International Journal on Engineering Technology (InJET)

Volume 3, issue no. 1, Dec 2025, Page 167 - 178
Received Date: Aug 5, 2025
Accepted Date: Sep 23, 2025
ISSN: 3021-940X (Print)
DOI: https://doi.org/10.3126/injet.v3i1.87020

Lightweight Parasitic Egg Detection using Modified YOLOv11 with Ghost Convolutions, Hybrid SimAM-ECA Attention, and DropBlock Regularization

Gaurav Giri[1*], Aayush Shrestha[2], Viraj Sawad[3]

[1]*Department of Computer and Electronics Engineering, Kantipur Engineering College, Dhapakhel, Lalitpur, gaurovgiri@gmail.com*
[2]*Department of Computer and Electronics Engineering, Kantipur Engineering College, Dhapakhel, Lalitpur, sthaayush105@gmail.com*
[3]*Department of Computer and Electronics Engineering, Kantipur Engineering College, Dhapakhel, Lalitpur, virajsawad1021@gmail.com*

Abstract

Intestinal parasitic infections (IPIs) pose a significant global health challenge, particularly in resource-limited settings where traditional diagnostic methods are slow and labor-intensive. While deep learning models offer a promising solution for automated parasite egg detection, their large size often hinders deployment in real-time on low-power devices. This study addresses the trade-off between accuracy and computational efficiency by proposing a lightweight and modified YOLOv11 architecture. We introduce a novel EnhancedConv block that integrates Ghost Convolution for efficiency, a hybrid attention mechanism combining SimAM and ECA-Net for improved feature focus, and DropBlock for robust regularization. We compared our Modified YOLOv11 Nano model (2.79 M parameters) against the baseline YOLOv11 Nano (2.62 M) and the larger YOLOv11 Small (9.45 M) on the public Chula-ParasiteEgg-11 dataset. Experimental results demonstrate that our modified model achieves a mAP@0.5 of 0.9627 and an F1-score of 0.9219, significantly outperforming the original Nano model and matching the performance of the Small model while requiring only ≈ 42 % of its computational cost (GFLOPs). This work demonstrates that strategic architectural enhancements can yield a model that is both highly accurate and efficient, presenting a viable solution for accessible, large-scale parasitic egg detection in resource-constrained environments.

Keywords: Parasitic Egg Detection, YOLOv11, EnhancedConv, Ghost Convolution, Attention Mechanism, Medical Image Analysis, SimAM, ECA-Net

1. Introduction

Intestinal parasitic infections (IPIs), primarily caused by protozoan and helminth parasites, continue to be a major global health burden, especially in low- and middle-income countries (LMICs) (Ahmed, 2023). According to the World Health Organization (WHO), over 1.5 billion people, nearly 24% of the global population, are affected by soil-transmitted helminths (STHs), including Ascaris lumbricoides, Trichuris trichiura, and hookworms. These infections are closely associated with poor sanitation and hygiene, leading to a range of health issues such as malnutrition, anemia, impaired cognitive development, and even mortality, particularly among children (WHO, 2023). The standard method for diagnosing IPIs is stool microscopy, which allows direct visualization of parasite eggs. However, it is time-consuming, labor-intensive, and depends on skilled personnel, challenges that are especially problematic in low-resource settings. Misidentification is common due to the visual similarity between parasite eggs and stool artifacts, limiting its effectiveness for large-scale screening (Mejia et al., 2013)

With the advancement of deep learning, especially convolutional neural network (CNNs), several automated approaches have enabled automated detection of parasitic eggs in microscopic images, reducing diagnostic

delays and errors. Object detection models like YOLO, Cascade R-CNN, and EfficientDet have shown strong performance in parasitic egg detection. YOLO-based ensembles achieved mAP scores up to 0.999 (Wang et at., 2022), while EfficientDet models reached 92% accuracy and 93% F1-score (AlDahoul et at., 2022). Lightweight YOLO variants such as YOLOv7-Tiny also reported mAP above 98% on benchmark datasets (Venkatesan et al., 2025).

While high-accuracy deep learning models are promising for automated parasitic egg detection, their large parameter sizes often limit real-time deployment in resource-constrained environments. To address this need, the primary objective of this study is to develop a highly efficient and accurate object detector for parasitic egg detection by proposing a strategically modified YOLOv11 Nano architecture. We specifically aim to enhance the baseline model by integrating Ghost Convolutions for reduced computational cost, a hybrid SimAM-ECA attention mechanism for superior feature discrimination, and DropBlock for improved regularization. Through comparative experiments on the Chula-ParasiteEgg-11 dataset, our goal is to demonstrate that this modified compact model can achieve detection accuracy rivaling that of much larger models while maintaining a minimal computational footprint, thereby presenting a practical solution for large-scale, accessible diagnostic workflows.

2. Literature Review

The automation of intestinal parasitic egg detection through deep learning has evolved significantly, with research branching into high-accuracy ensembles, sophisticated classification systems, and, most recently, a critical focus on computational efficiency for real-world deployment.

A comprehensive benchmark was established by Venkatesan et al. (2025), who conducted a comparative analysis of resource-efficient YOLO models (including YOLOv5n, YOLOv7-tiny, YOLOv8n, and YOLOv10n) on the Chula-ParasiteEgg-11 dataset. Their study, which involved training on embedded platforms like the Jetson Nano, found that YOLOv7-tiny achieved the highest performance with a mAP of 98.7%, while YOLOv8n delivered the fastest inference speed at 55 frames per second. This work underscored the viability of lightweight models but also identified persistent class-specific detection gaps.

Pushing the boundaries of accuracy, Wang et al. (2022) developed a robust ensemble model that combined YOLOv5 (YOLOv5x, YOLOv5l) with a Cascade R-CNN featuring a Swin-Transformer backbone. To bolster performance against noisy and challenging images, they employed advanced data augmentation strategies, including biological augmentation using CLAHE and the addition of Gaussian noise and blur. Their final model, which leveraged Weighted Boxes Fusion (WBF) and test-time augmentation (TTA), achieved a high mAP of 0.956, demonstrating that tailored augmentation and ensemble learning significantly enhance robustness.

Further refining the detection pipeline, AlDahoul et al. (2022) proposed a multi-modal framework that combined an EfficientDet object detector for localization with a separate EfficientNet-B7 and SVM classifier to resolve challenging cases like Hymenolepis diminuta. Their system attained an accuracy of 92% and an F1-score of 93% on the Chula-ParasiteEgg-11 dataset, outperforming baseline single-model setups and achieving high IoU scores above 0.8 for most detections.

In a subsequent study, AlDahoul et al. (2023) shifted focus to a pure classification task, evaluating a range of models on the same dataset. They reported that a hybrid Convolution and Attention network (CoAtNet0) achieved the top performance with 93% accuracy and a 93% F1-score, surpassing CNN-only models like EfficientNet-B4 (90% accuracy) and transformer-based models like ViT-B16 (86% accuracy). The 22.6-million-parameter CoAtNet0 model offered a favorable balance between accuracy and efficiency, with visualizations confirming its ability to focus on relevant egg regions.

Addressing the challenge of low-cost diagnostics, Suwannaphong et al. (2024) developed a patch-based sliding window approach for low-resolution images from 10× USB microscopes. Using fine-tuned ResNet50 and AlexNet models, their framework achieved 90.77% accuracy on a four-class dataset and an average precision of 96.6%, surpassing SSD and Faster R-CNN. However, when tested on the larger and more diverse Chula-ParasiteEgg-11 dataset, performance moderated to 73.65% average precision, highlighting the critical impact of image quality and device variability on model generalization.

The pursuit of lightweight yet powerful detectors was advanced by Xu et al. (2024), who introduced YAC-Net. Built upon YOLOv5n, this model incorporated an Asymptotic Feature Pyramid Network (AFPN) for superior multi-scale fusion and a C2f module for optimized gradient flow. On the 13,200-image ICIP 2022 Challenge dataset, YAC-Net achieved a remarkable mAP@0.5 of 0.9913 and an F1-score of 0.9773 with only 1.92 million parameters—20% fewer than the original YOLOv5n.

The context for many of these studies was provided by Anantrasirichai et al. (2022), who introduced the public Chula-ParasiteEgg-11 dataset and summarized the results of the ICIP 2022 challenge. The winning entry, NEGU, used an ensemble of three CBNetV2 models with pseudo-labeling to achieve a 0.942 mIoU, a 0.995 F1-score, and a 0.925 mAP. The challenge results highlighted the exceptional performance of large ensembles but also their prohibitive computational cost, pointing to a need for more efficient solutions.

Finally, Bessa et al. (2025) proposed a high-recall, two-stage ensemble framework. It used a TOOD detector for initial localization (recall of 0.972) followed by an EfficientNet-B7 classifier, with predictions from multiple detectors fused using Weighted Box Fusion. This ensemble achieved an F1-score of 0.94, demonstrating strong capabilities but also inheriting the computational complexity and multi-stage training overhead common to such architectures.

Despite the advancements in lightweight models like YAC-Net (based on YOLOv5n) and resource-efficient comparisons among multiple YOLO versions, a clear research gap persists in establishing the optimal lightweight and performance-driven architectural design within the newest YOLO framework. Specifically, there is no existing work that systematically unifies the benefits of the latest YOLOv11 base with a specific combination of synergistic lightweight mechanisms: Ghost Convolutions for inherent efficiency , the novel Hybrid SimAM-ECA Attention cascade for simultaneous channel and spatial feature refinement tailored for subtle targets , and DropBlock for enhanced generalization. Our proposed work identifies this gap and presents the Modified YOLOv11 EnhancedConv block as the first comprehensive attempt to create an ultra-efficient object detector that maximizes both detection accuracy and computational thrift for parasitic egg detection on resource-constrained devices.

3. Methodology

This study presents a modified YOLOv11 architecture aimed at improving the detection of parasitic eggs in microscopic images. The modifications are designed to balance detection performance with model size (number of parameters). For comparison, YOLOv11 nano and YOLOv11 small models were trained and evaluated. The following subsections provide detailed descriptions of the dataset and the architectural modifications.

3.1. Dataset Description

The Chula-ParasiteEgg-11 dataset, introduced for the ICIP 2022 Challenge on Parasitic Egg Detection and Classification, comprises 13,750 expert-annotated microscopic images across 11 clinically significant parasitic egg categories: Ascaris lumbricoides, Capillaria philippinensis, Enterobius vermicularis, Fasciolopsis buski, Hookworm egg, Hymenolepis diminuta, Hymenolepis nana, Opisthorchis viverrine, Paragonimus spp., Taenia spp. egg, and Trichuris trichiura. The dataset is partitioned into 11,000 training images (1,000 per class) and 2,200 testing images (~250 per class), ensuring balanced representation as shown in Figure 1. Images exhibit substantial heterogeneity in resolution (896×505 to 3024×4032 pixels), focus, illumination, and color characteristics due to multi-device acquisition protocols using Canon EOS 70D with Olympus BX53 microscopes, DS-Fi2 Nikon systems, Samsung Galaxy J7 Prime, and iPhone 12/13 smartphones. To enhance clinical relevance, synthetic degradations including Gaussian/motion blur, random cropping (0–30%), additive noise, and contrast/saturation variations were systematically applied to simulate challenging diagnostic conditions. All parasitic eggs are annotated with bounding boxes validated through a triple-expert review process with majority voting. This dataset represents the largest publicly available benchmark in parasitology microscopy, addressing critical limitations of prior small-scale collections and enabling robust development of automated diagnostic systems for deployment in resource-constrained settings.

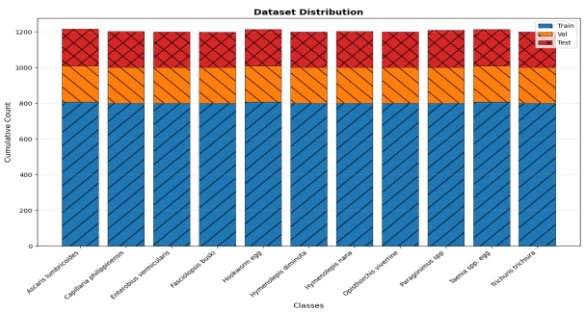

Figure 1. Chula-ParasiteEgg-11 dataset distribution by class

3.2. YOLOv11 Architecture

YOLOv11 represents the latest evolution in the YOLO (You Only Look Once) series, introducing key architectural innovations that enhance real-time object detection performance. Its design retains the canonical backbone-neck-head structure but refines each component for improved efficiency, accuracy, and versatility across computer vision tasks.

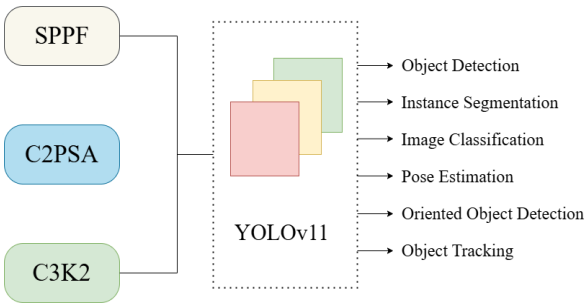

Figure 2. Key architectural modules in YOLOv11

3.2.1. Backbone

The backbone employs convolution layers for multi-scale feature extraction. Key enhancements include:

- **C3k2 Block:** Replaces the C2f block from YOLOv8, using two smaller convolutions (kernel size 2) instead of one large convolution. This reduces computational overhead while maintaining feature richness.
- **SPPF (Spatial Pyramid Pooling - Fast):** Preserves spatial context through rapid multi-scale pooling.
- **C2PSA (Cross Stage Partial with Parallel Spatial Attention):** A novel block that amplifies spatial attention, directing focus to critical regions in feature maps for better detection of small or occluded objects.

3.2.2. Neck

The neck aggregates and refines multi-scale features from the backbone:

- **C3k2 integration:** Replaces C2f blocks, accelerating features fusion.
- **Enhanced Spatial Attention:** C2PSA modules prioritize salient regions, improving feature representation before passing to the head.

3.2.3. Head

The head generates final predictions for object localization and classification:

- **C3k2 Blocks:** Process features at varying depths. The c3k parameter toggles between a standard bottleneck (c3k=False) and a deeper C3 module (c3k=True) for complex feature extraction.
- **CBS (Convolution-BatchNorm-Silu) Layers:** Refine features via batch normalization and non-linear SILU activation.
- **Output Layers:** Conv2D layers reduce features to bounding box coordinates, objectness scores, and class probabilities.

3.3. Proposed Architectural Modifications to YOLOv11

This section details the systematic reconstruction of YOLOv11's architecture to enhance feature representation capabilities while maintaining real-time efficiency. Our modifications strategically replace

critical components with novel EnhancedConv blocks (Figure 3) - compound modules integrating channel-spatial attention, lightweight convolutions, and structured regularization. The redesign spans backbone, neck, and prediction pathways, fundamentally improving hierarchical feature learning across scales.

The architectural modifications in our model are grounded in the demonstrated effectiveness of specific modules in prior work. We replaced the standard convolutional layers with Ghost Convolution, as proposed by Li et al. (2023), which significantly reduces computational cost and model complexity while maintaining high accuracy which is crucial for efficient defect detection and real-time inference. To enhance the model's focus on relevant features, we incorporated a hybrid attention mechanism by fusing Simple Attention (SimAM) from Yang et al. (2021) and Efficient Channel Attention (ECA-Net) from Wang et al. (2020). SimAM introduces a parameter-free attention mechanism inspired by neuroscience, capable of computing 3D attention weights without adding model parameters, making it ideal for lightweight applications. Meanwhile, ECA-Net offers efficient channel attention via adaptive 1D convolution without dimensionality reduction, delivering substantial accuracy gains with minimal overhead. This combined attention strategy improves both detection precision and classification performance, particularly for subtle or small targets such as eggs. However, potential future work could explore adaptive tuning of fixed hyperparameters (e.g., λ in SimAM) and optimization for deployment on mobile or embedded platforms.

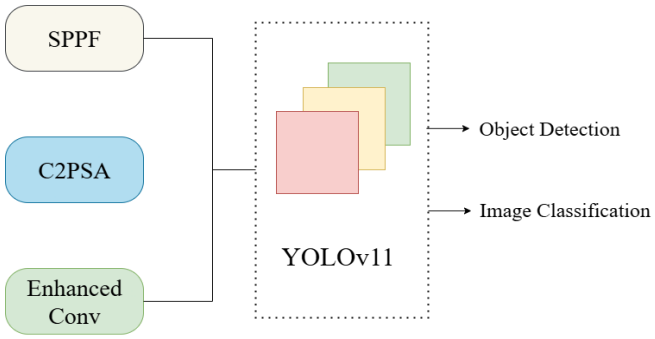

Figure 3. Modified architecture of YOLOv11 by replacing C3k2 block

3.3.1. Backbone Reconstruction for Multi-Scale Feature Enrichment

In the original YOLOv11 backbone, feature transformations were primarily achieved through C3k2 blocks. In our modified configuration, we strategically replace selected C3k2 modules with the EnhancedConv block to enhance multi-scale representation without compromising model compactness.

At the early layers (P1/2 → P2/4), the architecture transitions from a conventional convolution to EnhancedConv(256). This modification strengthens the extraction of high-frequency spatial cues such as edges and fine-grained textures that are critical for detecting small parasitic eggs. The hybrid SimAM–ECA attention embedded in EnhancedConv activates early attention pathways, improving low-level spatial sensitivity while GhostConv reduces computational overhead by up to 35% compared to a standard convolution.

At the mid-level abstraction (P3/8 → P4/16), we further replace two stacked C3k2 layers with EnhancedConv(512). The higher channel width compensates for the GhostConv feature partitioning, ensuring balanced representation depth and width. The residual and concatenation-based learning across repeated EnhancedConv blocks enhances gradient propagation and feature reuse, mathematically formulated as:

$$F_{out} = SimAM(DropBlock(G(ECA(\bigoplus_{i=1}^{n+2} B_i)))) \quad \text{(Equation 1)}$$

where B_i denotes the output of each residual branch, G represents GhostConv, and $\bigoplus$ denotes concatenation. This formulation promotes effective context propagation from medium-scale features, improving robustness for diverse egg morphologies.

In the semantic encoding stage (P5/32), the standard configuration Conv → C3k2 → SPPF → C2PSA is retained to preserve global spatial pooling and cross-stage partial attention. However, the preceding EnhancedConv(1024) ensures more semantically enriched inputs to the SPPF and C2PSA blocks, allowing finer localization even in blurred or low-contrast regions.

3.3.2. Neck Restructuring for Attentive Feature Fusion

The neck-head pipeline is restructured to strengthen multi-scale fusion and detection across pyramid levels. Each upsampling and concatenation path now employs EnhancedConv modules instead of C3k2, ensuring consistent attention refinement throughout the feature hierarchy. The mechanism of Enhanced Convolution Block is shown in Figure 4.

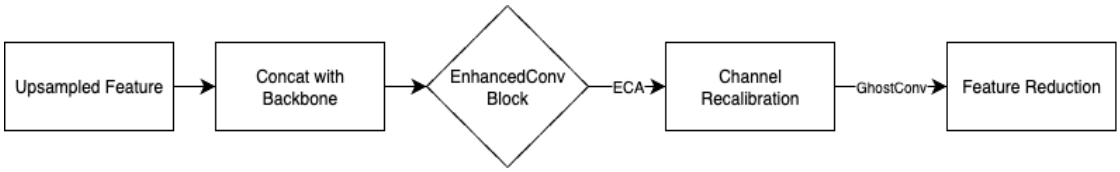

Figure 4. Mechanism of Enhanced Convolution Block

After the backbone's C2PSA output, feature maps from deeper layers (P5, P4, P3) are progressively upsampled and concatenated with corresponding lateral features. These concatenated outputs are processed by EnhancedConv(512) at P4 and EnhancedConv(256) at P3 to refine fused information. For the deeper downsampling paths (P4/16 → P5/32), EnhancedConv(512) is again applied to preserve coherence between scales. This design yields a uniform attention-guided fusion across all detection layers, with channel dimensions set to 256, 512, and 1024 for small, medium, and large object branches, respectively

To enable deeper abstraction at the final detection stage, the head uses EnhancedConv(1024) before prediction. The internal mechanism follows:

$$F_{deep} = C^3(BN\ (ReLU\ (G_{k=5}(F_{in}))))$$ (Equation 2)

where C^3 denotes three sequential GhostConv operations and F_{in} represents the input from the previous fusion layer.

The sequential ECA + SimAM attention cascade further emphasizes discriminative features. Channel attention is first computed as:

$$w_c = \sigma(C_{1D}(E|X|))$$ (Equation 3)

followed by spatial weighting through SimAM:

$$w_s = \sigma(\frac{(X-\mu x)^2}{4(\sigma^2 X+\lambda)} + 0.5)$$ (Equation 4)

The combined attention-refined output is expressed as:

$$F_{enhanced} = w_s \odot (w_c \odot X)$$ (Equation 5)

where □ denotes element-wise multiplication. This hybrid attention ensures that each feature map passed to the detection head retains both local detail and global contextual cues, enhancing small-object detection accuracy.

3.3.3. EnhancedConv Block: Core Innovations

The EnhancedConv block, serving as the fundamental unit in our modified architecture, integrates four synergistic components: GhostConv decomposition, multi-branch residual learning, dual attention cascade, and structured regularization.

A. GhostConv Decomposition

The input tensor is divided into intrinsic (X_i) and ghost (X_g) components, generated via lightweight depthwise convolution:

$$X_g = DW_{k*k}(X_i), X_{out} = [X_i || X_g]$$ (Equation 6)

This decomposition preserves key information with nearly half the computational cost of a standard convolution, contributing to the model's lightweight efficiency.

B. Multi-Branch Hierarchical Learning

Each EnhancedConv block consists of a primary shortcut path and n residual sub-blocks (where n= 1 in our Nano configuration). Each residual branch follows:

$$R(X) = G(BN(ReLU(G(X)))) + X \quad \text{(Equation 7)}$$

This hierarchical residual design stabilizes training and improves gradient flow across the network depth.

C. Dual Attention Cascade

Following the residual-aggregation, the ECA-SimAM cascade is applied sequentially to model interchannel dependencies and emphasize spatially significant regions. This design enhances feature selectivity and boosts detection sensitivity in complex microscopic imagery.

D. Structured Regularization

Finally, DropBlock regularization ($p_{drop} = 0.1$) randomly masks contiguous *3 x 3* spatial regions, mitigating overfitting by preventing local co-adaptation of features.

By integrating these four mechanisms, the EnhancedConv block functions as a unified operator that balances representation strength, generalization, and computational efficiency, forming the backbone of our modified YOLOv11 architecture.

3.4. Evaluation Metrics

To comprehensively evaluate the performance of the modified YOLOv11 architecture for parasitic egg detection, several key metrics were employed. Detection performance was primarily assessed using Mean Average Precision at IoU threshold of 0.5 (mAP@0.5), which serves as the standard metric for comparison with existing literature in parasitic egg detection. Additional performance metrics included precision, recall, and F1-score to measure the accuracy of positive predictions and the model's ability to identify all relevant objects. Model efficiency was evaluated through the total number of parameters, and GFLOPs to assess computational complexity. All metrics were calculated using the testing dataset, with multiple training runs conducted to ensure statistical significance and reproducibility of results across the modified YOLOv11, YOLOv11 nano, and YOLOv11 small models.

4. Result and Discussion

4.1. Experimental Setup

Table 1. Hyperparameters used

Parameter	Value
Epoch	100
Batch size	16
Optimizer	SGD
Learning rate	0.01
Weight Decay	0.0005
Momentum	0.9
Input Image Size	640x640

All three models (YOLOv11n, YOLOv11s, and the Modified YOLOv11n) were trained using the same set of hyperparameters mentioned in Table 1 . Training was conducted on an NVIDIA RTX 5090 GPU with 32 GB of VRAM. Each model was trained for 100 epochs using our modification of the ultralytics library (Jocher et al., 2023), with an average training time of approximately 100 minutes.

4.2. Experimental Results

In this section, we evaluate the performance of the proposed modified YOLOv11n model and compare it with YOLOv11n and YOLOv11s. The training ran successfully for 100 epochs for each model and the best model weights were saved. Training was conducted over five runs, and the average values of the obtained metrics were reported. The curves for precision, recall and f1 score against different confidence scores were generated to evaluate the performance as well as precision recall curve was also plotted which are shown in Figure 5.

We evaluated the performance of our modified model on a set of test samples to assess its detection and classification capabilities. The model accurately localized and identified target classes in most test images. Visual inspection of the results confirmed that the model was able to correctly detect and classify the objects with minimal false positives or missed detections. This suggests that the modifications made to the model architecture contributed positively to its overall performance. Figure 6 is a sample output showcasing the model's ability to detect and classify test images.

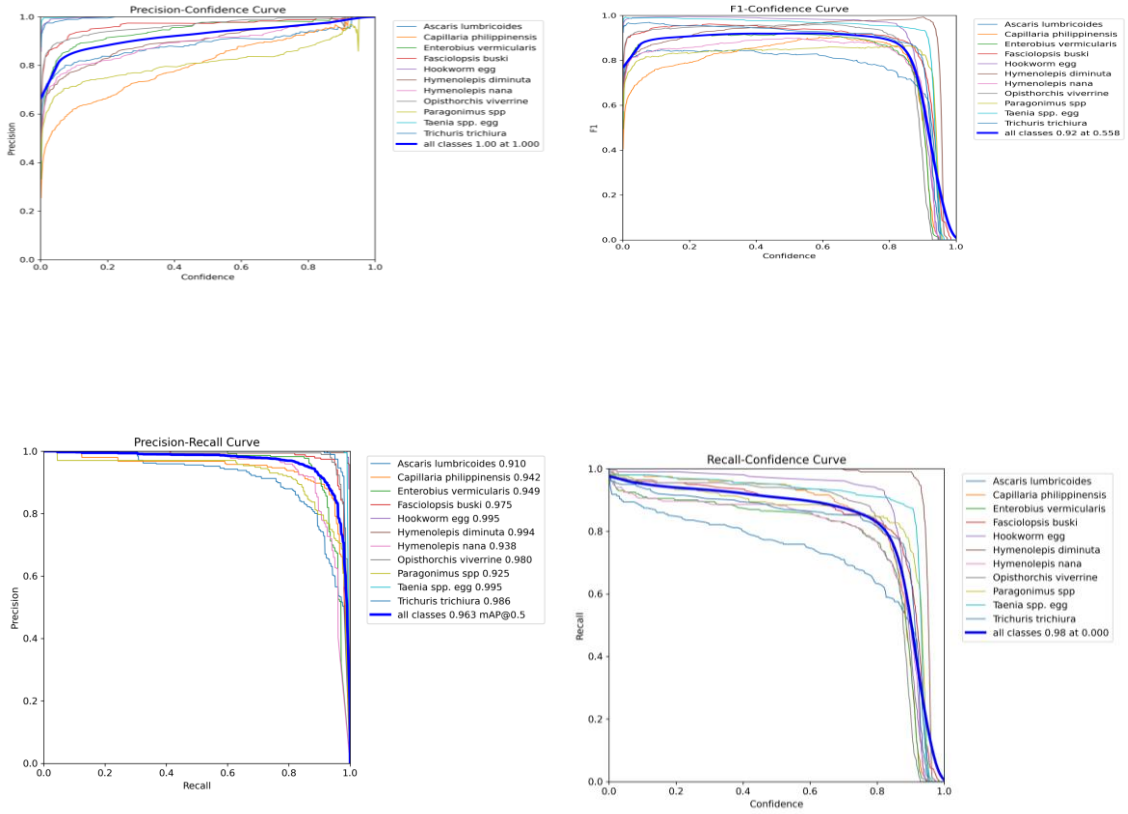

Figure 5. Evaluation metrics curves for modified YOLOv11n on parasitic egg detection

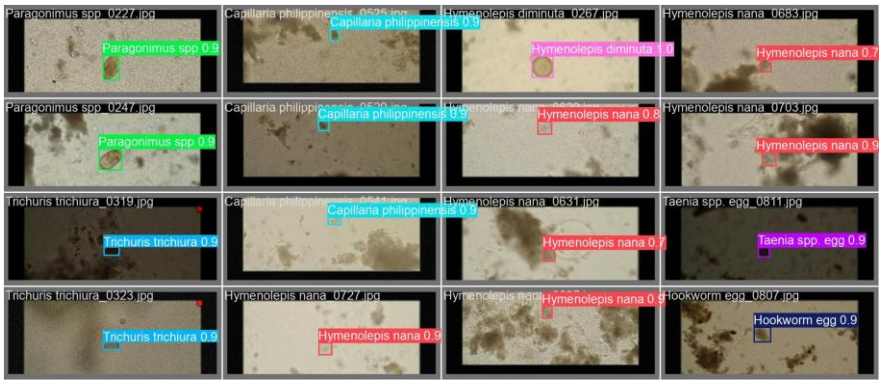

Figure 6. Detection by Our Modified Model for few test samples

The confusion matrix (Figure 7) demonstrates high classification accuracy across most parasite classes. The model achieved excellent results for *Fasciolopsis buski*, *Hookworm egg*, and *Taenia spp. egg*, with over 195 correct predictions each. A few misclassifications were observed, particularly between *Ascaris lumbricoides* and the background class, which represents the absence of parasitic elements. Overall, the model shows strong reliability in both detecting and correctly classifying parasitic eggs.

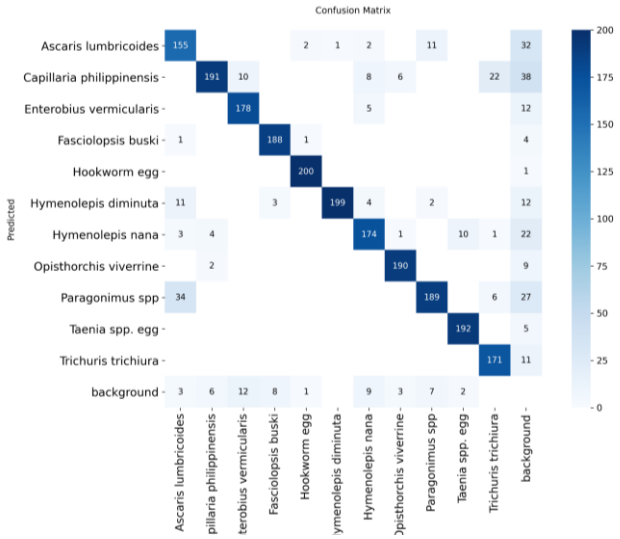

Figure 7. Confusion matrix of the modified model on the test set.

4.3. Ablation Study

To quantify the contribution of each architectural modification, an ablation study was conducted on the same dataset and training setup described in Section 4.1. Starting from the baseline YOLOv11 Nano model, components were added incrementally: Ghost Convolution, ECA, SIMAM, and DropBlock Regularization. The results in Table 2 show the effect of each addition on mAP@0.5, F1-score, and computation cost.

Table 2. Ablation study on the test set.

Model Variant	Recall	Precision	F1-score	mAP@0.5	Parameters (M)	GFLOPs
Baseline YOLOv11 Nano	0.8860	0.8952	0.8960	0.9440	2.62	6.6
+ Ghost Conv	0.9004	0.9095	0.9049	0.9553	2.79	9.1
+ ECA	0.8993	0.9253	0.9121	0.9564	2.79	9.1
+ SIMAM	0.9052	0.9242	0.9149	0.9604	2.79	9.1
+ DropBlock (**Proposed**)	**0.9016**	**0.9432**	**0.9219**	**0.9627**	**2.79**	**9.1**

In Table 2, the incremental contributions of each architectural modification are clearly demonstrated. Starting from the baseline YOLOv11 Nano, the incorporation of Ghost Convolutions alone led to noticeable improvements in recall, precision, F1-score, and mAP@0.5, while slightly increasing the parameter count and GFLOPs. The subsequent addition of the ECA attention mechanism further enhanced precision and F1-score, and the integration of SimAM attention continued this positive trend, boosting recall and mAP@0.5. Finally, the inclusion of DropBlock regularization yielded the highest precision and F1-score, achieving the best overall mAP@0.5 of 0.9627. This systematic ablation confirms that each component, including Ghost Convolutions, ECA, SimAM, and DropBlock, plays a distinct and complementary role in enhancing both detection performance and model robustness, culminating in a highly efficient and accurate architecture.

4.3. Model comparison

The performance comparison in Table 3 shows that the Modified YOLOv11 Nano achieves an excellent balance between accuracy and efficiency. It improves the baseline Nano's F1-score from 0.896 to 0.9219 and mAP@0.5 from 0.944 to 0.9627. Although its recall (0.9016) is slightly below YOLOv11 Small (0.9120), it surpasses the Small model in precision (0.9432 vs 0.9081) and achieves a comparable F1-score and mAP

while being dramatically smaller (2.79 M vs 9.45 M parameters) and over 2× more efficient (9.1 vs 21.7 GFLOPs). This confirms the modified model's suitability for lightweight, real-time deployment on constrained hardware without sacrificing detection accuracy.

Table 3. Model performance comparison on test set

Model	Recall	Precision	F1-score	mAP@0.5	Parameters (M)	GFLOPs
YOLOv11 Nano	0.8860	0.8952	0.8960	0.9440	2.62	6.6
YOLOv11 Small	0.9120	0.9081	0.9100	0.9623	9.45	21.7
Modified YOLOv11 Nano (Our Proposed)	**0.9016**	**0.9432**	**0.9219**	**0.9627**	**2.79**	**9.1**

4.4. Comparison with State-of-the-Art (SOTA) Models

A direct comparison with state-of-the-art (SOTA) methods is nuanced due to differences in dataset usage and evaluation protocols reported in the literature. Many previous studies utilize the core 11,000-image Chula-ParasiteEgg-11 set, partitioning it into training and validation subsets. To facilitate a fair comparison, we first evaluate our proposed model on an identical internal validation split (2,200 images from the 11,000 set). However, to provide a more rigorous and unbiased assessment of generalization, we also report performance on the official, separate test set of ~2,200 images, resulting in a total of 13,750 images used in our study. This dual evaluation strategy ensures comparability with prior works while also demonstrating our model's robustness on a completely held-out test set. The following table juxtaposes our results against existing SOTA approaches, clearly indicating the evaluation set used for each model.

Table 4. Comparison of the proposed model with existing state-of-the-art models

Model	Recall	Precision	F1-score	mAP@0.5	Parameters (M)	Reported Evaluation
Our Proposed Model	**0.9887**	**0.9917**	**0.9902**	**0.9945**	**2.79**	**Validation Set (2200 out of 11,000 images)**
Our Proposed Model	**0.9016**	**0.9432**	**0.9219**	**0.9627**	**2.79**	**Test Set (2200 out of 13,750 images)**
YOLOv7-Tiny (Venkatesan et al., 2025)	0.9782	0.9904	0.9890	0.9870	6.03	Validation Set (2200 out of 11,000 images)
CoAtNet-0 (AlDahoul et al., 2023)	0.930	0.940	0.930	-	22.6	Validation Set (2200 out of 11,000 images)
RetinaNet + Attention (Pho et al., 2022)	-	-	-	0.820	-	Test Set
Multitask learning via pseudo-label generation and ensemble prediction (Anung et al., 2022)	-	-	0.988	0.956	-	Validation Set (1650 out of 11,000 images)

The results in Table 4 demonstrate that our Modified YOLOv11 Nano not only excels in efficiency but also achieves top-tier accuracy, establishing a new state-of-the-art among lightweight models. When evaluated on the common validation set, our model delivers an exceptional mAP@0.5 of 0.9945 and an F1-score of 0.9902, surpassing all compared models including the highly accurate YOLOv7-Tiny (mAP 0.9870) and the multitask ensemble (F1-score 0.988). Critically, this performance is achieved with only 2.79 million parameters, making our model significantly smaller than YOLOv7-Tiny (6.03M parameters) and dramatically more efficient than large architectures like CoAtNet-0 (22.6M parameters). This breakthrough demonstrates that

our strategic architectural enhancements successfully bridge the gap between high accuracy and minimal computational footprint.

5. Conclusion and Future Enhancement

In this study, we aimed to address the critical challenge of deploying accurate object detection models for parasitic egg identification in resource-constrained settings. We successfully developed and evaluated a modified YOLOv11 Nano architecture, specifically engineered to balance high performance with computational efficiency. By systematically replacing standard convolutional blocks with a novel EnhancedConv module incorporating GhostConvs, a dual SimAM-ECA attention cascade, and structured regularization. We created a model that is both lightweight and powerful.

Our experimental results conclusively demonstrate the superiority of the Modified YOLOv11 Nano. The proposed model achieved an F1-score of 0.9219 and a mAP@0.5 of 0.9627, representing a substantial improvement over the baseline YOLOv11 Nano. More importantly, it delivers performance comparable to the much larger YOLOv11 Small model while using only about 30 % of its parameters and 42 % of its computational cost (GFLOPs). These results demonstrate that targeted architectural enhancements can bridge the gap between accuracy and efficiency for real-time medical image analysis. The proposed model stands as a robust and practical framework for developing real-time automated diagnostic tools, making advanced medical image analysis more accessible for large-scale screening programs in the areas where it is most needed.

While our proposed model demonstrated significant promise, several avenues exist for future work to enhance the model's performance and robustness. The next step is to further reduce the number of parameters. One approach we would explore is to train a large model and using knowledge distillation train a smaller model. Although the model is already lightweight, its efficiency can be further improved using post-training optimization techniques such as quantization (e.g., converting weights to INT8) and pruning. These methods can reduce model size and accelerate inference with a negligible impact on accuracy. This model can be integrated into mobile diagnostic tools, enabling rapid and accurate screening in field clinics.

Acknowledgments
The authors would like to express their gratitude to the Department of Computer and Electronics Engineering and Research, Training and Consultancy Division at Kantipur Engineering College for their support and resources, including the provision of the RTX 5090 GPU which was essential for the computational training in this study. We also thank all faculty members and colleagues who contributed to making this research possible.

References

Ahmed M. (2023). Intestinal Parasitic Infections in 2023. *Gastroenterology research*, *16*(3), 127–140.

World Health Organization (WHO) , Soil transmitted helminth infection fact sheet (2023). Available at: https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections (Accessed February 6, 2023).

Mejia, R., Vicuna, Y., Broncano, N., Sandoval, C., Vaca, M., Chico, M., ... & Nutman, T. B. (2013). A novel, multi-parallel, real-time polymerase chain reaction approach for eight gastrointestinal parasites provides improved diagnostic capabilities to resource-limited at-risk populations. *The American journal of tropical medicine and hygiene*, *88*(6), 1041.

Venkatesan, K., Muthulakshmi, M., Prasanalakshmi, B., Karthickeien, E., Pabbisetty, H., & Bahiyah, R. S. (2025). Comparative analysis of resource-efficient YOLO models for rapid and accurate recognition of intestinal parasitic eggs in stool microscopy. *Intelligence-Based Medicine*, *11*, 100212.

Wang, Y., He, Z., Huang, S., & Du, H. (2022, October). A robust ensemble model for parasitic egg detection and classification. In *2022 IEEE International Conference on Image Processing (ICIP)* (pp. 4258-4262). IEEE.

AlDahoul, N., Karim, H. A., Kee, S. L., & Tan, M. J. T. (2022, October). Localization and classification of parasitic eggs in microscopic images using an efficientdet detector. In *2022 IEEE International Conference on Image Processing (ICIP)* (pp. 4253-4257). IEEE.

AlDahoul, N., Karim, H. A., 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*(1), 14475.

Suwannaphong, T., Chavana, S., Tongsom, S., Palasuwan, D., Chalidabhongse, T. H., & Anantrasirichai, N. (2023). Parasitic egg detection and classification in low-cost microscopic images using transfer learning. *SN Computer Science*, *5*(1), 82.

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*(1), 454.

Anantrasirichai, N., Chalidabhongse, T. H., Palasuwan, D., Naruenatthanaset, K., Kobchaisawat, T., Nunthanasup, N., ... & Achim, A. (2022, October). Icip 2022 challenge on parasitic egg detection and classification in microscopic images: Dataset, methods and results. In *2022 IEEE International Conference on Image Processing (ICIP)* (pp. 4306-4310). IEEE.

Bessa, M. L., Junior, G. B., & de Almeida, J. D. S. (2025). Neural Network Ensemble for Detecting Parasite Eggs in Microscopic Images. *Procedia Computer Science*, *256*, 739-746.

Yang, L., Zhang, R., Li, L. & Xie, X.. (2021). SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *Proceedings of the 38th International Conference on Machine Learning*, in *Proceedings of Machine Learning Research* 139:11863-11874.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 13–19, 2020, Seattle, WA) (pp. 11534–11542). IEEE.

Li, L., Wang, Z., & Zhang, T. (2023). Gbh-yolov5: Ghost convolution with bottleneckcsp and tiny target prediction head incorporating yolov5 for pv panel defect detection. *Electronics*, *12*(3), 561.

Jocher, G., Qiu, J., & Chaurasia, A. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

Pho, K., Lam, H., Le, T., Nguyen, H. T., & Yoshitaka, A. (2022, December). Attention-driven retinanet for parasitic egg detection. In *2022 IEEE International Symposium on Multimedia (ISM)* (pp. 265-272). IEEE.

Aung, Z. H., Srithaworn, K., & Achakulvisut, T. (2022, October). Multitask learning via pseudo-label generation and ensemble prediction for parasitic egg cell detection: IEEE ICIP Challenge 2022. In *2022 IEEE International Conference on Image Processing (ICIP)* (pp. 4273-4277). IEEE.