Cost-Effective Smartphone-Based Computer Vision Pipeline for Actinidia Phenological Stage Mapping

Isabel Pinheiro

isabel.a.pinheiro@inesctec.pt

Institute for Systems and Computer Engineering of Porto https://orcid.org/0000-0001-5387-0980

Mário Cunha

University of Porto

António Valente

Universidade de Trás-os-Montes e Alto Douro

Filipe Neves dos Santos

University of Porto

Research Article

Keywords: Artificial pollination, BBCH scale, Deep Learning, GNSS, Kiwifruit, Phenological monitoring

Posted Date: September 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10934351/v1

Additional Declarations: The authors declare no competing interests.

Cost-Effective Smartphone-Based Computer Vision Pipeline for *Actinidia* Phenological Stage Mapping

Isabel Pinheiro[1,2,*], António Valente[1,2], Mário Cunha[1,3], Filipe Neves dos Santos[1,4]

[1]Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Porto 4200-465, Portugal

[2]School of Science and Technology, University of Trás-os-Montes e Alto Douro (UTAD), Vila Real 5000-801, Portugal

[3]Faculty of Sciences, University of Porto (FCUP), Porto 4169-007, Portugal

[4]Faculty of Engineering, University of Porto (FEUP), Porto 4200-465, Portugal

[*]Corresponding author: isabel.a.pinheiro@inesctec.pt

Preprint – September 4, 2026

Abstract

Dioecious crops face significant pollination challenges due to the asynchrony in flowering between male and female plants. This asynchrony varies spatially across orchards, requiring targeted interventions in zones where synchrony is lacking. Assisted pollination addresses this deficiency, albeit at a substantial operational cost that could be optimised through spatial phenological mapping. Manual assessment proves economically infeasible at commercial scales, while existing computer vision systems are unable to classify phenological stages and integrate geospatial information. This study presents the Mobile Phenological Mapping (MPM) Framework, which integrates GNSS-synchronised smartphone video with automated phenological detection to generate plant-level phenological distribution maps validated in commercial kiwifruit (Actinidia chinensis) orchards. This modular framework comprises (i) training and validation data acquisition, (ii) model optimisation, (iii) operational pipeline, and (iv) performance evaluation. MPM employs hierarchical deep learning across three stages: structure detection, gender classification, and phenological stage classification. Video frames are georeferenced through timestamp matching with GNSS metadata, enabling spatial phenological mapping. Operational validation across four commercial orchard zones demonstrated mean absolute percentage errors of 17.2% for structure detection and 20.1% for gender classification. The framework reduces monitoring time from 113 to 1.6 hours per hectare, decreasing labour costs from €2 060 (113 hours × 18.20 € per hour) to €29 (1.6 hours × 18.20 € per hour) per hectare based on the Portuguese hourly labour cost for minimum wage workers. When integrated with routine orchard operations, video acquisition incurs negligible additional cost. MPM provides growers with precision phenological maps for targeted pollination interventions. While validated in a kiwifruit orchard, the modular architecture can be adapted to other crops by replacing the training data.

Keywords: Artificial pollination, BBCH scale, Deep Learning, GNSS, Kiwifruit, Phenological monitoring

1 Introduction

Pollination is a crucial interaction that drives the reproduction of most flowering plants and is directly associated with global food security and ecosystem balance [Potts et al., 2016]. Approximately 75% of the world's food crops rely on pollinators, underscoring the importance of

pollination to agricultural sustainability [FAO, 2022]. In dioecious systems, in which male and female flowers are on separate plants, effective pollination depends on temporal synchronisation between male pollen release and female stigma receptivity. The long-established phenological patterns have become increasingly disrupted by climate change, resulting in reduced natural pollination services and requiring greater monitoring and management measures [S. Jha, L. Burkle, and C. Kremen, 2013].

Actinidia chinensis (kiwifruit) exemplifies the difficulties in managing pollination for dioecious fruit crops. Effective pollen transfer from male anthers to female stigmas on separate plants is crucial, as the number of fertilised seeds directly affects fruit size, shape, and quality [Spinelli et al., 2003, Castro et al., 2021, David et al., 2022]. Assisted pollination addresses insufficient natural pollination in dioecious crops but incurs high operational costs [Williams et al., 2020, Li et al., 2022b, Pinheiro et al., 2026b], requiring substantial labour investment over a short period that scales with orchard size. Systematic phenological monitoring provides essential data for creating effective pollination management strategies [Denny et al., 2014, Nordt et al., 2021, Richardson et al., 2013]. Identifying the timing differences between male and female flowering helps target assisted pollination to support insufficient natural pollination [Castro et al., 2021, David et al., 2022]. Improving these methods through precise phenological assessment could significantly reduce production costs while maintaining high-quality yields.

As climate change increasingly disrupts established phenological patterns and natural pollination services [S. Jha, L. Burkle, and C. Kremen, 2013], cost-effective monitoring technologies that enable precision management interventions become critical for maintaining production stability. However, current monitoring approaches face fundamental operational and technical limitations that prevent their deployment at commercial scales.

Manual plant assessment relies on labour-intensive counting and classification of flowering stages on sample plants by trained staff walking through orchard zones. Labour costs prevent comprehensive spatial coverage within the short pollination window, limiting sampling to small areas that fail to capture the spatial heterogeneity necessary for effective pollination management decisions. Existing computer vision systems address flower detection and localisation [Lim et al., 2020, Li et al., 2022b, Zhou et al., 2023] but lack integration with phenological stage classification and geospatial information, both of which are necessary for precision pollination management. Standardised frameworks, such as the BBCH (Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie) scale, provide consistent systems for describing developmental stages across species [Koch et al., 2007], including cultivar-specific adaptations in the phenology of *Actinidia chinensis* [Salinero et al., 2009]. For *Actinidia chinensis*, the morphological definitions of the key reproductive BBCH stages, including inflorescence emergence, flowering, and early fruit set, are detailed in [Pinheiro et al., 2026a], which provides photographic examples and stage consolidation for field annotation.

Although the BBCH classification system enables harmonised description of phenological events, modern monitoring increasingly relies on image-based observations that generate large volumes of complex visual data, requiring automated analysis methods for practical deployment [Correia et al., 2020]. Advances in Computer Vision (CV) and Deep Learning (DL) have substantially transformed phenological monitoring by enabling automated extraction of phenological stages across diverse image sources [Katal et al., 2022], including flowering detection [Kim et al., 2021], UAV-based phenology estimation [Yang et al., 2020], orchard-scale apple flower phenology mapping [Wang et al., 2021], and large-scale tree species classification using temporal phenological cues [Pearse et al., 2021, Wagner, 2021].

As a complement to automated phenological classification, integrating ground-based imagery with geospatial data has become a robust approach for spatial crop monitoring systems. Early agricultural research demonstrated that tractor-mounted video systems, combined with high-precision Global Navigation Satellite System (GNSS), could generate precise row-level maps [Downey et al., 2004, Sainz-Costa et al., 2011]. Building on this, geospatial video and imag-

ing aligned with GNSS coordinates have gained traction through dedicated spatiotemporal data models and indexing frameworks [Mills et al., 2010]. Recent developments have expanded into phenological monitoring: satellite and UAV-based products provide canopy-level temporal metrics [Zeng et al., 2020, Yang et al., 2020]. In parallel, georeferenced ground-level imagery has been used to describe crop development patterns, track flowering progression, and validate satellite-derived phenological timing, supporting operational crop mapping initiatives [d'Andrimont et al., 2022, Tsutsumida and Funada, 2023, Jiang et al., 2024, Liu et al., 2024]. Collectively, these studies suggest that coupling ground imagery with GNSS provides a practical and scalable method for capturing fine-resolution spatial heterogeneity in crop phenology, complementing UAV and satellite platforms.

In *Actinidia chinensis*, CV research has focused primarily on the detection and localisation of reproductive structures to support artificial or robotic pollination. Early work by Lim et al. [2020] demonstrated real-time YOLO-based flower detection under orchard conditions. In contrast, Li et al. [2022b] introduced a series of improvements for simultaneous bud–flower detection using YOLOv4, multi-class mapping of floral structures and their spatial distribution [Li et al., 2022a], and the estimation of individual flower poses for robotic manipulation [Li et al., 2023]. Additional advances include improved YOLO architectures for more efficient flower detection [Zhou et al., 2023] and automated flower gender classification [Pinheiro et al., 2024]. These systems demonstrate mature detection capabilities for individual reproductive structures. Although these approaches achieve strong detection performance, they do not incorporate BBCH-resolved phenological stages or integrate GNSS-synchronised imagery for orchard-scale mapping. This represents a critical gap, particularly in dioecious species such as *Actinidia chinensis*, where spatial variation in flowering progression strongly influences pollination timing and management decisions.

Despite progress in individual technologies, these advances remain unintegrated for operational orchard applications. Existing *Actinidia chinensis* CV systems achieve flower detection and gender classification but lack phenological stage discrimination, geospatial integration, and mobile accessibility for operational deployment. This limits the generation of spatial phenological maps, which are required for precision pollination management. Figure 1 illustrates the required conceptual integrated system, in which farmers capture georeferenced mobile videos during routine operations for automated processing into spatial phenological maps. Currently, no operational framework combines mobile accessibility, automated phenological stage detection, and geospatial mapping into a unified workflow. Such a system would eliminate the need for specialised equipment while providing actionable spatial data for targeted pollination interventions.

This study addresses this gap by testing three core hypotheses:

- H1: Spatial phenological mapping significantly improves the identification of pollination-relevant asynchronies compared to point sampling methods.
- H2: Hierarchical deep learning architectures reduce error propagation in multi-class phenological classification compared to single-stage approaches.
- H3: Mobile video acquisition integrated with GNSS provides sufficient spatial resolution for operational precision pollination management at commercial scales.

These hypotheses test whether accessible smartphone technology can bridge the gap between flower detection and spatially explicit phenological mapping for precision pollination management.

The main goal of this paper is to develop and validate a Mobile Phenological Mapping (MPM) Framework that enables large-scale, automated phenological monitoring of dioecious crops, with a focus on *Actinidia chinensis*. The specific goals are:

- G1: To acquire and structure training and validation data for automated phenological monitoring using the Multi-Modal *Actinidia chinensis* Phenology Dataset [Pinheiro, 2025,

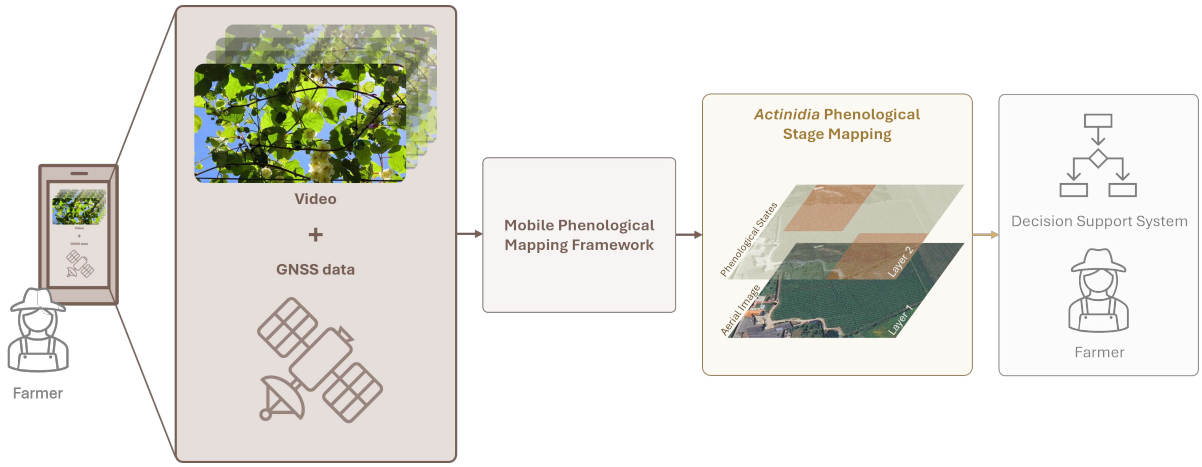

Figure 1: Conceptual framework for integrated phenological mapping showing system inputs (georeferenced mobile video with GNSS metadata), processing pipeline (computer vision-based phenological detection), and outputs (spatial density maps of phenological stages). Generated maps enable direct interpretation by farmers or integration with agricultural decision support systems for precision pollination management.

Pinheiro et al., 2026a].

G2: To optimise machine-learning models for detecting and staging flower phenological development within the MPM Framework.

G3: To implement an operational phenological mapping pipeline capable of large-scale deployment in commercial orchard environments.

G4: To perform detailed mapping of female flower phenology in Actinidia chinensis to determine flower receptivity and assess pollination success.

G5: To distinguish functional requirements between male and female flowers, applying flowering detection for male flowers to confirm pollen availability and precise phenological staging for female flowers.

G6: To evaluate the framework's performance and accuracy through statistical validation across multiple commercial orchard zones, demonstrating its practical applicability for precision pollination management.

2 Materials and Methods

This section presents the MPM Framework for automated phenological monitoring in *Actinidia chinensis* orchards. The phenological classification follows the hierarchical framework established in Pinheiro et al. [2026a], comprising 17 classes organised across three levels: structure (blue), gender (yellow), and phenological stage (orange) (Figure 2).

The MPM Framework comprises four components: (i) the Multi-Modal *Actinidia chinensis* Phenology Dataset [Pinheiro et al., 2026a], (ii) model optimisation, (iii) operational pipeline, and (iv) performance evaluation (Figure 3).

The methodology distinguishes between framework architecture and implementation. The architecture subsection describes the complete design and functional integration. The implementation subsection details the operational deployment for female flower phenological mapping, representing the subset of pathways instantiated during this development phase.

Figure 2: Hierarchical classification organisation from the Multi-Modal *Actinidia chinensis* Phenology Dataset [Pinheiro, 2025], comprising 17 classes organised across three levels: structure (blue), gender (yellow), and phenological stage (orange). Detailed morphological criteria for each class are provided in [Pinheiro et al., 2026a].

2.1 Mobile Phenological Mapping Framework - Architecture

The Multi-Modal *Actinidia chinensis* Phenology Dataset [Pinheiro et al., 2026a] provides labelled images for model training and georeferenced videos with manual counts for validation. Model optimisation uses the labelled images component to train the proposed models. The operational pipeline deploys trained models to process georeferenced field videos through phenological detection and probabilistic refinement, integrating GNSS metadata for mapping. Field-acquired videos support future dataset expansion by annotating additional images, enabling iterative model retraining and performance enhancement informed by operational evaluation feedback. The performance evaluation phase validates operational outputs by comparing generated phenological maps with manual field counts from the georeferenced videos component, thereby quantifying system accuracy and reliability for agricultural applications.

This modular architecture is intended to support continuous model refinement while maintaining deployment stability, producing phenological maps that benefit multiple stakeholders. Growers require spatial information for targeted pollination interventions, and agricultural systems require integration with decision support platforms for precision management.

Model optimisation comprises two phases: (i) a learning framework and (ii) performance evaluation. The learning framework incorporates task-specific dataset processing to generate training data optimised for different model tasks (Figure 3, green region).

The learning framework follows the methodology described by Pinheiro et al. [2024], with an additional processing stage that adapts the dataset configuration to accommodate models with different tasks. The labelled images component provides the foundation for model development. The labelled images component is stratified and balanced at the class level, and processed to generate training data suitable for the proposed models. K-fold cross-validation with iterative refinement produces k trained models. Performance evaluation assesses each trained model from the learning framework using test-set metrics. The model achieving the highest validation metrics

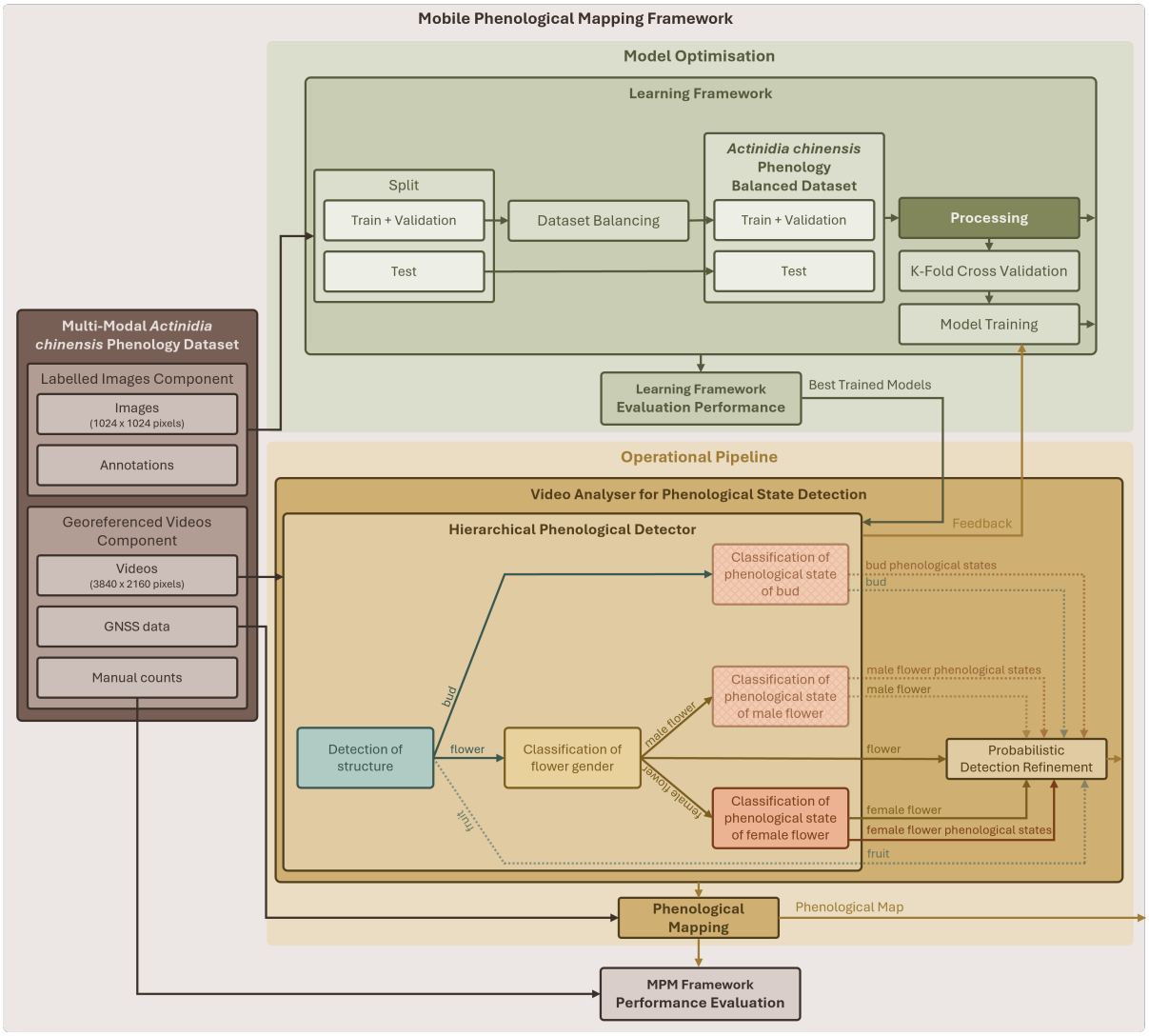

Figure 3: Mobile Phenological Mapping Framework Architecture showing the integration of Multi-Modal *Actinidia chinensis* Phenology Dataset [Pinheiro et al., 2026a], model optimisation, operational pipeline, and framework performance evaluation. Dotted borders indicate pathways designated for future implementation.

is selected for deployment in the operational pipeline.

The operational pipeline transforms georeferenced smartphone videos into spatial phenological maps through two integrated components: (i) a video analyser for hierarchical phenological detection and (ii) phenological mapping for spatial aggregation.

The video analyser incorporates the trained models from model optimisation to progressively classify phenological structures at multiple taxonomic levels (Figure 3, yellow region). The hierarchical detector employs a three-stage sequential detection process aligned with the colour-coded hierarchical organisation in the Multi-Modal *Actinidia chinensis* Phenology Dataset: (i) structure detection (blue), (ii) gender classification (yellow), and (iii) phenological stage classification (orange).

The probabilistic detection refinement method begins with a preprocessing step that adapts key parameters to the input data's characteristics, including video resolution, frame rate, and capture conditions. This step ensures consistent consolidation of object detections across multiple video frames by grouping them by unique identifier. Given the continuity of phenological stages, the final class for each object is determined by removing outliers and averaging classifications across frames, with greater weight given to the more detailed classes. Additionally, GNSS metadata is averaged across frames to yield the most reliable spatial coordinates, thereby enhancing the accuracy of georeferenced phenological maps. This approach improves both temporal consistency and spatial resolution, ensuring more reliable phenological stage mapping.

The phenological mapping integrates refined detection outputs with spatial positioning data to generate georeferenced distribution maps. Frame-level georeferencing uses linear interpolation between GNSS start and end positions recorded during video acquisition [Pinheiro et al., 2026a] and assigns geographic coordinates to each detection by matching timestamps. Detections are compiled across all georeferenced frames to produce a map of phenological stage distribution throughout the orchard. A visualisation interface presents georeferenced phenological data, enabling spatial interpretation of flowering patterns and identification of localised asynchronies to inform targeted pollination management decisions.

Performance evaluation validates the operational pipeline's outputs by comparing automated detections from georeferenced videos with manual field counts. Multi-scale validation enables assessment throughout the detection cascade.

Statistical analysis quantifies system performance through complementary metrics that capture different aspects of accuracy. For each hierarchical level (structure, gender, phenological stage), automated counts (y_i) are compared with manual ground-truth counts (x_i) across n validation zones. The validation dataset comprises four independent zones. This small sample limits the precision of formal statistical testing. The following metrics were computed descriptively:

- **Coefficient of determination (R^2)** measures the proportion of variance in manual counts explained by the automated system through linear regression. Values range from 0 to 1, where higher values indicate that automated counts better predict manual counts.
- **Mean Absolute Error (MAE)** quantifies the average magnitude of prediction errors in the original count units, calculated as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \tag{1}$$

  MAE provides an intuitive measure of average deviation in absolute terms (e.g., the number of flowers), but it does not account for the magnitude of the counts, which limits comparability across different detection scales.
- **Mean Absolute Percentage Error (MAPE)** normalises prediction errors relative to ground truth values, enabling comparison across different detection scales and hierarchical

stages:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{x_i - y_i}{x_i} \right| \tag{2}$$

MAPE expresses accuracy as a percentage, facilitating interpretation across hierarchical stages with different count magnitudes. Lower MAPE values indicate better relative accuracy, with values below 20% generally considered acceptable for operational agricultural monitoring systems [Wang et al., 2021].

These metrics offer complementary insights into system performance. R^2 describes the strength of association between automated and manual counts. MAE and MAPE quantify absolute and relative errors. The combination supports a descriptive evaluation. A high R^2 and low MAPE together indicate a strong association and accurate absolute predictions. Both aspects matter for operational decision support in precision pollination management.

2.2 Mobile Phenological Mapping Framework - Implementation

This subsection details the implementation of the hierarchical detection pathway for classifying female flower phenology. The implementation is grounded in the Multi-Modal *Actinidia chinensis* Phenology Dataset [Pinheiro et al., 2026a], which provides two complementary components that support distinct stages of the framework: the labelled images component for model optimisation, and the georeferenced videos component for operational deployment and performance evaluation. Full details on acquisition protocol, annotation methodology, and class distribution are provided in [Pinheiro et al., 2026a].

The labelled images component provides annotated training data for model development. Images were collected from five *Actinidia chinensis* orchards in northern Portugal, covering variations in age (5 to 30 years), variety (*deliciosa* and *chinensis*), cultivar ("Hayward", "Bo.Erika", "Dori"), and training system (pergola and T-bar). This multi-orchard sampling captured morphological variability in flower arrangement and canopy structure across diverse production conditions. From 3 104 acquired images, 1 665 were selected based on uniqueness calculations using the FiftyOne platform [Moore and Corso, 2020] to identify and remove duplicate or near-duplicate images, avoiding concept imbalance that could bias model learning. Selected images were standardised to 1 024 × 1 024-pixel resolution, with corresponding Pascal VOC annotations containing bounding box coordinates and class labels.

The georeferenced videos component provides field-acquired data for operational deployment and validation. This component comprises six videos per zone, each recorded at 3 840 × 2 160 pixel resolution and 30 fps, covering one inter-wire section of the pergola canopy. Videos were captured using a tripod-mounted smartphone positioned 1.80 m from the vegetative wall and oriented parallel to the canopy. The operator walked at approximately 0.25 m/s along each transect while recording GNSS waypoints using the Geo++ RINEX Logger smartphone application [Zangenehnejad and Gao, 2023]. GNSS data were post-processed to achieve centimetre-level positional accuracy. Frame-level georeferencing was performed through linear interpolation between recorded waypoints, assuming constant velocity along straight-line transects. This component also includes manual ground truth counts for each zone, enabling quantitative validation of automated detections.

Four spatial zones (A, B, C, D) were selected following the orchard owner's established sampling protocol for phenological assessment and yield estimation. This selection captured a range of canopy conditions, from sparse (7.0 structures per m^2) to dense (44.2 structures per m^2), allowing evaluation of framework performance under contrasting flowering densities. Each zone comprised 25 m^2 (5 × 5 m) of pergola-trained canopy, for a total of 100 m^2 across all validation areas. Zone A contained both male and female plants, while Zones B, C, and D contained exclusively female plants. Table 1 presents the phenological composition of each zone based on manual ground truth surveys.

Table 1: Phenological characteristics of the four validation zones based on manual ground truth counts. These zones represent the orchard farmer's established sampling framework for operational phenological monitoring and yield estimation.

Characteristic	Zone A	Zone B	Zone C	Zone D
Area (m^2)	25	25	25	25
Total structures	562	175	695	1 106
Density (structures per m^2)	22.5	7.0	27.8	44.2
Female flowers	527	175	695	1 106
Male flowers	36	0	0	0
Dominant class	flower_female_67	flower_female_67	flower_female_67	flower_female_67

Model optimisation instantiated the learning framework to train models for the three hierarchical stages required for female flower phenological mapping: (i) structure detection, (ii) flower gender classification, and (iii) female flower phenological stage classification. The present implementation utilises 9 of the 17 classes from the original dataset, focusing on the female flower phenological pathway, as the timing of female receptivity is crucial for pollination management decisions (Figure 4). For this purpose, the *Actinidia chinensis* Phenology Balanced Dataset [Pinheiro, 2026] was prepared as a stratified, class-balanced training subset, with detailed documentation on the splitting strategy, balancing methodology, and class distribution available in the dataset repository.

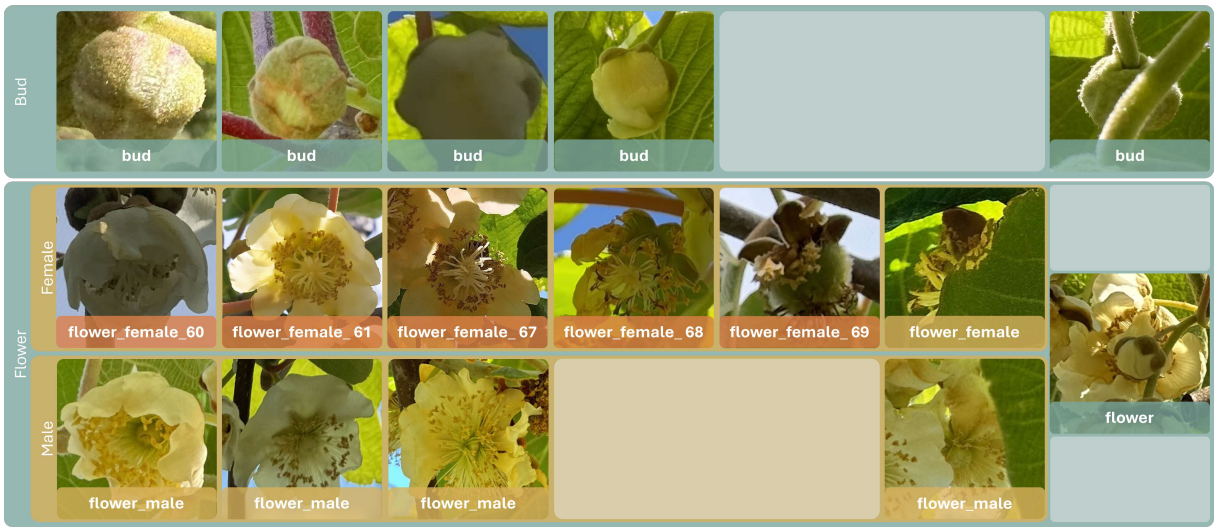

Figure 4: Subset of the hierarchical classification organisation implemented in this study. The implementation employs 9 classes organised across three levels: structure (blue), gender (yellow), and phenological stage (orange).

The learning framework evaluated multiple YOLO architectures to identify optimal performance across detection and classification tasks. YOLO architectures were selected for their proven effectiveness in agricultural object detection [Pinheiro et al., 2024, Li et al., 2022a, 2023], ease of implementation and training, computational efficiency suitable for mobile deployment, and established performance in flower detection tasks [Lim et al., 2020, Li et al., 2022b, Zhou et al., 2023]. Multiple versions were benchmarked to identify the optimal architecture for each hierarchical stage, as computational requirements differ between structure detection (which requires high throughput across dense imagery) and phenological classification (which requires fine-grained morphological discrimination). Structure detection models employed YOLOv8n [Jocher et al., 2023], YOLO11n [Jocher and Qiu, 2024], and YOLO12n [Tian et al., 2025]. Gender classification and phenological stage classification models used YOLOv8n and YOLO11n architectures, as YOLO12n is only available for object detection and does not support classification tasks.

For both detection and classification tasks, the architecture that achieved the best performance metrics on the test set was selected for operational deployment. Using 5-fold cross-validation, each dataset partition was created with stratified sampling to maintain class distributions across folds.

For each trained model, precision, recall, and F1-score were computed on the test set. For each hierarchical stage, the model demonstrating the highest validation metrics was selected for deployment in the operational pipeline.

The operational pipeline deployed trained models to process the georeferenced videos component, generating spatial phenological maps through hierarchical detection and probabilistic refinement.

The video analyser incorporated the selected trained models to progressively classify female flower phenological structures through three sequential processing stages (Figure 3, solid borders in yellow region):

- **Structure Detection** identified and localised flowers and buds within each video frame, generating bounding boxes with unique identifiers for detected regions.
- **Gender Classification** processed each detected flower region to classify specimens as background (false positives from structure detection), flower, female flower, or male flower.
- **Phenological Stage Classification** processed female flower detections to classify each instance as background (false positives from gender classification), flower_female, or BBCH phenological stages (flower_female_60, flower_female_61, flower_female_67, flower_female_68, or flower_female_69).

Structure detection used a 0.5 confidence threshold to ensure reliable flower localisation while limiting false positives. Progressive filtering reduced false positives at each stage while preserving classification accuracy for valid detections.

Probabilistic detection refinement begins with a preprocessing step that automatically detects the positions of the two wires delimiting each inter-wire section directly from the video frames. Detections whose identifiers fall outside these wire boundaries are removed, constraining the analysis to the spatially relevant flowering region. Since all videos share a consistent resolution and frame rate, no additional parameter adaptation is required. The method then employs the Ultralytics tracking module with the default BoT-SORT algorithm [Aharon et al., 2022] to assign persistent identifiers to detections across sequential frames. The tracker links detections between frames using an Intersection over Union (IoU) threshold of 0.5, ensuring that the same reproductive structure receives a consistent identifier as the camera moves. Detections sharing the same identifier were aggregated across all frames, and frequency-based outlier filtering removed classifications appearing in fewer than 10% of frames for each track. The most frequent phenological stage among remaining detections determined the final classification, preventing double-counting while improving classification robustness through temporal consensus.

Phenological mapping aggregated refined detection outputs with their interpolated geographic coordinates to generate distribution maps of female flower phenological stages across each validation zone. Interactive visualisation enabled spatial interpretation of phenological distributions through density-based rendering and stage-specific filtering.

Performance evaluation validated the operational pipeline outputs by comparing automated detections against the manual ground-truth counts provided in the georeferenced videos component. The validation protocol employed multi-scale assessment across three hierarchical detection levels: (i) structure (flower), (ii) gender (female flowers), and (iii) phenological stage (female flower phenological stages), enabling assessment throughout the cascade from structure detection to phenological stage classification.

For all hierarchical levels, statistical metrics (R^2, MAE, and MAPE) were computed across the four spatial zones. This approach enables identification of zone-dependent performance variation and stage-specific detection patterns.

3 Results

Model optimisation evaluated multiple YOLO architectures across three hierarchical detection stages using 5-fold cross-validation. (For detailed information about metrics results per class of each train iteration in each hierarchical stage, consult Supplementary Tables 10, 11, and 12).

The structure detection performance of the three selected YOLO architectures (YOLOv8n, YOLO11n, and YOLO12n) was evaluated across five independent training iterations (Table 2). All models consistently performed better at flower detection than at bud detection, owing to the larger size and more distinctive morphological features of flowers. YOLO11n achieved an optimal, balanced performance with an F1 score of 86% in train 3 and was selected for operational deployment.

Table 2: Structure detection performance metrics (precision, recall, F1-score) on the test set for the three YOLO architectures (YOLOv8n, YOLO11n, YOLO12n). The average shows the mean performance across five training iterations. Best indicates the highest-performing individual iteration.

	Model	**Train**	**Precision** (%)	**Recall** (%)	**F1 Score** (%)
Average	**YOLOv8n**	-	88	81	84
	YOLO11n	-	87	82	84
	YOLO12n	-	86	80	83
Best	**YOLOv8n**	4	88	82	85
	YOLO11n	3	88	83	86
	YOLO12n	2	88	80	84

Figure 5 shows the confusion matrix of YOLO11n in train 3, which reveals minimal inter-class confusion between bud and flower classes (1 to 2%), indicating strong morphological discrimination capability. The main source of error is background predictions (14.6 to 16.5%), which represent detection failures, either missed structures or low-confidence detections, rather than misclassification between structural classes. This pattern reflects the typical challenge in object detection where localisation failures manifest as false negatives. At the same time, the model demonstrates robust discrimination between the annotated classes when structures are successfully detected.

Gender classification evaluated two YOLO architectures (YOLOv8n and YOLO11n) for distinguishing between male and female flower morphologies (Table 3). Gender classification performance showed strong discrimination between the flower_male and flower_female classes, while the flower class showed lower metrics. The flower class serves as a filter, appropriately capturing morphologically ambiguous specimens without propagating uncertain false positives to phenological stage classification. YOLOv8n achieved optimal, balanced performance with an F1 score of 95% in train 4 and was selected for operational deployment.

Figure 6 shows the confusion matrix of YOLOv8n in train 4, which demonstrates exceptional gender discrimination with minimal inter-gender confusion. The flower class shows an asymmetric distribution bias toward male_flower (24.2%), indicating that male morphology is more ambiguous to classify, reflecting greater morphological variability across developmental stages. High background classification accuracy (93.8%) confirms the effective elimination of false positives inherited from the structure-detection stage, thereby validating the progressive filtering architecture.

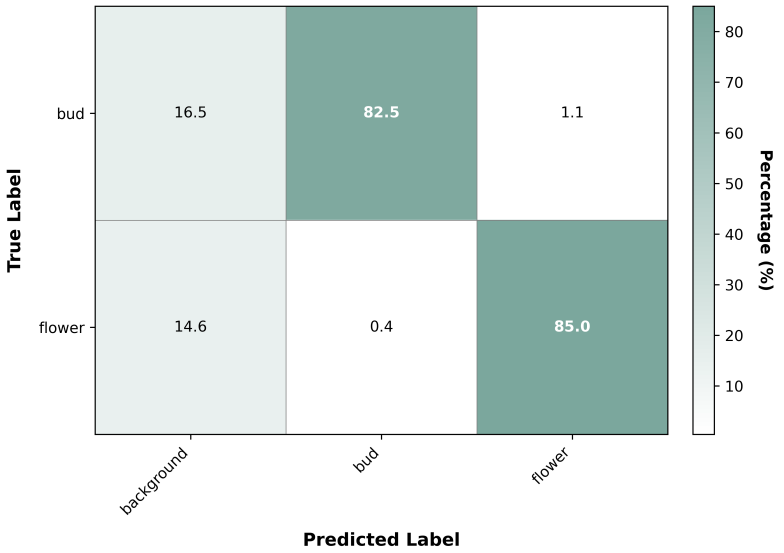

Figure 5: Normalised confusion matrix for structure detection with YOLO11n on train 3. Background predictions appear only as detection failures (no background class exists in ground truth annotations), while inter-class confusion between bud and flower remains minimal.

Table 3: Gender classification performance metrics (precision, recall, F1-score) on the test set for the two YOLO architectures (YOLOv8n and YOLO11n). The average shows the mean performance across five training iterations. Best indicates the highest-performing individual iteration.

	Model	**Train**	**Precision** (%)	**Recall** (%)	**F1 Score** (%)
Average	**YOLOv8n**	-	93	93	93
	YOLO11n	-	92	92	92
Best	**YOLOv8n**	4	95	95	95
	YOLO11n	1	93	93	93

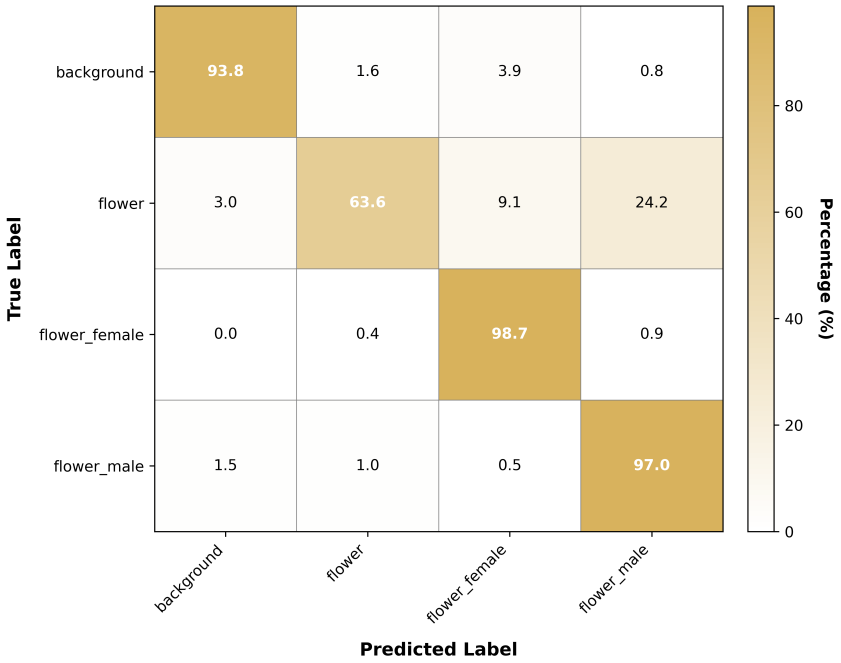

Figure 6: Normalised confusion matrix for gender classification with YOLOv8n on train 4, showing gender discrimination performance and confidence filtering.

Phenological stage classification for female flowers evaluated YOLOv8n and YOLO11n architectures (Table 4). Phenological stage performance ranged from the lowest accuracy on flower_female_68 to the highest on flower_female_69, reflecting morphological distinctiveness across developmental stages. The undifferentiated class effectively captured assignments with uncertainty without propagating low-confidence predictions. YOLOv8n achieved optimal, balanced performance with an F1 score of 89% in train 1 and was selected for operational deployment.

Table 4: Phenological classification performance metrics (precision, recall, F1-score) on the test set for the two YOLO architectures (YOLOv8n and YOLO11n). The average shows the mean performance across five training iterations. Best indicates the highest-performing individual iteration.

	Model	Train	Precision (%)	Recall (%)	F1 Score (%)
Average	**YOLOv8n**	-	87	87	87
	YOLO11n	-	87	87	87
Best	**YOLOv8n**	1	89	89	89
	YOLO11n	1	88	88	88

Figure 7 illustrates the confusion patterns of phenological stages in relation to the phenological development sequences. Flower_female_69 achieved the highest accuracy (92.3%) while flower_female_68 exhibited the lowest performance (65.4%) with 23.1% confusion with the adjacent flower_female_69, reflecting morphological overlap during late flowering to fruit development. Misclassifications occur predominantly between adjacent developmental classes rather than between distant ones, confirming that the model has learned sequential phenological patterns. The flower_female class effectively served its confidence-filtering role by capturing 59.4% of morphologically ambiguous specimens, with the remaining classifications skewed toward early flower_female_61 (21.9%), thereby preventing uncertain predictions from propagating to agricultural decision support. Sequential misclassification patterns confirm learned phenological relationships, ensuring reliability for pollination-critical stage mapping.

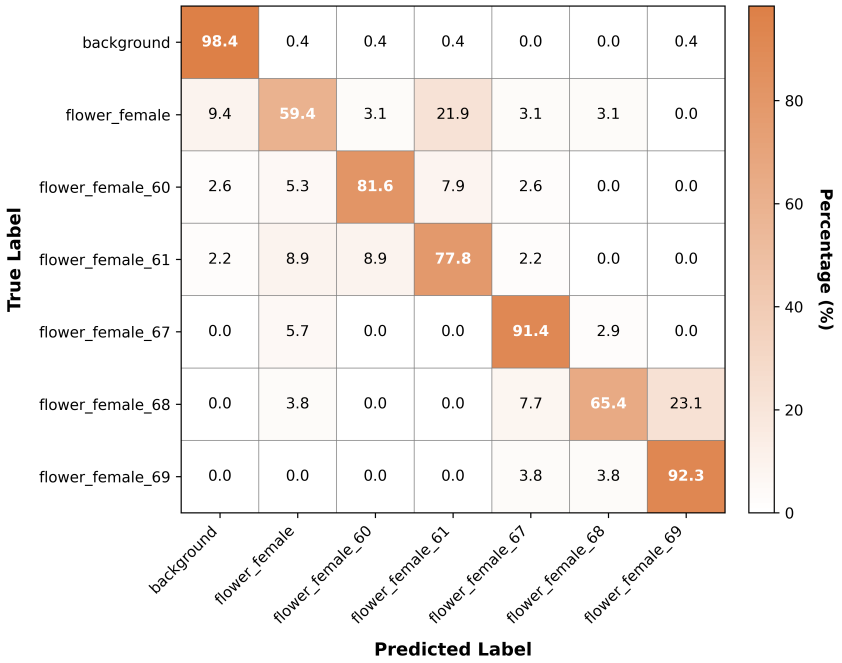

Figure 7: Normalised confusion matrix for phenological stage classification with YOLOv8n on train 1, showing stage-specific discrimination and sequential confusion patterns.

The operational pipeline deployed the optimised models to process georeferenced smartphone videos, generating spatially explicit phenological maps as the primary system output. Figure 8 presents representative video frames with detection bounding boxes from each validation zone, demonstrating the system's capacity to resolve individual flowers across varying canopy densities. In each frame, the two wire positions that delimit the inter-wire section are automatically detected and used to constrain detections to the spatially relevant flowering region, thereby excluding structures outside the canopy boundaries.

(a) Zone A

(b) Zone B

(c) Zone C

(d) Zone D

Figure 8: Representative video frames with detection bounding boxes from each validation zone. Wire boundaries (horizontal lines) define the spatially constrained flowering region used for analysis. The frames demonstrate the system's ability to resolve individual flowers within dense orchard canopies.

Figures 9 and 10 present the resulting phenological maps and transect heatmaps for each validation zone, displaying the spatial distribution of female flower developmental stages across the commercial orchard.

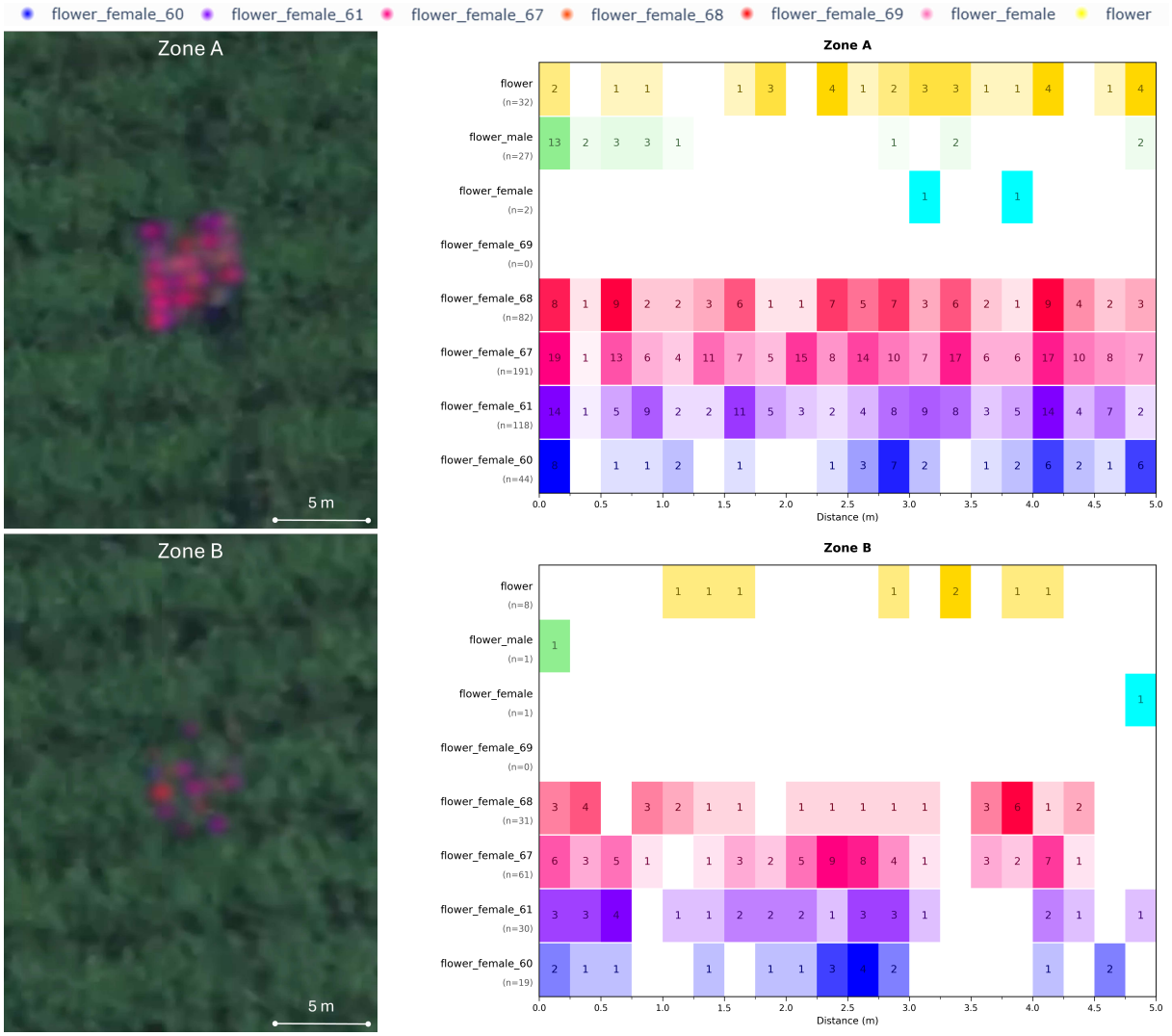

Figure 9: Phenological maps (left) and transect heatmaps (right) for validation Zones A and B. Phenological maps display the spatial distribution of female flower developmental stages across the orchard. Transect heatmaps show detection counts per 0.25-metre bin along 5-metre transects, with colour intensity proportional to detection frequency within each class.

The phenological maps reveal spatial heterogeneity in flowering development, with flower_female_67 as the dominant stage across all zones. The transect heatmaps visualise the distribution of phenological classes along 5-metre transects at 0.25-metre resolution, showing the spatial co-occurrence of developmental stages within each zone.

The performance of the automated detection pipeline was evaluated by direct comparison with manual ground-truth surveys across four orchard zones (A, B, C, and D). Table 5 presents detection counts at three hierarchical levels: (i) structure (flower), (ii) gender (flower_female), and (iii) phenological stage (flower_female_60 through flower_female_69).

Total structure detection showed variable performance across zones, with automated counts underestimating manual surveys by 11.7% to 21.2% in Zones A, B, and D, while overestimating by 22.0% in Zone C. Overall, the pipeline detected 2 366 flowers, compared to 2,538 manual counts, resulting in a 6.8% underestimation. Female flower classification showed 5.0% of detected structures retained in the generic flower class rather than receiving a confident gender assignment. This conservative behaviour varied across zones, from 1.8% in Zone C to 11.9% in Zone A. The automated female counts underestimated manual surveys by 10.2% overall, with Zone D showing the largest discrepancy (24.5% underestimation).

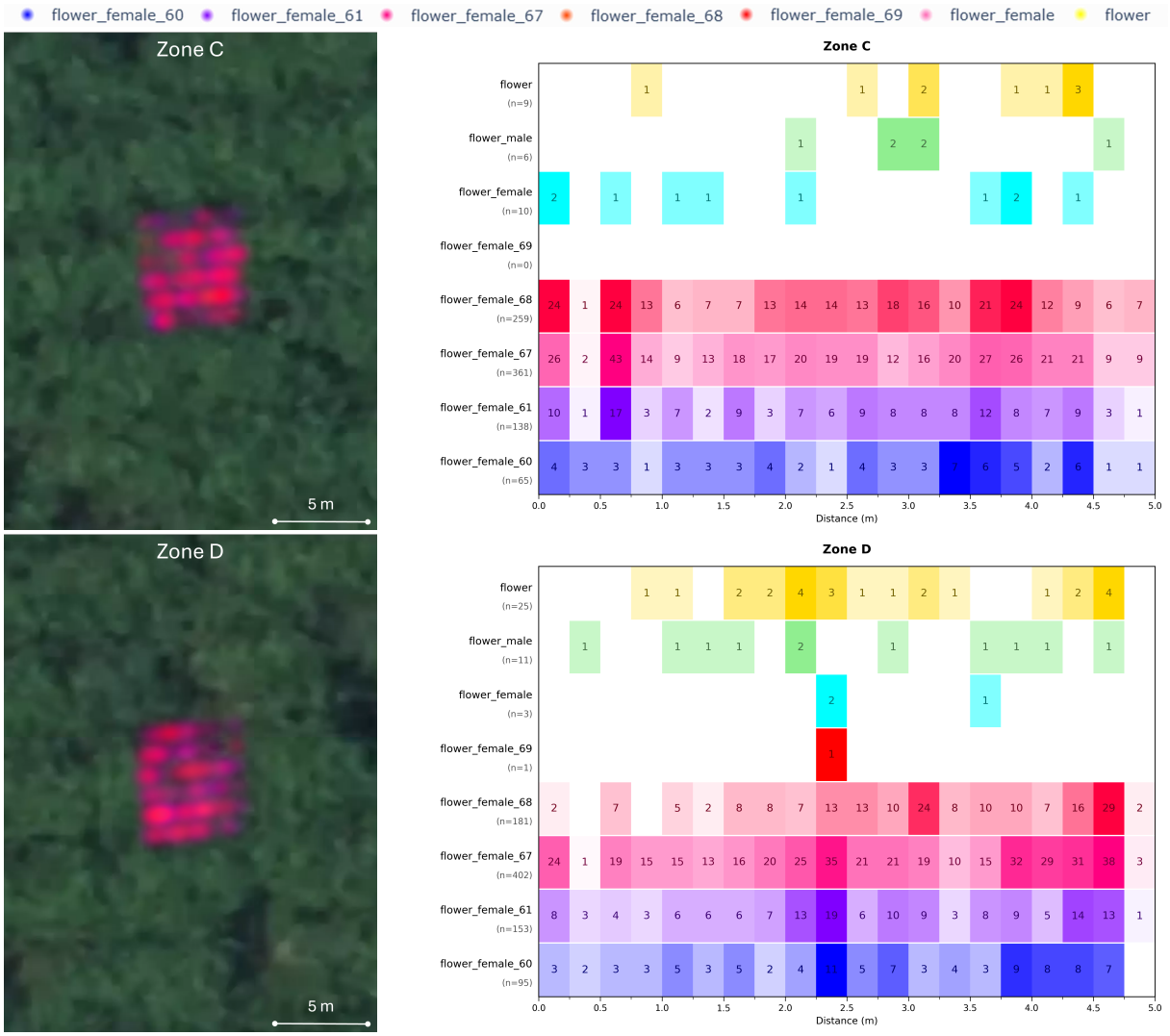

Figure 10: Phenological maps (left) and transect heatmaps (right) for validation Zones C and D. Phenological maps display the spatial distribution of female flower developmental stages across the orchard. Transect heatmaps show detection counts per 0.25-metre bin along 5-metre transects, with colour intensity proportional to detection frequency within each class.

Table 5: Complete validation dataset showing automated (A) detection counts compared against manual (M) ground truth surveys across four orchard zones (A, B, C and D) and across three hierarchical levels.

Zone	Method	flower	flower_female	flower_female_60	flower_female_61	flower_female_67	flower_female_68	flower_female_69
A	A	496	437	44	118	191	82	0
	M	562	527	92	196	233	6	0
B	A	151	142	19	30	61	31	0
	M	175	175	13	73	79	2	8
C	A	848	833	65	138	361	259	0
	M	695	695	41	256	375	23	0
D	A	871	835	95	153	402	181	1
	M	1 106	1 106	20	511	569	6	0

Phenological stage classification revealed systematic confusion between adjacent developmental stages. The pipeline underestimated flower_female_61 and flower_female_67, while overestimating flower_female_68. Combining flower_female_67 and flower_female_68 yields ratios close to unity across zones, indicating that detections are shifted from flower_female_67 to flower_female_68 rather than being lost. This pattern suggests that the primary confusion occurs in the late stages, where the morphological distinction between open flowers and petal senescence is difficult to discern in video frames. Similarly, the underestimation at flower_female_61 likely reflects misclassification toward flower_female_67, since early-opening flowers are assigned to more advanced developmental stages.

Table 6 presents validation metrics for flower detection across four orchard zones. Structure detection showed a high coefficient of determination, with an R^2 of 82.8%, between automated and manual counts. Given the small number of independent zones, this result is best interpreted as a descriptive indicator of the strength of association within this sample. The overall MAPE of 17.2% indicates acceptable relative accuracy for operational deployment. The system exhibited modest underestimation, with 2 366 automated counts against 2 538 manual counts. Zone-specific performance ranged from 11.7% error in Zone B to 22.0% in Zone C. These detection-stage discrepancies propagate through subsequent hierarchical classification levels. Confidence filtering mechanisms retain morphologically ambiguous specimens in parent classes to prevent error amplification.

Table 6: Flower detection validation comparing automated counts against manual ground truth across four orchard zones. Overall, R^2 was computed from a linear regression of automated counts on manual counts across the four validation zones. MAE and MAPE were computed from zone-level errors.

Zone	R^2 (%)	**MAE**	**MAPE** (%)	**Automated**	**Manual**
A	-	66.0	11.7	496	562
B	-	24.0	13.7	151	175
C	-	153.0	22.0	848	695
D	-	235.0	21.2	871	1106
Overall	82.8	119.5	17.2	2 366	2 538

Table 7 presents validation metrics for female flower detection following gender classification. Female flower detection showed a high coefficient of determination, with an R^2 of 81.1%. The pipeline detected 2 247 female flowers, compared with 2 503 manual counts. This corresponds to an overall underestimation of 10.2% and a MAPE of 20.1%. Zone-specific performance ranged from 17.1% to 24.5%. Three zones showed underestimation. Zone C exhibited overestimation. Overall, 5.0% of detected flowers were retained in the generic flower class rather than receiving confident gender assignments. This proportion ranged from 1.8% to 11.9%.

Table 7: Female flower classification validation comparing automated counts against manual ground truth across four orchard zones. Overall, R^2 was computed from a linear regression of automated counts on manual counts across the four validation zones. MAE and MAPE were computed from zone-level errors.

Zone	R^2 (%)	**MAE**	**MAPE** (%)	**Automated**	**Manual**
A	-	90.0	17.1	437	527
B	-	33.0	18.9	142	175
C	-	138.0	19.9	833	695
D	-	271.0	24.5	835	1106
Overall	81.1	133.0	20.1	2 247	2 503

Table 8 presents the performance of phenological stage classification across four orchard

zones. Class flower_female_67 achieved the highest R^2, at 92.2%, and a MAPE of 18.5%. This represents the most reliable phenological classification. Early-stage flower_female_60 showed a low R^2 of 1.3% and a high MAPE of 133.0%. This pattern indicates systematic overestimation with little association to manual counts. Stage flower_female_61 exhibited substantial underestimation, with a moderate R^2 of 68.0%. Late-stage flower_female_68 was systematically overestimated by a factor of 14.9×, with 553 automated counts against 37 manual counts. Flower_female_69 showed minimal detections, with 1 automated count against 8 manual counts. These patterns confirm the inter-stage confusion identified in Figure 7. Detections shift from stages 61 and 67 toward adjacent stages.

Table 8: Phenological stage classification validation comparing automated counts against manual ground truth across four orchard zones. For each class, R^2 was computed from a linear regression of automated counts on manual counts across the four validation zones. MAE and MAPE were computed from zone-level errors.

Class	R^2	**MAE**	**MAPE** (%)	**Automated**	**Manual**
flower_female_60	0.013	38.25	133.0	223	166
flower_female_61	0.680	149.25	53.7	439	1 036
flower_female_67	0.922	60.25	18.5	1 015	1 256
flower_female_68	0.754	129.00	1 664.9	553	37
flower_female_69	-[a]	2.25	-[b]	1	8

[a] R^2 not reported due to insufficient ground-truth representation of this class.
[b] MAPE not computed because manual ground-truth counts were zero in three of the four zones.

4 Discussion

The experimental results provide evidence for the three hypotheses proposed in this study:

H1: Spatial phenological mapping significantly improves the identification of pollination-relevant asynchronies compared to point sampling methods. The phenological maps (Figures 9 and 10) revealed within-orchard spatial heterogeneity, with structure density varying from 7.0 to 44.2 per m^2 across validation zones (Table 1). Although all zones shared the same dominant stage (flower_female_67), the proportional distribution of secondary stages differed substantially. Capturing equivalent spatial information through point sampling would require multiple spatially distributed samples per zone, substantially increasing labour requirements beyond the continuous mapping approach employed here.

H2: Hierarchical deep learning architectures reduce error propagation in multi-class phenological classification compared to single-stage approaches. Structure detection and gender classification achieved MAPE values of 17.2% and 20.1%. The gender classification value sits marginally above the cited 20% threshold. Phenological stage classification showed variable performance across stages. Class flower_female_67 reached the highest R^2, at 92.2%, with a MAPE of 18.5%. Earlier stages exhibited systematic confusion. Flower_female_61 was underestimated by 57.6%. Flower_female_68 was overestimated by a factor of 14.9×. These results indicate that confidence filtering mitigates classification uncertainty at the structure and gender levels. Morphological similarity between adjacent phenological stages still limits fine-grained discrimination.

H3: Mobile video acquisition integrated with GNSS provides sufficient spatial resolution for operational precision pollination management at commercial scales. Frame-level georeferencing through GNSS timestamp matching achieved plant-level spatial positioning

along orchard rows, enabling phenological stage mapping at sub-metre resolution. The transect heatmaps demonstrate that relative spatial patterns are preserved even where absolute stage quantification is limited, supporting the framework's utility for identifying spatial heterogeneity relevant to pollination management decisions.

4.1 Model Optimisation

Training data quality appears more influential than architectural selection, with F1 scores varying by only 1 to 3% across YOLOv8n, YOLO11n, and YOLO12n within each hierarchical stage. This finding suggests that investment in training data curation yields greater returns than architectural experimentation within the YOLO family.

Empirical results drove the decision to use hierarchical over single-stage detection. Preliminary experiments with a single-stage model trained on all 17 classes produced few valid detections on the test set (F1 score below 30%) despite acceptable validation performance (F1 score higher than 90%). This behaviour reflects the fine-grained morphological similarity between adjacent phenological stages, which a single model could not reliably discriminate. The three-stage architecture reduced inter-class similarity at each level, enabling robust generalisation.

The hierarchical detection results can be interpreted through the lens of *Actinidia chinensis* reproductive morphology. Structure detection achieved robust discrimination (F1 score of 86%) between buds and flowers, as expected given the fundamental morphological discontinuity. The 1 to 2% inter-class confusion observed in the confusion matrix reflects transitional specimens, bud_57 (balloon stage) to flower_female_60 or flower_male_60 (open flower), where petal separation initiates but complete flower opening has not occurred.

Gender classification exhibited asymmetric error patterns reflecting known sexual dimorphism in kiwifruit. Female flowers possess a prominent gynoecium with radiating stigmatic surfaces, surrounded by staminodes (non-functional stamens), which creates a distinctive central structure. Male flowers exhibit functional stamens arranged peripherally around a vestigial pistil. The observed 24.2% confusion rate for undifferentiated specimens exhibiting male-like features suggests that stamen-dominated visual signatures present greater classification ambiguity, particularly when flowers are captured at oblique angles where the central gynoecium or pistil region is occluded.

Phenological stage classification within the female pathway (F1 score of 89%) revealed error distributions consistent with the continuous nature of flower development. Adjacent-stage confusion dominated the error matrix, with the flower_female_68 to flower_female_69 transition accounting for 23.1% of misclassifications. The two structures, flower_female_68 (with most petals fallen and all pistils non-functional) and flower_female_69 (with visible fruit swelling and advanced stigma senescence) represent endpoints of a continuous petal abscission process rather than discrete morphological states. Individual flowers exhibit asynchronous petal loss, generating intermediate phenotypes that challenge categorical classification.

Direct performance comparison with prior kiwifruit flower detection systems requires methodological contextualisation. Lim et al. [2020] reported an F1 score of 89% for flower detection using controlled laboratory imaging with uniform backgrounds and standardised illumination. Li et al. [2022b] achieved mAP of 98% for flower and bud localisation from fixed-position orchard cameras with consistent capture geometry. Zhou et al. [2023] reported an F1 score of 90% using UAV imagery with nadir capture angles minimising occlusion. The MPM Framework operates under field constraints, including variable lighting, motion blur, and framing inconsistencies introduced by smartphone capture. Additionally, hierarchical classification expands from binary detection to multi-class staging.

4.2 Operational Pipeline

Structure detection achieved an R^2 of 82.8%, while gender classification achieved an R^2 of 81.1%. Both levels show a consistent degree of association between automated and manual counts. The confidence filtering mechanism retained 5.0% of detections in the undifferentiated flower class. This proportion ranged from 1.8% in Zone C to 11.9% in Zone A. This conservative behaviour prevents uncertainty from propagating across the classification hierarchy. However, this behaviour introduces a precision-recall trade-off. Excluded specimens may bias phenological distributions if the probability of exclusion correlates with developmental stage.

Zone-specific performance varied considerably, with MAPE ranging from 11.7% to 22.0% for flower detection and from 17.1% to 24.5% for female flower classification. Three zones showed systematic underestimation. Zone C exhibited overestimation at both hierarchical levels. This inversion follows from the same stage-level confusion pattern observed across all zones, rather than from a distinct error source in Zone C. Every zone shows underestimation at stage flower_female_61 and overestimation at stage flower_female_68, consistent with the sequential misclassification discussed above. The net zone-level bias depends on which of these two opposing effects dominates. In Zones A, B, and D, the stage 61 underestimation outweighs the stage 68 overestimation, producing net underestimation. In Zone C, the stage 68 overestimation of 236 flowers exceeds the stage 61 underestimation of 118 flowers, producing net overestimation. This difference in relative magnitude, rather than a separate zone-specific error mechanism, explains the reversal of direction in Zone C.

Phenological stage classification revealed systematic confusion between adjacent developmental stages. Class flower_female_67 achieved the highest R^2, at 92.2%, and a MAPE of 18.5%. Combining flower_female_67 and flower_female_68 yields ratios close to unity across zones. This pattern indicates that detections shift between adjacent stages rather than being lost entirely. The pattern reflects the subtle morphological distinction between flower_female_67, with petal fading and some pistils still fertile, and flower_female_68, with most petals fallen and all pistils non-functional. Pistil functionality is the primary differentiating criterion. This criterion is difficult to assess from video frames. The criterion requires careful inspection even during manual annotation.

4.3 Mobile Phenological Mapping Framework

The MPM Framework substantially reduced monitoring effort across the 100 m^2 validation area (Table 9), from 11.3 hours to 9.6 minutes of labour time. Using the 2024 Portuguese average hourly labour cost of €18.20 as a national labour-cost proxy [Eurostat, 2025], this corresponds to a reduction from approximately €206 to €3. Extrapolated using the orchard's established counting-zone sampling scheme, the corresponding monitoring requirements are approximately 113 hours and €2 060 per hectare for manual assessment, compared with 1.6 hours and €29 per hectare for the automated framework. This extrapolation assumes approximately linear scaling with the number of counting zones in the orchard sampling protocol.

The MPM Framework complements UAV-based phenology monitoring by trading canopy-scale coverage for flower-level resolution at substantially lower operational cost [Yang et al., 2020, Zeng et al., 2020]. UAV platforms provide canopy-scale temporal metrics but cannot resolve individual flower phenological stages, which are necessary for precision timing decisions. MPM provides BBCH-resolved staging that is directly aligned with management decisions, filling a critical gap. Ground-based robotic platforms [Pinheiro et al., 2026b] can achieve flower-level resolution but at substantially higher capital cost and operational complexity. MPM addresses the specific requirement for BBCH-resolved phenological staging at spatial resolutions relevant to pollination management decisions. The smartphone platform minimises capital investment, while the walking-path capture protocol efficiently covers linear orchard rows. Integration with robotic pollination systems represents a natural extension: spatial phenological maps generated

Table 9: Time and cost comparison between automated (A) detection and manual (M) ground truth surveys across four orchard zones (A, B, C, and D), each comprising 25 m^2 and totalling 100 m^2. Labour costs use the 2024 Portuguese average hourly labour cost of €18.20 as a national labour-cost proxy [Eurostat, 2025].

Zone	Method	Time (HH:MM:SS)	Labour Cost (€)
A	A	00:02:37	0.79
	M	03:17:00	59.76
B	A	00:02:18	0.70
	M	00:59:00	17.90
C	A	00:02:30	0.76
	M	03:05:00	56.12
D	A	00:02:10	0.66
	M	03:58:00	72.19
Average	A	00:02:24	0.73
	M	02:50:00	51.49
Total	A	00:09:35	2.91
	M	11:19:00	206.0

by MPM could provide the decision layer for variable-rate pollen application, targeting receptive zones while bypassing areas with asynchronous development.

4.4 Limitations and Future Work

Several limitations constrain the interpretation and generalisation of these findings. The validation dataset (4 zones of a single orchard in a single flowering season) limits the statistical power to estimate performance robustly, particularly for rare phenological stages. While the "Hayward" cultivar used in this study represents the dominant commercial variety, the similar floral morphology across green, gold, and red-fleshed cultivars suggests potential transferability, though this remains to be validated. The validation dataset (100 m^2 across four zones within a single orchard) captured a range of flowering densities but cannot confirm the representativeness of commercial *Actinidia* orchards more broadly. Additionally, conservative confidence thresholds that improve precision simultaneously reduce detection yield, potentially biasing phenological distributions and affecting downstream management decisions.

The validation relies on only four independent zones. This small sample limits the precision of the R^2 estimates reported throughout §3. Formal significance testing was not applied for this reason. The R^2, MAE, and MAPE values should be read as descriptive indicators of performance within this sample. These values do not establish a statistically confirmed relationship that generalises beyond the study orchard.

Despite these limitations, the framework retains practical value for precision pollination management. The systematic confusion between adjacent phenological stages does not preclude the identification of spatial heterogeneity: zones dominated by early-stage flowers can be distinguished from zones with advanced flowering, even when exact stage proportions are uncertain. For operational pollination decisions, this relative spatial information, identifying where intervention is most needed, may be more actionable than precise stage quantification.

Future work should address these limitations through multi-season validation across diverse orchard conditions and cultivar types. Additionally, developing temporal consistency constraints that use phenological progression patterns may improve classification accuracy for morphologically similar adjacent stages.

5 Conclusion

This work demonstrates a shift from point-based phenological assessment to spatially explicit phenological mapping oriented toward management decisions. The Mobile Phenological Mapping Framework introduces *operational spatial phenology* as a practical approach to precision pollination management. This concept encompasses phenological characterisation that is simultaneously spatially resolved, temporally repeatable, and directly actionable for agronomic decisions.

Phenological assessment in dioecious fruit crops has traditionally operated at the individual plant level, without systematic spatial aggregation across production areas. This plant-level perspective implicitly treats orchards as phenologically uniform, obscuring the localised asynchronies that determine pollination outcomes. The conceptual advance presented here reframes phenological state as a spatially distributed variable, treating spatial heterogeneity as the primary information layer for pollination management. When phenology is understood spatially, management transitions from uniform applications across entire fields to targeted interventions in specific zones, aligning resource expenditure with biological requirements.

Two strategic directions emerge for future development. First, integrating spatial phenological time series with environmental data would enable predictive modelling of optimal pollination windows, supporting anticipatory decisions on intervention timing and the need for supplementary applications. Second, validation of the framework's transferability across other dioecious crops and production regions, coupled with direct integration into agricultural decision-support systems, represents the pathway toward operational adoption.

This work establishes operational spatial phenology as a viable paradigm for precision pollination management, providing both the conceptual foundation and a validated implementation for spatially informed decision-making in cropping systems where pollination limits productivity.

Author Contributions

Conceptualisation: I.P. and F.N.d.S.; Methodology: I.P.; Software: I.P.; Validation: I.P.; Formal analysis: I.P.; Resources: M.C. and F.N.d.S.; Writing - original draft: I.P.; Writing - review & editing: A.V., M.C. and F.N.d.S.; Supervision: A.V., M.C. and F.N.d.S.

Declaration of Competing Interest

The authors declare no competing interests.

Funding

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Grant Agreement 101084265.

Data Availability

The Multi-Modal *Actinidia chinensis* Phenology Dataset is available on Zenodo under the persistent identifier [Pinheiro, 2025] and is described in the accompanying *Scientific Data* data descriptor [Pinheiro et al., 2026a].

The *Actinidia chinensis* Phenology Balanced Dataset is available on Zenodo under the persistent identifier [Pinheiro, 2026].

Acknowledgments

Isabel Pinheiro was funded by the Portuguese Foundation for Science and Technology (FCT) and European Social Fund (ESF) through the scholarship SFRH/2025.02081.BD.

References

Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*, 2022.

Helena Castro, Catarina Siopa, Vinicius Casais, Mariana Castro, João Loureiro, Hugo Gaspar, and Silvia Castro. Pollination as a key management tool in crop production: Kiwifruit orchards as a study case. *Scientia Horticulturae*, 290:110533, 12 2021. doi: 10.1016/j.scienta.2021.110533.

David LP Correia, Wassim Bouachir, David Gervais, Deepa Pureswaran, Daniel D Kneeshaw, and Louis De Grandpre. Leveraging artificial intelligence for large-scale plant phenology studies from noisy time-lapse images. *Ieee Access*, 8:13151–13160, 2020.

Raphaël d'Andrimont, Momchil Yordanov, Laura Martinez-Sánchez, and Marijn van der Velde. Monitoring crop phenology with street-level imagery using computer vision. *Computers and Electronics in Agriculture*, 196:106866, 2022. doi: 10.1016/j.compag.2022.106866. URL `https://doi.org/10.1016/j.compag.2022.106866`.

María A David, Alejandra Yommi, Enrique Sánchez, Analía Martinez, Natalia Murillo, Olga Marcellán, Osvaldo Atela, and María A Palacio. Strategic use of honey bees (apis mellifera l.) to increase the number and size of fruits in kiwifruit (actinidia chinensis var. deliciosa). *European Journal of Agronomy*, 133:126420, 2022.

Ellen G. Denny, Katharine L. Gerst, Abraham J. Miller-Rushing, Geraldine L. Tierney, Theresa M. Crimmins, Carolyn A. F. Enquist, Patricia Guertin, Alyssa H. Rosemartin, Mark D. Schwartz, Kathryn A. Thomas, and Jake F. Weltzin. Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. *International Journal of Biometeorology*, 58(4):591–601, 2014. doi: 10.1007/s00484-014-0789-5.

Daniel Downey, D. Ken Giles, and David C. Slaughter. Weeds accurately mapped using dgps and ground-based vision identification. *California Agriculture*, 58(4):218–221, 2004. doi: 10.3733/ca.v058n04p218. URL `https://doi.org/10.3733/ca.v058n04p218`.

Eurostat. Eu hourly labour costs ranged from eur 11 to eur 55 in 2024, 2025. URL `https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20250328-1`. Published 28 March 2025; source dataset: lc_lci_lev.

FAO. Pollination services for sustainable agriculture, 2022. URL `https://www.fao.org/pollination/en`. Accessed: [16-05-2025].

Xiaoqian Jiang et al. A vehicle imaging approach to acquire ground truth data for upscaling to satellite data: A case study for estimating harvesting dates. *Remote Sensing of Environment*, 300:113894, 2024. doi: 10.1016/j.rse.2023.113894. URL `https://doi.org/10.1016/j.rse.2023.113894`.

Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024. URL `https://github.com/ultralytics/ultralytics`.

Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. URL https://github.com/ultralytics/ultralytics.

Negin Katal, Michael Rzanny, Patrick Mäder, and Jana Wäldchen. Deep learning in plant phenological research: A systematic literature review. *Frontiers in Plant Science*, Volume 13 - 2022, 2022. ISSN 1664-462X. doi: 10.3389/fpls.2022.805738. URL https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.805738.

Tae Kyung Kim, Sukyung Kim, Myoungsoo Won, Jong-Hwan Lim, Sukhee Yoon, Keunchang Jang, Kye-Han Lee, Yeong Dae Park, and Hyun Seok Kim. Utilizing machine learning for detecting flowering in mid-range digital repeat photography. *Ecological Modelling*, 440:109419, 2021.

Elisabeth Koch, Ekko Bruns, Frank-M Chmielewski, Claudio Defila, Wolfgang Lipa, and Annette Menzel. Guidelines for plant phenological observations. *World Climate Data and Monitoring Programme*, 01 2007.

Guo Li, Longsheng Fu, Changqing Gao, Wentai Fang, Guanao Zhao, Fuxi Shi, Jaspreet Dhupia, Kegang Zhao, Rui Li, and Yongjie Cui. Multi-class detection of kiwifruit flower and its distribution identification in orchard based on yolov5l and euclidean distance. *Computers and Electronics in Agriculture*, 201:107342, 2022a.

Guo Li, Rui Suo, Guanao Zhao, Changqing Gao, Longsheng Fu, Fuxi Shi, Jaspreet Dhupia, Rui Li, and Yongjie Cui. Real-time detection of kiwifruit flower and bud simultaneously in orchard using yolov4 for robotic pollination. *Computers and Electronics in Agriculture*, 193:106641, 2022b.

Kai Li, Weixin Gong, Yinggang Shi, Li Li, Zhi He, Xinting Ding, Yinchu Wang, Li Ma, Wei Hao, Zhen Yang, et al. Predicting positions and orientations of individual kiwifruit flowers and clusters in natural environments. *Computers and Electronics in Agriculture*, 211:108039, 2023.

JongYoon Lim, Ho Seok Ahn, Mahla Nejati, Jamie Bell, Henry Williams, and Bruce A. MacDonald. Deep neural network based real-time kiwi fruit flower detection in an orchard environment, 2020. URL https://arxiv.org/abs/2006.04343.

Yin Liu, Chunyuan Diao, Weiye Mei, and Chishan Zhang. Cropsight: Towards a large-scale operational framework for object-based crop type ground truth retrieval using street view and planetscope satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 216:66–89, 2024. doi: 10.1016/j.isprsjprs.2024.07.025. URL https://doi.org/10.1016/j.isprsjprs.2024.07.025.

Jacqueline W. Mills, Andrew Curtis, Barrett Kennedy, S. Wright Kennedy, and Jay D. Edwards. Geospatial video for field data collection. *Applied Geography*, 30(4):533–547, 2010. doi: 10.1016/j.apgeog.2010.03.008. URL https://doi.org/10.1016/j.apgeog.2010.03.008.

B. E. Moore and J. J. Corso. Fiftyone. *GitHub. Note: https://github.com/voxel51/fiftyone*, 2020.

Birgit Nordt, Isabell Hensen, Solveig Franziska Bucher, Martin Freiberg, Richard Primack, Albert Dieter Stevens, Aletta Bonn, Christian Wirth, Desiree Dotter, Carolin Plos, Maria Sporbert, and Christine Römermann. The phenobs initiative: A standardised protocol for monitoring phenological responses to climate change using herbaceous plant species in botanical gardens. *Functional Ecology*, 35, 02 2021. doi: 10.1111/1365-2435.13747.

Grant D Pearse, Michael S Watt, Julia Soewarto, and Alan Tan. Deep learning and phenology enhance large-scale tree species classification in aerial imagery during a biosecurity response. *Remote Sensing*, 13(9):1789, 2021.

Isabel Pinheiro. Multi-modal actinidia chinensis phenology dataset (portugal, 2023-2024). Data set, 2025. URL https://doi.org/10.5281/zenodo.17371025. Available at https://doi.org/10.5281/zenodo.17371025.

Isabel Pinheiro. Actinidia chinensis phenology balanced dataset (portugal, 2025). Data set, 2026. URL https://doi.org/10.5281/zenodo.18224815. Available at https://doi.org/10.5281/zenodo.18224815.

Isabel Pinheiro, Germano Moreira, Sandro Magalhães, António Valente, Mário Cunha, and Filipe Neves Dos Santos. Deep learning based approach for actinidia flower detection and gender assessment. *Scientific Reports*, 14(1):24452, 2024.

Isabel Pinheiro, Pedro Moura, Leandro Rodrigues, Germano Moreira, Rui Manuel Coutinho, Francisco Terra, Ant'onio Valente, M'ario Cunha, and Filipe Neves dos Santos. A Multi-Modal Dataset for Automated Phenological Stage Mapping in *Actinidia chinensis*. *Scientific Data*, May 2026a. doi: 10.1038/s41597-026-07360-7. URL https://doi.org/10.1038/s41597-026-07360-7.

Isabel Pinheiro, Pedro Moura, Leandro Rodrigues, Abílio Pereira Pacheco, Jorge Teixeira, António Valente, Mário Cunha, and Filipe Neves dos Santos. Economic benchmarking of assisted pollination methods for kiwifruit flowers: Assessment of cost-effectiveness of robotic solution. *Agricultural Systems*, 231:104541, 2026b. ISSN 0308-521X. doi: https://doi.org/10.1016/j.agsy.2025.104541. URL https://www.sciencedirect.com/science/article/pii/S0308521X25002811.

Simon G Potts, Vera Imperatriz-Fonseca, Hien T Ngo, Jacobus C Biesmeijer, Thomas D Breeze, Lynn V Dicks, Lucas A Garibaldi, Rosemary Hill, Josef Settele, and Adam J Vanbergen. *The assessment report on pollinators, pollination and food production: summary for policymakers*. Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity, 2016.

Andrew D. Richardson, Trevor F. Keenan, Mirco Migliavacca, Youngryel Ryu, Oliver Sonnentag, and Michael Toomey. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, 169:156–173, 2013. ISSN 0168-1923. doi: https://doi.org/10.1016/j.agrformet.2012.09.012. URL https://www.sciencedirect.com/science/article/pii/S0168192312002869.

S. Jha, L. Burkle, and C. Kremen. Vulnerability of pollination ecosystem services. In Roger A. Pielke, editor, *Climate Vulnerability*, pages 117–128. Academic Press, Oxford, 2013. ISBN 978-0-12-384704-1. doi: https://doi.org/10.1016/B978-0-12-384703-4.00416-0. URL https://www.sciencedirect.com/science/article/pii/B9780123847034004160.

Nadir Sainz-Costa, Angela Ribeiro, Xavier P. Burgos-Artizzu, María Guijarro, and Gonzalo Pajares. Mapping wide row crops with video sequences acquired from a tractor moving at treatment speed. *Sensors*, 11(7):7095–7109, 2011. doi: 10.3390/s110707095. URL https://doi.org/10.3390/s110707095.

M.C. Salinero, P. Vela, and M.J. Sainz. Phenological growth stages of kiwifruit (actinidia deliciosa 'hayward'). *Scientia Horticulturae*, 121(1):27–31, 2009. ISSN 0304-4238. doi: https://doi.org/10.1016/j.scienta.2009.01.013. URL https://www.sciencedirect.com/science/article/pii/S0304423809000089.

R Spinelli, M Colangeli, et al. Importance of pollination for fruit size in kiwifruits. *Informatore Agrario*, 59(40):61–64, 2003.

Yunjie Tian, Qixiang Ye, and David Doermann. Yolo12: Attention-centric real-time object detectors, 2025. URL https://github.com/sunsmarterjie/yolov12.

Narumasa Tsutsumida and Shuya Funada. Mapping cherry blossom phenology using a semi-automatic observation system with street level photos. *Ecological Informatics*, 78:102314, 2023. doi: 10.1016/j.ecoinf.2023.102314. URL `https://doi.org/10.1016/j.ecoinf.2023.102314`.

Fabien H Wagner. The flowering of atlantic forest pleroma trees. *Scientific reports*, 11(1):20437, 2021.

Xu Annie Wang, Julie Tang, and Mark Whitty. Deepphenology: Estimation of apple flower phenology distributions based on deep learning. *Computers and Electronics in Agriculture*, 185:106123, 2021.

Henry Williams, Mahla Nejati, Salome Hussein, Nicky Penhall, Jong Yoon Lim, Mark Hedley Jones, Jamie Bell, Ho Seok Ahn, Stuart Bradley, Peter Schaare, et al. Autonomous pollination of individual kiwifruit flowers: Toward a robotic kiwifruit pollinator. *Journal of Field Robotics*, 37(2):246–262, 2020.

Qi Yang, Liangsheng Shi, Jingye Han, Jin Yu, and Kai Huang. A near real-time deep learning approach for detecting rice phenology based on uav images. *Agricultural and Forest Meteorology*, 287:107938, 2020.

Farzaneh Zangenehnejad and Yang Gao. GNSS observation generation from smartphone android location API: Performance of existing apps, issues and improvement. *Sensors*, 23(2):777, 2023. doi: 10.3390/s23020777.

Linglin Zeng, Brian D Wardlow, Daxiang Xiang, Shun Hu, and Deren Li. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, 237:111511, 2020.

Haili Zhou, Junlang Ou, Penghao Meng, Junhua Tong, Hongbao Ye, and Zhen Li. Reasearch on kiwi fruit flower recognition for efficient pollination based on an improved yolov5 algorithm. *Horticulturae*, 9(3):400, 2023.

Supplementary Information

Table 10: Structure detection performance metrics (precision, recall, F1-score) on the test set for the three YOLO architectures (YOLOv8n, YOLO11n, YOLO12n) across each train iteration with a confidence threshold of 50%.

Model	Metrics	Class	Train 1	Train 2	Train 3	Train 4	Train 5	Average
YOLOv8n	**Precision** (%)	bud	88	88	87	87	87	87
		flower	91	89	88	90	88	89
		all	90	88	88	88	88	88
	Recall (%)	bud	79	79	79	80	78	79
		flower	81	82	82	83	86	83
		all	80	80	80	82	82	81
	F1 score (%)	bud	83	83	83	83	82	83
		flower	86	86	85	87	87	86
		all	84	84	84	85	84	84
YOLO11n	**Precision** (%)	bud	80	86	86	81	86	84
		flower	89	90	90	90	90	90
		all	84	88	88	86	88	87
	Recall (%)	bud	84	80	82	81	81	82
		flower	84	82	84	81	83	83
		all	84	81	83	81	82	82
	F1 score (%)	bud	82	83	84	81	83	83
		flower	86	86	87	85	86	86
		all	84	84	86	83	84	84
YOLO12n	**Precision** (%)	bud	82	86	85	84	80	83
		flower	91	90	89	91	87	90
		all	86	88	87	88	84	86
	Recall (%)	bud	79	79	78	80	79	79
		flower	79	82	81	78	80	80
		all	79	80	80	79	80	80
	F1 score (%)	bud	81	83	82	82	80	82
		flower	85	86	85	84	83	85
		all	83	84	84	83	82	83

Table 11: Gender classification performance metrics (precision, recall, F1-score) on the test set for the two YOLO architectures (YOLOv8n and YOLO11n) across each train iteration.

Model	Metrics	Class	Train 1	Train 2	Train 3	Train 4	Train 5	Average
YOLOv8n	**Precision** (%)	background	97	94	97	97	97	96
		flower	68	64	71	81	61	69
		flower_female	93	92	92	96	93	93
		flower_male	95	94	94	95	95	95
		all	93	92	93	95	93	93
	Recall (%)	background	89	91	91	94	90	91
		flower	64	55	61	64	67	62
		flower_female	98	97	97	99	96	97
		flower_male	95	93	94	97	95	95
		all	93	92	93	95	93	93
	F1 score (%)	background	93	92	94	95	93	93
		flower	66	59	66	71	64	65
		flower_female	96	94	95	97	94	95
		flower_male	95	94	94	96	95	95
		all	93	92	93	95	93	93
YOLO11n	**Precision** (%)	background	94	93	96	94	95	94
		flower	79	65	56	72	56	66
		flower_female	93	92	91	92	91	92
		flower_male	95	96	96	95	94	95
		all	93	92	92	92	91	92
	Recall (%)	background	95	89	89	91	89	91
		flower	58	61	61	64	45	58
		flower_female	97	96	98	97	98	97
		flower_male	94	94	91	93	93	93
		all	93	92	91	93	91	92
	F1 score (%)	background	94	91	92	92	92	92
		flower	67	62	58	68	50	61
		flower_female	95	94	94	94	94	94
		flower_male	94	95	93	94	93	94
		all	93	92	92	92	91	92

Table 12: Phenological classification performance metrics (precision, recall, F1-score) on the test set for the two YOLO architectures (YOLOv8n and YOLO11n) across each train iteration

Model	Metrics	Class	Train 1	Train 2	Train 3	Train 4	Train 5	Average
YOLOv8n	**Precision** (%)	background	98	97	96	96	96	97
		flower_female	66	58	74	54	64	63
		flower_female_60	84	86	88	81	85	85
		flower_female_61	76	67	80	76	76	75
		flower_female_67	82	75	69	73	72	74
		flower_female_68	81	74	75	57	63	70
		flower_female_69	87	92	85	81	84	86
		all	89	87	88	84	86	87
	Recall (%)	background	98	97	98	97	98	98
		flower_female	59	44	72	41	56	54
		flower_female_60	82	66	76	79	76	76
		flower_female_61	78	84	73	76	76	77
		flower_female_67	91	86	89	91	89	89
		flower_female_68	65	77	46	46	46	56
		flower_female_69	92	90	90	85	90	89
		all	89	87	88	85	87	87
	F1 score (%)	background	98	97	97	96	97	97
		flower_female	62	50	73	46	60	58
		flower_female_60	83	75	82	80	81	80
		flower_female_61	77	75	77	76	76	76
		flower_female_67	86	80	78	81	79	81
		flower_female_68	72	75	57	51	53	62
		flower_female_69	90	91	88	83	87	88
		all	89	87	88	84	86	87
YOLO11n	**Precision** (%)	background	96	98	97	97	96	97
		flower_female	71	71	75	71	62	70
		flower_female_60	91	83	88	91	94	89
		flower_female_61	73	71	82	73	71	74
		flower_female_67	82	65	70	62	68	69
		flower_female_68	74	68	62	63	56	65
		flower_female_69	82	82	85	87	88	85
		all	88	87	88	87	86	87
	Recall (%)	background	98	98	97	97	98	98
		flower_female	53	53	75	53	56	58
		flower_female_60	82	79	74	82	79	79
		flower_female_61	80	78	82	82	82	81
		flower_female_67	89	86	91	80	91	87
		flower_female_68	54	50	62	46	35	49
		flower_female_69	94	87	77	90	83	86
		all	88	87	88	87	86	87
	F1 score (%)	background	97	98	97	97	97	97
		flower_female	61	61	75	61	59	63
		flower_female_60	86	81	80	86	86	84
		flower_female_61	77	74	82	77	76	77
		flower_female_67	85	74	79	70	78	77
		flower_female_68	62	58	62	53	43	56
		flower_female_69	87	84	81	89	85	85
		all	88	86	88	86	86	87