UAV-Based Mosquito Larval Detection with Minimal Training Data: From Controlled Environments to Turbid Field Conditions

Gaku Masuda *, Masaki Shuzo, Thomson Ngumbira, Dylo Foster Pemba, Hitoshi Kawada

Posted Date: 19 March 2026

doi: 10.20944/preprints202603.1543.v1

Keywords: mosquito larvae detection; larval source management; UAV imagery; deep learning; AutoML; minimal training data; object detection; malaria vector control

 doi:10.20944/preprints202603.1543.v1

Article

UAV-Based Mosquito Larval Detection with Minimal Training Data: From Controlled Environments to Turbid Field Conditions

Gaku Masuda [1,*], Masaki Shuzo [2], Thomson Ngumbira [3], Dylo Foster Pemba [4] and Hitoshi Kawada [5]

[1] School of Medicine, Tokyo Women's Medical University, Tokyo 162-8666, Japan
[2] Faculty of Information Science, Shonan Institute of Technology, Kanagawa 251-8511, Japan
[3] Biology Department, University of Malawi, Zomba, Malawi
[4] Vector and Disease Control Centre, Zomba, Malawi
[5] Institute of Tropical Medicine, Nagasaki University, Nagasaki 852-8523, Japan
* Correspondence: masuda.gaku@twmu.ac.jp

Highlights

What are the main findings?

- A functional mosquito larval detection model was constructed from only 8 UAV-captured training images using a free, no-code cloud AutoML platform, demonstrating that local vector control specialists can build detection models without programming expertise or large annotated datasets.
- The model generalised across imaging devices in controlled environments, but Precision collapsed in turbid field conditions, revealing that background diversity in training data—not model architecture or data volume—is the primary barrier to real-world deployment.

What are the implications of the main findings?

- Free cloud AutoML services lower the barrier for endemic-country practitioners to autonomously build and update larval detection models, making community-owned larval source management technologically feasible.
- The characterization of this minimum viable model provides a practical foundation for developing actionable guidelines on constructing minimum yet maximally efficient training datasets, enabling local specialists to achieve field-deployable performance with minimal annotated data.

Abstract

If water surfaces are left untreated, mosquito larvae can develop into adult vectors within a few weeks, posing a significant risk of infectious disease. Consequently, technology for weekly on-site detection of mosquito larvae is required to control such situations effectively. Using autonomous UAV-captured images based on water surface mapping, this study establishes a technological foundation to detect mosquito larvae using deep learning, targeting automated larval source management for vector control. To enable local entomologists to build and adapt models using free, no-code cloud services, determining the minimum viable training dataset is essential. Therefore, we built an object detection model using Google Cloud Platform's Vertex AI with an extremely limited annotated dataset comprising 10 images and 371 annotations in total, of which 8 images (300 annotations) were allocated for training, to evaluate feasibility under resource-constrained conditions. The model was tested on two external datasets: controlled indoor images (JIHS, Japan) and field photographs from natural breeding sites (Malawi). Results demonstrated that while the model achieved high Recall (75.0–100.0%) and Precision (56.3–57.1%) in controlled environments (F1 0.643–0.727), Precision dropped substantially to 9.1–13.3% in turbid field conditions due to false positives from environmental noise, such as floating particles. These findings indicate that dataset diversification incorporating diverse backgrounds is essential for real-world deployment.

 doi:10.20944/preprints202603.1543.v1

Keywords: mosquito larvae detection; larval source management; UAV imagery; deep learning; AutoML; minimal training data; object detection; malaria vector control

1. Introduction

Finding water surfaces such as puddles to treat with larvicide is a difficult task. Traditionally, Larval Source Management (LSM) has been considered logistically demanding and cost-prohibitive in widespread rural settings, often limited to areas where habitats are 'fixed, few, and findable' [1,2]. However, if these water surfaces are left untreated, mosquito larvae can develop into adult vectors within a few weeks, posing a significant risk of infectious disease transmission. Consequently, a technology capable of facilitating weekly, on-site detection of mosquito larvae is required to control such situations effectively.

While the identification of mosquito larval sources using surface mapping through Unmanned Aerial Vehicles (UAVs, drones) has garnered attention, most approaches remain indirect, relying on machine learning models for 'risk assessment' using landscape images that include vegetation information such as RGB and narrow-band multispectral NDVI. However, the correlation between vegetation indices and actual larval presence is limited, as breeding sites frequently occur in open, sparsely vegetated water bodies. Achieving large-scale automation therefore requires a two-stage approach: first, direct water surface identification using Normalized Difference Water Index (NDWI) analysis [4,5] to rapidly map water bodies across wide areas; and second, machine learning-based direct image detection of larvae within those identified surfaces. This combination enables an evidence-based, rapid response cycle for larval source management—a transition from indirect risk assessment to confirmed larval presence. In the context of Integrated Vector Management (IVM) [3], we have previously established the first stage of this pipeline, utilizing NDWI analysis to rapidly identify water surfaces scattered across a wide area (389 ha) and map them as georeferenced images (GeoTIFF) for navigation [6]. The present study addresses the second stage: prototyping the larval detection component.

In field settings within endemic countries, having experts label vast amounts of UAV aerial imagery is often unrealistic due to human and time costs, security and hygiene risks, and ethical considerations. Furthermore, from the perspective of sustainability and local ownership, the system must be adaptable by local human resources—specifically vector control specialists—who often rely on freely available cloud computing services with strict usage limits. Moreover, to enable low-cost local adaptation for detecting different vector species in other regions, it is essential that local stakeholders can construct models themselves. Consequently, obtaining insights into the minimum amount of data required is a critical prerequisite. Therefore, demonstrating the feasibility of establishing a functional image recognition model based on a "small dataset" is significant for ensuring the technology remains accessible within resource constraints.

Building on this foundation, the objective of this study is to prototype a deep learning AI detection model to automatically assess the presence and habitat density of mosquito larvae within breeding sources identified by NDWI. Specifically, envisioning rapid deployment in resource-constrained endemic areas, we verify the validity of building models using extremely limited training data and examine their characteristics across distinct geographical and environmental conditions (Japan and Malawi). The ultimate goal is to establish a technological foundation for detecting the location of mosquito larvae on water surfaces, enabling an integrated two-tier surveillance system in which periodic UAV mapping identifies candidate water bodies and community-operated smartphones confirm larval presence on-site. This aims to automate and streamline the source identification and infection control process [7–9].

The application of this technology is envisioned for precision larval source management using UAVs [10–12], with on-board edge processing enabling real-time assessment of habitat density during

 doi:10.20944/preprints202603.1543.v1

flight. The technical requirements for edge deployment, including model quantization and device selection, are discussed in Section 5.5.

To realize this vision in real-world scenarios, this study focuses on detecting larvae on turbid water surfaces, which are common in field settings [15–17]. Leveraging the characteristic behavior of mosquito larvae (particularly *Anopheles* species) floating horizontally on the water surface [18,19], we aim to construct a deep learning model suitable for computer vision (CV) detection, building on prior work in mosquito detection [20]. Consequently, at this stage, the technical objective prioritizes ensuring reproducibility in actual operation over absolute accuracy in precision.

2. Related Work

2.1. UAV-Based Larval Source Management and Habitat Mapping

UAVs have increasingly been integrated into malaria vector control workflows, with research spanning habitat mapping, larvicide application, and AI-assisted operational prioritization. Hardy et al. [11] demonstrated that low-cost consumer drones equipped with RGB cameras could reliably delineate potential *Anopheles* larval habitats in rural Tanzania, establishing a practical precedent for aerial reconnaissance in LSM programs. Stanton et al. [12] subsequently assessed the operational feasibility of UAV-based habitat identification in Malawi—the same field environment examined in the present study—concluding that drones could meaningfully augment ground-based surveys while reducing the labor burden on local vector control teams.

Beyond mapping, Mukabana et al. [13] demonstrated that commercial agricultural drones can be repurposed for aerial larvicide application in irrigated rice agro-ecosystems in Zanzibar. Their pilot trial showed that a single application of Aquatain Mosquito Formulation delivered by drone achieved larval density reductions exceeding 90%, with residual efficacy persisting for approximately five to six weeks—performance comparable to manual ground application while covering treatment areas at two to six times the speed. This finding substantiates the operational value of UAV platforms not merely for surveillance but for active larval source treatment in field settings.

More recently, Bokpin et al. [14] reported a field evaluation of drone-assisted LSM in Ghana's Eastern Region, in which aerial mapping was combined with an AI model that analyzed geospatial and morphological features of water bodies to prioritize high-risk larval habitats. The integrated system identified more than three times as many breeding sites as conventional manual scouting, reduced larvicide consumption by over 60%, and lowered worker requirements by approximately 50%, while maintaining malaria case trends comparable to control sub-districts. This study represents the most operationally complete example of drone–AI integration in LSM to date and directly motivates the detection component addressed in the present work.

Despite these advances, a critical gap remains: none of these systems incorporates real-time or near-real-time *in situ* detection of mosquito larvae on water surfaces. Habitat mapping with NDWI or visual inspection identifies candidate water bodies, and AI-assisted prioritization can rank these sites by risk, but confirmation of actual larval presence still requires ground-level dipper sampling by trained personnel. The present study addresses this gap by targeting the detection of larvae directly from images, positioning larval confirmation as an automated step within the broader UAV-based LSM pipeline.

2.2. Deep Learning for Mosquito Larvae Detection

Several deep learning-based approaches to mosquito larval detection have been reported in recent years. Feng et al. [21] systematically reviewed methods for small object detection, identifying low image resolution and complex backgrounds as the principal challenges—characteristics that directly apply to the detection of larvae on water surfaces, which constitutes a canonical small object detection problem. Sandhu et al. [22] applied deep learning-based semantic segmentation to detect and track *Aedes aegypti* larvae in turbid water tanks, demonstrating that high-precision detection is achievable even under controlled turbidity. However, their analysis was confined to indoor tank videos with

 doi:10.20944/preprints202603.1543.v1

precisely regulated imaging conditions, providing limited guidance for outdoor field deployment. Javed et al. [23] proposed LarvaeCountAI, a convolutional neural network tool for automatically counting *Culex annulirostris* larvae from standardized tray images, further confirming the feasibility of automated larval quantification but again within constrained laboratory conditions.

In contrast to these indoor-focused studies, the present work specifically targets larvae detected in outdoor puddles under uncontrolled environmental conditions, bridging the controlled-laboratory setting and the field deployments described in Section 2.1. The distinguishing characteristics of our approach—small training dataset, cross-device validation, and exposure to natural turbidity and background noise—reflect the practical constraints of resource-limited endemic settings that the preceding studies have not directly addressed.

2.3. Small Object Detection in UAV Imagery and Edge AI Deployment

The challenge of detecting small objects in drone-captured imagery has motivated substantial algorithmic development. Meng et al. [24] proposed SODCNN, a modified YOLOv7 architecture designed for dense small object detection in UAV imagery, incorporating early feature extraction, increased anchor box counts, an EIoU loss function, and an attention-based adaptive feature fusion module. On the VisDrone2019 benchmark, SODCNN achieved an mAP_{50} of 54.03%, surpassing contemporary YOLO-series and two-stage detectors. Their work illustrates that architectural adaptations specifically targeting small object detection in aerial images yield meaningful gains over general-purpose detectors—an insight that motivates the present study's interest in comparing Vertex AI AutoML with open-source YOLO-family models as a subsequent step [20].

Complementing detection research, Zhang [25] introduced the Simultaneous Learning Knowledge Distillation (SLKD) framework for compressing image restoration models intended for drone deployment. SLKD reduced model FLOPs and parameters by approximately 85% relative to teacher models while incurring only a marginal decrease in PSNR and SSIM, enabling real-time image restoration on NVIDIA Jetson edge platforms. While SLKD addresses image quality enhancement rather than larval detection per se, it underscores the feasibility and growing maturity of deploying computationally efficient deep learning pipelines on resource-constrained UAV hardware—a prerequisite for the edge processing vision outlined in the present study.

Taken together, the literature establishes that (1) UAVs can map, prioritize, and treat larval habitats at operational scale; (2) deep learning can detect mosquito larvae from images under controlled conditions; and (3) compact, edge-deployable models are technically achievable. The present study occupies the intersection of these three trajectories, asking whether a minimal-dataset detection model—built from only eight UAV-captured training images—can retain sufficient generalizability to serve as the detection layer in an integrated UAV-based LSM system.

3. Materials and Methods

3.1. Dataset and Annotation

We utilized Google Cloud Platform (GCP)'s Vertex AI to build an object detection model under two deliberate constraints designed to reflect the operational realities of vector control programs in endemic countries. First, we restricted ourselves to a no-code/low-code AutoML platform requiring no programming expertise, so that local public health practitioners—specifically vector control specialists without machine learning backgrounds—could replicate or adapt the workflow autonomously.[1] Second, we limited training to an extremely small dataset (8 images, 300 annotations) to assess the minimum viable data requirement for a functional detection model. Vertex AI's underlying transfer learning capabilities and automatic data augmentation strategies are well-suited to generalizing from such limited samples, making it an appropriate choice for this feasibility study. Vertex AI

[1] Vertex AI provides new users with a complimentary credit allowance (USD 300 at the time of writing; subject to change), after which standard pay-as-you-go rates apply. The training and inference costs incurred in this study fell within this initial allowance.

 doi:10.20944/preprints202603.1543.v1

AutoML Vision internally applies Neural Architecture Search (NAS) to select an optimized detection architecture from a search space that includes EfficientDet (with Weighted BiFPN) and SpineNet backbones, depending on the accuracy–latency trade-off. Training is initialized from weights pre-trained on large-scale datasets (ImageNet, OpenImages), with automatic data augmentation policies applied to maximize generalization from limited samples—making it particularly suited to the small dataset scenario investigated here.

Training Dataset (UAV-based): For model construction, we used images captured using an autonomous UAV (DJI Matrice 300 RTK equipped with a Zenmuse H20 camera) in artificially created outdoor environments in Japan. The camera was operated at a focal length of 120 mm (35 mm equivalent: 556 mm) with an additional digital zoom of 3.4–6.8×, achieving an effective focal length equivalent of approximately 1,890–3,780 mm, enabling ultra-telephoto close-up imaging of water surfaces from a flight altitude of 77 m above ground level (AGL). At the optical focal length of 120 mm and this altitude, the ground sampling distance (GSD) was approximately 4.0 mm/pixel, corresponding to an individual larval body occupying approximately 1–3 pixels in the image. The annotated dataset comprised 10 images and 371 annotations in total, of which 8 images (300 annotations) were allocated for training.

- Training: 8 images (300 annotations)
- Validation: 1 image (38 annotations)
- Test (Internal): 1 image (33 annotations)

Annotation Strategy and Quality Control: All bounding box annotations were implemented systematically based on the following rigorously defined rules:

- **Boundary Setting:** Bounding boxes were created by pushing to the exact contour (extreme) of the target object. This was done to avoid learning the background (such as containers) and based on the assumption that computational area expansion would be straightforward in post-processing.
- **Overlap Handling:** Even when two larvae overlapped, individual boxes were created for each with minimal area, intending accurate individual counting.
- **Inter-annotator Reliability:** All annotations were performed by a single expert entomologist with over 15 years of field experience in vector identification to ensure maximum consistency.

3.2. External Test Dataset (Smartphone-Based Cross-Device Validation)

To verify the generalizability of the trained model, we used two types of external test datasets independent from the training data. All test images were compressed from approximately 5 MB to around 1 MB prior to inference due to system upload constraints.

- **JIHS (Indoor Breeding Data):** Indoor smartphone photographs of mosquito larvae at the Japan Institute for Health Security. Uniform lighting conditions, minimal noise. This dataset enables evaluation of cross-device generalization—specifically, sensor transfer from UAV to smartphone—under optimal conditions.
- **Malawi (Field Data):** Smartphone photographs of mosquito larvae taken during field surveys in Malawi. Turbid water with diverse noise from natural environments.

3.3. Evaluation Metrics

For evaluating detection performance, we used Precision, Recall, and F1 Score. The ground truth total (GT) denotes the actual number of annotated larvae per image. A detection was considered a True Positive (TP) if the Intersection over Union (IoU) between the predicted bounding box and the ground truth annotation was 0.5 or higher (IoU $\geq$ 0.5). When ground truth annotations overlapped, individual bounding boxes were created for each larva; a predicted box was counted as a TP if it

 doi:10.20944/preprints202603.1543.v1

satisfied IoU ≥ 0.5 with any single ground truth box, and as a False Positive (FP) otherwise. Missed detections were recorded as False Negatives (FN). Precision, Recall, and F1 are defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{1}$$

4. Results

The quantitative performance of the model on the external test datasets is summarized in Table 1. Additionally, the Precision-Recall (PR) curve is presented in Figure 1.

Table 1. Performance Evaluation Results of Machine Learning Model for Mosquito Larvae Detection Using Vertex AI.

Dataset	GT	TP	FP	FN	Prec.	Recall	F1
JIHS 1	8	8	6	0	57.1%	100.0%	0.727
JIHS 2	24	18	14	6	56.3%	75.0%	0.643
Malawi 1	26	4	26	22	13.3%	15.4%	0.143
Malawi 2	3	3	30	0	9.1%	100.0%	0.167

GT = Ground Truth total (actual larval count); TP = True Positive (IoU ≥ 0.5); FP = False Positive; FN = False Negative; Prec. = TP / (TP + FP); Recall = TP / (TP + FN); F1 = 2 × Prec. × Recall / (Prec. + Recall).

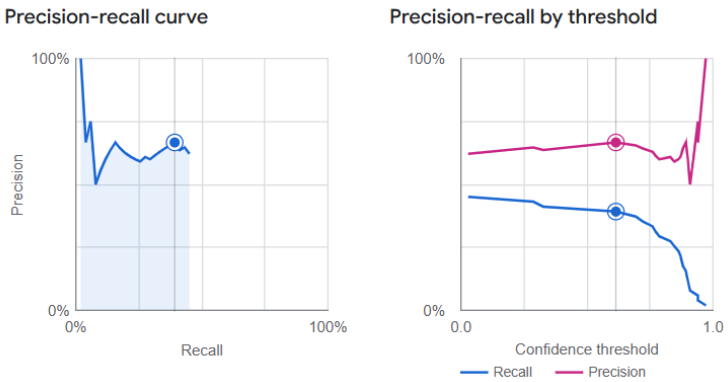

Figure 1. Precision-Recall (PR) Curve evaluating the detection performance. The curve illustrates the trade-off between precision and recall at varying confidence thresholds. A steep drop indicates limitations in model robustness against the validation dataset.

Visualizations of the detection results compared to ground truth annotations are presented in Figures 2–5.

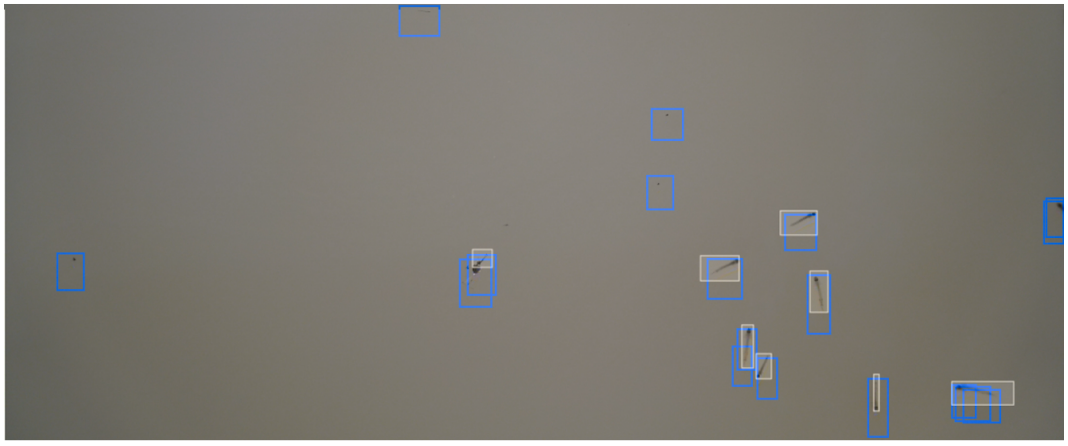

Figure 2. JIHS 1 dataset—Overlay of Ground Truth (base layer) vs. Prediction (semi-transparent overlay). The model demonstrates high sensitivity in controlled indoor environments with clear water and uniform lighting.

 doi:10.20944/preprints202603.1543.v1

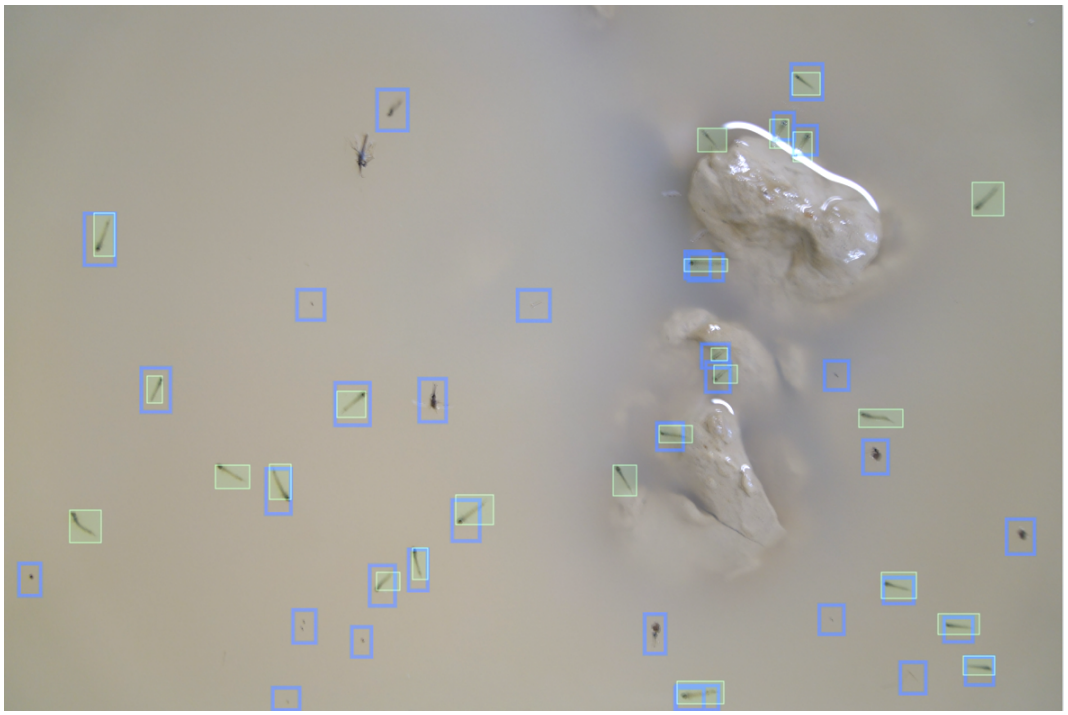

Figure 3. JIHS 2 dataset—Overlay of Ground Truth (base layer) vs. Prediction (semi-transparent overlay). The model maintains strong performance across different indoor breeding containers.

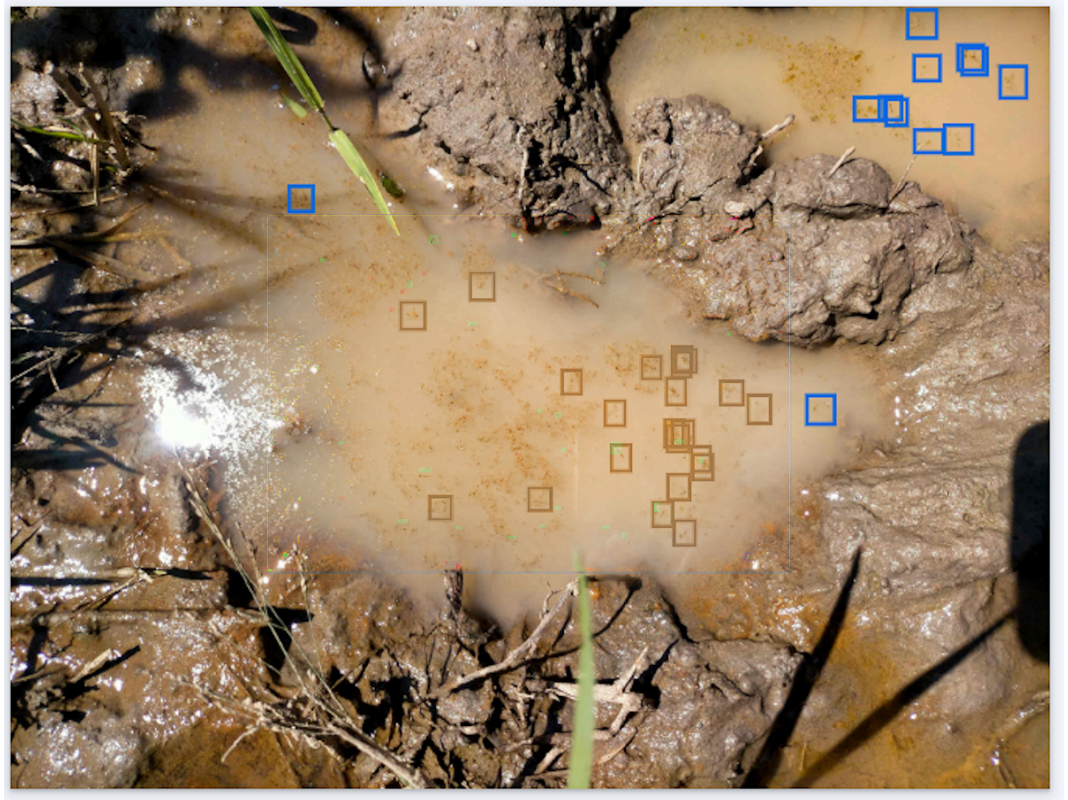

Figure 4. Malawi 1 field dataset—Prediction vs. Ground Truth. Green boxes indicate ground truth; brown/blue boxes indicate model predictions. Wide-range photography with numerous minute larvae in highly turbid water shows poor detection rates (Recall 15.4%, Precision 13.3%, F1 0.143). The large number of false positives on suspended particles and organic debris demonstrates the challenge of natural turbid field conditions.

 doi:10.20944/preprints202603.1543.v1

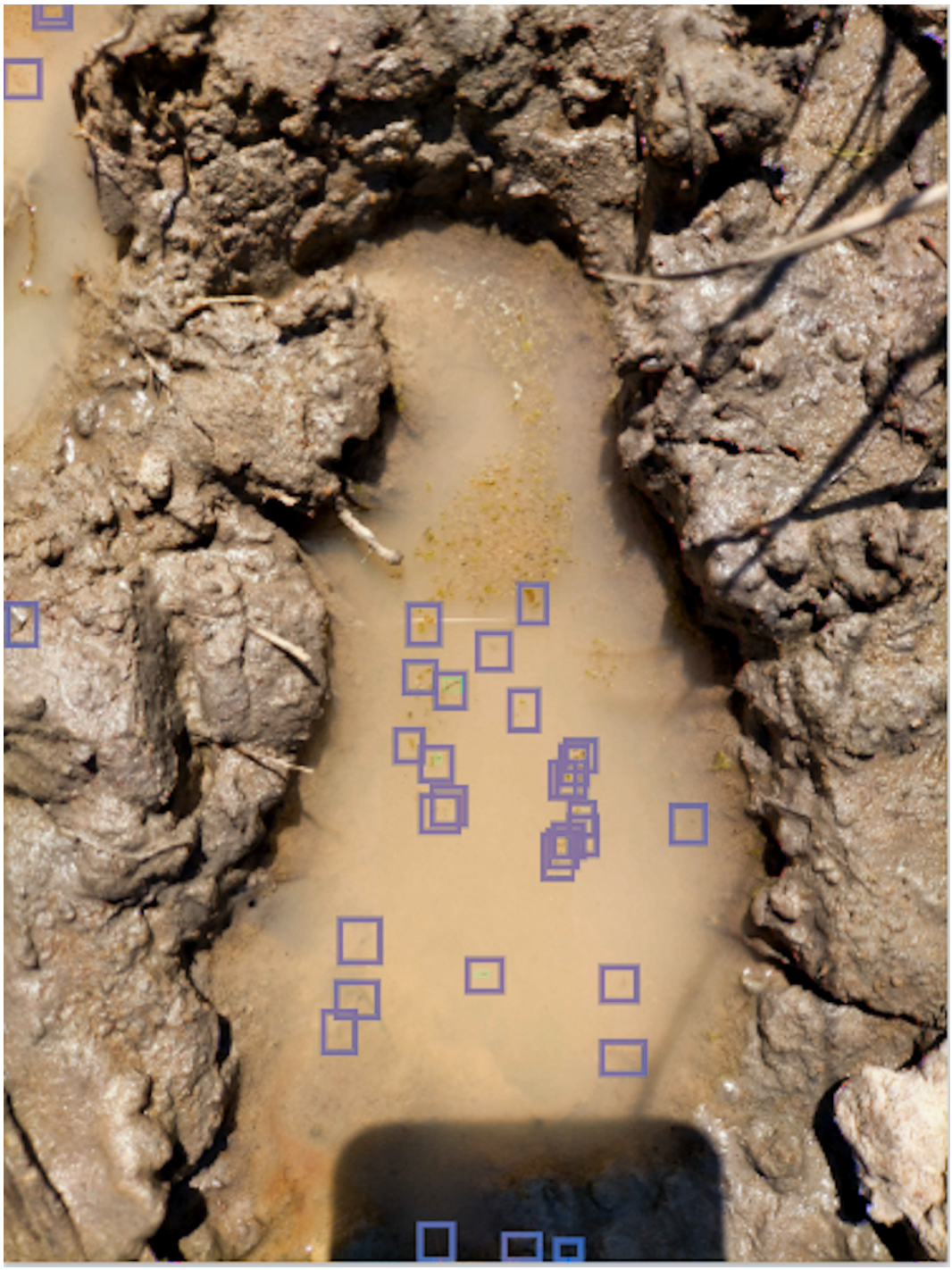

Figure 5. Malawi 2 field dataset—Prediction vs. Ground Truth. Green boxes indicate ground truth; purple/thick boxes indicate model predictions. Close-range photography of larger larvae achieves Recall of 100.0% but extremely low Precision (9.1%, F1 0.167) due to massive false positive generation from environmental noise in turbid water.

In the controlled JIHS datasets, the model demonstrated high sensitivity, achieving Recall of 100.0% and 75.0% with Precision of 57.1% and 56.3%, respectively, yielding F1 scores of 0.727 and 0.643. The strong performance in both JIHS images—captured with different containers and larval densities—suggests that the learned representations generalise across variation in indoor imaging conditions. In the unstructured Malawi field datasets, results diverged markedly between the two images. Malawi 1, comprising wide-range photographs of numerous minute larvae in highly turbid water, showed a Recall of only 15.4% and a Precision of 13.3% (F1 0.143); the absolute number of false positives equalled that of true positives (TP = FP = 4 and 26 respectively), indicating that the model

 doi:10.20944/preprints202603.1543.v1

could not discriminate larvae from suspended particles at this image scale. Malawi 2, photographed at closer range with larger apparent larvae, achieved a Recall of 100.0% but extremely low Precision (9.1%, F1 0.167), as 30 false positives were generated against only 3 true positives, demonstrating that environmental noise overwhelmed detection specificity even when larval size was adequate.

5. Discussion

5.1. Cross-Device Generalization

The results demonstrate that a model trained on an extremely small amount of homogeneous UAV aerial imagery (8 images, 300 annotations) shows remarkably high detection capability when applied to smartphone-captured photographs under controlled indoor conditions, achieving Recall of 75.0–100.0%, Precision of 56.3–57.1%, and F1 scores of 0.643–0.727. This cross-device performance provides evidence that the model has successfully learned universal morphological features characteristic of mosquito larvae—specifically, the distinctive elongated body shape, segmented structure, and characteristic floating posture—that transfer from UAV aerial imaging to ground-level smartphone photography.

This sensor-invariance finding has profound practical implications for developing scalable surveillance systems. The successful transfer from UAV aerial imaging (training domain) to ground-level smartphone photography (deployment domain) validates the two-tier surveillance architecture introduced in Section 1:

1. **Tier 1—Professional UAV Mapping:** Periodic large-area habitat mapping using NDWI-based water surface identification, generating prioritized target lists.
2. **Tier 2—Community Smartphone Verification:** Frequent on-site verification using smartphone applications powered by larvae detection AI, enabling immediate source elimination.

5.2. Environmental Dependency and Image Characteristics

The training images were captured at a GSD of approximately 4.0 mm/pixel (DJI Matrice 300 RTK with Zenmuse H20, 120 mm focal length with 3.4–6.8× digital zoom, 77 m AGL), meaning that individual larvae occupied only approximately 1–3 pixels. In contrast, the JIHS smartphone photographs were taken at close range, yielding larvae occupying 2 to 4 times more pixels than in the training images—a factor that facilitated feature extraction under controlled conditions.

For indoor data (JIHS), Recall showed high values of 75.0–100.0% with Precision of 56.3–57.1%. This is attributed to the similar characteristics of "homogeneous background and clear water" between training and test data, as well as larvae occupying 2 to 4 times more pixels than in the training images, facilitating feature extraction.

In contrast, for field data (Malawi), Precision dropped to 13.3% (Malawi 1) and 9.1% (Malawi 2). This indicates that the model, overfitted to the artificial environment, was mis-detecting diverse noise such as suspended micro-particles as larvae (False Positives).

The contrasting results between Malawi 1 and Malawi 2 highlight the critical role of apparent larval size in detection performance. In Malawi 1, wide-range photography meant that larvae occupied approximately one-sixth to one-half the pixel area of training images, reducing the distinctiveness of the morphological features the model relies upon for discrimination. In Malawi 2, close-range photography produced larvae of comparable or larger apparent size, enabling complete recall—but the absence of background-diverse training data meant that environmental noise was still classified as larvae at high rates. This suggests that an operational protocol of "capturing wide areas at high resolution and dividing them into partial images (tiles) without image quality degradation for inference" would be essential for stabilizing accuracy across varying field conditions.

A further methodological consideration concerns image compression applied prior to inference. All external test images were reduced from approximately 5 MB to approximately 1 MB due to system upload constraints (Section 3.2). This approximately fivefold reduction in file size entails a corresponding loss of fine spatial detail, which disproportionately affects the representation of small

 doi:10.20944/preprints202603.1543.v1

larvae whose distinguishing features—segmented body outline and surface posture—occupy only a few pixels in wide-area photographs. The compression artifact is therefore likely to have contributed to both the elevated false positive rate in turbid field conditions and the low detection rate for minute larvae in Malawi 1. Future evaluations should control for image resolution by standardizing capture distance and tile size prior to inference, rather than applying post-capture compression.

5.3. Impact of IoU Threshold Relaxation

To assess whether detection localization accuracy was a primary limiting factor, we conducted supplementary analysis using a relaxed IoU threshold criterion of 0.3 instead of the standard 0.5 threshold. The resulting optimized Precision-Recall curve is presented in Figure 6.

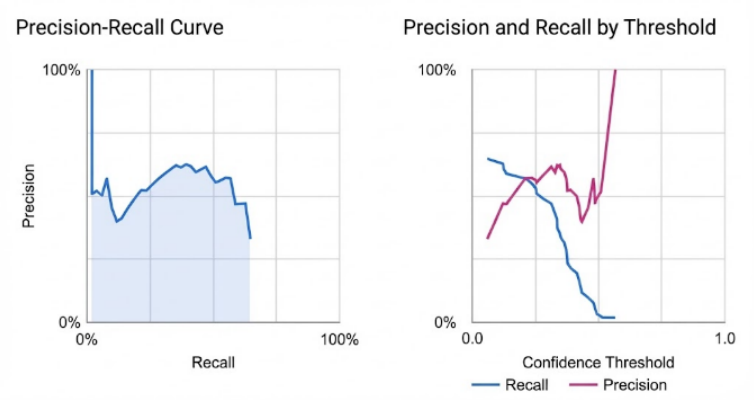

Figure 6. Optimized PR curve by IoU 0.3. The modest improvement indicates that localization precision is not the primary bottleneck; rather, the challenge lies in discriminating larvae from background noise in complex environments.

The relatively modest improvement in PR curve characteristics when lowering the IoU threshold from 0.5 to 0.3 indicates that bounding box localization accuracy is not the primary performance bottleneck. Instead, the fundamental limitation lies in the model's inability to reliably discriminate between actual larvae and visually similar background elements (high false positive rate) in complex natural environments.

5.4. Future Development Requirements

The objective of a highly robust model has not yet been achieved. The high number of false positives in field data undermines reliability in practical settings. It should be noted that this study focused on observing model behavior with limited training images; comparative analysis with standard object detection models such as YOLO is positioned as a subsequent step.

To address these challenges, the following steps are necessary:

- **Dataset Expansion:** Gradually increase training data, particularly including "turbid water" and "diverse backgrounds."
- **Fixed Independent Test Data:** Establish a fixed test set with uniform developmental stages for continuous verification.
- **Model Comparison:** Compare Vertex AI with open-source models such as YOLO [20,26] to clarify performance differences.

Framing the present model as a *minimum viable model* (MVM)—the simplest functional system that reveals real-world constraints—is instructive for program planning. The high Recall achieved in controlled conditions confirms that the detection capability exists within the learned representations; the variable Precision in field conditions (9.1–13.3%) identifies the specific gap (background diversity) that must be addressed before operational deployment. This characterization of MVM behavior provides local entomologists with a concrete, actionable roadmap: the bottleneck is not the platform,

 doi:10.20944/preprints202603.1543.v1

the annotation volume, or the model architecture, but the compositional diversity of the training backgrounds.

These findings suggest that practical guidelines for constructing minimum yet maximally efficient training datasets would be of substantial value to endemic-country practitioners. Such guidelines should specify the minimum number of images required, the types of backgrounds that must be represented (clear water, turbid water, diverse substrates), and the annotation density needed for robust field performance. Developing and validating such guidelines across different vector species and field environments is a priority for future work.

5.5. Prospects for Edge AI Implementation

For edge processing on UAVs and smartphones, model quantization becomes essential [27,28] due to computational resource constraints. By mounting a Neural Processing Unit (NPU) on the UAV and using a quantized (integer) model to process ultra-telephoto water surface images, real-time assessment of habitat density during flight becomes feasible.

This study underscores the importance of consistency in detectability over mere accuracy. Even with a lower detection rate, a constant bias allows for the estimation of habitat density through statistical correction. Future research should focus on identifying the optimal model size and computational load, ensuring a balance between latency and accuracy that aligns with the processing constraints of NPU-equipped edge devices (such as Coral TPU).

6. Conclusions

This study verified whether mosquito larvae on water surfaces could be detected using a deep learning model constructed from limited training images (8 UAV-captured images out of a total annotated dataset of 10 images), evaluating the feasibility of automating drone-based larval source management. Applying a model built from UAV imagery to indoor breeding environments (Japan) and turbid field environments (Malawi) revealed that while specific detection performance was achieved in indoor settings (Recall 75.0–100.0%, Precision 56.3–57.1%, F1 0.643–0.727), Precision dropped to 9.1–13.3% in turbid field conditions due to increased false positives. This clarified that while models built from small datasets show utility in homogeneous environments, dataset expansion reflecting background diversity is indispensable for real-world application.

Consequently, this study demonstrated both the feasibility of larval detection under small data constraints and the specific challenges that must be resolved for field implementation. Characterizing this *minimum viable model* is itself a meaningful contribution: it establishes that local vector control specialists can autonomously construct and update larval detection models using freely available, no-code cloud platforms with minimal annotated data—an essential prerequisite for sustainable, locally-owned larval source management in resource-constrained endemic settings. The evidence-based, rapid response cycle envisioned here—UAV water surface mapping via NDWI, followed by on-board or smartphone-based larval confirmation via the detection model, followed by targeted larvicide application—represents a technically grounded roadmap for fully automated larval source management. Furthermore, the minimum viable model characterized here provides a practical foundation for developing actionable guidelines on constructing minimum yet maximally efficient training datasets— specifying required image counts, background diversity, and annotation density—enabling local specialists to achieve field-deployable performance with minimal resources and without specialist machine learning expertise.

Author Contributions: Conceptualization, G.M. and H.K.; methodology, G.M. and M.S.; software, M.S.; validation, G.M., T.N., and D.F.P.; formal analysis, G.M. and M.S.; investigation, G.M., T.N., and D.F.P.; resources, H.K.; data curation, G.M. and T.N.; writing—original draft preparation, G.M.; writing—review and editing, G.M., M.S., and H.K.; visualization, G.M. and M.S.; supervision, H.K.; project administration, G.M. and H.K.; funding acquisition, H.K. All authors have read and agreed to the published version of the manuscript.

 doi:10.20944/preprints202603.1543.v1

Funding: This study was supported by SATREPS (Science and Technology Research Partnership for Sustainable Development) [Grant Number JP24jm0110032], a collaborative initiative of the Japan Science and Technology Agency (JST) and Japan International Cooperation Agency (JICA), under the research project titled "Sustainable Control of Zoonotic Malaria through an Integrated Approach."

Institutional Review Board Statement: The study was conducted in accordance with institutional and national research committee guidelines.

Informed Consent Statement: Not applicable for studies not involving direct human subjects.

Data Availability Statement: The training dataset and code developed for this study are available from the corresponding author upon reasonable request. Field data from Malawi are subject to ethical restrictions and collaborative agreements.

Acknowledgments: The authors would like to thank Dr. Yukiko Higa of the Japan Institute for Health Security (JIHS), Tokyo, Japan, for providing access to the insectarium to photograph live mosquito larvae. Malawi field images were obtained through joint research by Nagasaki University and Sumitomo Chemical Co. Ltd. We also acknowledge computational resources provided by Google Cloud Platform.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AGL	Above Ground Level
AI	Artificial Intelligence
BTi	*Bacillus thuringiensis israelensis*
GCP	Google Cloud Platform
GSD	Ground Sampling Distance
IoU	Intersection over Union
IVM	Integrated Vector Management
JIHS	Japan Institute for Health Security
LSM	Larval Source Management
MVM	Minimum Viable Model
NDWI	Normalized Difference Water Index
NPU	Neural Processing Unit
UAV	Unmanned Aerial Vehicle

References

1. Fillinger, U.; Lindsay, S.W. Larval source management for malaria control in Africa: myths and reality. *Malar. J.* **2011**, *10*, 353.
2. Wilson, A.L.; Courtenay, O.; Kelly-Hope, L.A.; Scott, T.W.; Takken, W.; Torr, S.J.; Lindsay, S.W. The importance of vector control for the control and elimination of vector-borne diseases. *PLoS Negl. Trop. Dis.* **2020**, *14*, e0007831.
3. World Health Organization. *Handbook for Integrated Vector Management*; WHO: Geneva, Switzerland, 2012.
4. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432.
5. McFeeters, S.K. Using the Normalized Difference Water Index (NDWI) within a Geographic Information System to Detect Swimming Pools for Mosquito Abatement: A Practical Approach. *Remote Sens.* **2013**, *5*, 3544–3561.
6. Masuda, G. Making maps of potential mosquito breeding sites by analyzing UAV aerial multispectral images. *J. Soc. Photogr. Imaging Jpn.* **2023**, *86*, 1–6. (In Japanese)
7. Alfred, R.; Obit, J.H. A review on spatial technologies for enhancing malaria control: Concepts, tools, and challenges. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2021**, *11*, 334–341.

 doi:10.20944/preprints202603.1543.v1

8. World Health Organization. *Larval Source Management: A Supplementary Measure for Malaria Vector Control*; WHO Press: Geneva, Switzerland, 2013.
9. Minakawa, N.; Seda, P.; Yan, G. Influence of host and larval habitat distribution on the abundance of African malaria vectors in western Kenya. *Am. J. Trop. Med. Hyg.* **2002**, *67*, 32–38.
10. Mechan, F.; Bartonicek, Z.; Malone, D.; Lees, R.S. Unmanned aerial vehicles for surveillance and control of vectors of malaria and other vector-borne diseases. *Malar. J.* **2022**, *21*, 15.
11. Hardy, A.; Makame, M.; Cross, D.; Majambere, S.; Msellem, M. Using low-cost UAVs to map malaria vector habitats. *Parasit. Vectors* **2017**, *10*, 29.
12. Stanton, M.C.; Kalonde, P.; Zembere, K.; Spaans, R.H.; Jones, C.M. The application of drones for mosquito larval habitat identification in rural environments: A practical approach for malaria control? *Malar. J.* **2021**, *20*, 244.
13. Mukabana, W.R.; Welter, G.; Ohr, P.; Tingitana, L.; Makame, M.H.; Ali, A.S.; Knols, B.G.J. Drones for area-wide larval source management of malaria mosquitoes. *Drones* **2022**, *6*, 180.
14. Bokpin, G.A.; Adzei, F.A.; Dadzie, S.; Umeda, M.; Kim, J. Field evaluation of drone and AI assisted larval source management in Ghana. *PLoS ONE* **2026**, *21*, e0340690.
15. Minakawa, N.; Mutero, C.M.; Githure, J.I.; Beier, J.C.; Yan, G. Spatial distribution and habitat characterization of anopheline mosquito larvae in Western Kenya. *Am. J. Trop. Med. Hyg.* **1999**, *61*, 1010–1016.
16. Paaijmans, K.P.; Wandago, M.O.; Githeko, A.K.; Takken, W. Unexpected high losses of *Anopheles gambiae* larvae due to rainfall. *PLoS ONE* **2007**, *2*, e1146.
17. Minakawa, N.; Sonye, G.; Mogi, M.; Yan, G. Habitat characteristics of *Anopheles gambiae* s.s. larvae in a Kenyan highland. *Med. Vet. Entomol.* **2004**, *18*, 301–305.
18. Clements, A.N. *The Biology of Mosquitoes, Volume 2: Sensory Reception and Behaviour*; CABI Publishing: Wallingford, UK, 1999.
19. Gimnig, J.E.; Ombok, M.; Otieno, S.; Kaufman, M.G.; Vulule, J.M.; Walker, E.D. Density-dependent development of *Anopheles gambiae* (Diptera: Culicidae) larvae in artificial habitats. *J. Med. Entomol.* **2002**, *39*, 162–172.
20. Kumar, N.; Nagarathna; Flammini, F. YOLO-based light-weight deep learning models for insect detection system with field adaption. *Agriculture* **2023**, *13*, 741.
21. Feng, Q.; Xu, X.; Wang, Z. Deep learning-based small object detection: A survey. *Math. Biosci. Eng.* **2023**, *20*, 6551–6590.
22. Sandhu, M.A.; Amin, A.; Tariq, S.; Mehmood, S. *Aedes aegypti* larvae detection and tracking in turbid videos by semantic segmentation using deep learning. *J. Intell. Fuzzy Syst.* **2024**, *46*, 2009–2021.
23. Javed, N.; López-Denman, A.J.; Paradkar, P.N.; Bhatti, A. LarvaeCountAI: A robust convolutional neural network-based tool for accurately counting the larvae of *Culex annulirostris* mosquitoes. *Acta Trop.* **2024**, *260*, 107468.
24. Meng, L.; Zhou, L.; Liu, Y. SODCNN: A convolutional neural network model for small object detection in drone-captured images. *Drones* **2023**, *7*, 615.
25. Zhang, Y. Simultaneous learning knowledge distillation for image restoration: Efficient model compression for drones. *Drones* **2025**, *9*, 209.
26. Jocher, G.; Stoken, A.; Borovec, J.; et al. YOLOv5 by Ultralytics, 2020. Available online: https://github.com/ultralytics/yolov5 (accessed on 15 January 2026).
27. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
28. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.