SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I10P101

Volume 13 Issue 10, 1-15, October 2026
© 2026 Seventh Sense Research Group®

Original Article

Rail-Guided IoT Imaging System for Leaf-Level Monitoring in Small-Scale Citrus Plots

José Fabián Huamaní Zegarra[1], German Alberto Echaiz Espinoza[2*], Daniel Domingo Yanyachi Aco Cardenas[2], Andrés Ortiz Salazar[3]

[1]*Escuela Profesional de Ingeniería Electrónica, Universidad Nacional de San Agustin de Arequipa, Arequipa, Perú.*
[2]*Departamento Académico de Ingeniería Electrónica, Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú.*
[3]*Department of Computer Engineering and Automation, Federal University of Rio Grande do Norte (DCA-UFRN), RN, Brazil.*

[*]*Corresponding Author: gechaiz@unsa.edu.pe*

Received: 11 August 2026 Revised: 15 September 2026 Accepted: 23 September 2026 Published: 28 October 2026

Abstract *- This paper presents a rail-guided IoT imaging system designed to overcome the difficulty of obtaining consistent, leaf-level citrus images for long-term agricultural monitoring by combining deterministic electromechanical positioning, embedded control, edge computing, and cloud-based data management. It incorporates a cable-suspended linear actuation mechanism together with two inclined cameras, motion control based on ESP32 microcontrollers, a Raspberry Pi 5 for processing, and AWS S3 for cloud services to automate the processes of image acquisition, synchronization, storage, and remote visualization. By limiting image capture to a fixed and repeatable path, the platform reduces the variation in viewpoint that is typically found with aerial and ground-mobile robots, thus allowing the systematic production of consistent image datasets for use in plant phenotyping and pest monitoring. The system was implemented and tested in a citrus orchard under real-world operating conditions. Experimental results demonstrate the system's autonomous and stable operation, reliable image acquisition, continuous synchronization with the cloud and repeatable monitoring over long periods of time with a low-cost hardware design suitable for research-scale setups. The platform provides a reproducible embedded imaging setup, which improves the consistency of agricultural image acquisition and enables the development and evaluation of computer vision algorithms for precision agriculture.*

Keywords *- Citrus crops, Rail-guided imaging, IoT monitoring system, Leaf-level imaging, Precision agriculture, Instance segmentation.*

1. Introduction

Pest outbreaks endanger the healthy development and growth of citrus crops around the world. Even though the recent focus on mitigating these effects has shifted toward the use of advanced technologies, much of the monitoring process remains labor-intensive, especially given the difficulty of detecting small pests or early-stage diseases. To address these broad challenges, precision agriculture is becoming more dependent on sensing technology, automation, and methods based on data, enhancing the way crops are managed and decreasing the use of resources [1-3]. However, it is especially difficult to carry out reliable monitoring of citrus crops since pests like aphids and whiteflies generally live on the abaxial side of the leaves, meaning that manual inspection is required [4]. Even though unmanned aerial vehicles and mobile ground robots have been extensively studied for the purpose of agricultural monitoring [5, 6], they usually have trouble achieving stable, repeatable, and close-up views of leaf surfaces because of varying angles, canopy blockages, and image blur caused by movement.

Alternatively, the development of IoT technologies has made it possible to deploy distributed, low-cost monitoring systems that can gather environmental and visual data in agricultural fields [7-9]. However, most current IoT systems focus on environmental sensing or on large-area classification rather than on controlled imaging at the leaf level [10, 11]. Moreover, there are no specialized systems that allow high-resolution images of citrus foliage to be taken repeatedly under natural field conditions. This shortcoming impedes the creation of realistic datasets for computer-vision research and limits studies that demand a consistent imaging geometry, such as phenotyping, analysis of pest progression or early-stage monitoring [12, 13].

To overcome these limitations, the present study introduces a rail-guided IoT imaging system designed for leaf-level monitoring in small-scale citrus fields. The system consists of a cable-driven rail, two inclined cameras, two motor controllers based on ESP32 chips, and a Raspberry Pi 5 (RPi5) processing unit, thus allowing for automated image

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

acquisition along a fixed and repeatable route. The images are collected and uploaded to a cloud storage service (AWS S3 Bucket) and are displayed via a custom web application, which supports remote inspection and research activities.

Unlike aerial or mobile agricultural platforms, the rail-based architecture ensures a highly controlled and repeatable imaging geometry, specifically configured to capture the abaxial sides of leaves. Although industrial automation systems assess performance using mechanical micro-positioning criteria, this platform establishes its research standard by achieving spatial reproducibility and generating high-fidelity datasets. By fixing the camera to a rigid linear vector, the system ensures that views recorded at each stop remain morphologically consistent over time. This controlled cycle directly addresses the realism gap by automatically collecting longitudinal in-field images under unconstrained environmental conditions. As a result, the system performance level is validated not by using expensive positional feedback hardware, but by its demonstrated ability to produce structured, longitudinal datasets that are optimized for deep learning instance segmentation models.

1.1. Objectives

The objective of this work is to design, implement, and evaluate an integrated low-cost remote monitoring system for citrus crops that utilizes edge computing and computer vision to detect early-stage pests and diseases:

1. Design and implement a linear motion system driven by cables so that it can move a cart fitted with two cameras along a track, which is 75 cm long, for the purpose of acquiring images at canopy level.
2. Implement a master-slave control system based on Raspberry Pi 5 and ESP32 microcontrollers to coordinate the motor movement and image capture cycles.
3. Fine-tune a lightweight YOLO11s-seg instance segmentation model so that it can identify the pest residues (the wax secretions of whiteflies) in citrus leaves, achieving a good balance between mean Average Precision (mAP) and the speed of edge inference.
4. Develop an automated pipeline for uploading processed images and associated metadata to an object-based cloud storage.
5. Develop a web-based interface by using the Astro framework to enable the remote review of detections and carry out a detailed inspection at the leaf level.

2. Literature Review

The development of sensing technology, automation, and embedded computing has greatly increased the use of Internet of Things (IoT) technologies in the field of agriculture. These IoT-based systems typically combine wireless sensor networks, microcontrollers, and cloud services in order to monitor environmental factors such as soil moisture, temperature, and humidity, which in turn enables data-driven decision making and better field management [2, 7-9, 14]. They also have the potential to offer low-cost, distributed systems for continuous monitoring in precision agriculture.

At the same time, computer-vision techniques for assessing plant health and for detecting pests or diseases have become more popular. Deep learning models (for example, YOLO variants, MobileNet classifiers, and transformer-based detectors) have demonstrated promising results on manually curated datasets under controlled imaging conditions [12, 13, 15].

However, many of these datasets are based on manually obtained close-up images, which are considerably different from those taken in actual field situations. This domain mismatch causes a decrease in the model's ability to generalize and highlights the need for more realistic and consistently acquired training data [6, 16, 17].

Field-deployable monitoring systems also have the extra advantage of making use of edge computing and embedded imaging systems. For example, Raspberry Pi devices enable light-weight image acquisition and processing directly on board, thus reducing the demand for high-bandwidth connectivity [10, 18]. Yet mobile ground robots and UAVs have trouble getting stable and repeatable leaf-level images because of differences in viewpoint, canopy occlusions, and motion-induced blur [1, 5]. Because of these difficulties, their usefulness is limited when it comes to longitudinal monitoring tasks, which require consistent imaging geometry.

Even though fixed-path imaging systems, including gantries and greenhouse phenotyping platforms, have shown the value of having control over image acquisition, they are seldom used with open-field citrus crops. The mechanical and optical limitations caused by field conditions, canopy structure, and the way the plants are spaced are very different from those in greenhouse environments. As a result, there are still no dedicated imaging systems capable of obtaining high-resolution, repeatable, leaf-level images of citrus foliage under natural outdoor conditions.

Literature as a whole shows considerable advances in the fields of IoT sensing, embedded vision, and deep learning as they apply to agricultural monitoring. Yet current systems either concentrate on environmental sensing, make use of wide-area UAV or mobile coverage, or depend on manually collected datasets that lack consistency.

This situation highlights the necessity of having a field-deployable imaging system that is specially designed to produce structured and high-quality images of citrus leaves in a controlled and repeatable way. The rail-guided IoT imaging system proposed meets this need by offering a stable, low-cost and research-focused approach to continuous leaf-level monitoring in small citrus plots.

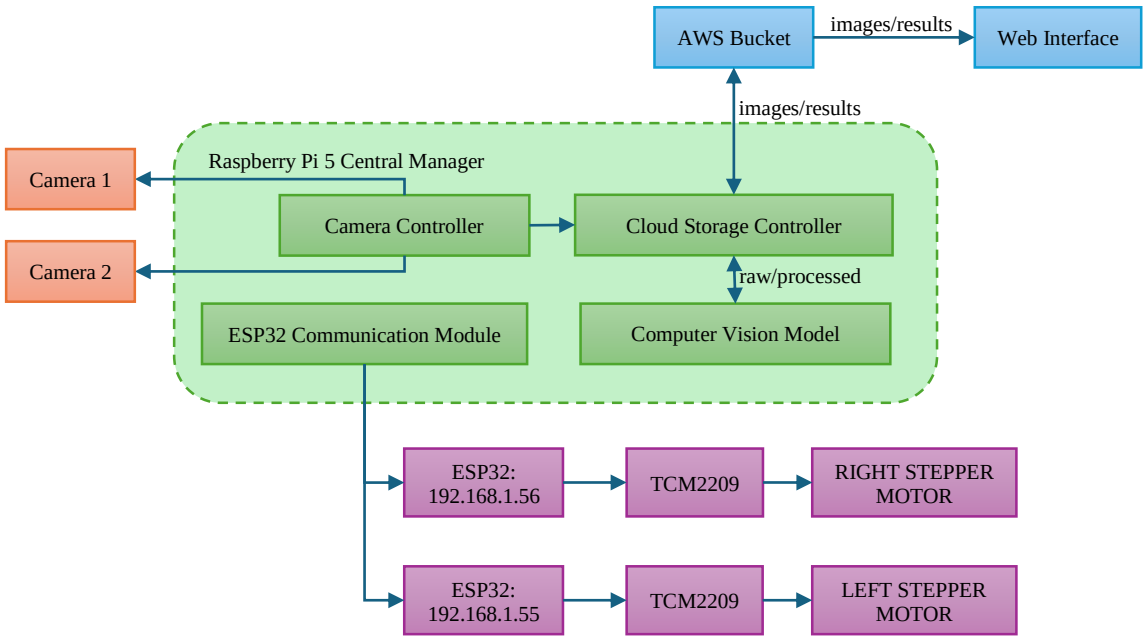


Fig. 1 Overview of the system architecture

2.1. Research Gap

Even though there has been considerable progress in the use of IoT for monitoring and in computer-vision techniques in agriculture, there are still a few gaps that need to be dealt with, particularly in the case of citrus crops.

To begin with, the datasets currently available for the purpose of detecting pests or diseases are mostly made up of manually obtained close-up images taken under controlled or laboratory conditions [16, 17].

These datasets do not match the problems associated with geometry, variations in lighting, and the way leaves are oriented that occur in actual citrus orchards. As a result, the existing models are trained on data that is unrealistic and does not reflect real field conditions, which in turn prevents them from generalizing and being used in practice.

Secondly, most IoT and embedded imaging systems are limited to environmental sensing or wide-area surveillance rather than carrying out controlled and repeatable image acquisition at the leaf level [2, 7, 11].

Although UAVs and ground robots offer mobility, their varying viewpoints and vulnerability to occlusions make them unsuitable as candidates for generating the consistent longitudinal datasets required for research or for early-stage monitoring studies [1, 5].

Greenhouse gantries and phenotyping platforms do provide controlled imaging, but their designs do not easily apply to open-field citrus crops, since the canopy structure, plant spacing, and exposure to the environment introduce additional constraints.

In conclusion, there is no specialized imaging hardware available for citrus leaves. Unlike other crops, citrus canopies have distinct characteristics in terms of leaf density, their vertical layering, and the way pests inhabit them, and this means that an imaging system is needed which can take high-resolution pictures of the lower and side surfaces of the leaves. Because there are no such dedicated systems, researchers are unable to investigate citrus-specific phenomena such as growth patterns, the way pests colonize, and the effects of microclimates.

The existence of these gaps shows that there is a need for an imaging system which is suitable for use in the field and which is both controlled and capable of being repeated, specifically one designed for citrus crops. The IoT imaging platform based on rails described in this study meets this need by allowing structured, high-resolution monitoring of leaves, a feature appropriate for research settings and small citrus farms.

3. Materials and Methods

3.1. General System Overview

The study describes the design of a low-cost monitoring system for citrus crops, which is specifically aimed at capturing the underside of the canopy since it is there that pests like whiteflies are found in the early stages. The system includes three main modules: an image acquisition procedure, the segmentation of whitefly wax, and web-based visualization (Figure 1).

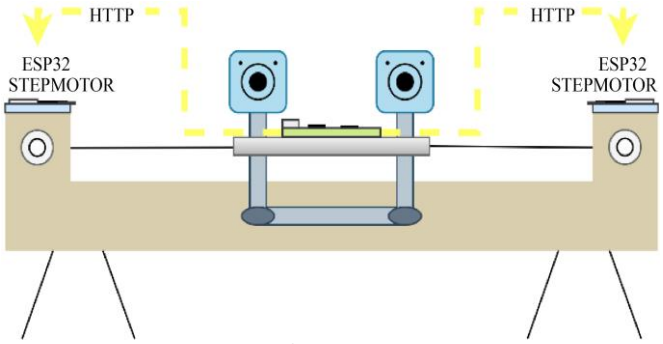


Fig. 2 Diagram of the mechanical structure

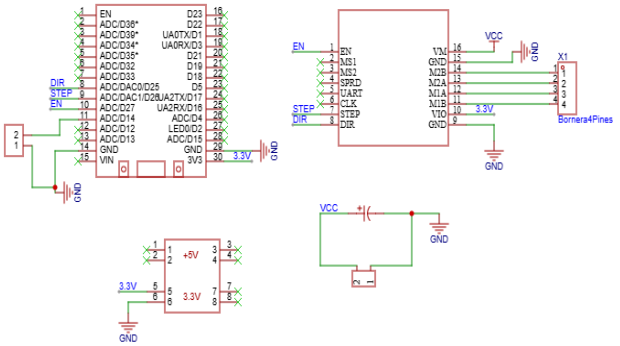


Fig. 3 Hardware diagram (ESP32, TMC2209, MB102)

3.2. Image Acquisition Workflow

The workflow for acquiring images makes use of a cable-driven mechanical system to enable linear movement of a trolley fitted with two leaf-level cameras. This trolley is controlled by a RPi5, which sends commands to two ESP32 microcontrollers. The system is built on a wooden track that is 75 cm long and is supported by two 50 cm legs, as illustrated in Figure 2.

Two motors, located at opposite ends of the track, are each controlled by a microcontroller which operates a web server to receive HTTP requests as control commands. Each node has a TMC2209 driver for the rotation of the stepper motor and a limit switch for homing; the circuit diagram corresponding to this is shown in Figure 3.

Figure 4 shows the sequence of the image acquisition workflow. The RPi5 first turns off "Motor B" and checks that "Motor A" is ready before starting the cycle, in order that "Motor A" may take on the main tension throughout the forward movement.

When the cart has not yet reached the limit switch, the RPi5 tells the ESP32 to move "Motor A" by a fixed step size. Once this movement has been carried out, the RPi5 takes images from both sides of the crop, carries out instance segmentation on them and uploads the original images together with the detection results to the cloud. The cycle then continues until the cart reaches the end of the rail, when a command is sent to return the cart to its homing position.

Images and metadata are stored using the directory format:

[date]/[hour]/[Track_ID]/[camera_orientation]/filename

These files are stored in an object-based cloud storage system that integrates with the web interface.

3.3. Computer Vision Pipeline for Whitefly Wax Segmentation

To be able to reliably monitor the early whitefly infestations on the abaxial side of citrus leaves, it is necessary to accurately define the irregular shape of the wax secretions produced by the whiteflies.

To detect these amorphous structures at the pixel level, an instance segmentation pipeline was developed, which allows for the precise localization of individual wax deposits under various imaging conditions. The computer vision framework that has been proposed consists of four main stages: dataset annotation, compositional data augmentation, model training, and deployment on the embedded edge-computing platform. Each of these stages is explained in the subsections that follow.

3.3.1. Dataset Annotation

High-quality pixel-level annotations are necessary if an instance segmentation model is to accurately depict the irregular morphology of the wax secretions of whiteflies. Since no publicly available dataset had adequately captured these features on citrus leaves, an annotation procedure was devised, making use of an internally acquired laboratory dataset originally focused on image acquisition and leaf classification methodology.

For this study, the laboratory image dataset was reused by producing instance-level segmentation annotations. Using Label Studio, polygon masks were created manually, each polygon accurately showing the visible boundaries of individual whitefly wax deposits. Representative examples of the resulting segmentation polygons are given in Figure 5. The images that had been manually annotated formed the initial ground-truth dataset for the purpose of developing the model.

A segmentation model was later trained and employed to produce polygon predictions for images which had not been labeled before. All the predicted masks were subjected to a thorough manual check, this involving the correction of inaccurate boundaries, the elimination of false-positive cases, and the verification that the annotations were complete. The corrected annotations were then added to the training dataset, and the model was retrained on the enlarged dataset. The process of annotation, training and verification was carried out repeatedly until the predicted masks consistently showed the complicated, cotton-like structure of the wax secretions of whiteflies.

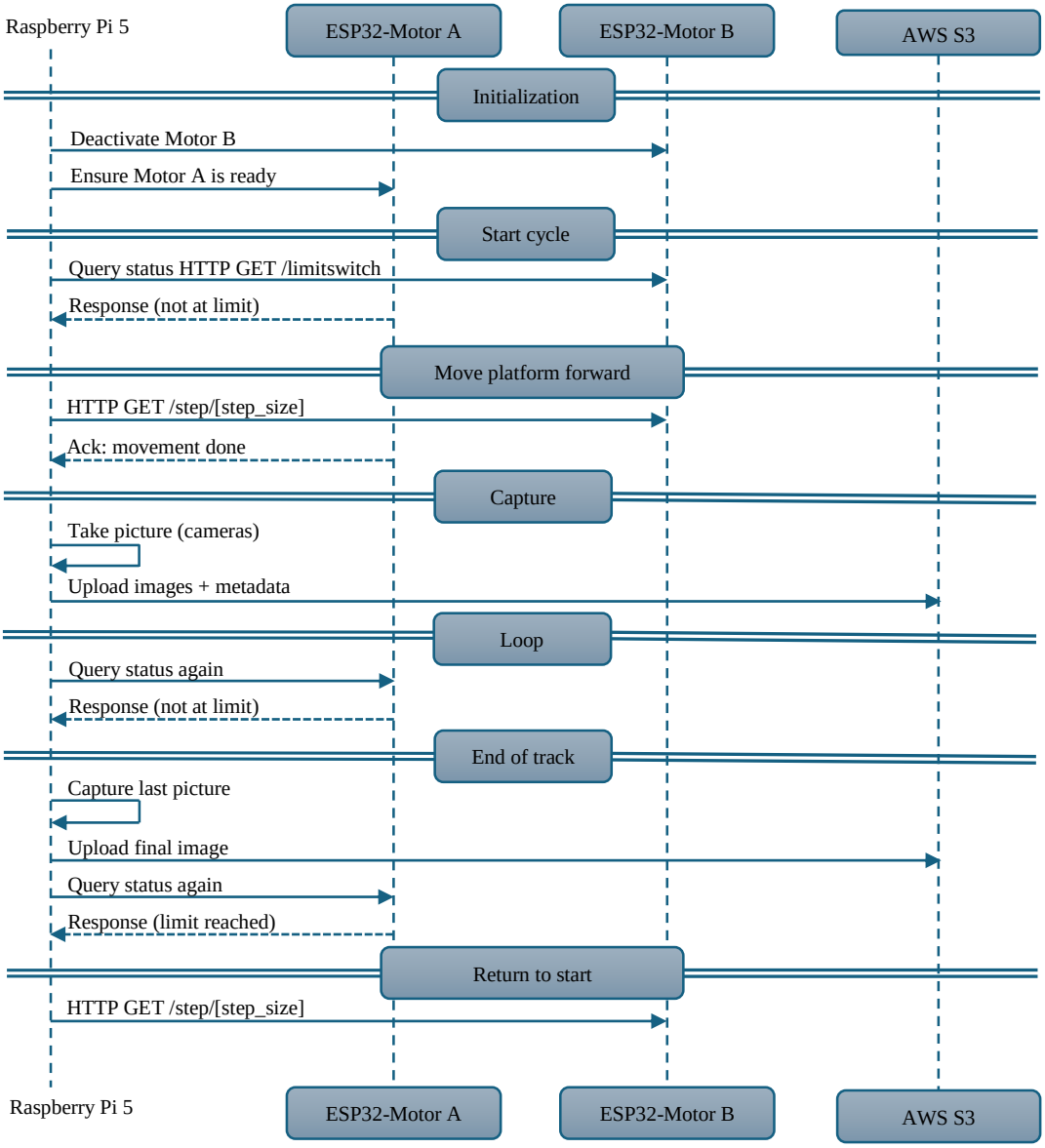


Fig. 4 Sequence diagram of the electronic control system

As a result of the iterative annotation process, a carefully selected dataset consisting of 344 images of citrus leaves was created, which included verified instance-level polygons. The use of this human-in-the-loop method greatly reduced the amount of manual annotation needed without compromising the level of annotation quality necessary for supervised instance segmentation.

3.3.2. Compositional Data Augmentation

Although the annotated lab dataset included accurate pixel-level segmentation masks, the controlled conditions under which the data were acquired meant that there was not much visual variability, which is necessary for reliable use in outdoor agricultural settings. To increase the diversity of the training data, a compositional data augmentation strategy was therefore created while keeping the accuracy of the manually annotated masks.



Fig. 5 Labeling examples

The citrus leaves that had been annotated were initially extracted from the laboratory images by means of the Segment Anything Model (SAM), which created accurate leaf masks while eliminating the uniform gray background. The extraction thus yielded individual leaf instances along with their respective segmentation annotations, which in turn enabled the subsequent image transformations without requiring any alteration of the ground-truth polygons.

To simulate the various environmental and optical conditions that occur in natural field situations, a special pipeline was created using the OpenCV and NumPy libraries to carry out random matrix and pixel-level operations directly on the individual leaf instances. The transformations were defined with the following specific limits:

- Geometric Rotation: Rotate the isolated leaf canvas continuously over a uniform distribution from 0° to 360°.
- Spatial Scaling: The area of the canvas was scaled dynamically such that the augmented leaf covered 40%-60% (scale factor 0.4-0.6) of the absolute target background patch dimensions.
- Contrast variations: Pixel intensities were scaled using a random multiplier between 0.7 and 1.3.
- Brightness shifting: Addition of a random offset between -30 and 30.
- Wax opacity modulation: The alpha blending of the whitefly wax overlays was dynamically varied (between 0.7 and 1.0) to accurately reflect how light passes through them in a real orchard environment.
- Gaussian blur: Random smoothing kernels with shapes 3×3, 5×5, or 7×7 and 5% activation probability.
- Ambient color correction: Applied Contrast Limited Adaptive Histogram Equalization (CLAHE) to the luminance with a blending factor of 0.6.
- Edge blending: To prevent the model from learning the high-frequency changes in edges, the outer edges of the polygon masks were smoothed during background compositing.

The realistic orchard backgrounds were acquired using the rail-guided monitoring platform. The 43 high-resolution field images were divided into patches each of 1024 × 1024 pixels, which produced 330 background sub-images. Those patches that had been confirmed to be free from whitefly infestations were chosen as the composition backgrounds. The augmented leaf examples were then placed onto these backgrounds to produce synthetic images that looked like citrus leaves in natural orchard settings.

The pipeline used for compositional augmentation generated 747 synthetic images, which greatly increased the visual variety of the training dataset without compromising the accuracy of the instance-level annotations. As a result, the dataset included a wide variety of background textures, different lighting conditions, and diverse leaf orientations, and this improved the model's capacity to generalize when faced with images taken in the field. An example of a synthetic image produced by this method is given in Figure 6.



Fig. 6 Image generation example

3.3.3. Dataset Partitioning

The dataset was divided in such a way as to maximize the diversity of the training data and at the same time, make it possible to carry out an unbiased evaluation of the segmentation model in real orchard conditions. Synthetic images were produced using the compositional augmentation pipeline, and the field images that were obtained naturally were included for the purpose intended.

Exactly 747 synthesized images were chosen as the main source of training data due to the high-quality instance annotations and the controlled method of image generation. The images provided a great variety in terms of leaf orientation, lighting, and background appearance, together with accurate polygon masks for supervised learning.

Alongside the synthetic dataset, 86 field image patches showing naturally occurring whitefly wax infestations were taken from images obtained by the rail-guided monitoring platform. To avoid data leakage, these images were split equally into two subsets. One of the subsets was added to the training dataset so that the model could become familiar with real outdoor imaging conditions, while the other subset was kept solely for the final field evaluation. As a result, no actual field images were used for testing during the development of the model.

The validation was carried out using only synthetic images to keep an eye on convergence as training progressed. The validation dataset included 74 composite images with annotated whitefly wax secretions and 14 background-only negative control images, making for a total of 88 validation samples. The validation metrics were intended to show the results of optimization on the synthetic compositional dataset, while the images held out from the dataset provided an

independent measure of the model's ability to generalize when used in real-world deployment situations.

The composition and split of the dataset are summarized in Table 1.

Table 1. Dataset composition

Dataset	Images	Purpose
Annotated laboratory	344	Initial annotations
Synthetic compositions	747	Model training
Synthetic validation	88	Training monitoring
Real field (training)	43	Domain adaptation
Real field (test)	43	Final field evaluation

3.3.4. Model Training Setup

The model was trained using the Ultralytics YOLO11s-seg framework since this framework is computationally efficient and can be used with embedded edge-computing devices. The network started with pre-trained weights and was then fine-tuned on the annotated dataset. A summary of the training setup is given in Table 2.

Table 2. Comprehensive hyperparameter for YOLO11s-seg

Parameter	Value
Model	YOLO11s-seg
Epoch Limit	100
Patience	15
Optimizer	AdamW
Base Learning Rate	0.002
Momentum	0.900
Weight Decay	0.0005

The training images were scaled to 640 by 640 pixels and then given to the network. The training process continued for up to 100 epochs using a batch size of 16. The choice of optimizer was made automatically by the Ultralytics framework, and an early stopping rule with a patience level of 15 epochs was put in place so that training would stop when the performance on the validation set ceased to improve, to reduce the risk of overfitting.

The training dataset consisted of the synthetic compositional images together with a selected subset of real field images; yet, validation of the model was carried out using only the synthetic validation dataset. The network was trained simultaneously for object localization, classification, and pixel-level segmentation, which enabled the detailed delineation of whitefly wax secretions under different imaging conditions.

To evaluate the pixel-level instance segmentation performance of the fine-tuned YOLO11s-seg model with mathematical rigor, the formulations for Mask Precision (P), Mask Recall (R), and mean Average Precision (mAP) are defined based on True Positive (TP), False Positive (FP), and False Negative (FN) pixel allocations relative to the bounding polygon mask intersection:

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

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

$$mAP = \frac{1}{C}\sum_{1}^{C}\int_{0}^{1} P_c(R)\, dR \quad (3)$$

The checkpoint of the model which showed the best results in validation was chosen for use in the rail-guided monitoring system and then put through experimental evaluation in real-life conditions.

3.3.5. Training Dataset Ablation Study

An ablation study was carried out to isolate and measure the effect of the proposed compositional data augmentation strategy. Four different training configurations for YOLO11s-seg were set up and assessed using the same frozen validation and field test sets:

1. Pure Synthetic: Trained exclusively on the 747 artificially generated compositional images.
2. Pure Field: Trained exclusively on the 43 real orchard training images.
3. Non-Compositional Hybrid: Trained on real field images combined with isolated pest leaves, but excluding the structural copy-paste augmentation and background blending pipeline.
4. Proposed Configuration: Trained on the complete dataset, integrating both the blended compositional augmentations and real field images.

This study systematically assesses the network's dependence on simulated background diversity for field generalization by keeping the testing data the same in all configurations.

3.3.6. Architectural Baseline Comparison

In order to provide a mathematical rationale for choosing pixel-level instance segmentation instead of traditional object detection, a comparative study of the two architectures was carried out.

The standard bounding-box detection model (YOLO11s) was trained on the same compositional dataset and with the same hyperparameter settings as the proposed segmentation

model (YOLO11s-seg). The two architectures were then assessed on the frozen field test set to measure the different performances associated with rectangular localization and exact morphological masking when dealing with amorphous biological structures.

3.3.7. Edge Deployment

The instance segmentation model that had been trained was put into use on the Raspberry Pi 5 embedded computer, which is part of the rail-guided monitoring platform. When the device was being used in the field, the images taken by the onboard camera were processed locally to identify and segment the whitefly wax secretions before they were stored and then visualized via the monitoring system.

The high-resolution images that were captured were first divided up into non-overlapping sections of 1024 × 1024 pixels and then passed on to the YOLO11s-seg inference pipeline. For each instance that was detected, the model produced a confidence score, a bounding box, and a pixel-level segmentation mask showing the spatial area of the whitefly wax secretion.

These outputs were then subjected to post-processing in order to create visualization images in which the predicted masks were overlaid on the original photographs, enabling a qualitative evaluation of the infestations that had been detected.

The original image and the associated segmentation result were sent to the cloud storage system as a result of the inference process. After that, the processed images could be accessed via the web-based monitoring application, which enabled users to examine the acquisition sessions from a distance and compare the original field images with the relevant segmentation outputs.

3.4. Web Visualization Interface

The web visualization interface allows users to monitor the images in the field from a distance and enables them to examine and review the data that has been collected. The web application is based on the Astro framework and uses Tailwind and React for its front-end design, as well as Clerk for user authentication.

Users can go through the different dates, tracks, camera orientations and time blocks to find images. The platform has two dedicated tools for examining images: one which highlights detections for the purpose of showing the model's outputs, and another which allows for a detailed zoom in on an individual leaf level.

The interaction between the user, the web interface, and cloud storage is illustrated in Figure 7, which describes the flow of user login, date and time selection, image requests, and data retrieval.

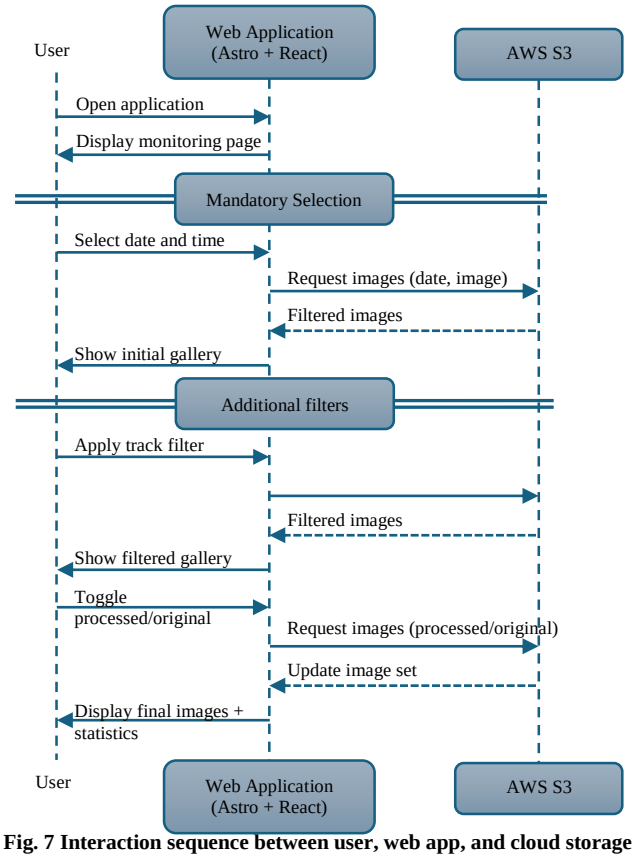


Fig. 7 Interaction sequence between user, web app, and cloud storage



Fig. 8 System installation in the crop field

4. Results

4.1. System Field Deployment

The system was validated by carrying out field experiments at the lemon orchard of the Santa Paula School in Arequipa, Peru. The tests were carried out over a period of five

days, from 8:00 AM to 12:00 PM. Since the orchard had poor signal connectivity, a Wi-Fi repeater was set up close to the rail to enhance communication between the RPi5 and ESP32 microcontrollers. The installation needed only a standard power outlet and involved only minor structural preparations, as illustrated in Figure 8.

4.2. Mechanical Performance

Initial testing of the cable-driven mechanism showed reliable linear movement over the entire length of the rail. The end-point limit switches homing sequence provided a reliable reference point for each operation cycle. However, the natural environment of the orchard presented elements that impacted the long-term persistence of the operation.

Certain trees had branches that stuck out, and these had to be cut by hand so that the cart could pass through without hindrance; furthermore, dust accumulated on the mechanical components, causing increased friction and, as a result, higher power consumption.

The rise in current demand produced a great deal of Electromagnetic Interference (EMI), which led to the limit switches being erroneously activated, thus interrupting the acquisition cycle. To assess the extent of these disturbances, a particular mechanical stress test was carried out, the results of which, as shown in Table 3, indicate that 85% of the cycles were prematurely terminated because of signal noise caused by EMI.

Table 3. Mechanical stress test evaluation

Operational Test Day	Successful Full Passes	Premature Terminations (EMI)	Failure Rate (%)
Day 1 (Baseline Clean)	18	2	10.0%
Day 2 (Dust Accumulation)	12	8	40.0%
Day 3 (High Friction / EMI)	5	15	75.0%
Day 4 (Post-Pruning / High Load)	4	16	80.0%
Day 5 (Optical & Mechanical Stress)	3	17	85.0%

4.3. Image Acquisition Workflow

A time-lapse of the images taken at 11:00 AM over five consecutive days at the fifth rail stop is shown in Figure 9; this sequence shows a high degree of visual stability and spatial consistency.

The degree of consistency of the images was determined by means of a comparison with the images from the first day, which were used as the baseline, employing Histogram Cosine Similarity (HCS).

The HCS is calculated between the baseline target vector (u) and the next daily observation vector (v) to mathematically quantify the long-term longitudinal consistency and viewpoint stability of images taken at similar stop intervals across different days:

$$HCS(u,v) = \frac{u.v}{\|u\|\|v\|} \quad (4)$$

Figure 10 shows that the three days following those shown had high similarity scores above 0.89, while the final day saw a sharp decrease in similarity to 0.4388. This drop corresponds directly to dust building up on the camera lens, a phenomenon which is also clearly visible in the image timeline.

An example of the system's output is given in Figure 11. During the entire period of the deployment, the obtained images always showed the following features: clear visibility of the leaf structure and surface texture, which is necessary for accurate feature extraction; sufficient high-quality detail to allow for future manual annotation and for expanding the dataset; natural variations in illumination caused by the changing sunlight during the morning; and small deviations in the camera's orientation between the different operational runs.



Fig. 9 Visual timeline of leaf alignment

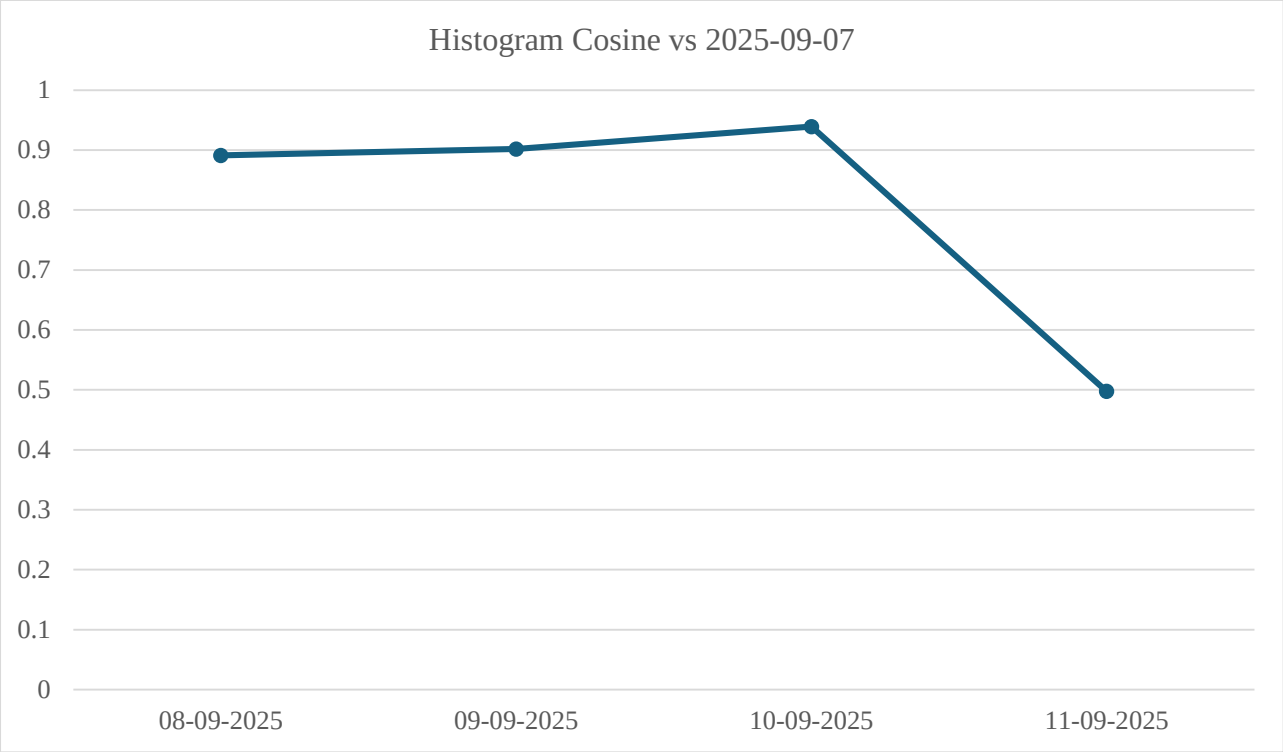


Fig. 10 Longitudinal stability analysis using HCS

4.4. Dataset Evaluation, Segmentation Performance, and Domain Gap Analysis

The segmentation performance of the fine-tuned YOLO11s-seg model was assessed by carrying out a quantitative analysis of the performance metrics in two different experimental settings: an internal validation dataset, consisting of synthetic leaf compositions (74 images with positive instances and 14 negative background control images) and an external field test dataset, which comprised raw, unconstrained images collected in-situ by the rail-guided IoT cart in the lemon orchard (43 images containing 126 natural pest instances).



Fig. 11 Image captured by the camera

The segmentation model showed good convergence on the internal validation set, with a bounding box mAP50 of 0.576 and an instance mask mAP50 of 0.575, as indicated in Table 4; Mask Precision (P) reached 0.588, together with a mask Recall (R) of 0.571 and a strict mask mAP50-95 of 0.294. These figures demonstrate that within a controlled compositional domain, the network had succeeded in mastering the complex textural boundaries and the varied spatial distributions of the whitefly wax deposits.

On the other hand, when the system was deployed on the actual, uncontrolled orchard field test dataset, a clear quantitative domain gap was observed. The field mask mAP50 dropped to 0.209, the box mAP50 fell to 0.172, and mask recall declined to 0.310. Yet, a qualitative visual assessment of the inference results in the field indicated that the model was able to successfully segment and localize most of the visible whitefly wax colonies, thus offering considerable value for use in agricultural monitoring in practice. The discrepancy between the numerical intersection metrics and the qualitative success in the field is due to the model's attempt to track the amorphous boundary geometry of the target pest residues, as shown in Figure 12.

Above all, the model achieved a high level of field mask precision, amounting to 0.494. In actual Integrated Pest Management (IPM) procedures, precision is given greater importance than absolute recall since it ensures that the automated messages sent to farm managers relate to confirmed infestations and thus prevents the use of chemicals based on false-positive alerts.

Table 4. Quantitative performance metrics across evaluation domains

Domain	Dataset Profile	Images	Instances	Mask (P)	Mask (R)	Mask (mAP50)	Mask (mAP50-95)
Internal Validation	Synthetic leaves	88	340	0.588	0.571	0.575	0.294
Field Testing	Rail captured	43	126	0.494	0.310	0.209	0.0823

Table 5. Ablation study of dataset configurations evaluated on the field test set

Training Configuration	Validation Mask (mAP50)	Test Mask (P)	Test Mask (R)	Test Mask (mAP50)
1. Pure Synthetic	0.528	0.111	0.0079	0.0007
2. Pure Field	0.0918	0.172	0.0794	0.0205
3. Non-Compositional Hybrid	0.410	0.347	0.262	0.165
4. Proposed Configuration	0.575	0.494	0.310	0.209

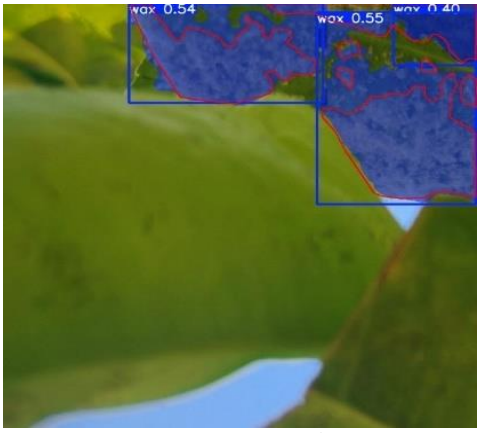


Fig. 12 Qualitative segmentation outputs on field images, with red being the ground truth and blue being the predictions

4.5. Training Dataset Ablation Study Results

To empirically establish the necessity of the compositional data augmentation pipeline, the performance of the four separate training configurations was compared using the same frozen field test set. The quantitative results of this ablation study are given in Table 5.

The ablation results clearly show the serious "Realism Gap" that exists in agricultural computer vision. In the case of Configuration 1 (Pure Synthetic), convergence was good in the internal validation (mAP50 = 0.528), but there was a complete collapse when the model was tested under uncontrolled field lighting (mAP50 = 0.0007), which proves that models trained solely on clean laboratory backgrounds are unable to generalize to real orchard environments. On the other hand, Configuration 2 (Pure Field) failed to converge (mAP50 = 0.0205) because there were not enough positive pest examples needed to train a deep segmentation model from scratch. The results from field testing indicated that Configuration 3 (Non-Compositional Hybrid) achieved a partial recovery (mAP50 = 0.165) of the network when it was presented with real orchard backgrounds. Yet this configuration still experienced a considerable precision penalty (P = 0.347).

This demonstrates that it is not enough merely to combine real field images with non-augmented, individual pest crops; the network will always tend to overfit the artificial edges and the uncorrected luminosity of the pasted leaves. It was only with the proposed configuration that the domain gap was successfully addressed, resulting in a maximum field precision of P = 0.494, which is appropriate for use in automated pest monitoring.

4.6. Web Interface Performance

To assess the computational overhead involved in deploying instance segmentation on an edge device, the proposed YOLO11s-seg model was compared with the standard YOLO11s bounding-box object detection architecture. The two models were trained under the same conditions and tested on the unconstrained orchard test set. The comparative metrics are given in Table 6.

The mathematical results support a move to pixel-level segmentation. Although the conventional bounding-box detector (YOLO11s) did achieve basic convergence, it performed very poorly in real-world situations, obtaining a mAP50 of just 0.0826. Since the wax secretions of whiteflies have an amorphous structure and lack any rigid geometric shapes, the traditional rectangular bounding boxes include a

large amount of healthy, green leaf tissue. This inclusion of background material disrupts the network's ability to extract features and therefore causes both lower recall and lower precision. The new architecture, which uses instance segmentation (YOLO11s-seg), separates out the exact boundaries of the biological deposits and, as a result, achieves more than a twofold increase in mAP50 (0.209) together with a robust field precision (P = 0.494) that is suitable for reliable IPM monitoring.

Table 6. Comparative performance of object detection vs. instance segmentation on the field test set

Architecture Type	Model	Test (P)	Test (R)	Test (mAP50)
Bounding Box Detection	YOLO11s	0.408	0.159	0.0826
Instance Segmentation	YOLO11s-seg	0.494	0.310	0.209

4.7. Web Interface Performance

The custom web visualization interface accurately reflected the state of the S3 bucket, giving a correct display of the dataset that had been collected. Figure 13 illustrates an example of the image navigation view. It was able to retrieve the folder structures from S3, show the thumbnails when required, offer smooth sequential navigation, and allow for quick examination of the image sequences. The discontinuities that were visible in the navigation timeline matched precisely those network dropout areas that had been seen during acquisition. This proves that the system conforms to its non-buffering architecture, with the web interface acting as a direct, real-time view of the dataset that has been successfully preserved without hiding the transmission losses.

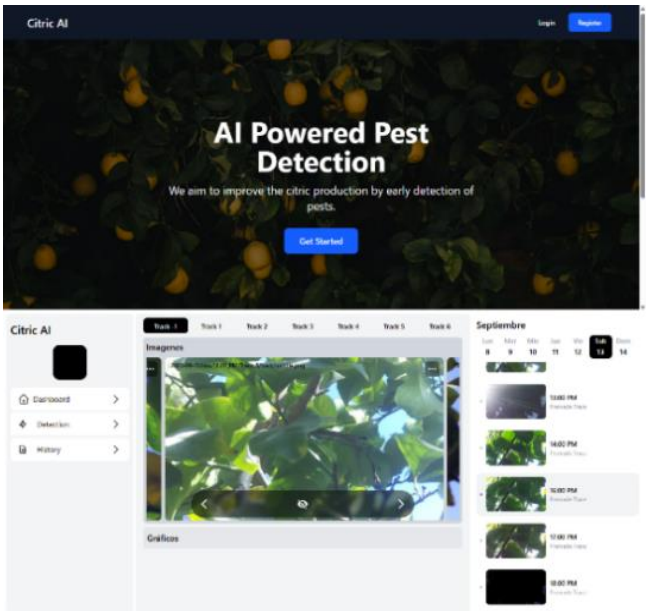


Fig. 13 Web interface image navigation view

5. Discussion

5.1. Comparative Performance Analysis

When compared to recent state-of-the-art agricultural monitoring systems, the proposed platform demonstrates a distinct advantage in maintaining consistent imaging geometry under unconstrained field conditions. For instance, while Zhang et al. [17] achieved high accuracy in detecting aphid clusters in the wild, their approach relied on manually acquired images, which lack the spatial reproducibility (HCS > 0.89) required for continuous longitudinal monitoring. Similarly, edge-computing platforms like AICropCAM [10] and enhanced YOLOv8 models [15] present robust theoretical mean Average Precision scores for crop monitoring and pest detection, but transferring these models to free-navigating field robots frequently exposes them to perspective shifts and motion-induced blur caused by unstructured agricultural terrain and platform vibration [1, 5].

The competitive field precision (P = 0.494) reported in this study was achieved specifically because the deterministic, cable-driven ESP32 rail system mechanically eliminates these navigational disturbances. This stability is precisely what allows the lightweight YOLO11s-seg model to accurately outline irregular wax secretions directly on the edge, overcoming the physical imaging limitations commonly reported in mobile agricultural setups.

5.2. Model Performance and the Realism Gap

A major achievement in this research was the successful acquisition and carrying out of pixel-level segmentation of actual, real whitefly (Aleurothrixus floccosus) infestations using the specially designed rail equipment. The shift from bounding box object detection to instance segmentation with YOLO11s-seg was crucial for separating the amorphous wax secretions.

Although the quantitative measures showed a gap between validation using synthetic data (mAP50 = 0.575) and testing in uncontrolled orchard fields (mAP50 = 0.209), failure mode analysis showed that this difference is mainly due to the strict Intersection over Union (IoU) penalties applied to the fuzzy biological boundaries, not because the localization failed.

The system can correctly identify the presence of the pest without producing many false alarms, thus providing an important operational safety measure for Integrated Pest Management (IPM) workflows, with a precision level of P = 0.494.

This method was checked by means of a comparative study based on architectural design. While rectangular bounding-box architecture represents the current state of the art for most agricultural detection tasks, it was not suitable for this specific agricultural application, since it resulted in a severely reduced field mAP50 of just 0.0826 and a precision value of 0.408.

Since rigid rectangular boxes are naturally driven to include a large amount of healthy leaf tissue together with the irregular, cotton-like whitefly wax, the presence of this background material interferes with the network's ability to extract features and causes a higher number of false positives under complex lighting conditions. In the case of YOLO11s-seg, however, the network was able to disregard the surrounding leaf textures and identify the exact contours of the biological deposits, thus more than doubling the mAP50 and directly justifying the computational cost associated with edge-based segmentation.

Furthermore, the ablation study shows that the compositional data augmentation pipeline is essential. When models were trained only on synthetic laboratory images, they suffered a severe domain shift upon evaluation in unconstrained field conditions (mAP50 = 0.0007), whereas training purely on field data failed to converge because of the extreme underrepresentation of certain classes. Crucially, the non-compositional hybrid method, which involves just placing individual pest crops onto orchard backgrounds without any structural blending, demonstrated a high level of false positives (P = 0.347). This indicates that without photometric correction and boundary smoothing, the deep segmentation models overfit on the synthetic paste edges and luminance differences instead of the real morphology of the whitefly wax. By incorporating structured edge-blurring and color correction, the proposed pipeline causes the model to disregard background artifacts and concentrate on the genuine biological features, which in turn allows the field accuracy to be restored to P = 0.494.

5.3. Mechanical Stability and Environmental Attenuation

Unlike ground rovers that struggle with rough terrain or UAVs blocked by dense canopies, the proposed rail-guided platform restricts image capture to a fixed, deterministic linear track. By eliminating the unpredictable vibrations and perspective shifts inherent to free-navigating systems, the platform maintains a constant working distance and upward viewing angle relative to the abaxial leaf surface. This controlled mechanical trajectory explains how and why the system achieves the High Spatial Consistency (HCS > 0.89) necessary for repeatable, long-term monitoring at the leaf level. Although there were some variations in camera angle and occasional motion blur when the cable tension changed, such effects are normal in low-cost cable-driven designs and could be dealt with in future versions through better tension control or by using encoder-based feedback.

Tests carried out over a period on the equipment to assess its susceptibility to the build-up of ambient particles showed that the similarity scores had dropped significantly to HCS = 0.4388 by the fifth day of operation. This decrease is due to dust accumulating on the optical lens arrays and not because of any deviations in the mechanical path, which therefore shows that it will be necessary to have automated lens cleaning mechanisms or protective covers when the system is next deployed.

5.4. Electromagnetic Interference (EMI) and Hardware Mitigation

Over the course of extended longitudinal testing, the proportion of premature mechanical terminations caused by the embedded control architecture increased gradually, rising from 10.0% on Day 1 to 85.0% by Day 5. Although optical degradation was caused by dust in the surrounding environment, the physical system stoppages were due to Electromagnetic Interference (EMI) in the motor control circuits. In uncontrolled field conditions, long and unshielded cabling functions as a parasitic antenna, picking up the high-frequency switching noise from the Pulse-Width Modulated (PWM) motor drivers and the inductive voltage spikes produced by the DC motor brushes and coupling it into the low-voltage sensor lines.

To decouple mechanical actuation from electrical instability and to gain industrial-grade reliability, the next versions require three specific mitigations at the hardware level:

1. Optical Limit Switches: Standard mechanical limit switches need to be replaced by optical limit switches (photo interrupters). Optical end-stops eliminate mechanical contact bounce and offer physical immunity to the inductive voltage spikes that are often coupled into microcontroller logic rails by mechanical switches.
2. Conducted Emissions and Power Isolation: Localized decoupling capacitors at the motor terminals themselves, plus inline ferrite beads on the power supply rails, will suppress high-frequency commutator arcing and transient logic sags.
3. Signal Routing and Optoisolation: All sensor lines will be routed to Shielded Twisted-Pair (STP) cabling with a single point earth ground. Hardware optocouplers (e.g., PC817) will also be used on all digital sensor inputs to physically isolate the low-voltage ESP32 logic circuitry from the high-voltage motor control domains, filtering out transient noise to prevent erroneous system halts.

5.5. Embedded Latency and Data Retention

The performance test built into the RPi5 offered strong evidence showing that in-field edge computing is feasible. The central processing unit took about eight seconds to carry out YOLO11s-seg instance segmentation inference on each 1024 × 1024 pixel image patch. The entire operational cycle, including the acquisition of images from multiple angles,

slicing the images, carrying out the local inference, and packaging the cloud metadata, took an average of thirty-one seconds per cycle.

Since the wireless upload stage is what mostly determines the cycle latency, the local weakening of the Wi-Fi signal through dense canopy foliage still acts as the main obstacle to communication. In the present implementation, because there is no local storage of images, the continuity of the sequence is completely reliant on continuous real-time wireless connectivity. To avoid missing any frames when carrying out extended deployments, future modifications to the architecture should include temporary onboard caching or a store-and-forward synchronization approach, thus ensuring reliable retention of the data while at the same time sending the lightweight classification metadata directly to the remote dashboard.

6. Conclusion

The study showed that it was technically feasible to use a rail-guided IoT imaging system for automated and repeatable monitoring of leaves on citrus trees on a per-leaf basis. By having the Raspberry Pi 5 control the movement of the sensor along a fixed linear path with the aid of two ESP32 motor controllers, the physical setup was able to avoid the problems of perspective variation and canopy obstruction, which are usual with mobile ground and aerial robots, thus achieving a high level of spatial reproducibility (HCS > 0.89) during the first few observations.

Moreover, the computer vision workflow confirmed the effectiveness of a compositional data augmentation pipeline that used the Segment Anything Model (SAM) to adapt images from laboratory settings to unconstrained field conditions. Although testing in the field revealed a numerical gap in the case of strict boundary intersections (mAP50 = 0.209) because of the complicated lighting in the orchards and the irregular shape of the pests, the edge-optimized YOLO11s-seg architecture managed to achieve a robust field precision of P = 0.494. The findings show that the proposed workflow offers reliable positive detections for use in remote monitoring applications, while at the same time indicating that segmentation accuracy is still affected by complex illumination conditions, irregular pest morphology, and variations in the natural background.

The main contribution of this study does not lie in the presentation of a new segmentation algorithm, but rather in the development and real-world testing of an integrated embedded imaging system that combines deterministic image acquisition, IoT-based data management, and edge artificial intelligence into a single monitoring platform. The proposed system, by minimizing variability in image acquisition and allowing for standardized data collection under operational conditions, offers a reproducible experimental platform for the development and evaluation of future computer vision methods in the field of precision agriculture.

There are several limitations that need to be recognized. The first is that the platform was tested on a rather small-scale citrus setup and in only a limited variety of environmental conditions. The second point is that the decline in segmentation performance under field conditions shows that further work is necessary to enhance domain generalization between the synthetic datasets obtained in the laboratory and those from natural orchards. Lastly, this study only included an evaluation of imaging performance, the implementation on embedded systems, and the ability to detect whiteflies, and did not directly examine wider agronomic results such as reductions in pesticide use or improvements in Integrated Pest Management (IPM) decision-making.

Future work will involve making step-by-step improvements to the hardware platform and the computer vision pipeline. In the short term, the research will concentrate on methods for tackling the domain gap by using expanded field datasets, improving the augmentation techniques, and developing better segmentation models, with performance being assessed by means of standard quantitative metrics. In the medium term, the hardware platform will be improved by enhancing electromagnetic compatibility, increasing mechanical robustness, adding automated lens-cleaning mechanisms, and improving communication reliability when operating outdoors for long periods. In the long term, these developments will enable extended multi-season deployments over larger orchards and among a variety of crop species to examine the long-term operational reliability, scalability, and integration into broader precision agriculture and Integrated Pest Management workflows.

Funding Statement

This research was funded by Universidad Nacional de San Agustín de Arequipa (UNSA) covered by project “Desarrollo e Implementación de un Sistema Avanzado de Detección Temprana de plagas para la Protección de Cultivos de Cítricos”, contract PI-01-2024-UNSA.

Acknowledgments

The authors wish to express their thanks to the Professional School of Electronic Engineering at the Universidad Nacional de San Agustín de Arequipa for having made available to them the laboratories needed for the development of the project.

Data Availability Statement

The full dataset, the source code for the web application, the microcontroller firmware, and the trained model checkpoints that support the findings of this study are available for reproducibility in a public GitHub repository at: https://github.com/Jusepepe/citrus-monitoring.

References

[1] Andrea Botta et al., "A Review of Robots, Perception, and Tasks in Precision Agriculture," *Applied Mechanics*, vol. 3, no. 3, pp. 1-25, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Vu Khanh Quy et al., "IoT-Enabled Smart Agriculture: Architecture, Applications, and Challenges," *Applied Sciences*, vol. 12, no. 7, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Showkat Ahmad Bhat, and Nen-Fu Huang, "Big Data and AI Revolution in Precision Agriculture: Survey and Challenges," *IEEE Access*, vol. 9, pp. 110209-1102222, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Oscar E. Liburd, and Teresia W. Nyoike, "Biology and Management of Aphids in Sustainable Field Production of Cucurbits," *Electronic Data Information Source*, vol. 2008, no. 2, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[5] Nadia Delavarpour et al., "A Technical Study on UAV Characteristics for Precision Agriculture Applications and Associated Practical Challenges," *Remote Sensing*, vol. 13, no. 6, pp. 1-25, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yuzhen Lu, and Sierra Young, "A Survey of Public Datasets for Computer Vision Tasks in Precision Agriculture," *Computers and Electronics in Agriculture*, vol. 178, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] M. Dhanaraju et al., "Smart Farming: Internet of Things (IoT)-based Sustainable Agriculture," *Agriculture*, vol. 12, no. 10, pp. 1-26, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Muhammad Shoaib Farooq et al., "Role of IoT Technology in Agriculture: A Systematic Literature Review," *Electronics*, vol. 9, no. 2, pp. 1-41, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Marwa Abdulla, and Ali Marhoon, "Deep Learning and IoT for Monitoring Tomato Plant," *Iraqi Journal for Electrical and Electronic Engineering*, vol. 19, no. 1, pp. 70-78, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Nipuna Chamara, Geng Bai, and Yufeng Ge, "AICropCAM: Deploying Classification, Segmentation, Detection, and Counting Deep-Learning Models for Crop Monitoring on the Edge," *Computers and Electronics in Agriculture*, vol. 215, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Gianfranco Gagliardi et al., "An Internet of Things Solution for Smart Agriculture," *Agronomy*, vol. 11, no. 11, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Jun Liu, and Xuewei Wang, "Plant Diseases and Pests Detection based on Deep Learning: A Review," *Plant Methods*, vol. 17, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Nazmun Nessa Moon, "IFSG: Intelligence Agriculture Crop-Pest Detection System using IoT Automation System," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 2, pp. 1091-1099, 2021. [Google Scholar]

[14] Younsuk Dong et al., "Implementation of an in-Field IoT System for Precision Irrigation Management," *Frontiers in Water*, vol. 6, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Bruno Pinheiro de Melo Lima et al., "A Lightweight and Enhanced Model for Detecting the Neotropical Brown Stink Bug, Euschistus Heros (Hemiptera: Pentatomidae) based on YOLOv8 for Soybean Fields," *Ecological Informatics*, vol. 80, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Jorge Parraga-Alava et al., "Lelephid: An Image Dataset for Aphid Detection and Infestation Severity on Lemon Leaves," *Data*, vol. 6, no. 5, pp. 1-7, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Tianxiao Zhang et al., "Aphid Cluster Recognition and Detection in the Wild Using Deep Learning Models," *Scientific Reports*, vol. 13, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Michael Kölling, "Educational Programming on the Raspberry Pi," *Electronics*, vol. 5, no. 3, pp. 1-17, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[19] Glenn Jocher, Ultralytics YOLO11, version 11.0.0, Ultralytics, 2024. [Online]. Available: https://github.com/ultralytics/ultralytics