

Article

An Automated AI-Based Vision Inspection System for Bee Mite and Deformed Bee Detection Using YOLO Models

Jeong-Yong Shin ¹, Hong-Gu Lee ¹, Su-bae Kim ² and Changyeun Mo ^{1,3,4,*}

¹ Program in Smart Agriculture, Department of Interdisciplinary, Kangwon National University, Chuncheon 24341, Republic of Korea; kvhffh@kangwon.ac.kr (J.-Y.S.); hgl@kangwon.ac.kr (H.-G.L.)

² Department of Agricultural Biology, National Institute of Agricultural Sciences, Wanju 55365, Republic of Korea; subaekim@korea.kr

³ Department of Biosystems Engineering, Kangwon National University, Chuncheon 24341, Republic of Korea

⁴ Terramolab Ltd., Chuncheon 24341, Republic of Korea

* Correspondence: cymoh100@kangwon.ac.kr; Tel.: +82-33-250-6494; Fax: +82-33-255-6406

Abstract

Varroa destructor (Bee mite) and Deformed Wing Virus are primary causes of honeybee colony collapse. This study developed an automated AI-based vision inspection system for detecting bee mites and deformed bees using the YOLO algorithm. The system integrates an RGB camera, a beecomb rotation motor, and an image transmission module to enable automated dual-sided image acquisition of the beecomb. The image characteristics of normal bees, bee mites, and deformed bees were analyzed, and YOLO-based object detection models were developed to classify them. Six YOLO models—based on YOLOv8 and YOLOv11 architectures across three model sizes (nano, small, and large)—were evaluated on 405 test images (6441 objects). The proposed system reduced the inspection time from 240 s required for manual method to 20 s per beecomb, achieving 12-fold efficiency improvement. Comparative analysis showed model-task specialization: YOLOv8l excelled in detecting small bee mites (F1: 92.5%, mAP[0.5]: 92.1%), while YOLOv11s achieved the highest performance for morphologically diverse deformed bees (F1: 95.1%). Error analysis indicated that detection performance was influenced by morphological characteristics. Deformed bee detection errors correlated with overlap in wing-to-body ratio: DB Type II exhibited 18.6% miss rate, while DB Type III achieved perfect detection. In bee mite detection, a sensitivity–specificity trade-off was observed: YOLOv11l had the lowest false negatives (2.5%) but highest false positives, while YOLOv8l demonstrated superior discrimination. These results demonstrate the practical potential of the proposed system for field deployment in apiaries, supporting early pest diagnosis and improved colony health management. The model-task specialization framework provides guidance for architecture selection based on object characteristics. Future work will focus on multi-location validation and real-time monitoring integration.

Academic Editor: Maciej Zaborowicz

Received: 25 February 2026

Revised: 19 March 2026

Accepted: 8 April 2026

Published: 10 April 2026

Keywords: smart beekeeping; bee mite detection; wing deformity monitoring; YOLO deep learning model; edge AI; automated beecomb inspection

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Honeybees (*Apis mellifera*, Bees) are economically and ecologically vital pollinators [1], which produce beekeeping products, such as honey, propolis, pollen, and bee venom.

In 2022, Colony Collapse Disorder (CCD) decimated bee populations across South Korea. In CCD, the bee population is rapidly reduced and the colony collapses, and it is characterized by the inability of bees to return to the hive [2–5]. Several factors contribute to CCD, including climate abnormalities caused by global warming, improper managements, insufficient nectar sources, and infestations by pests such as *Varroa destructor* (bee mite) [6,7].

Bee mites parasitize both adult bees and larvae by consuming their fat reserves, resulting in reduced body weight and lifespan [8]. Additionally, beehives are weakened, and in severe cases, can lead to complete hive collapse [9]. Bee mites are vectors that transmit viruses and pathogens to bees [10]. Deformed wing virus (DWV), which is transmitted by bee mites [11,12] is related to CCD. Infected worker bees exhibit impaired development and reduced flight ability [13,14]. Therefore, effectively controlling bee mites is crucial for the beekeeping industry.

Traditional inspection methods remain labor-intensive and highly dependent on the beekeeper's expertise, making large-scale monitoring impractical [15]. Approximately 120 beehives are managed by the average beekeeper in Korea [16]. The direct inspection of all beehives is hindered by the aging of beekeepers because 52.6% of the agricultural population in Korea is over 65 years old [17]. Furthermore, bee mites are reddish brown in color, making them difficult to distinguish from bees, and extremely small (1.6 mm (width) × 1.1 mm (length)) [7,18,19]. Detecting, by eye, deformed-wing virus–infected bees (deformed bees) or bee mites in a beehive of roughly 2000 bees within one minute is extremely challenging, and the accuracy of such diagnosis hinges on the beekeeper's level of expertise [20]. Therefore, a technology that can rapidly and objectively identify bee mites and deformed bees is required [21].

Recent studies have explored image-based computer vision techniques for beekeeping application. Various systems, including real-time bee monitoring using cameras [22], video monitoring units (VMU) with reflective surfaces [23], and Image Acquisition Systems (IAS) [24], have been developed. However, these methods suffer from limitations such as restricted monitoring of external bees, duplicate detections, manual intervention requirements, and instability in image acquisition.

Mrozek et al. [22] developed a real-time bee monitoring system using external cameras. While this system provided continuous monitoring of bee activity outside the hive, it was unable to inspect internal hive conditions. Similarly, Bjerger et al. [23] introduced the VMU, which used acrylic windows and reflective surfaces to capture bee abdomen images as they passed through a designated passage. Although this system allowed non-invasive monitoring, it was limited to external bees entering and exiting the hive, leading to potential duplicate detections and incomplete colony assessments.

Lee et al. [24] developed the IAS with a fixed RGB camera to capture images of the beehive for directly detecting mite. Unlike the VMU, this system enabled direct inspection of bee mites on the beehive. However, manual intervention was required to rotate the beehive after capturing one side, and motion-induced instability affected image quality. Additionally, Kongsilp et al. [25] designed an observation hive using an RGB-based image acquisition system to analyze bee behavior patterns. This system incorporated a shading door for non-contact measurements but was limited by its inability to inspect the back of the hive, restricting its applicability for comprehensive colony health monitoring.

With the advancements in artificial intelligence (AI)-based object detection, deep learning models have been increasingly applied in agricultural and beekeeping research [24,26]. Unlike traditional visual inspection, AI-based object detection improves accuracy and objectivity. However, a significant challenge in training these models is the inherent class imbalance in datasets, where pests like bee mites are far less frequent than bees. Recent research has demonstrated that employing strategies such as data augmentation and

stratified sampling can effectively mitigate this issue, thereby enhancing the detection performance of YOLO models for bee mites [27]. Beekeeping environments require multi-object and small-object detection due to the presence of various elements on the beecomb. In this study, You Only Look Once -version 8 (YOLOv8) and -version 11 (YOLOv11), the state-of-the-art object detection algorithm known for its superior performance in detecting small and overlapping objects, were utilized [28,29].

This study aims to develop and validate an automated AI-based vision inspection system for detecting honeybee pests and diseases by YOLO-based object detection models. The primary contribution of this work is the development of an integrated edge-based inspection system that combines specialized dual-sided image acquisition hardware with a tiling-based inference pipeline, enabling high-resolution beecomb analysis on resource-constrained edge devices while preserving small-object details such as bee mites.

The image characteristics of normal bee, bee mites, and deformed bees were analyzed to construct a structured detection dataset. Furthermore, six YOLO models—based on YOLOv8 and YOLOv11 architectures across three model sizes (nano, small, and large)—were systematically evaluated to identify architectures suitable for practical apiary deployment and to investigate the relationship between morphological characteristics and detection errors in small-object recognition.

2. Materials and Methods

2.1. Design of an Automated Beecomb Inspection System

An automated AI-based vision inspection system was developed to identify normal bees, bee mites, and deformed bees, as illustrated in the system architecture and hardware configuration diagrams (Figure 1). The proposed system consists of three main components: (1) a beecomb rotation unit, (2) an image acquisition unit, and (3) a control unit with hierarchical architecture.

2.1.1. Beecomb Rotation Unit

The beecomb rotation unit was designed to stabilize the beecomb and enable precise rotation during inspection. A beecomb supporter frame (508 mm × 290 mm, width × height) held the beecomb securely (Figure 1b). Automated 180° rotation was achieved by integrating a stepping motor (NK245E, Motor Bank, Seoul, Korea) with the following specifications: 24 V rated voltage, 54 Nm torque, 0.04 gear ratio, and 1.8° step angle. The stepping motor was operated with a microstepping setting of 32, providing a rotational resolution of 160,000 pulses per revolution (Pulse/rev) at the output shaft and positional accuracy of 0.00225° per pulse. Motor control was managed through a motor driver (MSD-224, Motor Bank, Seoul, Republic of Korea), and a photosensor (PM-L24, Panasonic, Osaka, Japan) was installed to establish the initial reference position. When a reference protrusion on the stepping motor was detected by the photosensor, a signal was transmitted to an Arduino microcontroller, which then determined the initial position of the beecomb rotation unit based on this feedback.

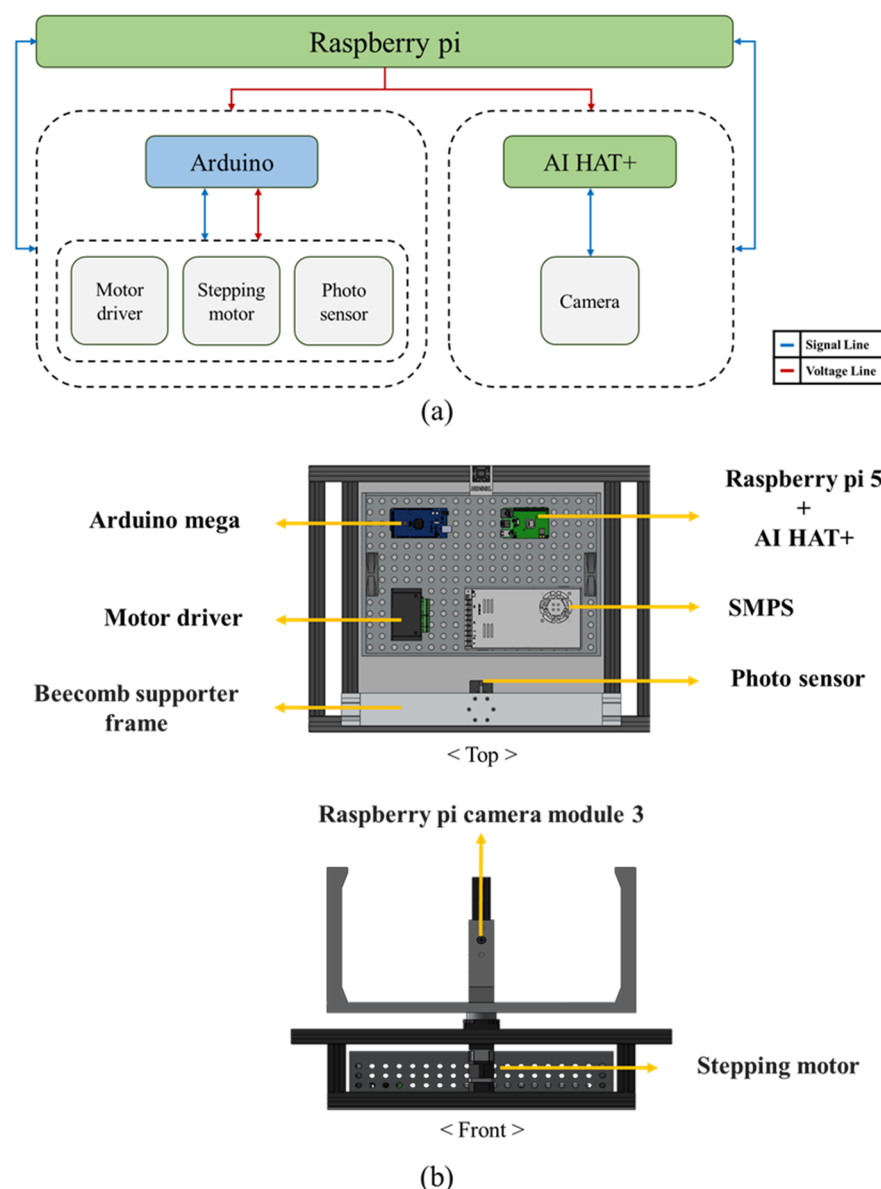


Figure 1. (a) Signal flow diagram of the automated AI-based vision inspection system for dual-side beecomb; (b) Top and front views illustrating main structural components.

2.1.2. Image Acquisition Unit

High-resolution image acquisition was achieved using a Raspberry Pi Camera Module 3 (Raspberry Pi Foundation, Cambridge, UK) with the following specifications: 4608×2592 pixel resolution, $1.4 \mu\text{m} \times 1.4 \mu\text{m}$ pixel size, 66° horizontal field of view (FOV), and 41° vertical FOV. These specifications enabled clear capture of small objects such as bee mites.

2.1.3. Control Unit with Hierarchical Architecture

A hierarchical control structure was adopted to manage the system's complexity and enhance overall stability and real-time responsiveness (Figure 1a). The upper-level control layer was managed by a Raspberry Pi 5 (16 GB RAM, Raspberry Pi Foundation, Cambridge, UK), which performed overall computational processing and operation scenario management. To accelerate high-load AI model inference, an AI HAT+ (13 TOPS, Raspberry Pi Foundation, UK) neural processing unit (NPU) was connected directly to the Raspberry Pi 5 via a Flexible Printed Circuit (FPC) cable on the PCIe port. The integration

of the AI HAT+ enabled the execution of complex YOLO models directly on the edge device. By offloading the computational burden of high-resolution image inference from the CPU to the NPU, the system achieved the real-time responsiveness required for field operations, processing dual-sided images of the beehive without the need for high-performance external servers or cloud connectivity. The Raspberry Pi 5 CPU, freed from inference tasks, concentrated on system control, sensor data aggregation, and communication tasks, which improved overall system efficiency and responsiveness.

For real-time hardware control requiring precise timing, an Arduino Mega (Arduino LLC, Monza, Italy) served as the lower-level controller, receiving high-level commands from the Raspberry Pi 5 via serial communication. The Arduino Mega independently executed low-level tasks such as stepping motor control and photosensor signal detection. A regulated power supply unit (SMPS, KO ELECTRONIC, Seoul, Korea) provided stable voltage and current to all system components according to their rated specifications. For field deployment and operational mobility, the entire system was powered by a portable battery pack (power station, 324,000 mAh capacity) (DLP8092C, PHILIPS, Amsterdam, The Netherlands), enabling autonomous operation without dependence on external AC power sources and facilitating flexible deployment across apiary locations.

2.2. Optical System Design and Motor Positioning Accuracy

2.2.1. Optimal Imaging Distance Calculation

To capture the entire beehive in a single image, the optimal distance between the beehive and camera was determined using the camera's field of view and target dimensions (Figure 2). The minimum required distance was calculated as:

$$l = a \times (2 \times \tan(\theta \times 2^{-1}))^{-1} \quad (1)$$

where l is the minimum distance between beehive and camera, a is the camera's horizontal measuring range (set to 508 mm, matching the beehive supporter frame width), and θ is the camera's horizontal angle of view (66°).

The vertical measuring range was determined as:

$$v = 2 \times l \times \tan(\theta' \times 2^{-1}) \quad (2)$$

where v is the vertical measuring range and θ' is the vertical angle of view (41°). The calculated vertical measuring range was verified to encompass the entire beehive-supporter frame height (Figure 2b,c).

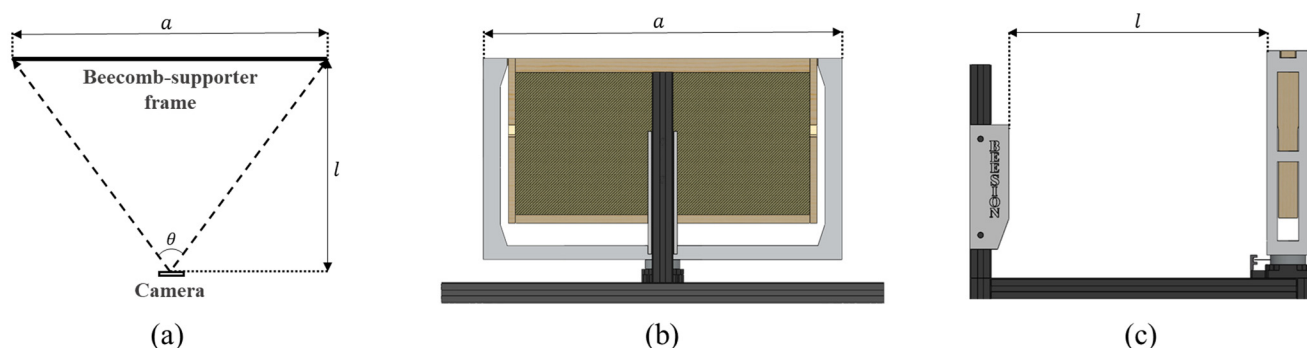


Figure 2. Optical design and geometric configuration for the automated beehive imaging system: (a) Geometric principles for calculating the imaging distance based on the camera's field of view (FOV); (b) Frontal view indicating the physical dimensions of the beehive supporter frame; (c) Side view illustrating the minimum distance between the beehive and the camera.

To validate the suitability of deep learning-based object detection at this fixed distance, the pixel dimensions of bee mites in this study were compared with previously reported values using pixel-scaling Equations (3)–(5) [23].

$$BMP_1 = BMP_2 \times BP_1 \times (BP_2)^{-1} \quad (3)$$

$$BMP_1 = BMP_{x1} \times BMP_{y1} \quad (4)$$

$$BMP_{x1} : BMP_{y1} = BMP_{x2} : BMP_{y2} \quad (5)$$

where BMP_1 is pixel values of bee mites measured in previous study*; BMP_2 is pixel values of bee mites measured by the system developed in this study; BP_1 is pixel values of bees measured in previous study*; BP_2 is pixel values of bees measured by the system developed in this study; BMP_{x1} is horizontal pixel values of bee mites measured in previous study*; BMP_{y1} is vertical pixel values of bee mites measured in previous study*; BMP_{x2} is horizontal pixel values of bee mites measured by the system developed in this study; BMP_{y2} is vertical pixel values of bee mites measured by the system developed in this study* [23].

2.2.2. Automated Beecomb Rotation Mechanism

To address the limitations of manual rotation, an automated beecomb rotation mechanism was designed to maintain the beecomb in a vertical orientation and rotate it 180° around its vertical axis (Figure 3). The stepping motor provided high-precision rotation, with positional repeatability not exceeding 15 arcseconds (0.004166°), backlash less than 1 arcminute (0.0167°), and output table parallelism and concentricity controlled within 0.01 mm.

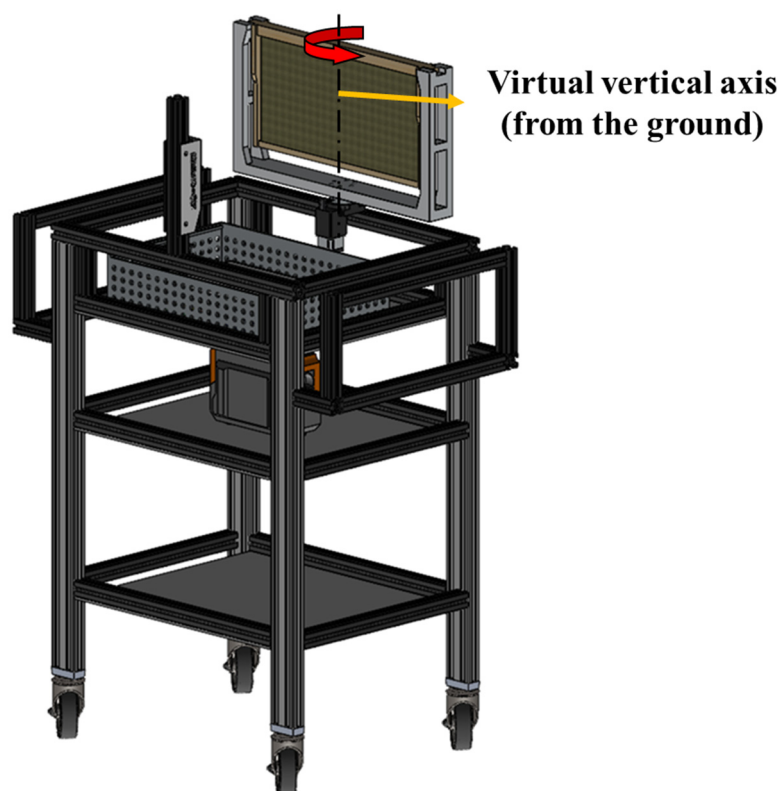


Figure 3. Rotation of the beecomb. Position of the stepping motor and the conceptual vertical axis used for rotation in the proposed system.

Motor control was implemented using a trapezoidal acceleration–deceleration profile. For an 80,000 pulse 180° rotation command, the acceleration phase (first 2000 pulses) used a 0.15 ms step interval delay, the cruise phase (middle 76,000 pulses) used a 0.015 ms interval, and the deceleration phase (final 2000 pulses) returned to 0.150 ms. This control strategy minimized mechanical vibrations and ensured smooth rotation throughout the operational cycle.

Absolute position accuracy was ensured through a photosensor-based reference point correction system installed on the rotation table. After each 360° rotation, the photosensor re-detected the reference marker to remove cumulative errors from mechanical backlash or slip and to maintain high positional repeatability.

2.2.3. Image Acquisition Workflow

The overall image acquisition workflow for dual-sided beecomb inspection is presented in a flowchart (Figure 4). Upon signal activation, the image counter was initialized and the beecomb supporter frame moved to the initial reference position. After the operator placed a beecomb on the beecomb supporter frame, the system acquired the i -th image of the front side, rotated the beecomb 180°, and then acquired the $(i + 1)$ -th image of the back side. The image index was then increased by 2 and the system returned to standby mode until the next acquisition trigger.

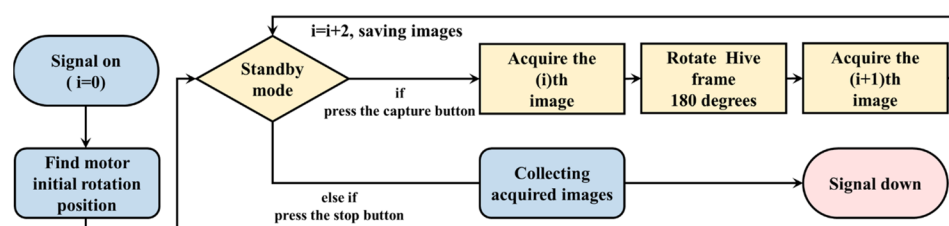


Figure 4. Image acquisition procedure for obtaining both sides of the beecomb.

2.3. Dataset Construction

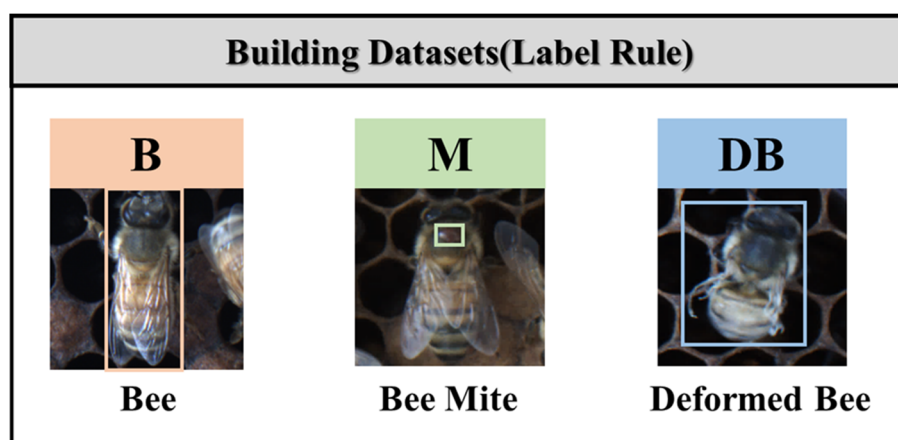
The dataset for model development was constructed using images acquired by the automated inspection system. Three target object classes were defined: Normal Bee (B), Bee Mite (M), and Deformed Bee (DB). Images containing target objects were captured in 2023, 2024, and 2025 at the National Institute of Agricultural Sciences apiary in Wanju-gun, Jeollabuk-do, and an additional apiary in Gimje-si, Jeollabuk-do, Republic of Korea. Imaging parameters were fixed at a beecomb-to-camera distance of 400 mm, lens diopter of 2.50 D, and exposure time of 12.5 ms.

The data acquisition was performed within an apiary integrated with shading film to minimize the impact of natural lighting fluctuations, such as direct solar radiation and cloud-induced shadows. By stabilizing the ambient light environment through this structural shielding, consistent image acquisition was achieved without additional illumination. This setup allowed the system to operate under semi-outdoor conditions while maintaining stable imaging quality for detecting small objects such as bee mites.

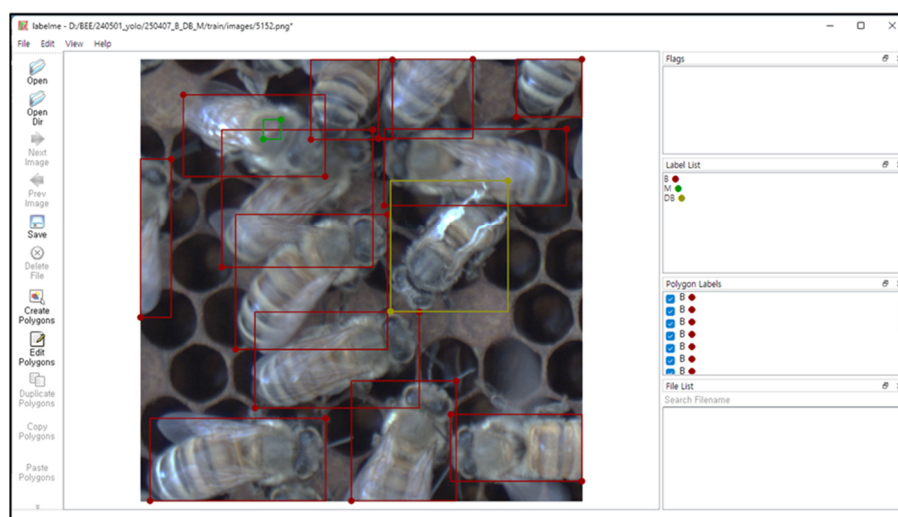
Representative labeling criteria for each class are illustrated in example images (Figure 5). Object pixel counts were calculated using Python 3.9.19 and opencv-python 4.9.0.80 to confirm that acquired images provided sufficient resolution for reliable detection of bee mites and deformed bees.

A total of 1294 raw images were acquired. To preserve the original pixel resolution and maintain the visibility of small-scale features such as bee mites, 640 × 640 pixel regions of interest (ROIs) were manually extracted from the high-resolution raw images (4608 × 2592 pixels). The ROI extraction was performed by cropping image regions containing bee mites or deformed bees so that the target objects were included within the 640 × 640 frame

while maintaining sufficient surrounding context. Because the locations of pests and deformed bees varied naturally across the beecomb surface, the resulting ROIs contained naturally varying object positions within the frame.



(a)



(b)

Figure 5. Overview of the dataset annotation. (a) Visual labeling rules for the ‘B’ (Bee), ‘M’ (Bee Mite), and ‘DB’ (Deformed Bee) target classes; (b) Example annotation interface using labelme software (ver. 5.2.1) with YOLO-format labeling. Among the bounding boxes, the red box is the bee (B) class, the green box is the bee mite (M) class, and the yellow box is the deformed bee (DB) class.

This object-centered ROI extraction strategy also helped mitigate the severe class imbalance inherent in full-frame beecomb images. Because regions were cropped based on the presence of bee mites or deformed bees, the relative representation of these target classes in the dataset was increased compared with randomly sampled image patches. This approach allowed the models to learn target features under realistic field conditions while partially balancing the dataset distribution.

This approach was prioritized to provide the models with high-quality morphological features while ensuring that the detection capability remains robust across various spatial contexts within the input frame, which is essential for identifying microscopic targets. This cropping method prevented the loss of critical morphological details—such as the fine structures of bee mites and the distinct shapes of deformed wings—that would typically occur if the original images were downscaled to the model’s input size. From the initial 1294 raw images, a total of 2018 ROIs containing the target classes were selected to

construct the final dataset for model development. Annotations were created using labelme (ver. 5.2.1) and saved in YOLO-formatted txt files. The resulting dataset contained 33,864 labeled objects, with class-wise distributions summarized in Table 1.

The dataset was partitioned into training, validation, and test sets in a 7:2:1 ratio. To eliminate the risk of data leakage, which is critical for ensuring the validity of performance metrics, a source-aware split strategy was strictly enforced. While the partitioning occurred at the ROI level, we ensured that all ROIs originating from the same primary high-resolution image were assigned to the same subset (train, validation, or test). This isolation by raw image source prevents the model from memorizing specific backgrounds, honey residues, or lighting conditions unique to a particular beehive frame, thereby ensuring that the evaluation stages reflect the model's true ability to detect bee mites and deformed bees in novel contexts.

The independent test set (10%) was completely isolated from the training and hyperparameter tuning processes, enabling a reliable evaluation of the model's generalization capability under unseen field conditions. All models were trained under identical experimental conditions to ensure fair and reproducible comparison.

Table 1. Number of objects for each class in the training, test, and validation datasets.

Class	Number of Objects			
	Train (1412) ¹	Test (405) ¹	Validation (201) ¹	Total (2018) ¹
Normal bee	22,393 (92%) ²	5883 (91%) ²	2831 (91%) ²	31,107 (92%) ²
Bee mite	1512 (6%) ²	437 (7%) ²	200 (6%) ²	2149 (6%) ²
Deformed bee	420 (2%) ²	121 (2%) ²	67 (2%) ²	608 (2%) ²
Total	24,325	6441	3098	33,864

¹ Number of images (640 × 640 pixel ROI crops), Total objects: 33,864 across 2018 images. ² Distribution of objects by class within each group (train, test, validation).

2.4. Development of YOLO-Based Deep Learning Models

2.4.1. Model Selection and Architecture Overview

Six object detection models were developed to systematically evaluate the effects of model architecture and model scale on detection performance. The selected models consisted of YOLOv8 nano, small, and large (YOLOv8n/s/l), and YOLOv11 nano, small, and large (YOLOv11n/s/l), which represent widely used real-time object detection architectures suitable for edge deployment. YOLO-based detectors were selected due to their proven balance between detection accuracy and computational efficiency, making them particularly suitable for resource-constrained edge devices. In addition, YOLOv11 represents a recent architectural advancement over YOLOv8, enabling an evaluation of how updated backbone and attention mechanisms influence detection performance in apiculture imaging tasks.

YOLOv8 employs C2f (Concat-to-Fuse) modules in the Backbone and Neck, which split input features into parallel pathways—one applying repeated Bottleneck blocks and the other preserving skip connections—before fusing them through concatenation operations. This design improves hierarchical feature propagation and enhances small-object localization capability. YOLOv11 improves upon YOLOv8 by adopting C3k2 (Cross Stage Partial with kernel size 2) blocks to reduce FLOPs and parameters through consecutive smaller kernel operations, and C2PSA (Cross Stage Partial with Parallel Spatial Attention) blocks after the SPPF module to enhance spatial attention on multi-level features. These improvements allow YOLOv11 to achieve comparable accuracy with approximately 20% fewer parameters than YOLOv8 at equivalent model scales (e.g., YOLOv11s: 9.4M vs. YOLOv8s: 11.2M parameters).

Furthermore, three model scales (nano, small, and large) were selected to investigate the trade-off between computational efficiency and detection accuracy, which is critical for practical edge-based deployment. This experimental design enabled direct comparison of: (1) architectural differences between YOLOv8 and YOLOv11, and (2) model-scale effects (nano vs. small vs. large) on bee mite and deformed bee detection performance.

2.4.2. Hyperparameters and Training Configuration

To ensure the reproducibility of the study and maintain a fair comparison between the architectures, all YOLO models were trained using identical hyperparameters and environment settings derived from the experimental logs. The models were initialized with COCO-pretrained weights to leverage transfer learning and were trained on two Nvidia RTX A6000 GPUs within a Python 3.10.13 environment.

The training was conducted for a maximum of 1000 epochs, incorporating an early stopping (patience = 50 epochs) applied to prevent overfitting. The input resolution was fixed at 640×640 pixel, and a batch size was set to 16 for all models.

For the optimization process, an auto-selected optimizer (transitioning between SGD and AdamW) was utilized with an initial learning rate of 0.01, a momentum of 0.937, and a weight decay of 0.0005. A warm-up phase of 3 epochs was applied to stabilize the initial training process. The loss function configurations were defined with weights of 7.5 for Box loss, 0.5 for Class loss, and 1.5 for Distribution Focal Loss. All models were trained using a Non-Maximum Suppression (NMS) Intersection over Union threshold of 0.7.

Crucially, data augmentation techniques such as Mosaic, MixUp, HSV adjustment, flipping, and geometric transformations were not activated during the training process. This decision was made for two primary reasons. First, the objective of this study was to evaluate the intrinsic capability of the YOLO architectures to extract authentic morphological features from high-quality field images without introducing artificial distortions. Because the detection targets—bee mites and wing-deformed bees—are characterized by subtle morphological differences, aggressive augmentation could potentially alter the visual patterns relevant to diagnosis. Second, although each 640×640 pixel ROI was extracted based on the presence of bee mites or deformed bees, each frame contained a substantially higher number of normal bee instances. Applying global image augmentation would replicate these dominant background objects and potentially amplify the inherent class imbalance within the dataset. Therefore, augmentation was intentionally disabled to preserve the original object distribution of the field data and allow the models to learn from authentic morphological characteristics of the target pests.

2.4.3. Performance Evaluation Metrics

Model performance was evaluated on images not used during training using Accuracy, Precision, Recall, F1 score, and mAP[0.5], as defined in Equations (6)–(10).

$$Accuracy = \frac{TP}{TP + FP + TN + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{All\ Detection} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{Ground\ Truth} \quad (8)$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k \quad (10)$$

where TP is the True Positive (correctly detected objects); FP is the False Positive (incorrectly detected non-objects as objects); FN is the False Negative (missed detections of actual objects); TN is the True Negative (correctly identified background regions as non-objects); n is the number of classes (3 in this study: Normal Bee, Bee Mite, Deformed Bee); AP_k is the Average Precision of class k , calculated as the area under the Precision–Recall curve for that class [30,31].

F1-confidence curves were generated to identify optimal confidence thresholds that maximize F1-score for each class. Precision–Recall (PR) curves were plotted to evaluate the trade-off between precision and recall across different confidence thresholds and to calculate $mAP[0.5]$ values. To thoroughly understand model limitations and failure modes, a comprehensive manual inspection was conducted on all test dataset images. Every instance of False Negative (FN, missed detection) and False Positive (FP, incorrect detection) was individually examined and categorized by failure type.

To qualitatively interpret model decision-making processes and validate that models attend to biologically relevant image regions, Occlusion Sensitivity Maps (OSM) were computed for representative examples from each class. OSM systematically occludes different regions of input images and measures the resulting reduction in class confidence scores, thereby highlighting which spatial regions are most critical for the model's predictions. This analysis confirmed whether models focused on biologically meaningful features (e.g., wing structures for deformed bees, mite body shapes for bee mites) rather than spurious background correlations.

2.5. Tiling-Based Inference Pipeline and Deployment

To analyze high-resolution images (4608×2592 pixels) while preserving fine-scale visual details of small objects such as bee mites, a tiling-based inference strategy was implemented. Directly resizing the full-frame image (4608×2592 pixels) to the standard YOLO input resolution (640×640) would reduce the linear resolution by a factor of 7.2. Because a bee mite occupies only about 7×7 pixels in the original image, such resizing would compress the mite representation to approximately 1 pixel, resulting in the loss of critical morphological information required for reliable detection. To address this limitation, a tiling-based inference strategy was implemented. Each raw image is divided into 32 image tiles, with each tile measuring approximately 576×648 pixels. These tiles are then resized to 640×640 pixels, which corresponds to the fixed input resolution required by the YOLO models.

The inference process is executed on a Raspberry Pi AI HAT+ equipped with a Hailo NPU accelerator, enabling hardware-accelerated processing of each tile. After tile-level inference, Non-Maximum Suppression (NMS) is applied to manage overlapping detections at tile boundaries, and the tiles are merged to reconstruct the original full-resolution output to ensure that features as small as bee mites are accurately preserved.

3. Results

3.1. Performance Evaluation of Automated Beecomb Inspection System

3.1.1. System Configuration

The automated beecomb inspection system developed in this study consists of three main hardware modules: control unit, beecomb rotation unit, and image acquisition unit (Figure 6).

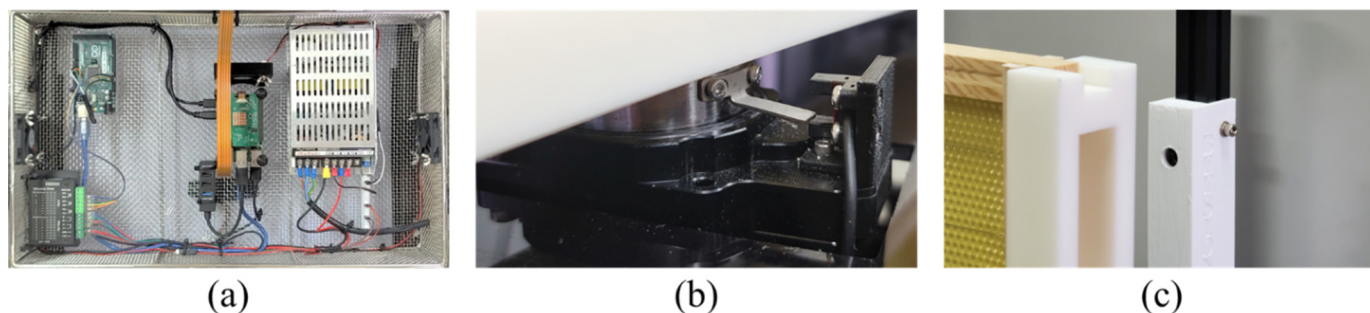


Figure 6. Hardware components of the automated beecomb inspection system. (a) The control unit, featuring the Raspberry Pi 5, AI HAT+, Arduino Mega, SMPS, and motor driver; (b) The beecomb rotation unit, showing the stepping motor and photo sensor; (c) The image acquisition unit.

The image acquisition unit was configured at a camera-to-beecomb distance of 400 mm, as determined from Equation (1) in Section 2.2.1. At this distance, the vertical measurement range calculated from Equation (2) was 299 mm, which encompasses the 240 mm height of the beecomb supporter frame. The overall dimensions of the system are 835 mm (width) × 490 mm (depth) × 1310 mm (height).

3.1.2. Inspection Time Efficiency Comparison

The proposed system significantly reduced inspection time compared to traditional manual methods and previous automated systems (Figure 7).

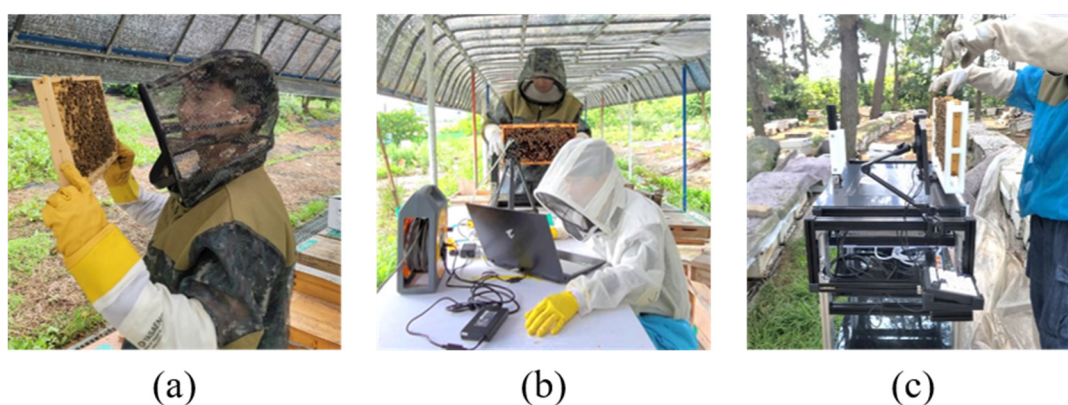


Figure 7. On-site beecomb inspection methodologies. (a) Traditional visual inspection method; (b) IAS developed by Lee et al. [24]; (c) Proposed automatic vision inspection system developed in this study.

While a beekeeper typically requires approximately 4 min (240 s) to manually inspect both sides of a single beecomb (Figure 7a), the proposed system accomplished the same task in approximately 20 s, achieving a 12-fold increase in inspection throughput.

Additionally, the 20 s inspection cycle was measured across three distinct phases to evaluate operational bottlenecks: approximately 10 s for manual handling (including hive opening and placement on the beecomb supporter frame), 5 s for dual-sided image acquisition, and 3–5 s for the integrated AI-based computational analysis. For further validation, system performance was compared with the IAS (Image Acquisition System) developed by Lee et al. [24] (Figure 7b). The IAS required approximately 180 s per beecomb, whereas the proposed device (Figure 7c) completed inspection in one-ninth the time by consolidating the imaging process into a single automated rotation cycle, thereby eliminating the redundancy of manual multi-shot acquisition used in previous systems.

Beyond throughput improvements, the edge-side computational performance was evaluated to determine system responsiveness. For a single high-resolution image, the computational breakdown using the YOLO-large model was precisely measured as follows: segmentation of the full-resolution image into 32 tiles took approximately 45.1 ms, NPU-accelerated inference for all 32 tiles required 2253.9 ms, and the subsequent Non-Maximum Suppression (NMS) combined with image reconstruction and counting consumed only 30.9 ms. The transition delay between images was negligible at 0.001 ms. Because the NMS and tile-merging process introduces a minimal overhead of approximately 31 ms per image ($\approx 1.3\%$ of total processing time), it does not constitute a computational bottleneck. Furthermore, these measurements represent the computational load of the parameter-rich large model. Downgrading to medium or small architectures further reduces the required inference time, providing operational flexibility without compromising the mechanical workflow. The end-to-end computational latency consistently averaged between 3 and 5 s per beecomb image pair when executed on a Raspberry Pi AI HAT+ equipped with a Hailo Neural Processing Unit (NPU). The measured inference latency corresponds to approximately 90–150 ms per tile during NPU inference. This near real-time processing capability enables rapid feedback during inspection while maintaining the overall 20 s inspection cycle, making the system suitable for routine field deployment in apiary monitoring.

Moreover, the system's operational scalability and field endurance were evaluated to assess its practical applicability in large-scale apiary management. The integrated hardware—comprising a Raspberry Pi 5, AI HAT+, Arduino Mega, stepper motor, and a cooling system with three 12 V fans—exhibited an average power consumption of approximately 18.9 W during operation. The system was powered by a 324,000 mAh high-capacity battery pack, corresponding to approximately 1019 Wh of usable energy. Under continuous operation, this configuration provided an estimated operational autonomy of approximately 53.9 h, enabling multi-day field deployment without recharging. To evaluate practical deployment scenarios, a representative commercial apiary consisting of 120 hives was considered. Assuming 7 beecombs per hive and an average transition time of 3 min between hives, a full inspection of the entire apiary required approximately 10.67 h. Throughout this continuous operation, the system demonstrated exceptional thermal stability. Supported by copper heatsinks and the active cooling system, the hardware maintained a baseline temperature of approximately 34 °C, peaking at only 39 °C during NPU-accelerated inference. This effective thermal management ensures consistent inference latency and confirms the system's operational readiness. No evidence of thermal throttling was observed during extended continuous operation. Under these conditions, the inspection process consumed only about 19.8% of the total battery capacity, indicating that a single battery charge could theoretically support inspections of up to 606 hives. These results demonstrate that the proposed system provides sufficient operational autonomy for multi-day field deployment, enabling practical large-scale monitoring in precision apiculture.

A key factor contributing to this time reduction is the difference in image acquisition strategy. The IAS manually acquired nine high-resolution images per side (total 18 images per beecomb), while the proposed system automatically acquired only one image per side, totaling two images per beecomb. Despite this simplification, resolution and object detection performance were maintained, as demonstrated in Section 3.2.

3.2. Image Resolution and Morphological Analysis

3.2.1. Morphological Characteristics and Quantitative Classification

Three distinct object classes were identified and annotated in the acquired images: normal bees, bee mites, and deformed bees (Figures 8–10). Normal bees, the most frequent object in the dataset, exhibited three typical postures on the beecomb: (a) folded wings with visible abdominal stripes through semi-transparent wings; (b) spread wings, which often reflected ambient light and appeared more opaque; and (c) partial body exposure, where bees were partially inside cells cleaning or feeding (Figure 8).

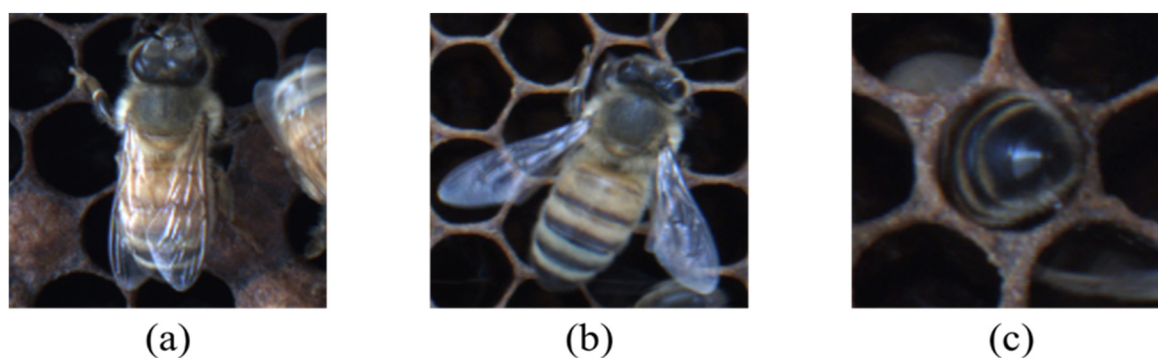


Figure 8. Representative images of normal bees categorized by posture. (a) folded wings; (b) spread wings; (c) partially inside a cell for cleaning or feeding.

Bee mites were identified in four characteristic locations: on bee's backs with clearly distinguishable outlines due to color contrast (Figure 9a); obscured by wings appearing blurred and less distinct (Figure 9b); on bee eyes with shapes resembling the eye structure (Figure 9c); or within cells, where outline clarity depended on cell background color (Figure 9d).

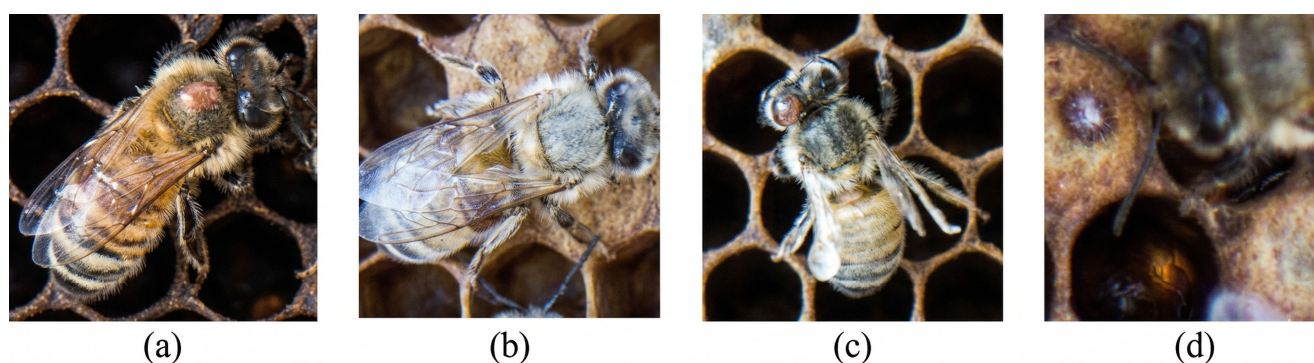


Figure 9. Representative images of bee mites categorized by location. (a) on the bee's back; (b) partially obscured by the wing; (c) on the bee's eye; (d) on the surface of the cell.

Deformed bees, though fewer in number, represented the most critical diagnostic target. Based on the appearance and structure of their wings, five morphological types were defined: DB Type I—one wing shortened by approximately 40%; DB Type II—one wing visibly contracted with reduced surface area; DB Type III—both wings shortened

symmetrically; DB Type IV—both wings contracted and appearing opaque rather than transparent; DB Type V—both wings contracted and split into branch-like structures (Figure 10).

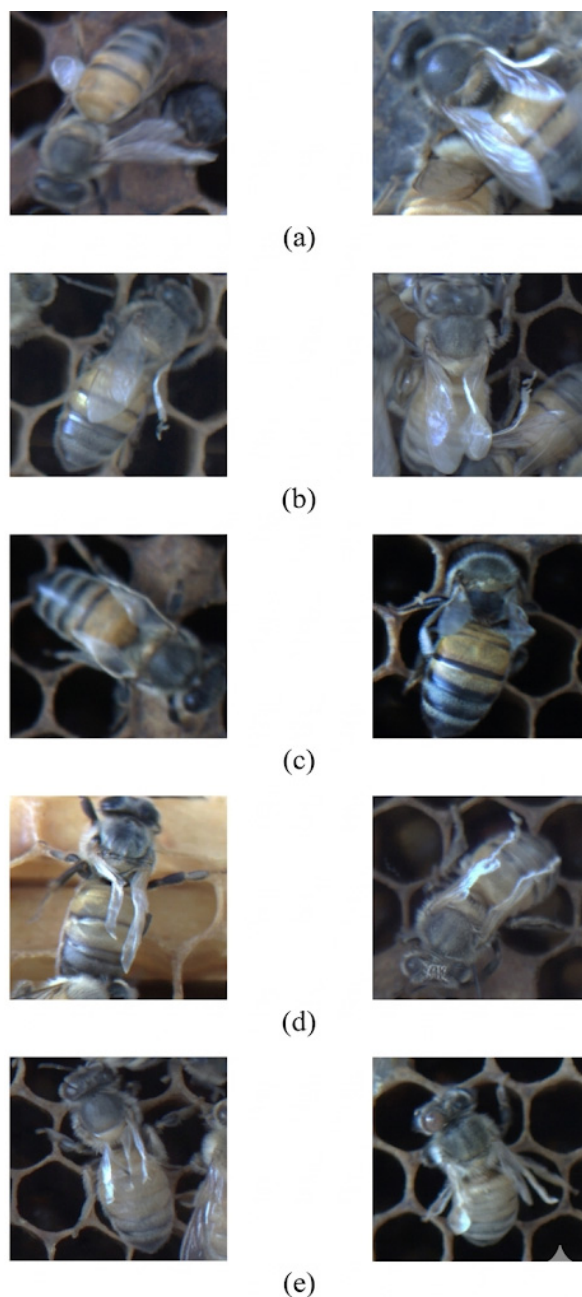


Figure 10. Representative images of deformed bees categorized by wing deformation type. (a) one wing shortened (DB Type I); (b) one wing contracted (DB Type II); (c) both wings shortened (DB Type III); (d) both wings contracted and opaque (DB Type IV); (e) both wings split into multiple branches (DB Type V).

To quantify morphological differences and establish objective classification criteria, wing-to-body area ratios were calculated using image masking techniques (Figure 11, Table 2). Normal bees with folded wings (NB I) exhibited a ratio of 47.0%, while those with spread wings (NB II) showed 35.8%. Deformed bees exhibited characteristic patterns correlated with deformity severity: single-wing deformities (DB Type I: 31.8%, DB Type II: 37.5%) yielded ratios similar to or overlapping with normal spread-wing bees (35.8%),

whereas bilateral deformities (DB Type III–V: 13.3–16.7%) showed substantially reduced ratios distinct from normal bee morphology.

This quantitative morphological characterization reveals a critical finding: DB Type II, with a wing-to-body ratio (37.5%) nearly identical to normal spread-wing bees (35.8%), presents inherent visual ambiguity, while bilateral deformities provide unambiguous morphological signals for automated detection. This quantitative baseline provides essential context for understanding subsequent model detection performance patterns, particularly for identifying which deformity types present fundamental classification challenges due to morphological overlap with normal variation.

Table 2. Wing-to-body area ratio of normal and deformed bees.

Type	Wing-to Body Ratio [%]
Normal Bee I (NB I, folded wings)	47.0
Normal Bee II (NB II, spread wings)	35.8
¹ DB type I	31.8
² DB type II	37.5
³ DB type III	16.1
⁴ DB type IV	13.3
⁵ DB type V	13.7

¹ DB type I: Bee with one wing shortened, ² DB type II: Bee with one wing contracted, ³ DB type III: Bee with both wings shortened, ⁴ DB type IV: Bee with both wings contracted, ⁵ DB type V: Bee with both wings contracted and split into several branches.

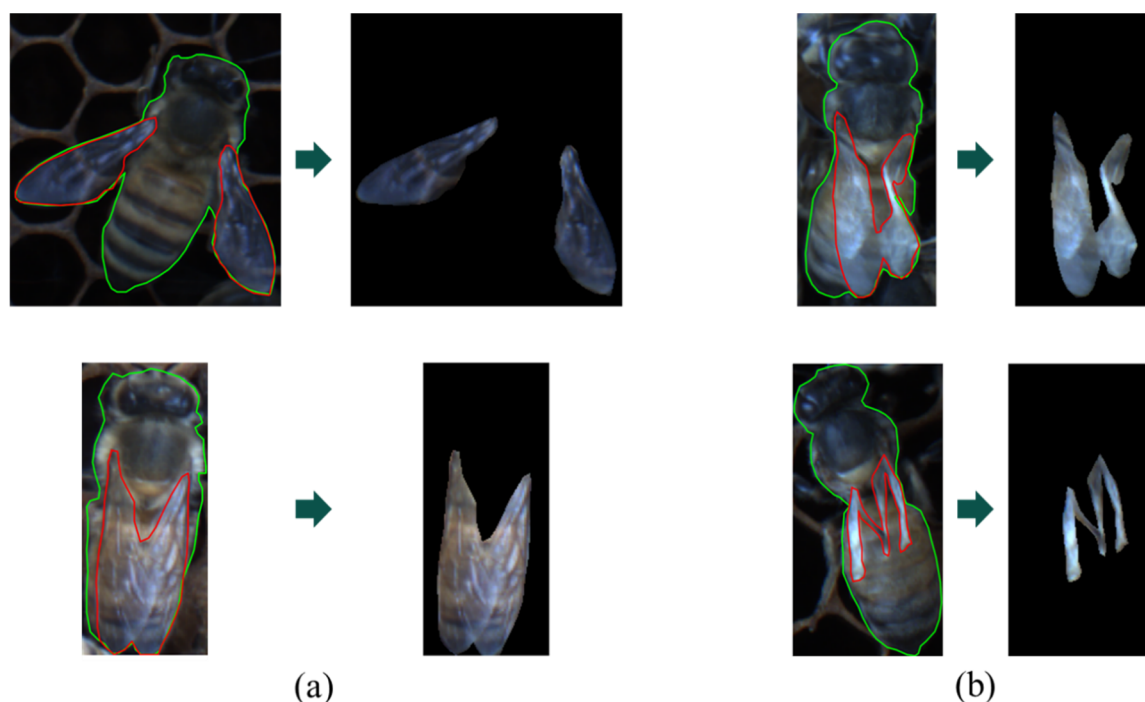


Figure 11. Example of the wing masking process for calculating the wing-to-body area ratio. The green line represents the pixel values of the entire bee area, including the wings, and the red line represents only the pixel values of the wing area. (a) Normal bee; (b) Deformed bee.

3.2.2. Image Resolution Validation for Small Object Detection

To evaluate the system's suitability for identifying small objects such as bee mites, pixel-level analysis of detected mite regions was conducted. Based on the camera working distance of 400 mm (Equation (1)) and typical dimensions of bees (10–11 mm) and bee

mites ($1.6 \text{ mm} \times 1.1 \text{ mm}$), estimated pixel values were approximately 4800 pixels for a bee and 154 pixels for a bee mite.

Actual pixel area of bee mites in the acquired images yielded an average of 143.8 pixels, with a maximum of 256 pixels (16×16) and a minimum of 49 pixels (7×7) (Figure 12). These values were compared with the resolution benchmarks established by Bjerger et al. [23], who achieved robust detection performance (88% precision, 93% recall, and 91% F1-score) for bee mites occupying approximately 41 pixels (7×6 pixel).

The average bee mite pixel value of 143.8 obtained in this study represents a 3.5-fold increase in pixel resolution (143.8 pixels vs. 41 pixels), providing over the 41-pixel reference threshold validated for reliable CNN-based detection. This substantial resolution advantage confirms that the proposed system's optical configuration provides more than adequate image quality for deep learning-based small object detection. The higher resolution may support improved feature learning, enabling models to capture finer-grained textural and morphological details critical for distinguishing bee mites from similar-appearing background features.

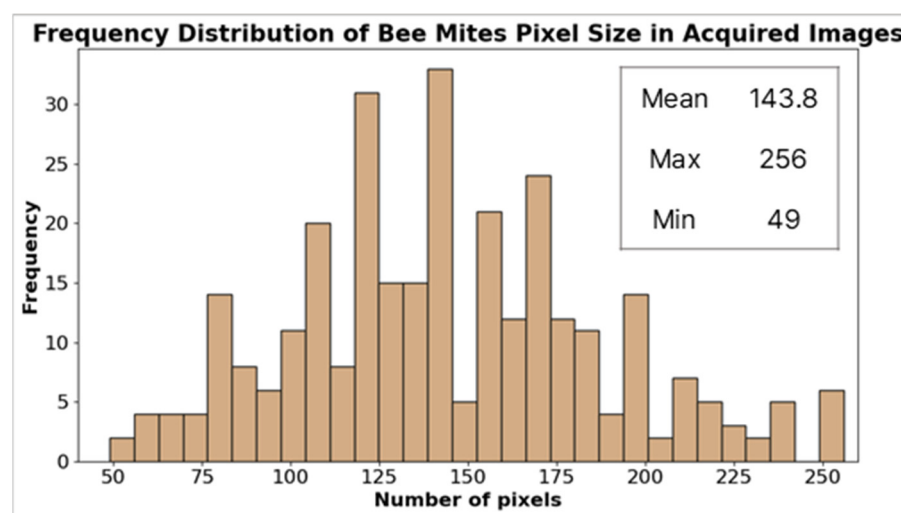


Figure 12. Histogram of bee mite pixel dimensions used to assess small-object detection performance.

3.3. YOLO Model Detection Performance

3.3.1. Overall Performance Comparison

Six YOLO models (YOLOv8/v11 $\times$ nano/small/large) were evaluated on the test dataset (405 images, 6441 objects), with comprehensive performance metrics summarized in Table 3, F1-confidence curves in Figure 13, and Precision–Recall curves in Figure 14. All models demonstrated high accuracy across the three object types, though notable performance differences emerged in specific metrics.

Table 3. Performance metrics of YOLO models on the original dataset for all classes.

Object	YOLO Version	Parameter Size (Early Stop Epoch)	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	mAP[0.5] (%)
All classes	8	n (214)	91.7	91.0	91.7	91.2	93.6
		s (150)	93.5	94.4	91.8	93.0	94.5
		l (86)	92.9	95.8	92.8	94.3	95.7
	11	n (368)	91.6	93.8	92.5	93.2	94.8
		s (140)	94.8	95.2	92.4	93.8	94.8
		l (96)	92.8	91.9	92.4	92.2	95.5

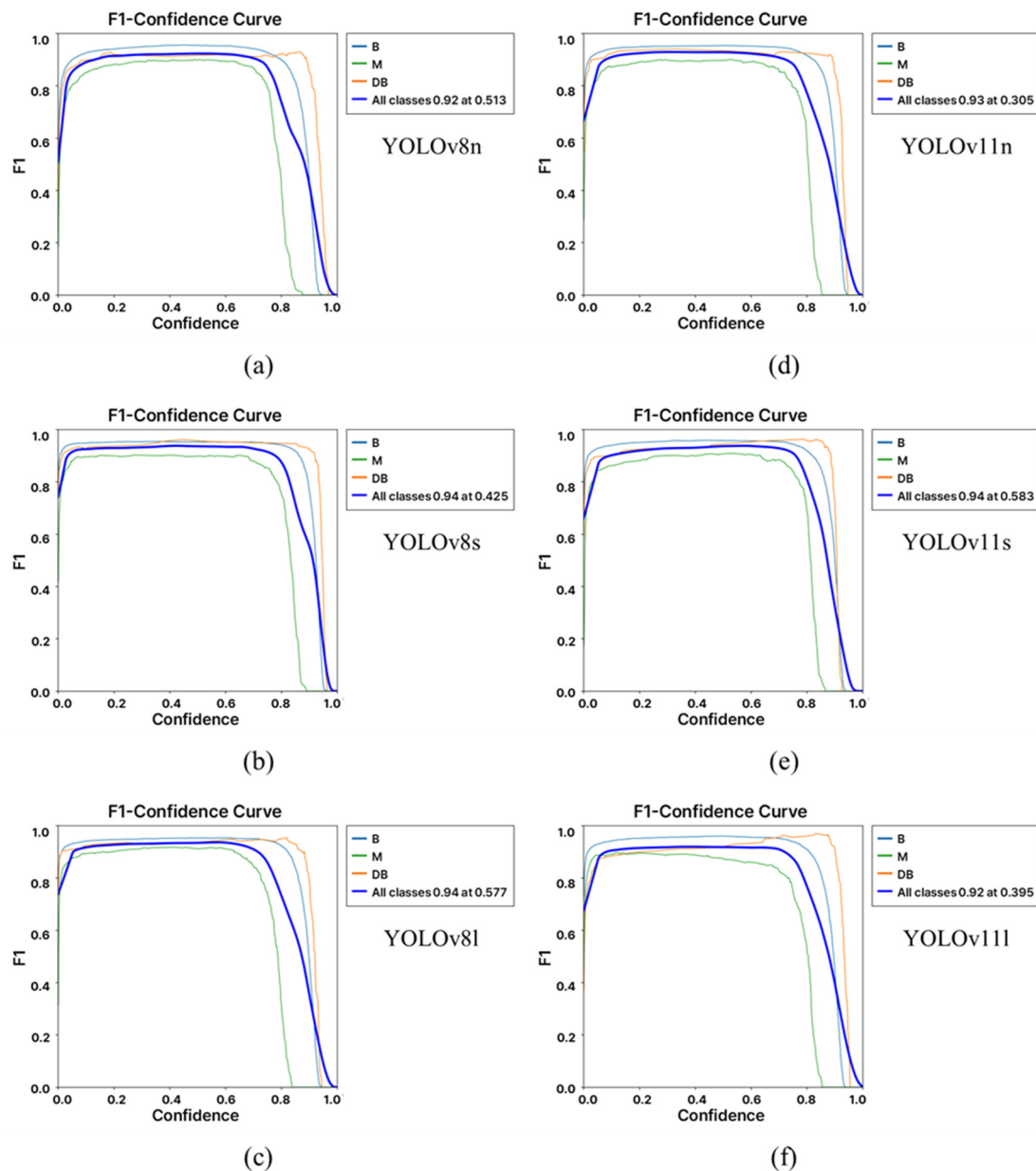


Figure 13. F1-Confidence curves on the original dataset for beekeeping pest and disease identification models. (a) YOLOv8n; (b) YOLOv8s; (c) YOLOv8l; (d) YOLOv11n; (e) YOLOv11s; (f) YOLOv11l.

YOLOv11s achieved the highest overall F1-score of 93.8%, with an F1-confidence curve showing a peak of 0.94 at confidence threshold 0.583 (Figure 13e), reflecting the most balanced classification performance. The optimal threshold of 0.583 (vs. default 0.5) indicates that slightly higher confidence filtering reduces false positives while maintaining recall, optimizing F1-score. In contrast, YOLOv8l achieved the highest mAP[0.5] of 95.7%, with peak F1-score of 0.94 at confidence 0.577 (Figure 14c), excelling in both classification and localization precision. Precision–Recall curves further revealed class-specific performance patterns (Figure 14): all models maintained high precision (>0.95) for normal bees across the entire recall range (Figure 14a), whereas YOLOv8l demonstrated superior precision at high recall levels for bee mite detection (Figure 14b), and YOLOv11s exhibited the most balanced PR curve for deformed bee detection (Figure 14c).

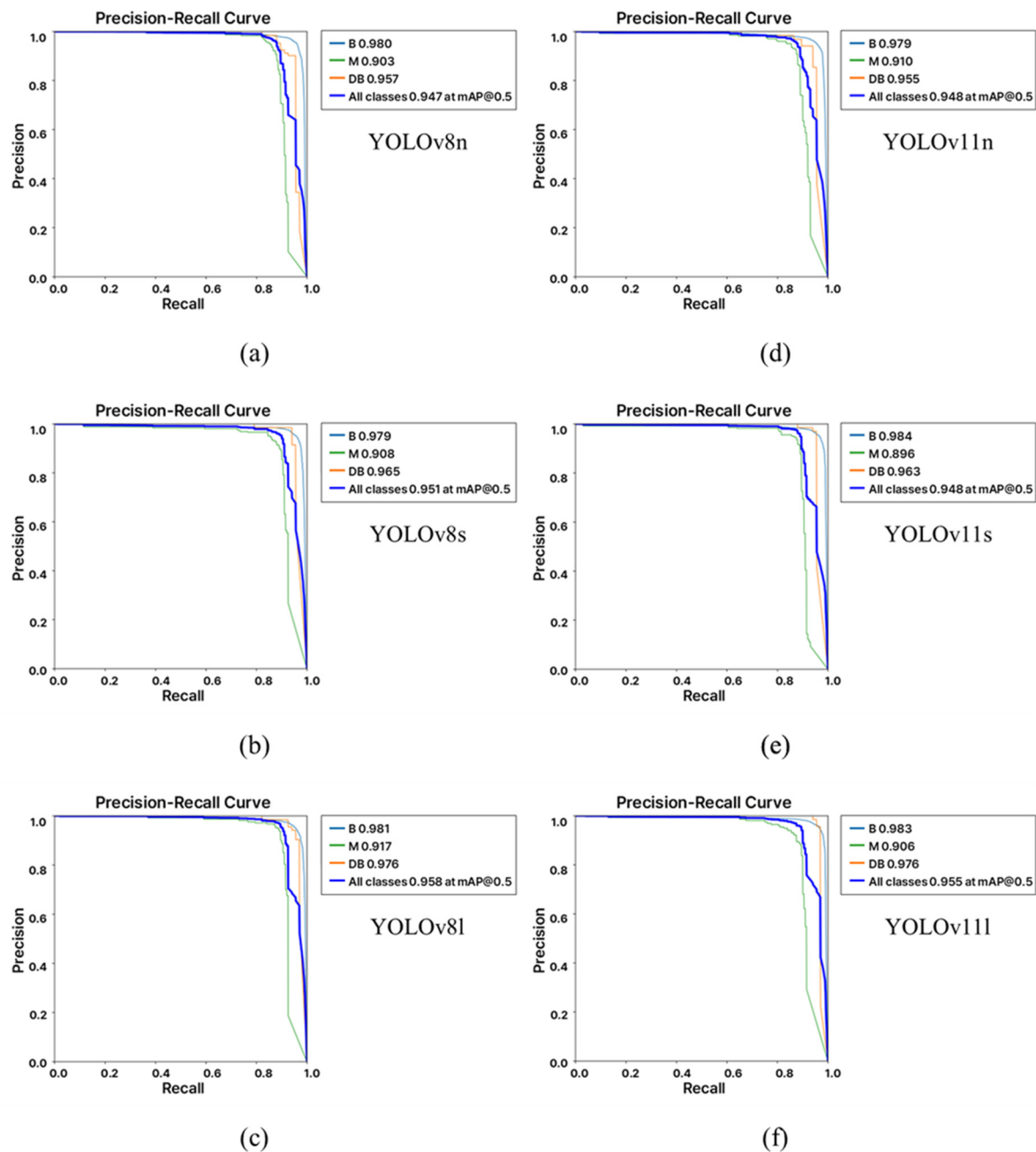


Figure 14. Precision–Recall curves on the original dataset for beekeeping pest and disease identification models. (a) YOLOv8n; (b) YOLOv8s; (c) YOLOv8l; (d) YOLOv11n; (e) YOLOv11s; (f) YOLOv11l.

This discrepancy between top-performing models for F1-score versus mAP[0.5] reveals an important architectural trade-off. F1-score primarily measures classification correctness (whether the object category was correctly identified), whereas mAP[0.5] evaluates both classification accuracy and bounding box localization precision (IoU ≥ 0.5 with ground truth). YOLOv11s, with 9.4 million parameters, demonstrated superior classification ability while maintaining balanced precision–recall trade-offs, consistent with its design emphasis on efficient feature extraction through C3k2 and C2PSA modules. Conversely, YOLOv8l, with 43.7 million parameters, excelled at precise boundary delineation, indicating that greater model capacity enabled learning of nuanced spatial features advantageous for accurate bounding box regression.

3.3.2. Class-Specific Performance Analysis

Normal Bee Detection: All models demonstrated exceptional performance, with F1-scores consistently exceeding 95% (Table 4). YOLOv11s and YOLOv11l achieved the highest performance at 95.7%, reflecting both architectural refinement and sufficient model capacity. The consistently high performance across all model sizes indicates that morphological features of healthy bees—including body shape, wing patterns, and coloration—are relatively homogeneous and easily learned by deep learning models regardless of architectural complexity.

Table 4. Performance metrics of YOLO models on the original dataset for normal bee.

Object	YOLO Version	Parameter Size (Early Stop Epoch)	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	mAP[0.5] (%)
Normal bee	8	n (214)	97.6	93.4	93.4	93.4	93.4
		s (150)	97.4	95.5	95.5	95.5	95.5
		l (86)	95.6	97.4	97.4	97.4	97.4
	11	n (368)	96.8	93.9	93.9	93.9	93.9
		s (140)	97.4	96.2	96.2	96.2	96.2
		l (96)	97.4	94.7	94.7	94.7	94.7

Bee Mite Detection: Performance differences were most pronounced in bee mite detection, revealing critical insights into model capacity requirements for small object detection. YOLOv8l unequivocally outperformed all other models, achieving the highest F1-score of 92.5% and mAP[0.5] of 92.1% (Table 5).

Table 5. Performance metrics of YOLO models on the original dataset for bee mite.

Object	YOLO Version	Parameter Size (Early Stop Epoch)	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	mAP[0.5] (%)
Bee mite	8	n (214)	86.5	93.4	84.5	88.7	88.3
		s (150)	87.5	95.1	85.0	89.8	89.3
		l (86)	90.5	96.2	89.0	92.5	92.1
	11	n (368)	88.5	93.5	87.0	90.1	91.0
		s (140)	91.5	94.8	86.5	90.5	89.6
		l (96)	85.5	94.9	83.5	88.8	90.6

These results represent substantial improvements over previous bee mite detection studies. Bjerger et al. [23] reported 88% precision, 93% recall, and 91% F1-score using CNN-based detection with 41-pixel mite images. The superior performance of YOLOv8l (92.5% F1-score) compared to these previous benchmarks can be attributed to higher image resolution (143.8 pixels vs. 41 pixels) and the advanced hierarchical design of the network.

Specifically, the necessity of the parameter-rich YOLOv8l (43.7M parameters) for small-object localization became evident when compared to lighter architectures. For “micro-targets” such as a 7×7 pixel mite, the finer spatial resolution required for accurate detection benefits more from a deeper backbone that preserves high-dimensional features than from the specific architectural refinements (such as C3k2 or C2PSA) found in lighter models like YOLOv11s. While lighter models prioritize computational efficiency, they may suffer from information loss during aggressive downsampling. In contrast, the deeper layers of YOLOv8l provide the necessary capacity to maintain and extract the subtle morphological cues essential for differentiating mites from complex background artifacts. This observation suggests that, for micro-target detection tasks, preserving fine

spatial features through deeper backbones can be more critical than adopting lightweight architectural refinements.

OSM analysis suggests that YOLOv8l learned discriminative features specific to mite morphology (Figure 15). High-intensity activation was concentrated on the mite's body region, matching its actual size and shape (Figure 15a). Under partial wing occlusion (Figure 15b), the model maintained focused attention on the mite's correct position, though with reduced confidence. Additionally, strong attention was observed on bee eye regions during normal bee classification. This indicates context-aware differentiation between visually similar features (mite bodies vs. bee eyes), demonstrating sophisticated scene understanding beyond simple color or size matching.

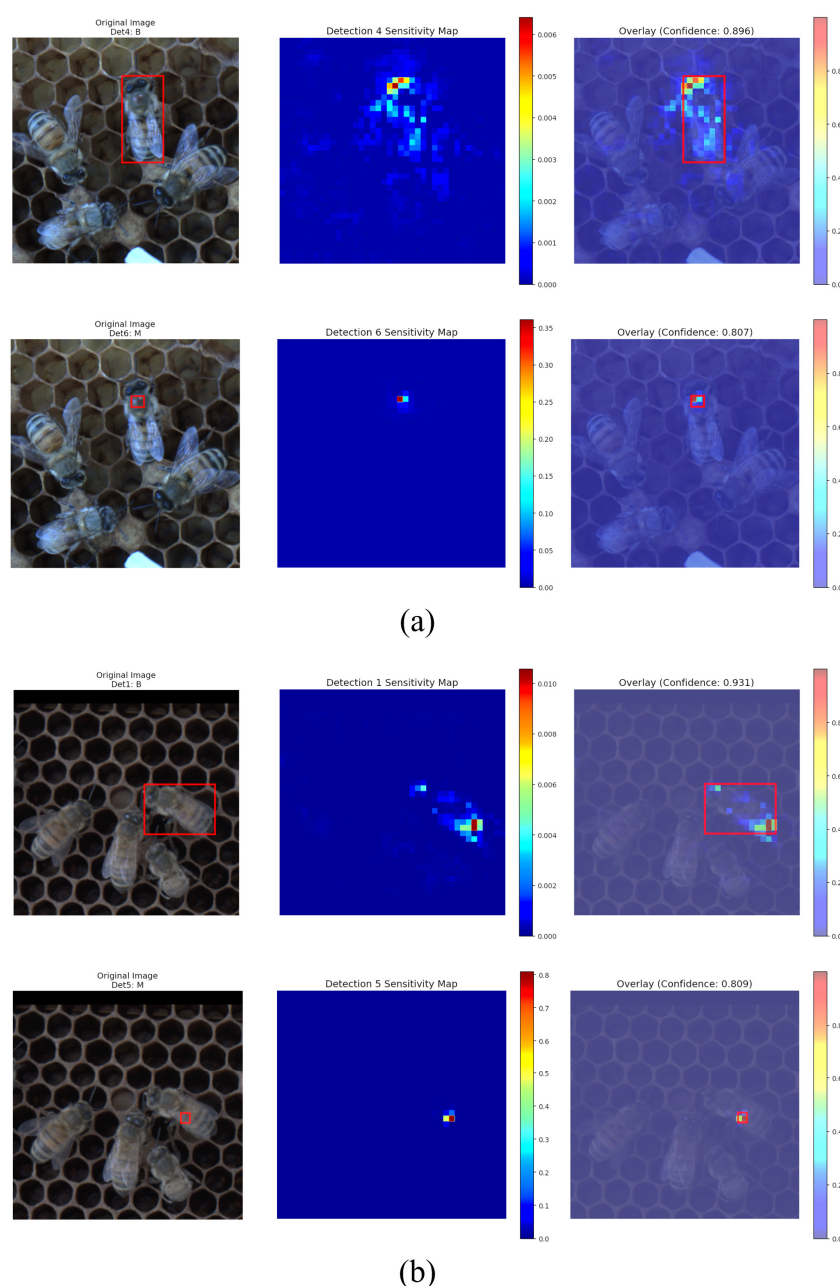


Figure 15. Occlusion Sensitivity Map (OSM) analysis for 'Bee' (B) class detections in the presence of 'Bee mite' (M). The red box is the class that performs OSM analysis. (a) OSM result when the bee mite is clearly visible (non-occluded); (b) OSM result when a bee mite is obscured by the bee's wings.

Deformed Bee Detection: YOLOv11s delivered the best performance with the highest F1-score of 95.1% (Table 6). This clearly demonstrates that architectural refinements in YOLOv11—specifically improved feature fusion via C3k2 blocks and enhanced spatial attention through C2PSA modules—are particularly effective for classifying complex and morphologically diverse abnormalities associated with wing deformities. Notably, the small-size model (YOLOv11s, 9.4 million parameters) outperformed the large model (YOLOv11l: F1 = 91.2%, 25.3 million parameters), suggesting that YOLOv11s achieved an optimal balance between architectural sophistication and model size for this specific classification task, avoiding potential overfitting or excessive parameter redundancy that may have hindered YOLOv11l.

OSM analysis demonstrated that YOLOv11s systematically focused on deformed wing regions across all pathology types (Figure 16). This validates the C2PSA module’s effectiveness for spatially distributed abnormalities, explaining YOLOv11’s superior performance in detecting morphologically diverse wing deformities.

Table 6. Performance metrics of YOLO models on the original dataset for deformed bee.

Object	YOLO Version	Parameter Size (Early Stop Epoch)	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	mAP[0.5] (%)
Deformed bee	8	n (214)	91.0	86.2	94.0	89.9	94.5
		s (150)	95.5	92.7	95.5	94.1	96.1
		l (86)	92.5	93.9	95.5	94.7	96.8
	11	n (368)	89.6	93.9	94.0	94.0	95.5
		s (140)	95.5	94.7	95.5	95.1	96.3
		l (96)	95.5	86.1	97.0	91.2	97.6

3.4. Error Analysis and Model Limitations

3.4.1. Bee Mite Detection: Sensitivity–Specificity Trade-Offs

False positive (FP) and false negative (FN) analyses revealed critical trade-offs between sensitivity and specificity (Table 7). YOLOv11l exhibited the lowest FN rate, missing only 11 out of 437 mites (2.5%), demonstrating high sensitivity for small object detection. In contrast, YOLOv8n missed 31 mites (7.1%)—nearly four times as many. However, YOLOv8l achieved the lowest FP rate (11 errors), while YOLOv11l produced 27 FPs—more than double that of YOLOv8l.

Table 7. Misclassification analysis of bee mite detection using YOLO models on the test dataset.

YOLO Version	Bee Mite	
	False Negatives	False Positives ¹ (Unlabeled Object ² /Background ³)
8n	31	16 (10/6)
8s	24	10 (5/5)
8l	26	12 (5/7)
11n	30	9 (7/2)
11s	21	8 (6/2)
11l	29	11 (7/4)

False positives are reported as “Total (Unlabeled object/Background)”. ¹ False positive: The overall number of false positive detections. ² Unlabeled object: Valid target detections without corresponding ground-truth annotations, potentially due to annotation limitations or oversight. ³ Background: Detections corresponding to pure background artifacts, image noise, or beecomb structures mistakenly identified as objects.

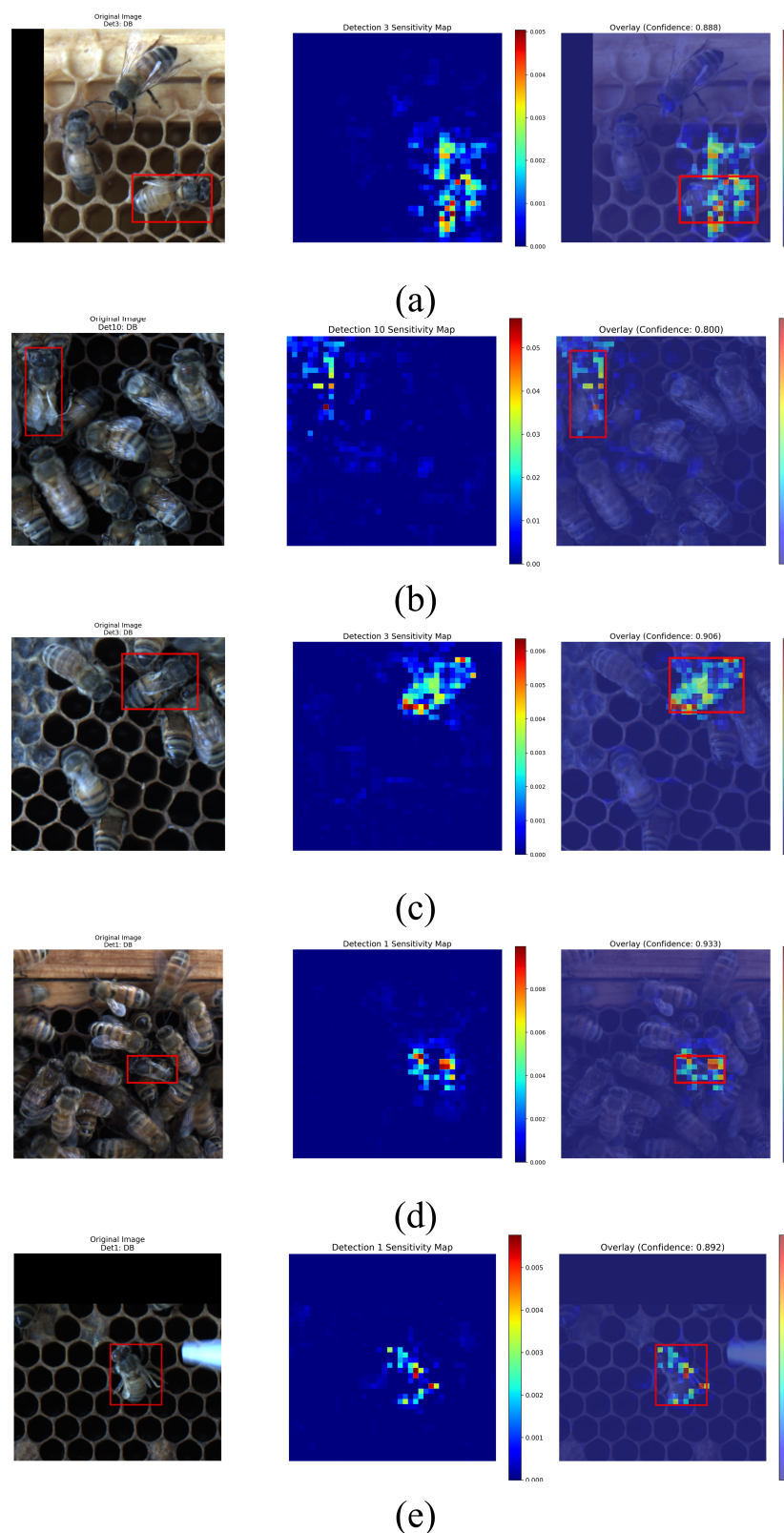


Figure 16. Occlusion Sensitivity Map (OSM) analysis for the ‘Deformed Bee’ (DB) class, categorized by deformity type. The red box is the class that performs OSM analysis. (a) OSM result for DB Type I; (b) OSM result for DB Type II; (c) OSM result for DB Type III; (d) OSM result for DB Type IV; (e) OSM result for DB Type V.

The “Unlabeled” category represents model detections do not correspond to any annotated ground-truth objects. Manual review indicated that most of these detections

corresponded to background artifacts or image noise rather than missed annotations. A small number of detections occurred in visually ambiguous regions where mites were partially obscured by bee wings, making consistent manual annotation challenging. However, such cases were rare and did not significantly affect the reported precision or false-positive analysis.

This sensitivity–specificity trade-off reflects fundamental differences in model behavior. YOLOv11l’s C2PSA spatial attention mechanisms enhance sensitivity to small features but frequently misidentify debris or body spots as mites, reducing specificity. Conversely, YOLOv8l demonstrates superior discrimination, achieving higher detection reliability at the cost of occasional missed detections. Optimal model selection depends on application context: YOLOv11l is advantageous for early warning systems prioritizing outbreak prevention, while YOLOv8l’s low FP rate is essential for research applications requiring accurate infestation quantification.

3.4.2. Deformed Bee Detection: Morphological Ambiguity and Detection Failure

Deformity-type-specific analysis revealed systematic associations between morphological distinctiveness—quantified by wing-to-body area ratio—and detection accuracy (Table 8). Detection failure rates correlated strongly with the degree of ratio overlap between deformed and normal bees.

Table 8. Misclassification analysis of deformed bee detection using YOLO models on the test dataset.

YOLO Version	DB Type I	DB Type II	DB Type III	DB Type IV	DB Type V	DB
	False Negatives (Missed/Total)					False Positives
8n	0/9	7/26	0/10	2/45	1/31	12
8s	0/9	4/26	0/10	2/45	1/31	8
8l	0/9	6/26	0/10	2/45	1/31	14
11n	0/9	5/26	0/10	3/45	0/31	10
11s	0/9	4/26	0/10	2/45	0/31	10
11l	0/9	3/26	0/10	1/45	0/31	17

DB: Deformed bee.

DB Type I (31.8% ratio): Larger models (YOLOv11s/l, YOLOv8l) achieved 100% recall, while lightweight models each missed one case. DB Type II (37.5% ratio): Most problematic class, with YOLOv8n missing 7/26 cases (26.9%) and YOLOv11l missing 3 cases (11.5%). Critically, the wing-to-body ratio (37.5%) nearly matches normal spread-wing bees (35.8%), representing only a 1.7 percentage point difference. This minimal morphological deviation provides extremely limited visual cues, fundamentally challenging model discrimination regardless of architectural sophistication.

DB Type III (16.1% ratio): All models achieved 100% recall (0/10 missed). The ratio deviates substantially from normal bees (35.8%), representing a 19.7 percentage point difference—more than 10-fold greater than DB Type II. DB Types IV and V (13.3–13.7% ratios): High detection reliability (0–2 missed cases) due to severe ratio reductions and distinctive structural features.

Quantitative correlation: DB Type II, with minimal ratio deviation (1.7%), exhibited 23.1% miss rate, while DB Type III, with maximal deviation (19.7%), achieved 0% miss rate. This confirms that morphological ambiguity, objectively quantified by wing-to-body ratio overlap with normal variation, directly determines detection feasibility. These findings suggest that DB Type II cases require targeted synthetic augmentation strategies emphasizing subtle textural features beyond gross morphology, or alternative multi-view imaging approaches for reliable identification.

4. Discussion

Comprehensive performance analysis revealed a clear phenomenon of model-task specialization: no single model emerged as universally optimal across all detection tasks. YOLOv8l specialized in detecting small, low-contrast bee mites (F1: 92.5%, mAP: 92.1%), leveraging its 43.7M parameters for robust small-object feature learning. Conversely, YOLOv11s specialized in classifying morphologically diverse deformed bees (F1: 95.1%), demonstrating the effectiveness of its C3k2 and C2PSA architectural refinements for handling varied morphological patterns.

These findings indicate that optimal model selection should prioritize target object characteristics rather than assuming newer architectures universally outperform predecessors. For small objects requiring precise localization, parameter-rich models offer advantages; for classification tasks with diverse morphologies, architecturally refined models provide superior efficiency and accuracy. Lightweight models (YOLOv8n, YOLOv11n) consistently underperformed (bee mites miss rate: 9.3%), indicating insufficient capacity for operational apiculture applications where detection reliability directly impacts colony health management.

These specialization patterns were observed under controlled conditions that prioritized data consistency. Specifically, all image acquisition was conducted within a rain-shielding apiary equipped with shading film. This structural environment effectively minimized external variables such as direct solar radiation and fluctuating shadows from cloud movement, allowing for high-quality image capture with a fixed exposure time of 12.5 ms without supplemental lighting. While this controlled setup ensured the reliability of the training dataset, it may limit the generalizability of the system in completely open-field apiaries where lighting conditions are more variable. Therefore, further validation under diverse environmental conditions will be necessary to confirm the robustness of the proposed system.

Furthermore, the system's optical configuration was designed to accommodate the three-dimensional structure of the beehive. By using a 2.50 diopter lens at a fixed imaging distance of approximately 400 mm, the system provides a depth of field that is sufficient to cover the typical beehive cell depth of approximately 12 mm. This optical setup allows bee mites located at different vertical positions within the cells to remain within an acceptable focus range during image acquisition. However, in practical conditions, particularly under high bee density, the detection of bee mites deep inside the hexagonal cells is often limited by physical occlusion from surrounding bees. As a result, the primary objective of the proposed system for rapid field screening is to detect visible bee mites on the surface of adult bees, which serves as a reliable indicator of the overall colony infestation level.

5. Conclusions

This study developed and validated an automated AI-based vision inspection system for detecting bee mite and deformed bee, achieving a 12-fold reduction in inspection time compared to manual methods (20 s vs. 240 s per beehive). By integrating a motorized beehive rotation mechanism with YOLO-based deep learning models, the system enables rapid, dual-sided image acquisition and demonstrates significant potential for operational deployment in commercial apiaries.

To ensure the reliability of the performance metrics, the dataset was partitioned at the raw image level before tiling, effectively eliminating data leakage and ensuring spatial independence. The implementation of the tiling-based strategy was instrumental in preserving the fine-scale morphological features of bee mites, which would otherwise be mathematically undetectable through standard image resizing. Comprehensive

evaluation of six YOLO models showed a model-task specialization pattern: YOLOv8l achieved the highest performance for small, low-contrast bee mites (F1: 92.5%, mAP[0.5]: 92.1%), while YOLOv11s excelled in classifying morphologically diverse wing deformities (F1: 95.1%). These results indicate that optimal model selection should be determined by the specific visual characteristics of target objects rather than by architectural generation alone.

Granular error analysis further demonstrated systematic relationships between morphological features and detection accuracy. For deformed bees, misclassification rates were associated with overlap in wing-to-body ratios relative to normal bees. DB Type II, which exhibited only a 1.7% difference from normal wing ratios, showed an 18.6% miss rate, whereas DB Type III, with a 19.7% difference, achieved perfect detection. This 1.7% threshold is explicitly identified as a “critical limit of detection” for single-view RGB systems. This boundary justifies the necessity for future research into multi-view imaging or the utilization of specific textural features, such as wing venation patterns, rather than relying solely on gross morphology to differentiate these subtle cases. This quantitative relationship suggests that morphological similarity to normal bees significantly influences detection difficulty.

While this study utilized shading films and fixed exposure settings to establish a stable imaging baseline, further research is required to evaluate system robustness under unconstrained outdoor lighting conditions. Future work will include multi-location dataset collection and field validation across diverse apiary environments. Expanding the dataset to include different geographical regions, seasonal conditions, and honeybee populations will be essential for improving the model’s generalization capability. To further improve environmental robustness, future implementations will integrate an LED illumination system synchronized with the rotation motor to reduce sensitivity to ambient lighting variations. To maintain imaging consistency and prevent motion blur, the LED system will employ indirect illumination, thereby minimizing harsh reflections on the comb surface. Furthermore, the lighting control will be integrated with the motor driver to activate exclusively when the comb is stationary for image acquisition. During the 5 s 180° rotation phase, the illumination will be turned off. This synchronized operation ensures uniform exposure across both sides of the becomb while optimizing power consumption for long-term field deployment.

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

Funding: This study was supported by the Rural Development Administration as part of the Cooperative Research Program for Agriculture Science and Technology Development, grant number [RS-2023-00232224].

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hung, K.-L.J.; Kingston, J.M.; Albrecht, M.; Holway, D.A.; Kohn, J.R. The worldwide importance of honey bees as pollinators in natural habitats. *Proc. R. Soc. B Biol. Sci.* **2018**, *285*, 20172140. <https://doi.org/10.1098/rspb.2017.2140>.
2. Johnson, R. *Honey Bee Colony Collapse Disorder*; Congressional Research Service: Washington, DC, USA, 2010; Volume 444.
3. Spleen, A.M.; Lengerich, E.J.; Rennich, K.; Caron, D.; Rose, R.; Pettis, J.S.; Henson, M.; Wilkes, J.T.; Wilson, M.; Stitzinger, J. A national survey of managed honey bee 2011–12 winter colony losses in the United States: Results from the Bee Informed Partnership. *J. Apic. Res.* **2013**, *52*, 44–53. <https://doi.org/10.3896/IBRA.1.52.2.07>.
4. Lee, K.; Steinhauer, N.; Rennich, K.; Wilson, M.; Tarpy, D.; Caron, D.; Rose, R.; Delaplane, K.; Baylis, K.; Lengerich, E. A national survey of managed honey bee 2013–2014 annual colony losses in the USA. *Apidologie* **2015**, *46*, 292–305. <https://doi.org/10.1007/s13592-015-0356-z>.
5. Seitz, N.; Traynor, K.S.; Steinhauer, N.; Rennich, K.; Wilson, M.E.; Ellis, J.D.; Rose, R.; Tarpy, D.R.; Sagili, R.R.; Caron, D.M. A national survey of managed honey bee 2014–2015 annual colony losses in the USA. *J. Apic. Res.* **2022**, *54*, 292–304. <https://doi.org/10.1080/00218839.2016.1153294>.
6. Buczek, K. Honey bee colony collapse disorder (CCD). *Ann. Univ. Mariae Curie-Skłodowska Med. Vet.* **2009**, *64*, 1–6.
7. Kim, H.-K. The Effect of Honey Bee Mites on the Winter Colony Losses. *J. Apic.* **2022**, *37*, 291–299. <https://doi.org/10.17519/api-culture.2022.09.37.3.291>.
8. Ramsey, S.D.; Ochoa, R.; Bauchan, G.; Gulbranson, C.; Mowery, J.D.; Cohen, A.; Lim, D.; Joklik, J.; Cicero, J.M.; Ellis, J.D. *Varroa destructor* feeds primarily on honey bee fat body tissue and not hemolymph. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1792–1801. <https://doi.org/10.1073/pnas.1818371116>.
9. Noël, A.; Le Conte, Y.; Mondet, F. *Varroa destructor*: How does it harm *Apis mellifera* honey bees and what can be done about it? *Emerg. Top. Life Sci.* **2020**, *4*, 45–57. <https://doi.org/10.1042/ETLS20190125>.
10. Choi, Y.; Lee, M.; Lee, M.; Lee, K. Detection of Seven Bee Viruses from *Varroa destructor* Mite. *Korean J. Apic.* **2008**, *23*, 171–176.
11. Gisder, S.; Aumeier, P.; Genersch, E. Deformed wing virus: Replication and viral load in mites (*Varroa destructor*). *J. Gen. Virol.* **2009**, *90*, 463–467. <https://doi.org/10.1099/vir.0.005579-0>.
12. De Miranda, J.R.; Genersch, E. Deformed wing virus. *J. Invertebr. Pathol.* **2010**, *103*, S48–S61. <https://doi.org/10.1016/j.jip.2009.06.012>.
13. Chen, Y.P.; Siede, R. Honey bee viruses. *Adv. Virus Res.* **2007**, *70*, 33–80. [https://doi.org/10.1016/S0065-3527\(07\)70002-7](https://doi.org/10.1016/S0065-3527(07)70002-7).
14. Wilfert, L.; Long, G.; Leggett, H.; Schmid-Hempel, P.; Butlin, R.; Martin, S.; Boots, M. Deformed wing virus is a recent global epidemic in honeybees driven by *Varroa* mites. *Science* **2016**, *351*, 594–597. <https://doi.org/10.1126/science.aac9976>.
15. Meixner, M.D.; Pinto, M.A.; Bouga, M.; Kryger, P.; Ivanova, E.; Fuchs, S. Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *J. Apic. Res.* **2013**, *52*, 1–28. <https://doi.org/10.3896/IBRA.1.52.4.05>.
16. Statistics Korea. Farm Households Raising Livestock/Total Head. 2015. Available online: https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1AG15501&conn_path=I2&language=en (accessed on 8 April 2026).
17. Statistics Korea. Agriculture, Forestry and Fishery Survey in 2023. In *Farm, Forest and Fishery Household and Population 2024*; Statistics Korea: Daejeon, Republic of Korea, 2024. Available online: https://ko-stat.go.kr/board.es?mid=a20102040000&bid=11709&act=view&list_no=431430 (accessed on 8 April 2026).
18. Park, Y.; Lee, M. Notes on the Mitochondrial Haplotype of *Varroa destructor* Anderson and Trueman (*Acari: Varroidae*) infested on Honeybees in Korea and New Zealand. *Korean J. Apic.* **2004**, *19*, 117–124.
19. Jack, C.J.; Ellis, J.D. Integrated pest management control of *Varroa destructor* (*Acari: Varroidae*), the most damaging pest of (*Apis mellifera* L. (Hymenoptera: Apidae)) colonies. *J. Insect Sci.* **2021**, *21*, 6. <https://doi.org/10.1093/jisesa/ieab058>.
20. Delaplane, K.S.; Van Der Steen, J.; Guzman-Novoa, E. Standard methods for estimating strength parameters of *Apis mellifera* colonies. *J. Apic. Res.* **2013**, *52*, 1–12. <https://doi.org/10.3896/IBRA.1.52.1.03>.
21. Zacepins, A.; Kviesis, A.; Ahrendt, P.; Richter, U.; Tekin, S.; Durgun, M. Beekeeping in the future—Smart apiary management. In *Proceedings of the 2016 17th International Carpathian Control Conference (ICCC)*; IEEE: New York, NY, USA, 2016; pp. 808–812.
22. Mrozek, D.; Górny, R.; Wachowicz, A.; Małysiak-Mrozek, B. Edge-based detection of varroosis in beehives with IoT devices with embedded and tpu-accelerated machine learning. *Appl. Sci.* **2021**, *11*, 11078. <https://doi.org/10.3390/app112211078>.
23. Bjerger, K.; Frigaard, C.E.; Mikkelsen, P.H.; Nielsen, T.H.; Misbi, M.; Kryger, P. A computer vision system to monitor the infestation level of *Varroa destructor* in a honeybee colony. *Comput. Electron. Agric.* **2019**, *164*, 104898. <https://doi.org/10.1016/j.compag.2019.104898>.
24. Lee, H.G.; Kim, M.-J.; Kim, S.-b.; Lee, S.; Lee, H.; Sin, J.Y.; Mo, C. Identifying an image-processing method for detection of bee mite in honey bee based on keypoint analysis. *Agriculture* **2023**, *13*, 1511. <https://doi.org/10.3390/agriculture13081511>.

25. Kongsilp, P.; Taetragool, U.; Duangphakdee, O. Individual honey bee tracking in a beehive environment using deep learning and Kalman filter. *Sci. Rep.* **2024**, *14*, 1061. <https://doi.org/10.1038/s41598-023-44718-y>.
26. Bhat, S.A.; Huang, N.-F. Big data and ai revolution in precision agriculture: Survey and challenges. *IEEE Access* **2021**, *9*, 110209–110222. <https://doi.org/10.1109/ACCESS.2021.3102227>.
27. Lee, H.-G.; Shin, J.-Y.; Kim, S.-B.; Kim, M.-J.; Kim, M.S.; Lee, H.; Mo, C. Enhancing Bee Mite Detection with YOLO: The Role of Data Augmentation and Stratified Sampling. *Agriculture* **2025**, *15*, 1221. <https://doi.org/10.3390/agriculture15111221>.
28. Lou, H.; Duan, X.; Guo, J.; Liu, H.; Gu, J.; Bi, L.; Chen, H. DC-YOLOv8: Small-size object detection algorithm based on camera sensor. *Electronics* **2023**, *12*, 2323. <https://doi.org/10.3390/electronics12102323>.
29. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors* **2023**, *23*, 7190. <https://doi.org/10.3390/s23167190>.
30. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. <https://doi.org/10.1007/s11263-009-0275-4>.
31. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *European Conference on Computer Vision 2014*; Springer: Berlin/Heidelberg, Germany; pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.