components: the backbone, neck and detection head. The backbone is responsible for hierarchical feature extraction from input images. It learns low level features such as edges and textures as well as high level semantic representations that are crucial for distinguishing parasite eggs with subtle morphological differences. The neck aggregates multi scale features using feature fusion mechanisms, enabling the model to detect objects of varying sizes. This is particularly important in microscopic images, where parasite eggs appear at different scales depending on magnification and slide preparation. The detection head predicts bounding box coordinates, objectness scores and class probabilities. YOLOv12 incorporates refined bounding box regression strategies including Distribution Focal Loss (DFL), which improves localization precision by modeling bounding box coordinates as probability distributions rather than fixed points. YOLOv12 is available in multiple variants (nano to medium and beyond) allowing flexible deployment depending on computational constraints. Smaller variants prioritize speed and low memory usage, while larger variants improve accuracy through increased model capacity [9, 27].

In this study, YOLOv12 was employed to automatically detect and classify STH and schistosomiasis eggs in microscopic images. The model was trained on annotated datasets containing multiple parasite classes, using data augmentation techniques to improve robustness to variations in egg orientation, illumination, and background noise. Among the evaluated variants, YOLOv12m demonstrated the best tradeoff between accuracy and efficiency. Its enhanced feature extraction and multiscale representation capabilities enabled effective detection of small and morphologically similar parasite eggs, which are challenging to distinguish using traditional image processing methods. The inclusion of DFL contributed to more accurate bounding box localization, especially for partially occluded or low-contrast eggs [27].

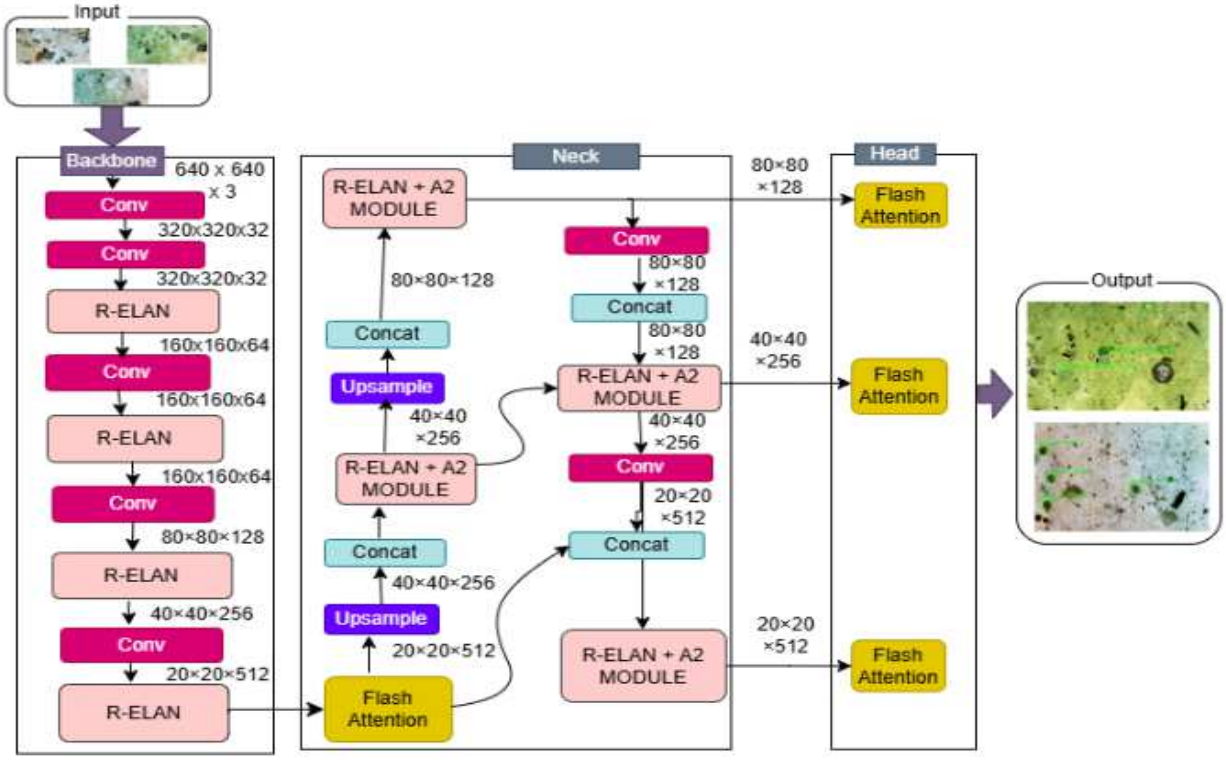


Figure 4: Architecture of the YOLOv12 model, featuring a backbone with R-ELAN blocks for multi-scale feature extraction, a neck enhanced with A2 modules and Flash Attention for cross-scale fusion, and a head utilizing full attention mechanisms for final prediction.

4. Model Training and Selection

4.1. Model Training and Evaluation

Model training and evaluation were performed in a Jupyter Notebook environment using the Ultralytics 8.3.237 framework, Python 3.12.12 and PyTorch 2.9.0. CUDA 126 drivers were used for GPU acceleration on NVIDIA Tesla T4 GPU (16GB VRAM) to ensure consistent and reproducible timing measurements. YOLOv11m model was chosen as the baseline because it provides a strong balance between detection performance and computational efficiency. To identify optimal training settings, a grid search with fivefold cross validation was performed using 20% of the training data. This strategy helped limit overfitting and enabled a fair comparison among the different model variants [1]. Key hyperparameters including the learning rate, momentum, weight decay, and batch size, were systematically evaluated. The parameter set that produced the highest mAP@0.5 (Table 2) was then consistently applied across all YOLOv11 and YOLOv12 variants, from nano to medium. Model evaluation was carried out using mAP@0.5, precision, and recall [1, 8].

Table 2: Hyperparameter settings used for model training, including 50 epochs, batch size of 16, AdamW optimizer, initial learning rate of 0.00125, momentum of 0.9, weight decay of 0.0005, and a 3-epoch warmup period.

Hyperparameters	Value	Purpose
Epochs	50	Controls the number of training iterations
Batch Size	16	Balances memory usage and gradient stability
Optimizer	AdamW	Improves convergence with weight regularization
Initial Learning Rate	0.00125	Controls step size during optimization
Momentum	0.9	Stabilizes gradient updates
Weight Decay	0.0005	Prevents overfitting by regularization
Warmup Epochs	3	Gradually increases learning rate for stable training

4.2. Evaluation Metrics

The study used standard object detection metrics such as precision, recall, F1 score and mean average precision (mAP) at specified intersection over union (IoU) thresholds, including mAP@50 and mAP@50:95 to assess model performance [1, 8]. These metrics are derived from the confusion matrix, which is defined by four core elements: true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs). A TP indicates a correct detection of a positive instance, an FP occurs when the model incorrectly flags a negative instance as positive, a TN represents a correct rejection of a negative instance, and an FN signifies a failure to detect a positive instance.

4.2.1. Precision

Precision measures the proportion of correctly predicted positive cases among all cases predicted as positive. It is especially important when FPs carry significant consequences [1, 8]. In medical diagnosis, high precision ensures that individuals identified as ill truly have the disease, minimizing the risk of misclassifying healthy patients. In this study, precision was calculated using Eq. (1).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

4.2.2. Recall (Sensitivity)

Recall measures a model's ability to identify all actual positive cases. High recall is crucial in medical diagnosis, as it minimizes the chances of missing sick patients [1, 8]. In this study, recall (also referred to as sensitivity) was calculated using Eq. (2)

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (2)$$

4.2.3. F1 score

The F1 score offers a balanced metric that considers both FPs and FNs. It is particularly valuable in medical diagnosis, where both precision and recall are critically important [1, 8]. In this study, the F1 score was calculated using Eq. (3).

$$\text{F1 score} = \frac{2x(Precision+Recall}{Precisoin\ X\ Recall} \quad (3)$$

4.2.4. Mean Average Precision (mAP)

Mean Average Precision (mAP) is a comprehensive metric that evaluates a model's performance in both detecting objects and precisely localizing them by averaging precision values across multiple recall levels [1, 10]. A higher mAP score indicates superior detection and localization accuracy with fewer errors. Following common practice in object detection, mAP@50 and mAP@50:95 was reported to evaluate performance at different IoU stringencies [1, 8]. The overall mAP was calculated as per Eq. (4),

$$\text{mAP} = \frac{1}{\text{N}} \sum AP_{\text{i}} \quad (4)$$

Where AP represents the average precision calculated for each individual class denotes the total number of classes.

4.3. Model Selection

The best-performing model was identified using stratified fivefold cross validation to obtain reliable and generalizable results across the various YOLOv11 and YOLOv12 configurations. The available training and validation images were combined, randomly shuffled and divided into five nearly equal subsets. Each subset was prepared in YOLO format, with one-fold used for validation and the remaining four used for training in each iteration. Every model variant was trained and evaluated across all five folds and performance indicators namely mAP@0.5, precision and recall were calculated for each fold. To compare models statistically a one-way ANOVA was applied to the mAP@0.5 results followed by Tukey's post hoc analysis to determine pairwise differences. A significance threshold of $p < 0.05$ (95% confidence level) was used [1, 8]. The model that achieved the highest average mAP@0.5 and demonstrated statistically significant performance differences was selected as the final model.

5. Results and Discussion

5.1 Statistical evaluation of experimental results

Table 3 presents the performance analysis of YOLOv11 and YOLOv12 model variants evaluated using precision, recall, mAP@0.5, mAP@0.5–0.9, and inference speed (FPS). The results demonstrate clear performance variations across model scales, reflecting the trade-off between detection accuracy and computational efficiency. YOLOv12 models exhibit improved detection accuracy compared to their YOLOv11 counterparts, particularly in medium-scale configurations. Among all evaluated models, YOLOv12m achieved the strongest overall detection performance, recording a precision of 0.938, recall of 0.878, mAP@0.5 of 0.94, and mAP@0.5–0.9 of 67.4%, indicating robust object localization and classification across multiple IoU thresholds. These results suggest that the architectural refinements introduced in YOLOv12 contribute to enhanced feature representation and improved generalization [27].

The YOLOv11 variants also demonstrated competitive performance with YOLOv11s and YOLOv11m achieving high mAP@0.5 values of 0.933 and 0.927 respectively, while maintaining higher inference speeds. Lightweight models such as YOLOv11n and YOLOv12n achieved the highest FPS making them suitable for real time or resource constrained deployment scenarios albeit with comparatively lower detection accuracy.

Table 3: Performance comparison of YOLOv11 and YOLOv12 model variants evaluated by precision, recall, mAP@50, mAP@50–95 and FPS.

Model	Precision (μ ± σ)	Recall (μ ± σ)	Map 0.5 (μ ± σ)	Map 0.5-0.9(μ ± σ)	FPS
Yolo 11n	0.88±0.003	0.832 ±0.006	0.900±0.006	0.624±0.006	154

Yolo 11s	0.915±0.004	0.856+0.004	0.933±0.003	0.660±0.003	98
Yolo 11m	0.909±0.004	0.866±0.004	0.927±0.002	0.674±0.002	122
Yolo 12n	0.821±0.003	0.795±0.005	0.843±0.004	0.586±0.005	159
Yolo 12s	0.884±0.002	0.859±0.003	0.909±0.002	0.641±0.001	84
Yolo 12m	0.938±0.003	0.878±0.004	0.94±0.001	0.674±0.003	57

5.2. Experimental results of the YOLOv12m model

Figure 5 illustrates the training and validation performance of the best-performing YOLOv12m model. The loss curves for both training and validation show a consistent downward trend, indicating stable convergence and effective learning. The bounding box loss decreased progressively from 3.50 to 0.90, demonstrating improved localization accuracy. Likewise, the classification loss declined from 3.40 to 0.44, reflecting enhanced class prediction capability. In addition, the Distribution Focal Loss (DFL) which improves bounding box regression precision was reduced from 2.90 to 1.11 over the training epochs. Validation losses closely followed the training losses, demonstrating good generalization without signs of overfitting. The box loss dropped from 1.90 to 0.20 indicating accurate bounding box predictions on unseen data. Classification loss decreased substantially from 1.80 to 0.21 showing the model's robustness in classifying parasite eggs. Similarly, the DFL decreased from 1.90 to 1.10 further confirming improved localization accuracy on the validation set. Regarding evaluation metrics, precision increased from 0.45 to 0.91 indicating effective minimization of FPs. Recall improved significantly from 0.10 to 0.85, demonstrating enhanced ability to detect schistosomiasis and STH eggs and reduce FNs. Moreover, mAP@50 rose sharply from 0.10 to 0.90, eventually stabilizing near 0.85, highlighting high detection accuracy for schistosomiasis and STH eggs. The mAP@50–95 metric also improved gradually from 0.08 to 0.69, reflecting the model's robustness across various object scales and occlusion levels, though it shows some challenges at stricter IoU thresholds. Overall, the YOLOv12m model exhibits strong learning efficiency, good generalization, and stable performance, as evidenced by consistent improvement in evaluation metrics and corresponding decreases in training and validation losses.

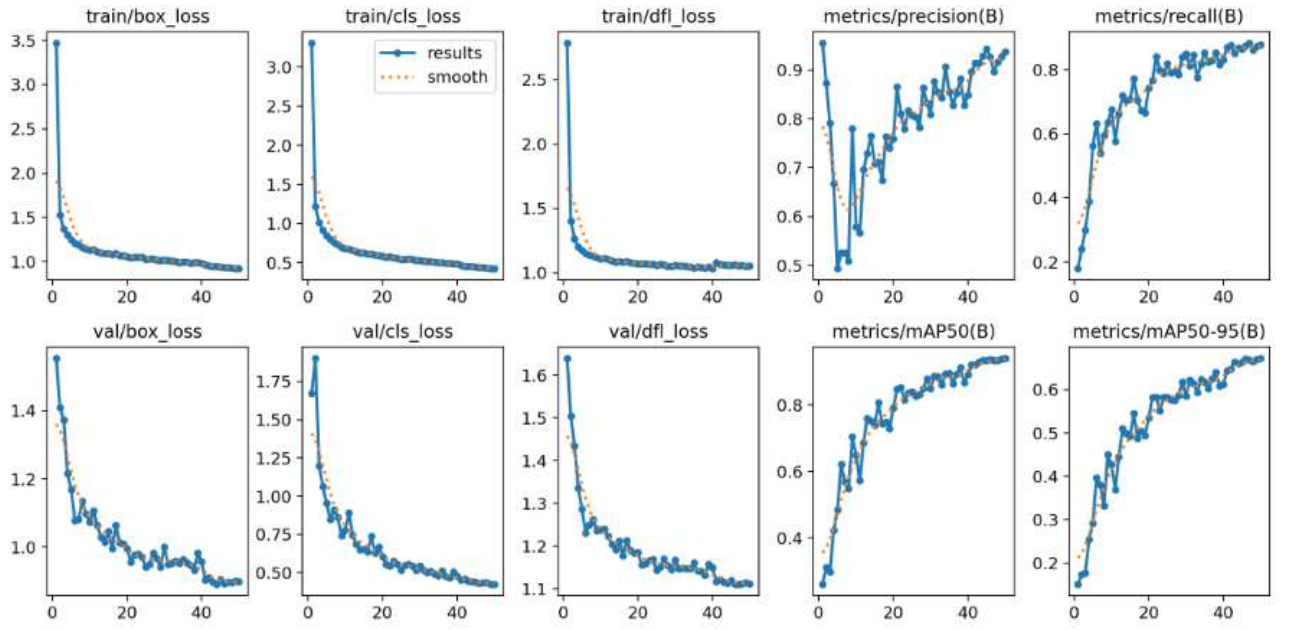


Figure 5: Training and validation performance metrics for the YOLOv12m model variant. A model illustrating the effective reduction of training and validation losses alongside steady improvements in performance metrics like precision, recall, and mean average precision across 50 epochs.

Figure 6 presents the confusion matrix of the YOLOv12m model, the best-performing variant in this study, illustrating its classification performance across Ascaris lumbricoides, Hookworm, Schistosomiasis, and Trichuris trichiura. The diagonal values indicate strong class-wise performance, with correct classification rates of 0.98 for Ascaris, 0.88 for Hookworm, 0.89 for Schistosomiasis, and 0.86 for Trichuris, demonstrating the model's overall robustness in distinguishing between helminth species.

Misclassification was most evident for Ascaris, where 5% of instances were incorrectly predicted as Hookworm and 4% as Schistosomiasis, suggesting visual similarities among certain egg morphologies. Despite these errors, confusion between other class pairs remained limited, indicating effective interclass separation. Overall, the confusion matrix confirms that YOLOv12m achieves high discriminative performance across all target parasite classes with minor misclassifications likely attributable to overlapping morphological features. These findings highlight the model's strong diagnostic potential while also indicating opportunities for further improvement through enhanced feature learning or class specific optimization strategies.

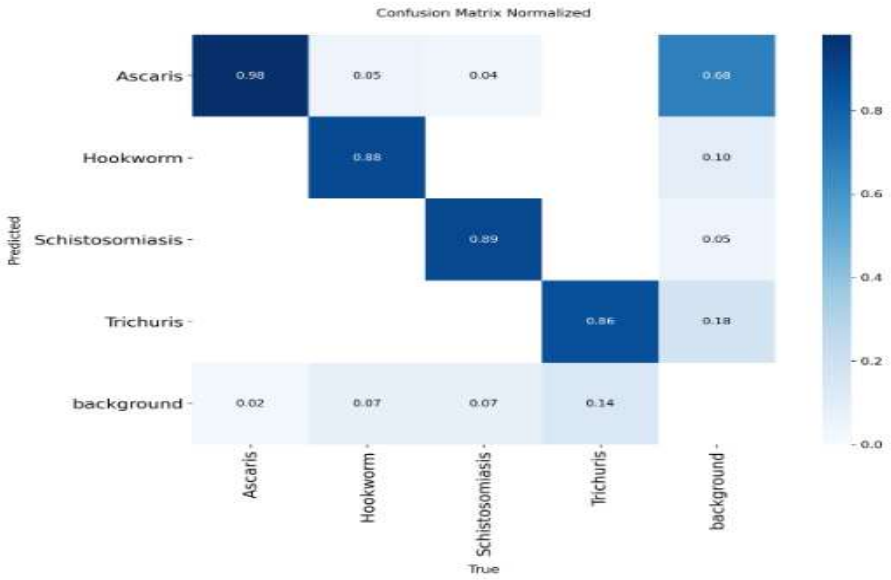


Figure 6: Confusion matrix of YOLOv12m model. This normalized confusion matrix illustrates the classification performance across five classes: four parasitic helminths (Ascaris, Hookworm, Schistosomiasis, Trichuris) and a background class.

Figure 7 depicts the training and validation loss curves of the YOLOv12m model across 50 epochs. All training loss components exhibit a steady decline, indicating effective optimization of both localization and classification tasks. Specifically, the training classification loss decreased from 3.35 to 0.36, the box loss reduced from 3.49 to 0.88, and the Distribution Focal Loss (DFL) declined from 2.75 to 1.20.

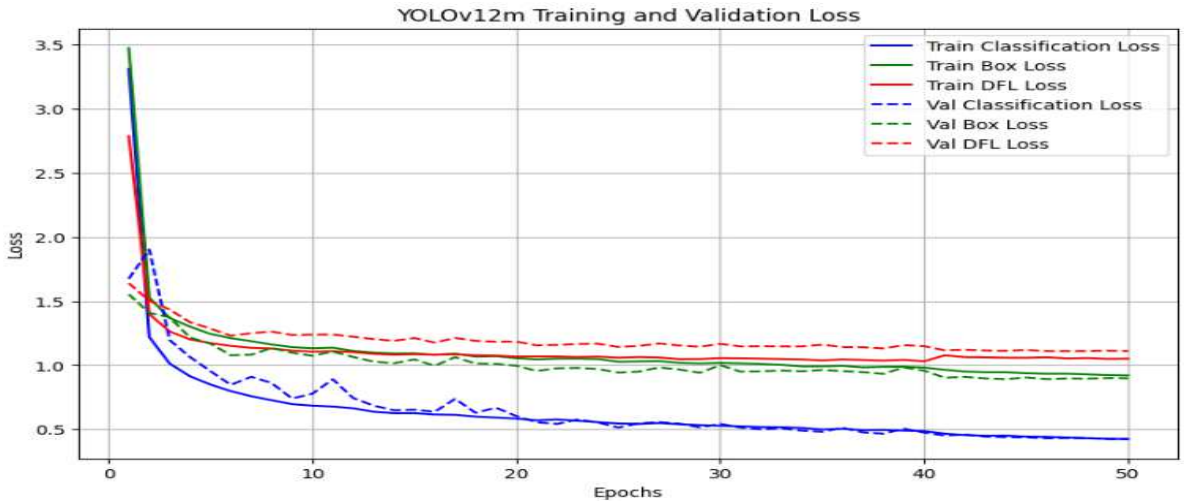


Figure 7: Training and validation loss curves for YOLOv12m over 50 epochs, showing classification, bounding box (Box), and Distribution Focal Loss (DFL) components for both training and validation phases.

A similar trend was observed for the validation losses, which showed notable improvement throughout training. The validation classification loss dropped from 1.75 to 0.36, the box loss decreased from 1.59 to 0.87, and the DFL reduced from 1.60 to 1.40. The close correspondence between training and validation loss trajectories suggests strong generalization capability and minimal overfitting. Moreover, the substantial reduction in validation classification and DFL losses highlights the model's robustness in accurately detecting soil-transmitted helminths and Schistosomia species. Overall, these results demonstrate the stability and effectiveness of the YOLOv12m model for automated parasite detection, supporting its applicability in resource-limited diagnostic settings.

Figure 8 illustrates the performance curves of the YOLOv12m model, including the precision–recall (PR) curve, F1-score curve, precision curve, and recall curve. The PR curve demonstrates strong detection capability, achieving an overall mAP@0.5 of 0.940, which indicates a favorable balance between precision and recall across confidence thresholds. The F1-score curve peaks at 0.900 at a confidence threshold of 0.580, reflecting the optimal trade-off between precision and recall. The recall curve reaches a maximum value of 0.970 at a confidence threshold of 0.000, indicating the model's high sensitivity in detecting target parasite classes. Conversely, the precision curve attains a maximum value of 1.000 at a confidence threshold of 0.994, demonstrating the model's ability to produce highly accurate predictions at higher confidence levels. Collectively, these curves confirm the robustness and reliability of the YOLOv12m model for parasite detection across varying confidence thresholds.

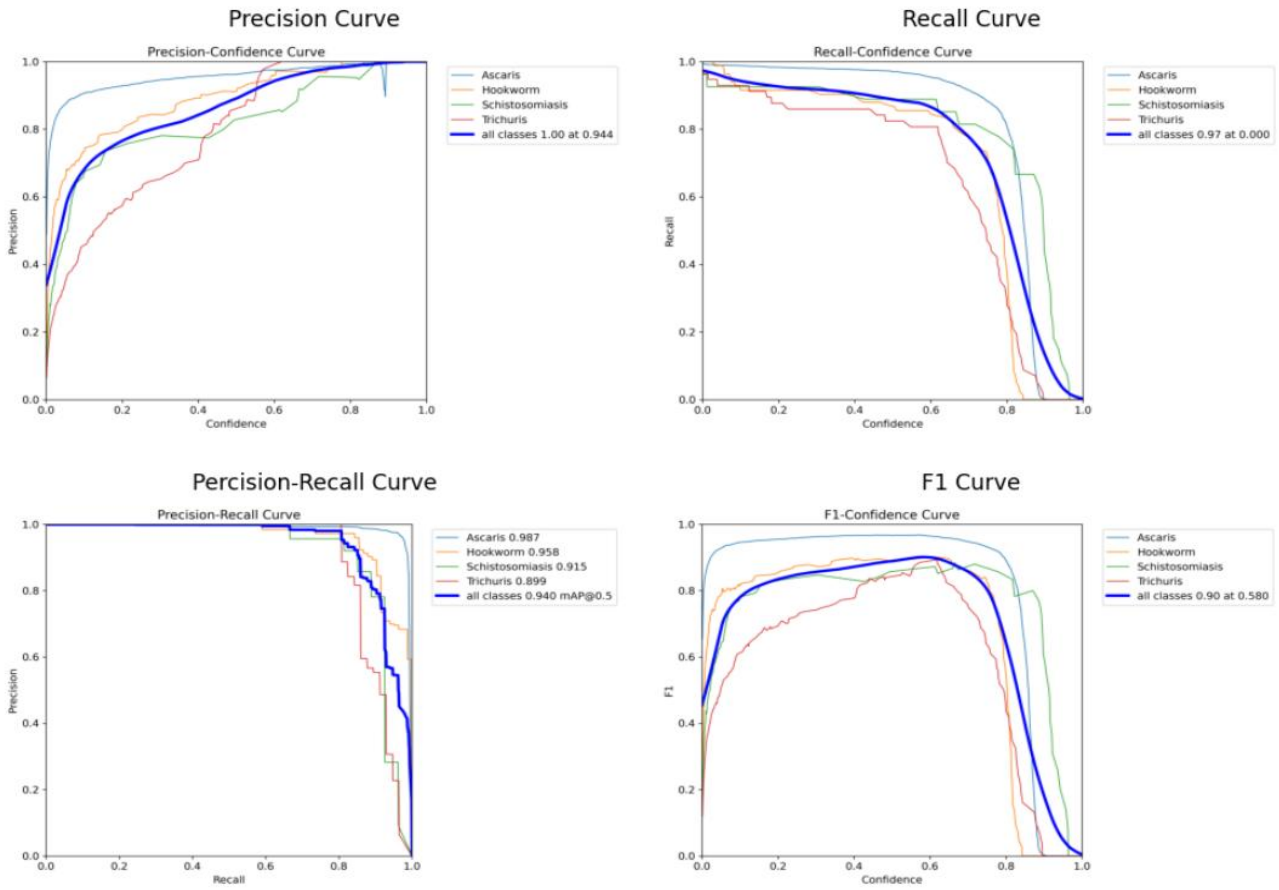


Figure 8: Precision, Recall, F1- score, and precision-recall curves of Yolo v12m model. Precision-Confidence, Recall-Confidence, Precision-Recall (PR), and F1-Confidence curves for four helminth classes (Ascaris, Hookworm, Schistosomiasis, Trichuris). The solid blue line represents the overall metrics for all classes. The PR curve's mAP@0.5 is 0.845, and the F1 curve peaks at an optimal confidence threshold of 0.890. If this figure is reproduced or adapted from another source, a copyright attribution statement is required at the end of this note.

Figure 9 illustrates representative inference outputs produced by the YOLOv12m model on parasitological microscopy images. Each detected helminth species is labeled with a confidence score normalized between 0 and 1, which reflects the model's estimated probability for the predicted class. In the displayed samples, Trichuris was detected with a confidence score of 0.86, Schistosomia with 0.93, Ascaris with 0.88, and Hookworm with 0.77. These confidence values indicate the degree of certainty associated with each prediction based on the learned feature representations of the model. Scores closer to 1 represent stronger model confidence in the assigned class label. The annotations demonstrate that YOLOv12m is capable of simultaneously localizing and classifying multiple helminth species within a single inference pass. The normalized confidence scores provide an interpretable measure of prediction reliability and support the qualitative assessment of detection results.

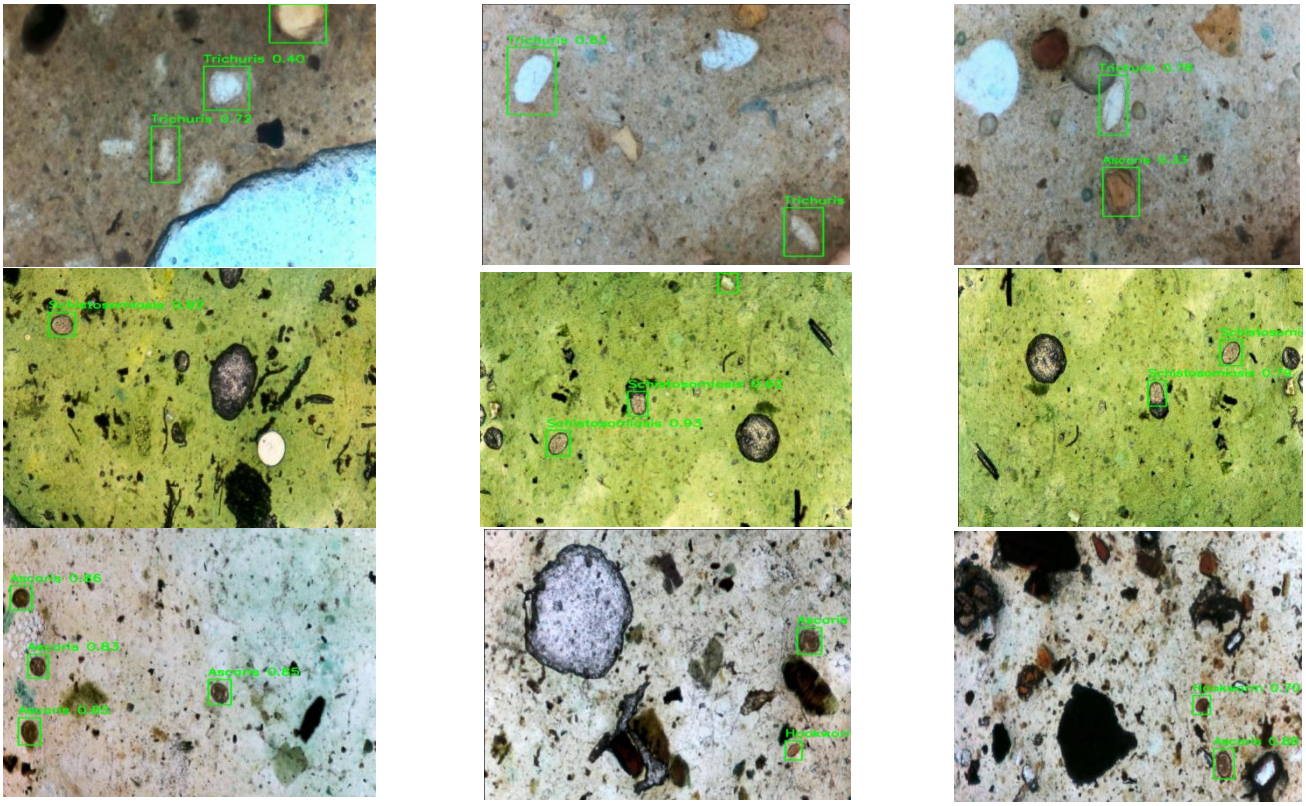


Figure 7: Sample inference outputs of the YOLOv12m model on microscopic parasite images with green detection result in number and a green box around the eggs of the parasites.

5.3. Discussion and Limitations.

This study demonstrates the feasibility and effectiveness of yolo based deep learning models for the automated, real-time detection and classification of soil transmitted helminths (STHs) and Schistosoma eggs in microscopy images. By evaluating YOLOv11 and YOLOv12 architectures across Nano, Small and Medium variants, this work provides a practical framework for balancing accuracy, speed and computational cost a critical consideration for deployment in resource constrained settings. The YOLOv12m model emerged as the most accurate detector, achieving a mAP@0.5 of 94% and a mAP@0.5–0.9 of 67.4%. This superior performance can be attributed to its advanced architectural refinements, particularly its enhanced Area Attention mechanisms and refined Residual Efficient Layer Aggregation Networks (RELARN) [27]. The Area Attention mechanism allows the model to dynamically focus on multi scale spatial regions within the dense and heterogeneous context of a fecal smear, improving its ability to distinguish morphologically similar and often overlapping parasite eggs [1, 6]. The RELARN block promotes more efficient gradient flow and hierarchical feature fusion, enabling the model to extract and aggregate discriminative fine-grained features crucial for accurate classification especially for smaller or subtly textured eggs like those of hookworm or Schistosoma [6, 7]. These innovations collectively enhance both localization precision and classification confidence, setting a new benchmark for automated parasitological diagnosis.

In contrast, the YOLOv11m model achieved the highest recall (88.6%) indicating a superior ability to minimize false negatives a clinically critical metric to avoid missing infections. This suggests that YOLOv11m's architecture, possibly through its balanced feature pyramid and attention mechanisms is exceptionally sensitive to the presence of parasite eggs even in suboptimal image conditions. For screening purposes where sensitivity is paramount YOLOv11m presents a compelling option. The YOLOv12n variant offers a distinct advantage for real time edge device deployment with the lowest computational cost (6.3 GFLOPs) and the fastest inference speed (159 FPS). While its sensitivity is reduced compared to the medium variants its efficiency makes it suitable for high throughput initial screening scenarios where computational resources are severely limited [9, 20].

When benchmarked against prior studies on AI-assisted parasite detection our YOLOv12m model shows competitive or superior performance. While direct comparisons are challenging due to differences in datasets and evaluation metrics the achieved mAP@0.5 of 94% surpasses many reported results in the literature for multiclass STH detection which often range between 85–92% mAP. The systematic use of 5-fold cross validation and statistical significance testing further strengthens the reliability and generalizability of our findings.

Despite the promising results, several limitations of this study must be acknowledged. First, the dataset, though carefully annotated was sourced from a limited geographical scope. This may introduce bias related to local staining techniques, microscope models and egg morphological variations potentially affecting model generalizability to other regions. Future work should incorporate multicenter, multi-protocol datasets to enhance robustness. Second, while the models perform well on clear images, performance can degrade with heavy debris, staining artifacts, or poor focus common challenges in field microscopy. Integrating pre-processing steps for background suppression or employing data augmentation strategies simulating these artifacts could improve resilience. Third, this study focused on detection accuracy and speed but did not validate the models in a live clinical workflow. Future research must involve field trials to assess practical impact, usability and integration with existing laboratory information systems.

Conclusion

This study successfully demonstrated the effectiveness of YOLO-based object detection models for automated identification and classification of soil-transmitted helminth and schistosomiasis eggs in microscopic images. The YOLOv12m model emerged as the best-performing variant, achieving a precision of 0.938, recall of 0.878, and mAP@0.5 of 0.94, significantly outperforming other model configurations while maintaining a reasonable inference speed of 57 FPS. The systematic evaluation across multiple YOLOv11 and YOLOv12 variants revealed a clear trade-off between detection accuracy and computational efficiency, with medium-scale models providing the optimal balance for diagnostic applications. The model demonstrated strong discriminative performance across all target parasite classes with confusion matrix analysis revealing only minor misclassifications primarily attributable to overlapping morphological features between certain species. The stable convergence and good generalization observed during training validate the model's reliability for deployment in clinical settings.

Practical impact of this research, YOLOv12m should be deployed as a decision support tool in diagnostic laboratories particularly in high throughput or resource limited settings. A hybrid workflow is recommended where the AI performs initial screening and experts verify uncertain predictions. For field use in remote areas the lightweight YOLOv12n variants should be adapted for mobile deployment [9, 27]. Establishing a continuous feedback loop for model retraining using misclassified clinical cases will ensure sustained improvement. Future work should include multicenter validation studies and expansion of the model to detect a broader range of parasite species integrated with existing laboratory information systems to streamline the diagnostic pipeline [17, 18].

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

Abrham Adamu: conceptualization, methodology, validation, formal analysis, experimentation, investigation, data curation, writing (original draft, review, and editing), visualization, and project administration. Etefa Belachew: methodology, formal analysis, data curation, experimentation, writing (review and editing), supervision, and advising. Kris Calpotura: resources, data curation, visualization, writing – review and editing, supervision and advising. Aregawi Gebremedhin: resources, data curation, visualization, writing – review and editing, supervision and advising.

Funding

No funding was received for this manuscript

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author upon request.

References

[1] Mohammed, M. A., Anzaku, E. T., Ward, P. K., Levecke, B., Krishnamoorthy, J., De Neve, W., & Van Hoecke, S. (2025). Detecting soil-transmitted helminth and Schistosomia mansoni eggs in Kato-Katz stool smear microscopy images: A comprehensive evaluation of YOLOv7 variants. PLOS Neglected Tropical Diseases. https://doi.org/10.1371/journal.pntd.0013234

[2] AI-supported versus manual microscopy of Kato-Katz smears for diagnosis of soil-transmitted helminth infections. Scientific Reports, 15(1), Article 73109. https://doi.org/10.1038/s41598-025-07309-7

[3] Lemma, W., et al. (2025). Constructing a predictive model for STH and schistosomiasis classification from microscopic images. BioMed Research International, [volume(issue)], [page range]. https://doi.org/10.1155/bmri/8074581

4] Oyibo, I., et al. (2025). Deep learning-based automated detection and multiclass classification of soil-transmitted helminths and Schistosomia mansoni eggs. BioMed Research International, [Volume(Issue)], pages. https://doi.org/10.1038/s41598-025-02755-9

[5] Deng, W., Wang, S., Wang, L., Lv, C., Li, Y., Feng, T., Qin, Z., & Xu, J. (2022). Laboratory evaluation of a basic recombinase polymerase amplification (RPA) assay for early detection of Schistosomia japonicum. Pathogens, 11(3), 319. https://doi.org/10.3390/pathogens11030319

[6] He, W., Zhu, H., Geng, J., Hu, X., Li, Y., Shi, H., Wang, Y., Zhu, D., Wang, H., Xie, L., Yang, H., & Li, J. (2024). Recognition of parasitic helminth eggs via a deep learning-based platform. Frontiers in Microbiology, 15. https://doi.org/10.3389/fmicb.2024.1485001

[7] AlDahoul, N., Abdul Karim, H., Momo, M. A., Escobar, F. I. F., Magallanes, V. A., & Tan, M. J. T. (2023). Parasitic egg recognition using convolution and attention network. Scientific Reports, 13, Article 41711. https://doi.org/10.1038/s41598-023-41711-3

[8] Mirzaei, O., Ilhan, A., Guler, E., Suer, K., & Sekeroglu, B. (2025). Comparative evaluation of deep learning models for diagnosis of helminth infections. Journal of Personalized Medicine, 15(3), 121. https://doi.org/10.3390/jpm15030121

[9] Xu, W., Zhai, Q., Liu, J., Xu, X., & Hua, J. (2024). A lightweight deep-learning model for parasite egg detection in microscopy images. Parasites & Vectors, 17, 454. https://doi.org/10.1186/s13071-024-06503-2

[10] Terven, J., Cordova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and beyond. arXiv preprint arXiv:2304.00501. https://arxiv.org/abs/2304.00501

[11] Jocher, G., et al. (2024). Ultralytics YOLO Docs. https://docs.ultralytics.com

[12] Wang, C., et al. (2024). YOLOv12: Official Implementation and Benchmark. https://github.com/

[13] Villamizar-Monsalve, M. A., Sánchez-Montejo, J., López-Abán, J., Vicente, B., Marín, M., Fernández-Ceballos, N., Peláez, R., & Muro, A. (2024). Development and application of an in vitro drug screening assay for Schistosomia mansoni schistosomula using YOLOv5. Biomedicines, 12(12), 2894. https://doi.org/10.3390/biomedicines12122894

[14] Degarege, A., Erko, B., Negash, Y., & Animut, A. (2022). Intestinal helminth infection, anemia, undernutrition and academic performance among school children in northwestern Ethiopia. Microorganisms, 10(7). https://doi.org/10.3390/microorganisms10071353

[15] Scavuzzo, C., Scavuzzo, J., Campero, M., Anegagrie, M., Aramendia, A., Benito, A., & Periago, V. (2022). Feature importance: Opening a soil-transmitted helminth machine learning model via SHAP. Infectious Disease Modelling, 7(1), 262–276. https://doi.org/10.1016/j.idm.2022.01.004

[16] Lee, C., Huang, P., Yeh, Y., Li, P., Chiu, C., Cheng, W., & Tang, P. (2022). Helminth Egg Analysis Platform (HEAP): An open platform for microscopic helminth egg identification and quantification based on the integration of deep learning architectures. Journal of Microbiology, Immunology and Infection, 55(3), 395–404. https://doi.org/10.1016/j.jmii.2021.07.014

[17] Oyibo, P., Meulah, B., Bengtson, M., van Lieshout, L., Oyibo, W., Diehl, J., Vdovine, G., & Agbana, T. (2023). Two-stage automated diagnosis framework for urogenital schistosomiasis in microscopy images from low-resource settings. Journal of Medical Imaging, 10(4), 044005. https://doi.org/10.1117/1.JMI.10.4.044005

[18] Savitha, V., & Karthick, M. (2023). Parasitic egg detection from microscopic images using convolutional neural networks. Tamjeed Journal of Healthcare

Engineering and Science Technology, 1(1), 24–34. https://doi.org/10.59785/tjhest.v1i1.3

[19] Singh, Y., Farrelly, C., Hathaway, Q., Leiner, T., Jagtap, J., Carlsson, G., & Erickson, B. (2023). Topological data analysis in medical imaging: Current state of the art. Insights into Imaging, 14(1). https://doi.org/10.1186/s13244-023-01413-w

[20] Mosiichuk, V., Sampaio, A., Viana, P., Oliveira, T., & Rosado, L. (2023). Improving mobile-based cervical cytology screening: A deep learning nucleus-based approach for lesion detection. Applied Sciences, 13(17). https://doi.org/10.3390/app13179850

[21] Ally, O., Kanoi, B., Ochola, L., Nyanjom, S., Shiluli, C., Misinzo, G., & Gitaka, J. (2024). Schistosomiasis diagnosis: Challenges and opportunities for elimination. PLOS Neglected Tropical Diseases, 18(7), e0012282. https://doi.org/10.1371/journal.pntd.0012282

[22] World Health Organization. (2023). Schistosomiasis and soil-transmitted helminthiases: Progress report and strategic updates. https://www.who.int

[23] Meulah, B., Oyibo, P., Bengtson, M., van Lieshout, L., Diehl, J., Vdovine, G., & Agbana, T. (2022). A review on innovative optical devices for the diagnosis of human soil-transmitted helminthiasis and schistosomiasis: From research and development to commercialisation. Parasitology, 149(12), 1473–1488. https://doi.org/10.1017/S0031182022001664

[24] Naing, K. M., Boonsang, S., Chuwongin, S., Kittichai, V., Tongloy, T., Prommongkol, S., Dekumyoy, P., & Watthanakulpanich, D. (2022). Automatic recognition of parasitic products in stool examination using object detection approach. PeerJ Computer Science, 8, e1065. https://doi.org/10.7717/peerj-cs.1065

[25] Parra, C., Grijalva, F., Núñez, B., Núñez, A., Pérez, N., & Benítez, D. (2022). Automatic identification of intestinal parasites in reptiles using microscopic stool images and convolutional neural networks. PLOS ONE, 17(8), e0271529. https://doi.org/10.1371/journal.pone.0271529

[26] Khanam, R., & Hussain, M. (2024). YOLOv11: An overview of the key architectural enhancements (arXiv:2410.17725) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2410.17725

[27] Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention centric real time object detectors (arXiv:2502.12524) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2502.12524