Automatic detection of phenological stages in *Rosa* spp. using YOLOv8 convolutional neural networks

Detección automática de estados fenológicos en *Rosa* spp. mediante redes neuronales convolucionales YOLOv8

Oscar Hernán Franco Montoya[1*], José Leonardo Franco Montoya[1], and Luis Joel Martínez Martínez[1]

ABSTRACT

This study evaluated the performance of YOLOv8 convolutional neural network models for the automatic detection of phenological stages in greenhouse-grown cut roses (*Rosa* spp.). Image acquisition was conducted in a commercial greenhouse in Tocancipá, Colombia, using a ground-based mobile platform equipped with RGB cameras, thereby avoiding the operational limitations of unmanned aerial vehicles (UAVs) in enclosed environments. Images were collected during five sampling periods using a Nikon camera mounted on the mobile platform across five hydroponic benches, each divided into five 6.4-m plots, for a total of 25 plots. In total, 2,000 images and 4,653 annotated objects were obtained across 9 classes (8 phenological and 1 multipurpose). Model performance was evaluated using precision, recall, F1-score, mAP50, and mAP50–95. Individual models outperformed the multipurpose model, with the *C_stage* model achieving an F1-score of 0.87 in validation and 0.84 in testing. The multipurpose model required extending training to 200 epochs to achieve convergence, resulting in improved performance (F1-score = 0.75 and Precision = 0.78 in validation; F1-score and Precision = 0.72 in testing), indicating its potential for simultaneous multi-stage detection under greenhouse conditions. Correlation analysis showed that object size was the main factor influencing model performance ($r \geq 0.90$). At the same time, the number of labeled samples per class had only a weak relationship with the metrics. This explained the higher accuracy in phenological stages with larger and more distinctive floral structures (*C_stage*, *S_color*) and the lower performance in early stages (rice, chickpea), whose buds occupied less than 0.3% of the image area.

Keywords: computer vision, deep learning, artificial intelligence in agriculture, plant phenology, greenhouse crops.

RESUMEN

Este estudio evaluó el desempeño de modelos de redes neuronales convolucionales YOLOv8 para la detección automática de estados fenológicos en rosas de corte (*Rosa* spp.) cultivadas bajo invernadero. La captura de imágenes se realizó en un invernadero comercial en Tocancipá, Colombia, utilizando una plataforma móvil terrestre equipada con cámaras RGB, evitando así las limitaciones operativas del uso de vehículos aéreos no tripulados (VAT) en espacios cerrados. Las imágenes se obtuvieron durante cinco periodos de muestreo, empleando una cámara Nikon montada en la plataforma a lo largo de cinco bancos hidropónicos, cada uno dividido en cinco parcelas de 6,4 m, para un total de 25 parcelas. En total, se recolectaron 2.000 imágenes y 4.653 objetos etiquetados en nueve clases (8 fenológicas y 1 clase multipropósito). El desempeño de los modelos se evaluó mediante precisión, exhaustividad (recall), F1-score, mAP50 y mAP50–95. Los modelos individuales superaron al modelo multipropósito, destacándose *C_stage*, que alcanzó un F1-score de 0,87 en validación y 0,84 en prueba. El modelo multipropósito requirió extender el entrenamiento hasta 200 épocas para lograr la convergencia, obteniendo un mejor desempeño (F1-score = 0,75 y Precisión = 0,78 en validación; F1-score y Precisión = 0,72 en prueba), lo que indicó su potencial para la detección simultánea de múltiples etapas fenológicas bajo condiciones de invernadero. El análisis de correlación mostró que el tamaño del objeto fue el principal factor que determinó el desempeño del modelo ($r \geq 0,90$), mientras que el número de etiquetas por clase presentó una relación débil con las métricas. Esto explicó el mejor rendimiento en etapas con estructuras florales más grandes (*C_stage*, *S_color*) y el desempeño limitado en etapas tempranas (rice, chickpea), cuyos botones ocuparon menos del 0,3% del área de la imagen.

Palabras clave: visión computacional, aprendizaje profundo, inteligencia artificial aplicada a la agricultura, fenología de plantas, cultivos bajo invernadero.

Received for publication: August 14, 2025. Accepted for publication: December 5, 2025. Doi: 10.15446/agron.colomb.v43n3.122164

[1] Universidad Nacional de Colombia, Facultad de Ciencias Agrarias, Bogotá (Colombia).
[*] Corresponding author: ohfrancom@unal.edu.co

Introduction

The cultivation of cut roses (*Rosa* spp.) is one of the most important agricultural export sectors in Colombia, representing a strategic component of the country's exports and recognized for its high level of specialization in agronomic management under controlled greenhouse conditions. Currently, Colombia is the leading flower exporter in Latin America and the second worldwide (ICA, 2024). One of the main challenges for producers is accurately estimating the phenological stage of stems and projecting harvests, tasks traditionally carried out through manual field walks, stem-by-stem counting, and physical data recording. This process, in addition to being time-consuming and labor-intensive, is prone to human error and lacks the immediacy required to feed predictive models. Therefore, the automation of phenological monitoring through computer vision tools emerges as a promising alternative.

In recent years, multiple successful applications of deep learning models have been documented for counting and classifying structures in agricultural crops. In soybean, for example, the effectiveness of YOLO-X models is demonstrated for the automatic counting of reproductive structures such as pods, achieving an accuracy greater than 96% ($R^2 = 0.967$), MAE of 4.18, MAPE of 10%, and RMSE of 6.48, even in scenarios with high density and overlap, validating their applicability as an alternative to manual methods (Xiang *et al*., 2023). In blueberry and lingonberry, YOLOv5 is used for the automatic detection of fruits at different phenological stages (flowering, green fruits, and ripe fruits), reaching a precision of 0.6 and a recall of 0.55, demonstrating its feasibility for optimizing harvest density estimation processes, despite some limitations in detecting individual flowers (Pajula, 2022).

Additionally, in peach orchards (*Prunus persica*), where images exhibit high floral density, background interference, and severe occlusion, a recent study evaluates YOLO architectures (v5, v7, and v8) alongside a density-map approach based on a multi-column deep neural network. The density-map model achieves the best counting performance (MAE = 39.13; RMSE = 69.69; %Err = 9.98), substantially outperforming detection-based methods, among which YOLOv7x yields the highest accuracy (MAE = 152.7; RMSE = 212.9; %Err = 29.7). This research highlights the challenges posed by densely populated floral scenes for detection algorithms and underscores the potential of deep learning models to enhance phenological monitoring in fruit crops (Estrada *et al*., 2024). In cotton (*Gossypium barbadense*), the use of YOLOv8 (extra-large) achieves an accuracy of 99.81%, with high recall and F1-score values, validating its capacity for automatic counting of dense fruits from images and integrating it into web applications aimed at precision agriculture (Ballena *et al*., 2025).

Recent advances demonstrate the potential of computer vision and deep learning for phenological and floral monitoring across various crops and ecosystems. In rapeseed (*Brassica napus* L.), Li *et al*. (2023) integrate YOLOv5 + CBAM on UAV imagery, achieving an $R^2 > 0.96$ for flower counting and mAP > 92% for localization, while also reporting a positive correlation between detected inflorescences and crop yield. In an ecological context, John *et al*. (2024) combine crowd-sourced images with deep learning models such as Mask R-CNN, RetinaNet, and YOLOv5 to estimate floral species richness in alpine meadows, with Mask R-CNN reaching the highest performance (mAP = 0.67). In fruit crops, Wang *et al*. (2021) develop DeepPhenology, based on VGG-16, to estimate apple flower distribution, achieving average Kullback-Leibler (KL) divergences of 0.23-0.27 and outperforming YOLOv5. Similarly, Mann *et al*. (2022) automate flower phenology monitoring in Arctic species using time-lapse cameras and CNNs, achieving precision = 0.918 and recall = 0.907, while Qi *et al*. (2021) propose Fusion-YOLO, a lightweight model capable of detecting chrysanthemum flowering stages under complex field conditions. Finally, Zhou *et al*. (2023) introduce S-YOLO, a transformer-based IoT-integrated model for apple flowering monitoring, reaching an accuracy of up to 91.95%.

Collectively, these studies highlight the growing adoption of advanced detection architectures—from traditional CNNs to transformer-based models—for automated analysis of flowering and phenology, consolidating computer vision as a key tool for precision agriculture and digital ecology.

These studies demonstrate that convolutional neural networks (CNNs) enable the automatic detection and classification of objects from images without the need for manual selection of explanatory variables. However, their application to highly specialized crops such as greenhouse-grown cut roses remains incipient, particularly for stem phenological stages. In the specific case of rose cultivation, previous studies mainly focus on nutritional monitoring using VIS-NIR spectroscopy (Franco Montoya & Martínez Martínez, 2024) and multivariate models (Franco Montoya & Martínez Martínez, 2025) to predict foliar nutrient concentrations, such as manganese. Nonetheless, the application of computer vision techniques for the

automatic detection of stem phenological stages represents a novel approach with the potential to improve efficiency in agronomic and logistical processes within the sector. This advancement, however, posed considerable technical challenges, as image acquisition in greenhouse crops is constrained by the production system's infrastructure, including plastic coverings, metal structures, wooden frames, wires, and other physical barriers that hinder the use of Unmanned Aerial Vehicles (UAVs).

For this reason, a ground-based mobile platform was chosen, enabling the circumvention of these restrictions and facilitating the systematic acquisition of field data. Although producers have adopted UAV-based technologies, no prior studies have examined their implementation in greenhouse systems, which reinforces the relevance and innovation of the adopted solution.

This study makes a novel contribution by demonstrating, for the first time, the application of YOLOv8 to the automatic detection of phenological stages in cut roses grown under greenhouse conditions—an ornamental crop of high commercial value for which no previous studies of this kind are reported. The findings provide a methodological foundation for integrating computer vision systems into the productive management of ornamental crops, promoting process automation and data-driven decision-making. Based on this context, the objective of this research was to evaluate the performance of YOLOv8 models trained on RGB images for the automatic detection of phenological stages in cut rose stems, comparing both individual models for each stage and a multipurpose model. This approach aimed to identify the most accurate and operationally feasible alternative to support harvest estimation processes under real greenhouse production conditions.

Materials and methods

Study area

The study was conducted in a greenhouse located in Tocancipá, Cundinamarca, Colombia, at an altitude of 2,605 m a.s.l. (4°58'40.1" N, 73°59'06.6" W). The plant material consisted of *Rosa* spp. (cv. Freedom) grown under a fully mechanized production system, with standardized agronomic management practices typical of the floriculture sector for greenhouse-grown crops.

Image sampling design

Although this work did not aim to evaluate agronomic treatments, the images used for model training and validation were organized in a randomized complete block design (RCBD). A total of 25 plots (five per block) were used as spatial reference units. In addition, five temporal sampling campaigns were conducted throughout the floral cycle to ensure representation of all nine phenological classes. This structure allowed capturing the spatial and temporal variability present within the greenhouse, thereby enhancing the robustness of model training and evaluation.

Image capture

The images were captured using a portable platform equipped with an RGB camera (Nikon) with sensors in the blue, green (550 nm), and red (660 nm) bands. The camera was mounted on a mobile structure that moved along the hydroponic benches via a motorized rail system, maintaining a constant height of 1.2 m above the plant canopy and uniform movement. The longitudinal overlap between images was 60%, ensuring complete coverage of each plot without loss of relevant information (Fig. 1).

We collected a total of 2,000 RGB images and 4653 annotated objects, corresponding to nine phenological classes of *Rosa* spp. under greenhouse conditions. The distribution of labeled objects per class was as follows: *Rice* (n = 147), *chickpea* (n = 461), *D_Chickpea* (n = 1014), *S_Color* (n = 1650), *S_Opening* (n = 381), *S_Sepals* (n = 365), *C_Stage* (n = 298), and *T_Bud* (n = 337). This class distribution reflects the natural imbalance typical of commercial rose production, in which early and intermediate phenological stages are more frequent than advanced harvest stages. The phenological stage description is given in Table 1.

We captured images between 10:00 a.m. and 2:00 p.m. under natural light conditions inside the greenhouse, taking advantage of the most stable lighting of the day. The Nikon cameras were set to automatic shooting mode, capturing one image every 0.4 m. A uniform number of pictures per plot was ensured, maintaining an even distribution among replicates and guaranteeing adequate spatial representation of the data throughout the experiment.

Image processing and training environment

The images were organized, classified, and manually labeled using the LabelMe software, identifying in each case the different phenological stages of the stems as categories of interest for model training. Subsequently, we converted the annotations into the format required by YOLOv8 using the *labelme2yolov8* package. All images, initially captured at 4000×3000 pixels, were resized to 640×640 pixels to standardize the dataset and meet the YOLOv8 model requirements.

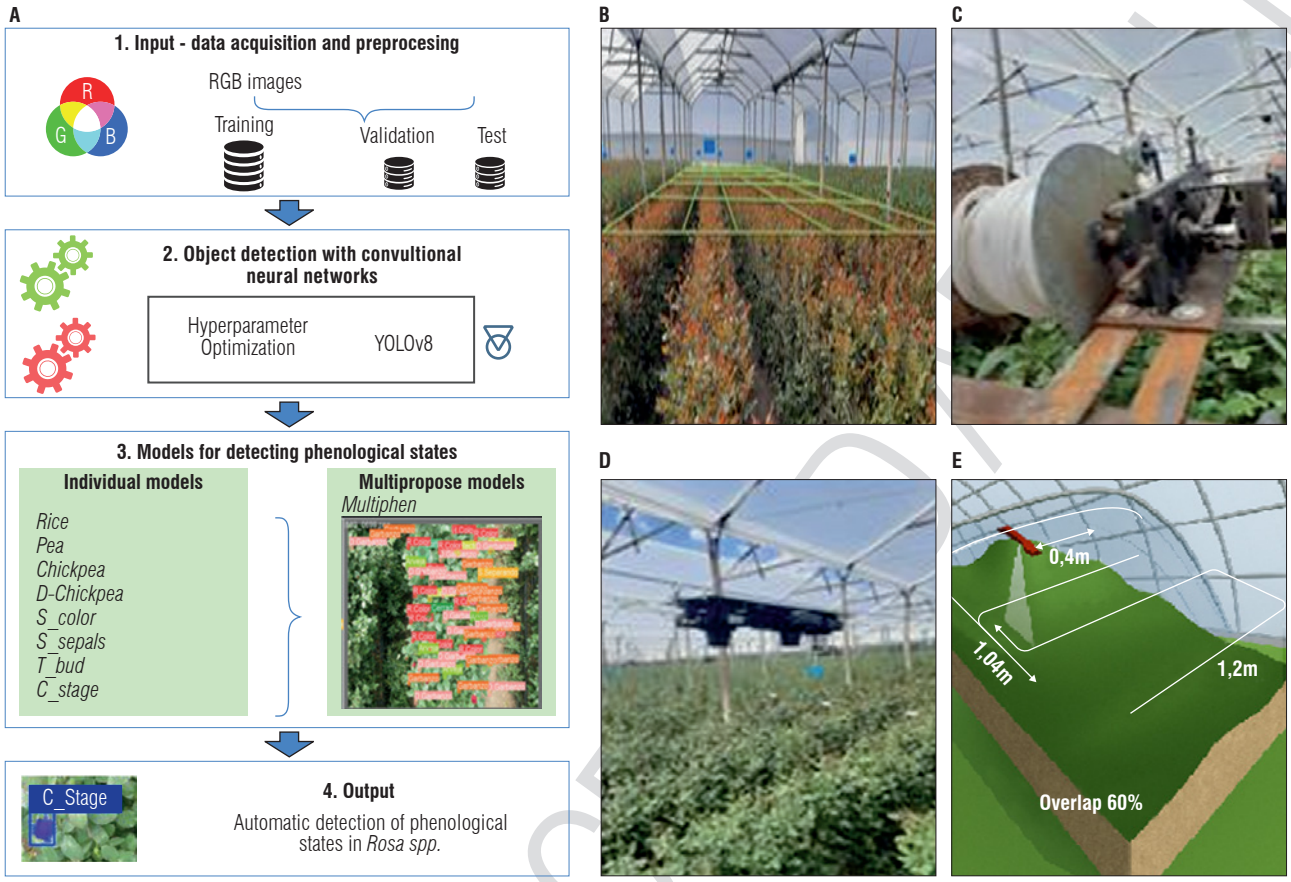

FIGURE 1. Workflow of convolutional neural networks and image acquisition: A) CNN workflow, B) plots (RCBD), C) motor providing traction to the platform, D) platform with adapted cameras, and E) diagram of image capture inside the greenhouse.

We trained in a virtual environment configured with Python 3.9 (64-bit), using the following versions: torch 2.0.1+cu117, ultralytics, numpy 1.24.4, and opencv-python 4.7.0.72, all compatible with each other. The *labelme2yolov8* package was used for annotation conversion.

Training and model definition

For automatic classification of phenological stages, models were trained for each stage, and a multipurpose model combining all stages. The training images were previously labeled, ensuring the correct phenological assignment to each visible stem (Fig. 2).

The individual models enabled analysis of each stage, while the multipurpose model aimed to simulate a real-world joint classification. Each model was assigned an abbreviated identifier (Tab. 1) to facilitate tracking in the results section.

To contextualize the scale of the detected objects, the average size of floral structures was estimated based on

TABLE 1. Models used for the automatic detection of phenological stages in *Rosa* spp.

Model	Phenological stage	Stage description
Rice	rice	Small-sized floral bud, completely closed
Chickpea	chickpea	Globose bud, beginning to show expansion, still closed
D _ chickpea	Double chickpea	Larger bud, multiple layers of petals visible
S _ color	Scratch color	Petals begin changing color
S _ opening	Sepal opening	Sepals begin to open, without fully exposing the flower
S _ sepals	Straight sepals	Sepals fully open and straight, the flower is still not completely opened
T _ bud	Tight bud	Flower about to open but still closed; ready for near-term cutting
C _ stage	Cutting stage	Ideal commercial stage for harvest, petals partially open
Multiphen	Multipurpose	Multipurpose model capable of identifying the nine phenological stages

FIGURE 2. Representative examples of the eight phenological classes of *Rosa* spp. used for model training. A) *Rice,* B) *Chickpea,* C) *D_chickpea*, D) *S_color*, E) *S_opening*, F) *S_sepals*, G) *T_bud*, H) *C_stage*, and I) Multipurpose model, which integrates all phenological classes into a single detection task.

the dimensions of the bounding boxes relative to the total image area (4000×3000 pixels). On average, the objects accounted for 0.2% and 2.6% of the total area, depending on the phenological class. Additionally, the mean diameters measured directly in the field from a top-view perspective were 0.75 cm (rice), 1.2 cm (chickpea), 2.2 cm (Double chickpea), 2.75 cm (Scratch Color), 2.95 cm (Sepal Opening), 3.2 cm (Straight Sepals), 3.6 cm (Tight Bud), and 4.2 cm (Cutting Stage). These values provide context on the spatial scale variability of the objects analyzed within the dataset.

For automatic phenological stages, we used the YOLOv8 (nano) architecture. This model, based on convolutional neural networks (CNNs), was trained independently for each RGB image. We split the dataset into 70% for training, 15% for validation, and 15% for testing, ensuring an adequate balance during training.

Model training, configuration, and evaluation

Model training was performed in the PyCharm 2023.2 development environment using Python 3.9. The YOLOv8 architecture was configured with 100 training epochs to balance between processing capacity and learning stability. Training was performed on a workstation equipped with an NVIDIA GeForce GTX 960M GPU (4 GB VRAM), using a batch size of 4, adjusted to the available memory to ensure stable convergence.

The YOLOv8 models were trained using the default hyperparameter configuration provided by Ultralytics 2023 (Ultralytics Inc.). The optimizer was stochastic gradient descent (SGD) with a learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005. The input image size was 640 × 640 px, and the confidence threshold (*conf-thres*) and intersection over union value for non-maximum suppression (*IoU NMS*) were set to 0.25 and 0.45, respectively. No additional data augmentation techniques were applied, as the dataset already captured natural variability derived from different sampling dates and illumination conditions within the greenhouse.

The performance of the models was evaluated using standard object detection metrics: precision, recall, F1-score, mAP50, and mAP50-95.

Results and discussion

Figure 3 shows the evolution of F1-score and precision across representative phenological classes, revealing distinct learning dynamics among the models. A general stabilization trend was observed around epoch 40, indicating that the main spatial and spectral features were already being captured. Peak performance for the single-class models occurred between epochs 80 and 100. In contrast, the multipurpose model (*Multiphen*) required additional training to achieve convergence.

The *C_stage* model achieved the highest stability and accuracy (F1-score = 0.89; precision = 0.94), reflecting the larger object size and higher visual distinctiveness of fully developed flowers. The *S_sepals* model also showed early convergence (epoch 40), with moderate but consistent performance (F1-score = 0.61; precision = 0.54).

In contrast, the *Multiphen* model, which initially reached F1 = 0.61 and precision = 0.73 after 100 epochs, still showed a positive loss gradient—indicating incomplete convergence. Therefore, the training was extended to 200 epochs, leading to a marked improvement in all metrics (F1 = 0.75; and precision = 0.78). This confirms that the model retained learning potential and required additional iterations due to the complexity of handling nine phenological classes simultaneously.

These results highlight that model performance was influenced not only by dataset imbalance but also by the intrinsic learning dynamics of multi-class architectures. Training under real production conditions, with natural variability in floral stages (1,650 labels for *S_color* vs. 147 for rice), provided a realistic scenario that strengthened the practical applicability of the models for greenhouse phenological monitoring.

Figure 4 summarizes the average performance of YOLOv8 models trained for single and multipurpose tasks, evaluated using F1-score, precision, recall, mAP50, and mAP50-95. Models trained individually for advanced stages exhibited the best overall performance, particularly *C_stage*, which reached the highest mean scores across all metrics (F1-score = 0.87; precision = 0.89; recall = 0.86; mAP50 = 0.89; mAP50-95 = 0.51).

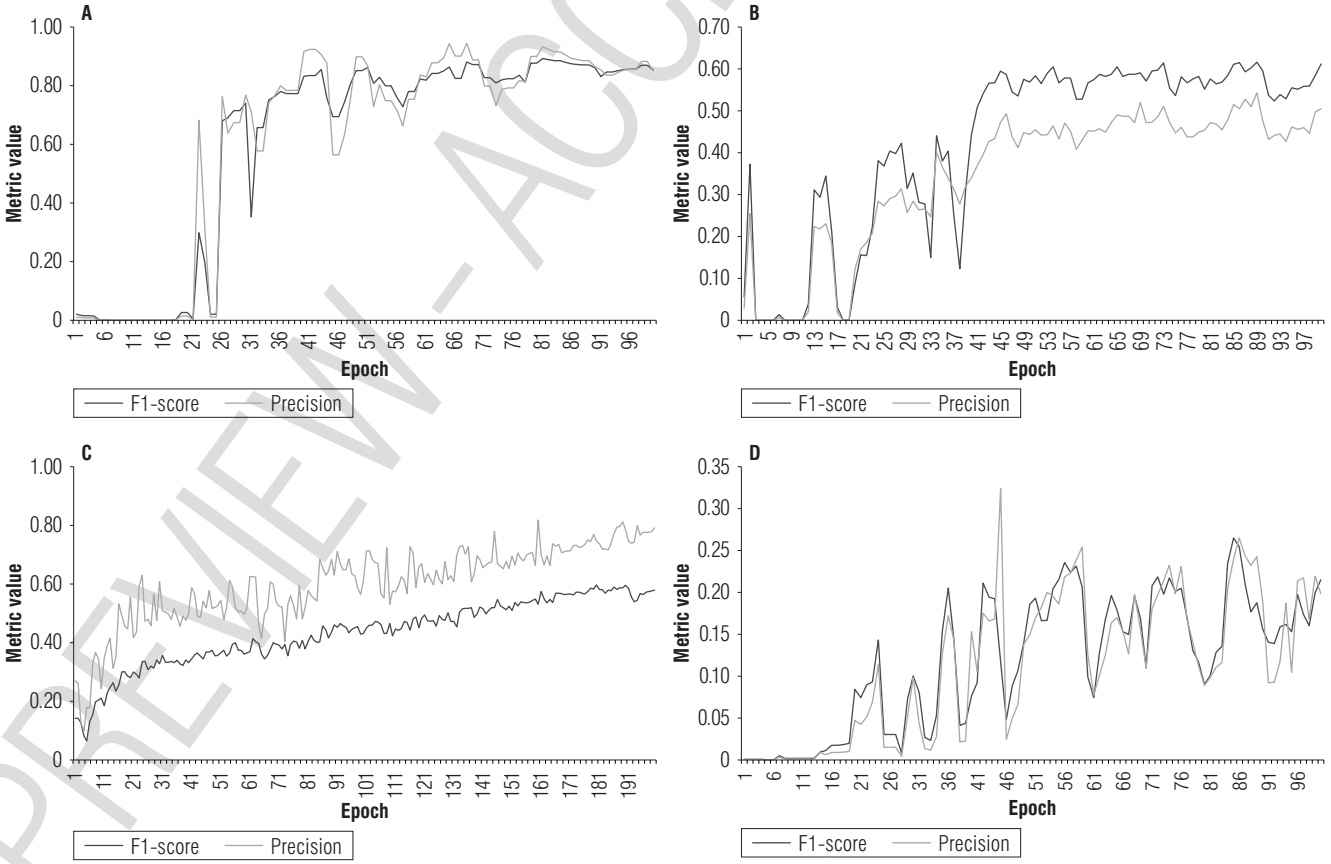

FIGURE 3. Evolution curves of epochs for F1-score and precision: A) *C_stage,* B) *S_sepals,* C) *Multiphen,* and D) *Chickpea*.

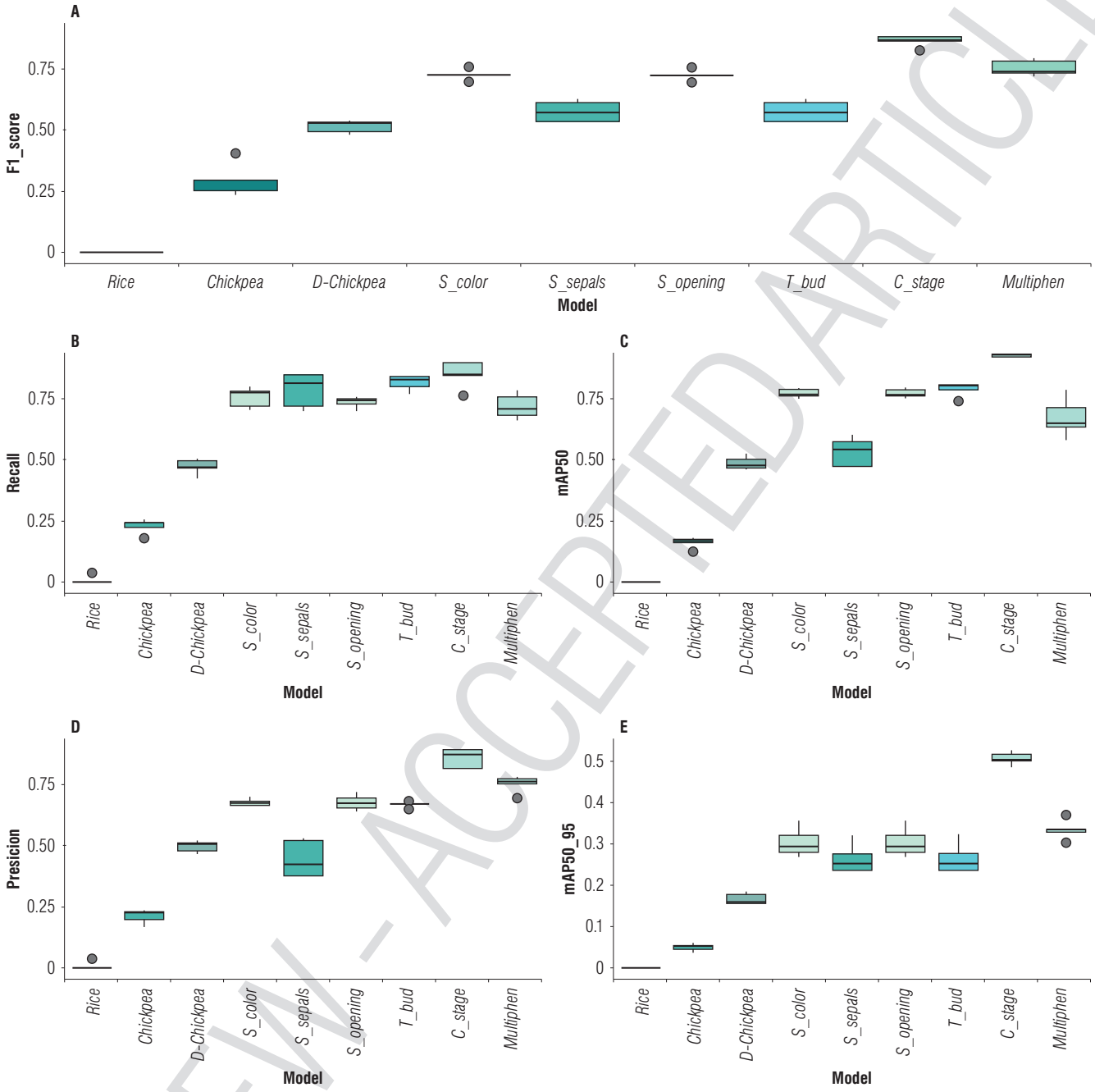

FIGURE 4. Performance metric distribution for detection models: A) F1-score, B) Recall, C) mAP50, D) Precision, and E) mAP50-95. Model description as in Table 1.

Models corresponding to visually distinctive and morphologically stable stages (*S_color* and *S_opening*) also performed well (F1 ≈ 0.72; precision ≈ 0.70). In contrast, early and small-structure stages *(Rice* and *Chickpea)* achieved lower accuracy due to their limited pixel representation and reduced number of labeled instances, confirming the influence of both scale and class imbalance. The *Rice* model, with only 147 labels, achieved F1 = 0.005 and precision = 0.003, indicating that it struggles to generalize under severe imbalance conditions.

The *Multiphen* model, trained jointly across all classes, initially exhibited lower convergence at 100 epochs. However, after extending the training to 200 epochs, its performance improved substantially (F1 = 0.75; precision = 0.78, approaching the results of the best single-class models. This

demonstrates that, despite its complexity, the multi-class approach is operationally feasible and particularly suitable for real-world production environments, where simultaneous identification of multiple stages reduces the need for various models and simplifies deployment.

The observed imbalance is consistent with natural production dynamics within the greenhouse, where intermediate stages such as *S_color* (1,650 labels) and *D_chickpea* (1,014) dominate during peak harvest periods. Similar effects of imbalance and class dominance have been reported by Sambasivam *et al.* (2021) in cassava disease detection and by Beloiu *et al.* (2023) in heterogeneous forest scenes, both of which note that minority classes yield lower metrics even under balanced training strategies. In agreement with these studies, our results confirm that model performance in *Rosa* spp. depends jointly on class representation, object scale, and the convergence strategy.

Figure 5 presents the normalized confusion matricerices of the segmentation models, including the multipurpose model (A) and the individual models *C_stage, S_opening, S_color,* and *Rice* (B–E). The multipurpose model showed strong consistency across classes, especially in *S_color, T_bud,* and *C_stage*, with diagonal values ≥ 0.80, indicating high agreement between predictions and reference annotations. These classes correspond to larger, more visually contrasting floral structures, which facilitates their detection and localization by the neural network.

In contrast, the classes *D_chickpea, Chickpea,* and *S_sepals* had higher confusion rates with adjacent stages or with the background, suggesting that the model had difficulty distinguishing morphologically similar structures, especially those occupying less than 0.2% of the image area. This confusion coincides with the influence of object size and the low representation of lower classes. These factors limit the discriminative ability of deep detectors.

In the individual models (B–E), there was better separation between the target class and the background, with accuracies above 0.90, indicating that specialization by class improves delimitation and reduced false positives. However, although these models achieved greater accuracy, their operational implementation was less efficient. In productive environments such as greenhouses, a single model capable of detecting multiple phenological stages—such as *Multiphen* retrained up to 200 epochs—represents a more scalable and economically viable alternative, even with moderate metrics. This trade-off between accuracy and applicability is also reported by Beloiu *et al.* (2023) in heterogeneous forest environments, highlighting the relevance of multi-class models for dynamic agricultural systems.

Table 2 summarizes the results obtained on the test set. Although a slight decrease in metrics is observed compared to the validation stage, the hierarchy among models remained consistent, confirming the stability and generalization capacity of the trained architectures. The model

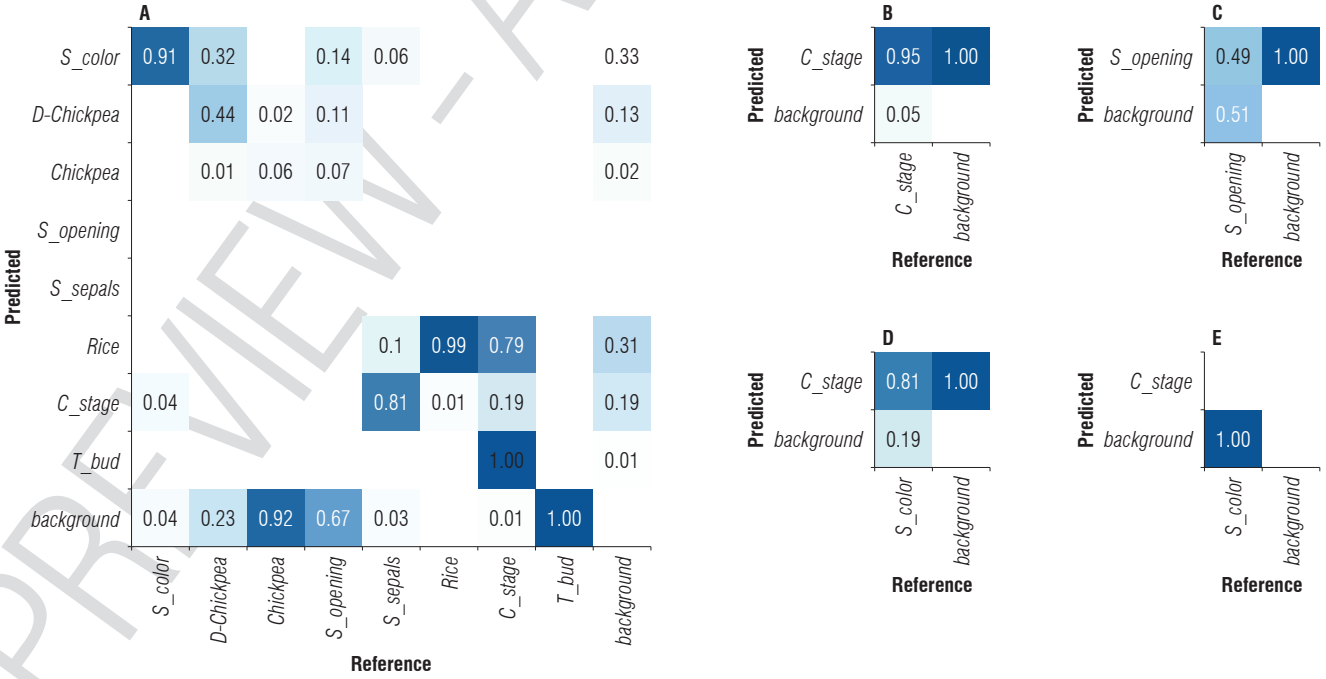

FIGURE 5. Confusion matrix for the models: A) *Multiphen,* B) *C_stage,* C) *S_opening,* D) *S_color,* and E) *Rice*.

corresponding to the *C_stage* retained the best overall performance (F1-score = 0.84, Precision = 0.85, Recall = 0.77), followed by *S_color*, *S_opening*, and *T_bud*, with F1-scores close to 0.70.

These results indicated that the models effectively generalize to new, unseen data, especially in the more morphologically defined floral stages. The *Multiphen* model achieved competitive performance (F1-score = 0.72, Precision = 0.72), reinforcing its potential for operational use in real greenhouse scenarios where simultaneous multi-class detection is desirable.

In contrast, *Chickpea*, *D_chickpea*, and especially *Rice* exhibited poor performance (F1 < 0.50), indicating the most significant classification difficulty. This behavior is mainly associated with the small size and subtle morphological features of early floral stages, which occupy less than 0.1% of the image area and are visually indistinguishable from the surrounding background. These limitations reflected real production conditions and highlighted the need for enhanced training strategies and improved imaging setups for early-stage detection.

The correlation analysis (Fig. 6) demonstrated that object size was the main factor influencing model performance. At the same time, the number of labeled samples per class had a weak relationship with both F1-score and mAP50-95. Specifically, object diameter in pixels was strongly correlated with F1-score ($r = 0.90$, $R^2 = 0.81$) and mAP50-95 ($r = 0.94$, $R^2 = 0.89$), whereas the number of labeled instances showed low correlations ($r = 0.32$, $R^2 = 0.10$; $r = 0.09$, $R^2 = 0.01$, respectively). These results indicated that

TABLE 2. Performance of the models by phenological class in the test set.

Model	F1-score	Precision	Recall	mAP50	mAP50-95
C _ stage	0.84	0.85	0.77	0.88	0.48
Multiphen	0.72	0.72	0.68	0.55	0.34
S _ color	0.70	0.69	0.71	0.72	0.27
S _ opening	0.69	0.67	0.71	0.72	0.27
S _ sepals	0.54	0.39	0.71	0.45	0.24
T _ bud	0.54	0.68	0.78	0.71	0.24
D _ chickpea	0.48	0.48	0.44	0.44	0.15
Chickpea	0.24	0.17	0.20	0.12	0.04
Rice	0.00	0.00	0.02	0.00	0.00

variation in model accuracy across phenological stages of *Rosa* spp. was primarily determined by the spatial scale and visual distinctiveness of the floral structures, rather than by class imbalance. Detecting early stages, such as rice and chickpea, which occupied less than 0.3% of the image area, proved considerably more challenging than identifying fully developed flowers with larger spatial footprints.

Similar patterns are reported in other crops and ecosystems. Li *et al*. (2023) achieve a mean average precision above 92% when counting rapeseed inflorescences with YOLOv5 + CBAM, yet note that small floral targets were the main limitation for precise localization. Likewise, John *et al*. (2024) find that flower abundance and small object size constrain detection accuracy in alpine meadows when using Mask R-CNN, RetinaNet, and YOLOv5. In line with those studies, our results confirmed that, even under relatively homogeneous greenhouse conditions, object

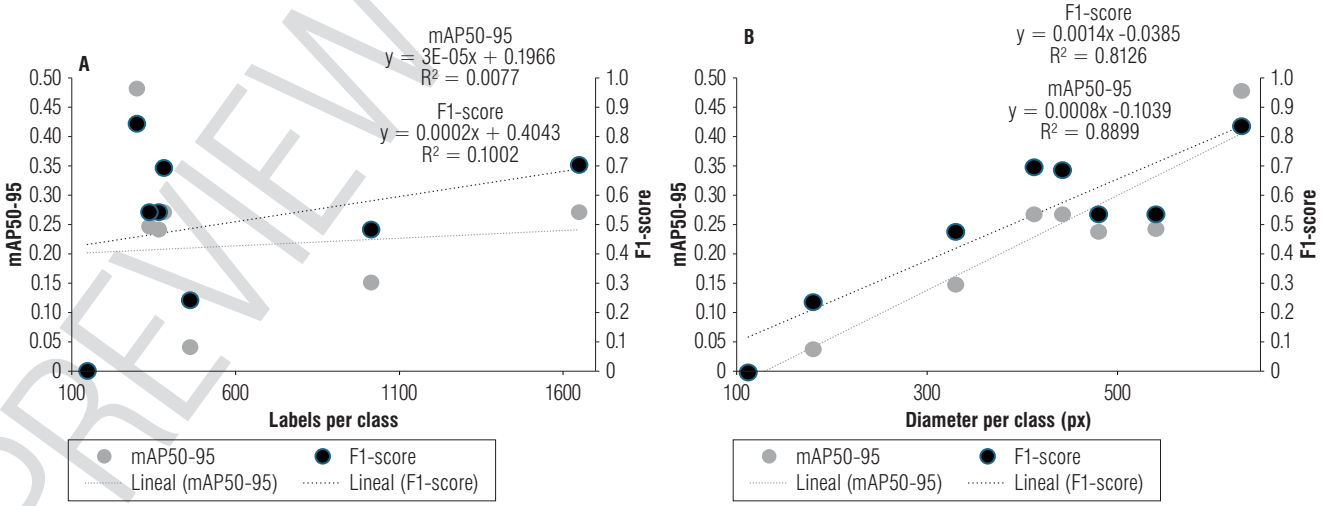

FIGURE 6. Relationship between model performance and object characteristics in *Rosa* spp. A) Correlation between performance metrics (F1-score and mAP50-95) and the number of labeled instances per class, B) correlation between metrics and average object size (diameter in pixels).

scale remained the dominant constraint for convolutional network performance in phenological stage detection of *Rosa* spp.

The quantitative results presented in Table 2 were complemented by visual examples from the test set. Figure 7 illustrates different scenarios in which the system detects and classifies the phenological stages of the stems, including cases with multiple objects per image and variations in lighting, leaf density, and flower position. These representations allowed us to appreciate how performance, as reflected in metrics such as F1-score, precision, and recall, translated into the visual identification of the different stages under real greenhouse conditions.

Overall, the results showed that individually trained models outperformed the multipurpose model across most metrics, with *C_stage* standing out for its consistency in validation and testing. However, the multipurpose model demonstrated stable performance and applicability in real greenhouse production environments. The detailed analysis by phenological stage highlighted challenges, including class imbalance and the physical limitations of image capture under greenhouse conditions, which were overcome using a mobile platform. These findings confirmed the potential of computer vision to optimize agronomic and logistical processes in rose production.

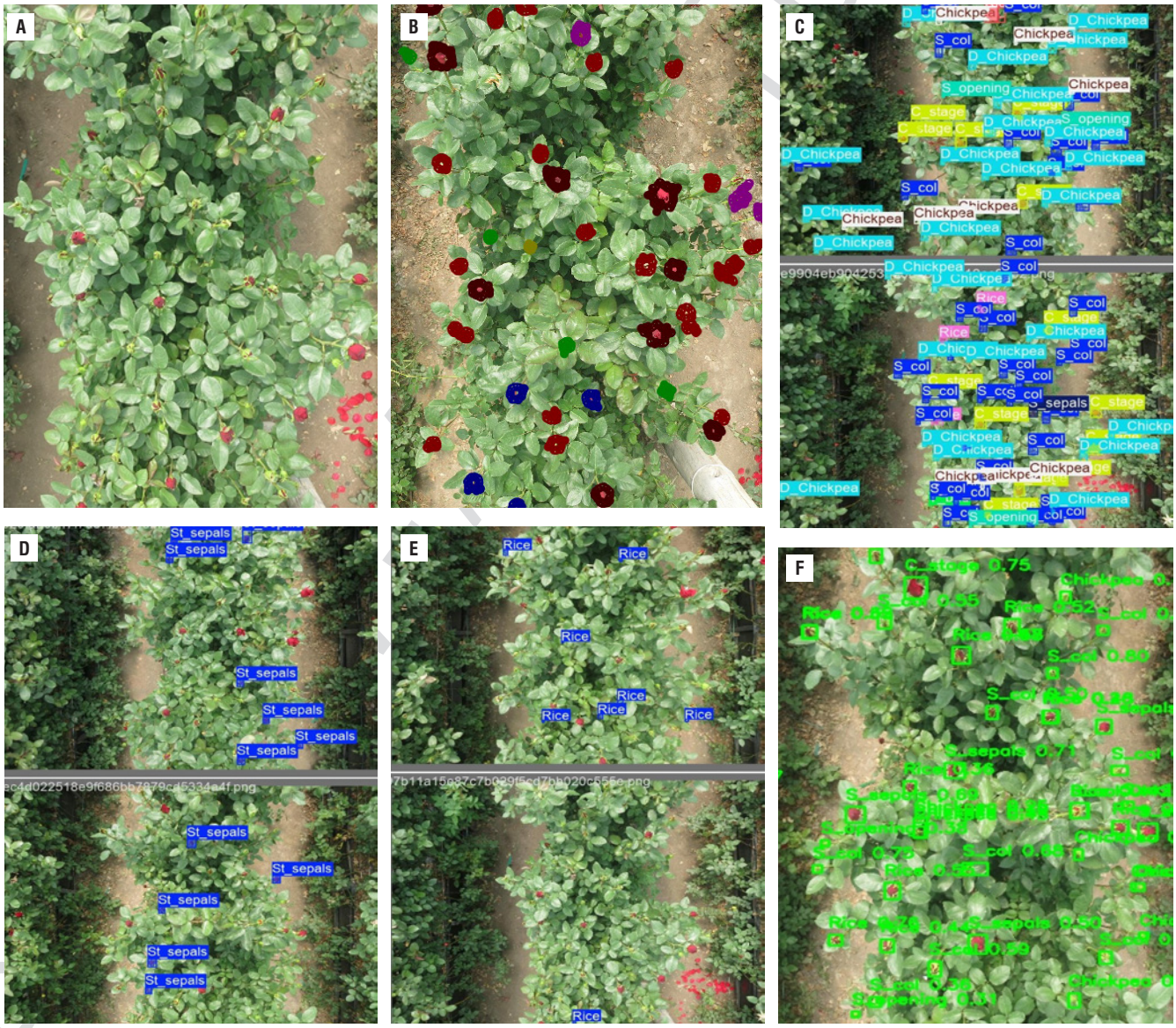

FIGURE 7. Detection by the trained convolutional neural network (CNN) in the test sites: A) RGB image; B) Manual identification of stages; C) *Multiphen* model; D) *S_sepals* model; E) *Rice* model; F) Test site image for illustrative purposes.

Conclusions

The implementation of YOLOv8 convolutional models for the automatic detection of phenological stages in greenhouse-grown roses demonstrated the feasibility of applying computer vision to complex ornamental production systems. The best results were obtained for advanced phenological stages such as *C_stage*, *S_color*, and *S_opening*, where larger, morphologically well-defined floral structures facilitated feature extraction and consistent model learning. In particular, the *C_stage* model achieved an F1-score of 0.87 on validation and 0.84 on test, confirming both robustness and stability across training.

In contrast, early stages such as rice and chickpea exhibited limited performance, mainly due to the tiny size of floral buds (less than 0.1% of the image area) and the low number of labeled instances. The correlation analyses indicated that object size exerts a stronger influence on model performance than class imbalance, highlighting a critical factor for future optimization efforts.

The multipurpose (*Multiphen*) model initially showed moderate metrics; however, after extending training to 200 epochs, its performance improved substantially, demonstrating the relevance of training depth in multi-class architectures. Although its precision remains lower than that of individual models, its operational versatility makes it a practical tool for simultaneous detection of multiple phenological stages under real greenhouse conditions.

From an applied perspective, this research provided a reference framework for integrating deep-learning-based phenological detection into production monitoring systems for ornamental crops. Future research should explore (i) new YOLO versions and alternative deep-learning architectures, (ii) optimization of hyperparameters to improve convergence and class separation, and (iii) innovative image acquisition strategies, including the use of mobile phones, panoramic or fixed cameras for block scanning, mobile ground platforms, or even UAV-based systems where feasible.

This study established a scientific and technological precedent for the use of deep learning in greenhouse floriculture, paving the way for scalable, non-destructive, and real-time phenological monitoring systems applicable to other ornamental and high-value crops.

Conflict of interest statement

The authors declare that there is no conflict of interests regarding the publication of this article.

Author's contributions

OFM designed the conceptual framework and defined the study objectives. OFM supervised the experiments, managed data collection and system maintenance, and coordinated funding. OFM and JLFM performed the statistical analyses and implemented the computer code. OFM and JLFM conducted the field and laboratory experiments and generated effective visual representations of the data and results. LJMM and OFM jointly developed the research methodology, including the data-acquisition strategy and equipment selection. They validated the accuracy and reliability through a rigorous review process. JLFM and OFM contributed to the development and implementation of the code used for model training. They participated in the model evaluation and validation. OFM prepared the initial draft of the manuscript. All authors participated in the critical review and approved the final version of the manuscript.

Literature cited

Ballena-Ruiz, J., Arcila-Diaz, J., & Tuesta-Monteza, V. (2025). Automated detection and counting of *Gossypium barbadense* fruits in Peruvian crops using convolutional neural networks. *AgriEngineering*, *7*(5), Article 152. https://doi.org/10.3390/agriengineering7050152

Beloiu, M., Heinzmann, L., Rehush, N., Gessler, A., & Griess, V. C. (2023). Individual tree-crown detection and species identification in heterogeneous forests using aerial RGB imagery and deep learning. *Remote Sensing*, *15*(5), Article 1463. https://doi.org/10.3390/rs15051463

Estrada, J. S., Vasconez, J. P., Fu, L., & Cheein, F. A. (2024). Deep learning based flower detection and counting in highly populated images: A peach grove case study. *Journal of Agriculture and Food Research*, *15*, Article 100930. https://doi.org/10.1016/j.jafr.2023.100930

Franco Montoya, O. H., & Martínez Martínez, L. J. (2024). Relationship between spectral response and manganese concentrations for assessment of the nutrient status in rose crop. *Agronomía Colombiana*, *42*(2), Article e110294. https://doi.org/10.15446/agron.colomb.v42n2.110294

Franco Montoya, O. H., & Martínez Martínez, L. J. (2025). Comparing spectral models to predict manganese content in *Rosa* spp. leaves using VIS-NIR data. *Agronomía Colombiana*, *43*(1), Article e118322. https://doi.org/10.15446/agron.colomb.v43n1.118322

ICA – Instituto Colombiano Agropecuario. (2024). *Con 700 millones de tallos, Colombia aporta variedad, color y belleza a la celebración de San Valentín*. https://www.ica.gov.co/noticias/ica-colombia-exporta-flores-san-valentin-2024

John, A., Theobald, E. J., Cristea, N., Tan, A., & Hille Ris Lambers, J. (2024). Using photographs and deep neural networks to understand flowering phenology and diversity in mountain meadows. *Remote Sensing in Ecology and Conservation*, *10*(4), 480–499. https://doi.org/10.1002/rse2.382

Li, J., Li, Y., Qiao, J., Li, L., Wang, X., Yao, J., & Liao, G. (2023). Automatic counting of rapeseed inflorescences using deep learning method and UAV RGB imagery. *Frontiers in Plant Science*, *14*, Article 1101143. https://doi.org/10.3389/fpls.2023.1101143

Mann, H. M., Iosifidis, A., Jepsen, J. U., Welker, J. M., Loonen, M. J., & Høye, T. T. (2022). Automatic flower detection and phenology monitoring using time-lapse cameras and deep learning. *Remote Sensing in Ecology and Conservation*, *8*(6), 765–777. https://doi.org/10.1002/rse2.275

Pajula, M. (2022). *Berry density estimation with deep learning - Estimating density of bilberry and lingonberry harvest with object detection* [Master thesis, Oulu University of Applied Sciences]. https://www.theseus.fi/bitstream/handle/10024/786112/mikko_pajula.pdf?sequence=2

Qi, C., Nyalala, I., & Chen, K. (2021). Detecting the early flowering stage of tea chrysanthemum using the F-YOLO model. *Agronomy*, *11*(5), Article 834. https://doi.org/10.3390/agronomy11050834

Sambasivam, G., & Opiyo, G. D. (2021). A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egyptian Informatics Journal*, *22*(1), 27–34. https://doi.org/10.1016/j.eij.2020.02.007

Wang, X. A., Tang, J., & Whitty, M. (2021). DeepPhenology: Estimation of apple flower phenology distributions based on deep learning. *Computers and Electronics in Agriculture*, *185*, Article 106123. https://doi.org/10.1016/j.compag.2021.106123

Xiang, S., Wang, S., Xu, M., Wang, W., & Liu, W. (2023). YOLO POD: a fast and accurate multi-task model for dense Soybean Pod counting. *Plant Methods*, *19*(1), Article 8. https://doi.org/10.1186/s13007-023-00985-4

Zhou, X., Sun, G., Xu, N., Zhang, X., Cai, J., Yuan, Y., & Huang, Y. (2023). A method of modern standardized apple orchard flowering monitoring based on S-YOLO. *Agriculture*, *13*(2), Article 380. https://doi.org/10.3390/agriculture13020380