pagepress

Journal of Agricultural Engineering

Journal of Agricultural Engineering 2026; LVII:2003
doi:10.4081/jae.2026.2003

Real-time banana harvest readiness prediction using mobile SE-enhanced YOLO classification

Preety Baglat,[1,2] Fábio Mendonça,[1,2] Sheikh Shanawaz Mostafa,[2] Sidharth Gupta,[3] Francisco Silva,[4] Helena Garcês,[4] Ruben Sousa,[4] Diana Côrte,[4] Fernando Morgado-Dias[1,2]

[1]Faculty of Exact Sciences and Engineering, University of Madeira, Funchal; [2]Interactive Technologies Institute (ITI/LARSyS and ARDITI), Funchal; [3]Spice P S.A. - SwissPost IT Lisbon Campus, Lisboa; [4]Gesba-Empresa de Gestão do Setor da Banana, Lda, Funchal, Portugal

Abstract

A digital banana harvesting solution was developed to improve the speed and consistency of banana harvesting by integrating real-time bunch detection with harvest-readiness classification into a mobile decision support system used directly in the field. The banana bunch detection module utilizes a You Only Look Once (YOLO) model trained on a custom dataset collected under real plantation conditions, enabling consistent performance across varied environments. Specifically, a YOLOv12n detector was used for banana bunch detection, achieving 93% AP50-test with an inference latency of 5.1 ms per image, making it suitable for mobile deployment in plantation environments. For the readiness of harvesting prediction, a second model was developed, based on a squeeze-and-excitation YOLO classifier, using annotated images gathered with guidance from harvesting experts. In this work, this SE-enhanced YOLO classifier is used as a lightweight, task-specific YOLO classification backbone for the binary "cut" vs "keep" decision, and this harvest-readiness classifier achieved 94% accuracy with an inference time of 2.8 ms per image. Then, an application was built using Flutter and Dart, which uses intuitive interfaces for both field operators and administrators, and includes integrated feedback mechanisms to collect user input and support continuous model refinement. Field testing across diverse lighting and environmental conditions, as well as usability assessments with expert harvesters and administrative staff, demonstrated reliable performance with potential to contribute to faster decision-making and reduced manual labour.

Key words: banana bunch detection; harvest readiness prediction; image analysis; precision agriculture; user feedback.

Correspondence: Preety Baglat, Faculty of Exact Sciences and Engineering, University of Madeira, 9000-082 Funchal, Portugal.
E-mail: preety.baglat@iti.larsys.pt

Introduction

Bananas are among the world's most widely consumed fruits and are of utmost relevance in global trade due to their year-round availability and nutritional value (Fu *et al*., 2020). Each ripening stage has unique characteristics due to changes in its nutritional and chemical composition, making accurate assessment of maturity an important aspect for quality control (Ni *et al*., 2020; Dewi *et al*., 2021).

Commercial growers typically harvest at the green stage of maturity when fruit firmness, colour, and chemical composition are optimal for transport and storage (Dadzie and Orchard, 1997). International standards such as the Codex Alimentarius (CXS 205-1997) mandate uniform green colouring, absence from serious defects, and a firm texture for dessert bananas (*Musa* spp. AAA) (FAO and WHO, 2022). Yet, harvesting operations remain vulnerable to factors such as weather fluctuations, labour shortages, and human fatigue that can compromise visual grading and lead to premature or delayed picks (Bac *et al*., 2014; Sa *et al*., 2016). In practice, harvesting decisions rely mainly on a small group of highly competent specialists, who have gained knowledge *via* years of field experience. However, this creates challenges because the number of such experts is steadily declining as younger generations are less inclined to enter the agricultural domain. Much of the expertise needed to judge the right cutting time comes from experience, and it is difficult to document, which makes it hard to transfer consistently to new workers. As a result, harvesting outcomes can vary when less experienced labour is involved, and it also leads to bad decisions and loss of crop. These problems highlight the need to digitalize expert knowledge in decision-support systems, so that banana production can benefit from more consistent, scalable, and reliable harvesting practices without depending heavily on a limited group of harvesting experts.

Recent works in computer vision have reported promising results for automated grading and sorting of banana crops (Altaf *et al*., 2020; Hayat *et al*., 2024). Building on these, this work presents a digital harvesting solution based on a mobile application that combines real-time detection of banana bunches with an expert-trained classifier to predict harvest readiness. For bunch localization, the application utilizes a You Only Look Once (YOLO) detector that is pre-trained and fine-tuned on a field-collected dataset (Baglat *et al*., 2025a), specifically designed for detection

tasks, achieving 93% $AP^{50\text{-test}}$ and an average inference time of 5.1 ms per image. For harvest-readiness classification, a separate dataset (Hayat *et al*., 2023), assembled with the guidance of harvesting experts, was used to train another YOLO model to identify the visual cues associated with harvesting decisions.

The primary objective of this study is, therefore, to assist farmers in making informed decisions about harvesting banana bunches. For this purpose, the specific contributions of this paper are:

i) Design and implement an end-to-end mobile decision support system that integrates banana bunch detection and harvest-readiness classification into a single workflow for use directly in the field.
ii) Proposed a Squeeze and Excitation (SE) YOLO classifier for the harvest-readiness task, trained on expert-labelled images and optimized for edge devices.
iii) Use two field-collected datasets for bunch detection and harvest readiness, both collected under real plantation conditions with input from harvesting experts.
iv) Field testing and usability evaluation with expert harvesters and administrative staff to assess the system's performance, usability, future improvements, and potential impact on harvesting decisions.

Literature review

Machine learning and mobile technologies have been increasingly applied in agriculture to automate tasks such as fruit detection, ripeness classification, and yield estimation. Specifically, multiple studies have examined the use of deep learning for banana crop management, focusing predominantly on the classification of ripeness stages, variety identification, and postharvest quality analysis.

Several works have investigated banana bunch detection using object detection models. YOLOv4 model for bunch detection, achieving a detection rate of 99.29% with an average precision (AP) of 99.95% and an average inference time of 171 ms per image, outperforming YOLOv3 (Fu *et al*., 2020). In a subsequent study (Fu *et al*., 2022b), proposed YOLOv-Banana, a lightweight model optimized for banana bunches and stalks detection, which reached a lower mean AP (mAP) of 92.19% but reduced inference time to 35.33 ms, making it more efficient for real-time use. The AP of the YOLO-Banana detection model on banana bunches is 98.4% and stalks is 85.98%. YOLOv3 (Zhang *et al*., 2021) study detect banana stalks and strings with AP values of 97.96% and 88.45% respectively, with improvements for small-object detection under dense canopy conditions. Wu *et al*. (2021) extended YOLOv3 to multi-target detection of banana fruits, buds, and inflorescence axes, achieving accuracies of 92.98% and 93.46% with execution times of 240 ms and 200 ms, respectively, while indicating robust performance under challenging lighting conditions. Other approaches include Fu *et al*. (2019), who applied traditional machine learning (support vector machine with histogram of oriented gradients and local binary patterns features) for bunch detection, achieving up to 92.55% accuracy under natural environments (Fu *et al*., 2019). Later, Fu *et al*. (2022) study returned to YOLOv4 for bunch and stalk detection, reporting AP values of 99.55% (banana) and 87.82% (stalk), with robustness under varying illumination and occlusion (Fu *et al*., 2022a). More recently, Baglat *et al*. (2025) presented a comparative evaluation of YOLO versions from v1 to v12 for banana bunch detection on a field-acquired dataset (Baglat *et al*., 2025a). Their analysis identified YOLOv12n as the most suitable model, achieving 93% $AP^{50\text{-test}}$ and 51% $AP^{50\text{-95test}}$, with an inference latency of 5.1ms.

Few studies address banana bunch harvesting readiness classification, which restricts direct comparisons. However, the ripeness stages have been widely studied and are also one of the criteria for banana harvesting readiness. Previously, Zhang *et al*. (2018) proposed a Convolutional Neural Network (CNN) based model for classifying banana ripeness and achieved 94.4% and 92.4% accuracy for 7 ripeness stages and 14 ripeness stages, respectively (Zhang *et al*., 2018). Their model used a fine-grained analysis of image features by capturing slight differences in colour and texture between adjacent ripeness stages through a combined softmax and triplet loss framework. Piedad *et al*. (2018) applied a random forest classifier to postharvest Musa acuminata 'Lakatan', reporting classification accuracy of 94.2% (Piedad *et al*., 2018). Chuquimarca *et al*. (2023) addressed data scarcity by augmenting small datasets with synthetic images, achieving 91.7% accuracy following transfer learning. More recently, Wang *et al*. (2025) study introduced an Enhanced YOLO v9 architecture incorporating attention mechanisms for drone-based detection and classification of banana bunches and ripeness levels.

In parallel, various datasets such as BananaImageBD (Ferdaus *et al*., 2025) and feature extraction approaches, including Vision Transformers (ViT) (Knott, Perez-Cruz and Defraeye, 2023) and support vector machine classifier using ViT embeddings (Ergün, 2025), have been explored for fine-grained classification of bananas. These studies reflect the ongoing progress in computer vision applications for banana-related classification tasks. However, existing literature mostly focuses on classification under controlled laboratory or postharvest conditions**,** without addressing the complexities of field-based, real-time harvest decision-making**.** Most prior work emphasizes ripeness prediction for postharvest quality management or consumer information, rather than supporting field workers during the harvest process itself.

In an earlier study (Hayat *et al*., 2024) on banana harvesting readiness classification, conventional CNN-based transfer learning models, including DenseNet121, MobileNetV2, VGG19, and others, were employed. These models were fine-tuned on a curated dataset of banana bunches categorized as ready and not ready to harvest, achieving 83% accuracy. Although the examined models reached promising results, they have relatively high inference latency and suboptimal accuracy under complex field conditions (*e.g.,* occlusion or varying maturity stages). Moreover, these CNN architectures were originally designed for general-purpose classification and did not leverage the spatial context provided by object detection pipelines. They also lacked attention mechanisms to prioritize key discriminative features. Furthermore, their work did not progress to real-time deployment, which this work intends, so that it can provide a practical tool for harvesters.

Despite advances in classification and detection models**,** no existing studies have presented an integrated mobile solution capable of real-time banana bunch detection, harvest-readiness prediction, and user feedback collection in the field. Such is of high relevance as accurate and timely harvest decisions directly influence logistics, product quality, and operational efficiency in banana production. Overall, from literature review, it was observed that YOLO based models reach a good performance in the type of problem under consideration and are known for their real-time inference capabilities. Furthermore, transformer-based models also showed promising results in computer vision, mostly due to the attention mechanisms. Thus, this study proposes a combined detection-classification pipeline**,** utilizing a YOLOv12n detection

model and a custom YOLOv11n classifier that uses a channel-wise attention mechanism based on SE, which is embedded within a mobile application (v12 was also tested but reached a lower performance than v11). The rationale for this solution is to have a tool for harvesters that can operate in real-time, in-field decision-making, and incorporate structured feedback mechanisms for continuous model refinement, aligning with the evolving needs of precision agriculture.

Materials and Methods

The mobile application was built in Flutter (using Dart), which provides cross-platform development capabilities that enable code reuse across different platforms, reducing development time and maintenance overhead. For this implementation, the application was deployed exclusively on Android devices to ensure consistent testing conditions and facilitate field deployment within the target environment. For bunch detection, the application integrates a pre-trained YOLOv12n model fine-tuned on a custom dataset of banana-bunch images collected under real plantation conditions. Once a bunch is detected and cropped, the classification pipeline uses an SE-enhanced variant of YOLOv11n to predict harvest readiness. The classification model was trained on a second expert-annotated dataset, where each image was labelled as harvestable or not. During field trials, harvesters captured images through the application, which logged inference times, prediction confidence scores, and user interactions. After harvesting sessions, a structured questionnaire recorded harvester feedback on perceived performance, responsiveness, and ease of use of the application. All telemetry (model timings and prediction outputs) and survey responses were synchronized back to a central server for analysis.

Dataset collection

Data were collected from four fields in Madeira Island, Portugal, in Santo António, Câmara de Lobos, Ponta do Sol, and Lugar de Baixo. Images were captured using mobile phones (Samsung Galaxy A12, Samsung Galaxy Note 9, and OnePlus 9) under various environmental conditions throughout the year. This approach was taken to better represent practical scenarios to improve the system's robustness. For the bunch detection model, a total of 2,841 samples (Baglat *et al*., 2025a) were taken under different light conditions, and from which 662 images were removed during data labeling due to multiple factors such as unclear images, obstacle covering banana bunches, too much light exposure, and banana bunches covered under plastic bags. For the harvesting dataset of 2,685 images (Hayat *et al*., 2023) of banana bunches were collected along with the banana harvesting experts' team. The research team visited the fields with the harvesting expert team 29 times, taking pictures of all bunches before they were harvested and marking those bunches, and later (on the same day) taking photographs of non-harvested bunches. Among the collected images, 1,143 were harvested (or "Cut" images), whereas 1,542 images of unharvested bananas were indicated as "Keep". To differentiate between images, a number plate system was used with "X" and "A" for "Cut" and "Keep", respectively. Figure 1 illustrates the banana bunch samples for harvesting prediction and bunch detection.

Object detection and classification model

Banana bunch detection and harvest-readiness prediction were automated using a two-stage deep learning pipeline based on YOLO models. The first stage is a banana bunch detection and filtering of the single banana bunch. In a real-world environment, when a picture is taken, there can be multiple banana bunches, and a filter is applied to select the bunch in focus, passing this single bunch to the classification model. Then the classification model is used to detect the harvesting readiness of the filtered bunch. A classification model based on YOLOv11n has been used, which includes SE blocks. Figure 2 shows the proposed framework for the banana harvesting prediction using YOLO models and a mobile application.

YOLOv11 refines the Ultralytics YOLOv8 CNN architecture by redesigning the backbone and neck to achieve higher mAP with

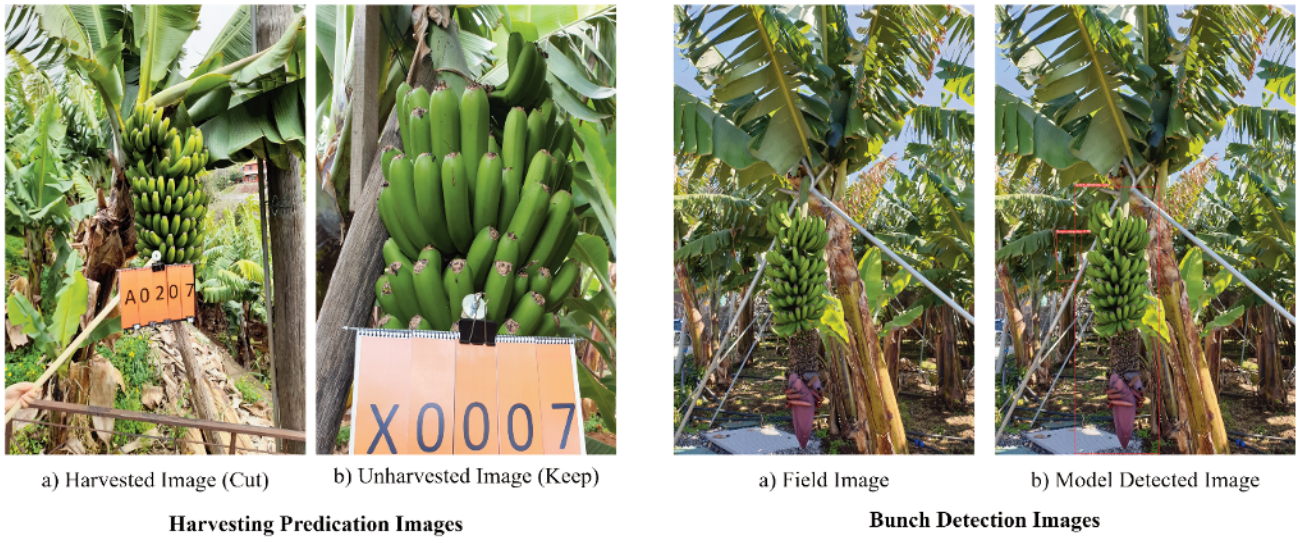

Figure 1. Banana bunch samples for harvesting prediction and bunch detection.

fewer parameters and FLOPs, and substantially faster inference, which is advantageous on resource-constrained edge hardware. In particular, YOLOv11 replaces the Cross Stage Partial C2F blocks used in YOLOv8 with more efficient Cross Stage Partial with kernel size 2 (C3K2) backbone blocks and employs a Spatial Pyramid Pooling Fast (SPPF) neck with Cross-Stage Partial Position-Sensitive Attention (C2PSA) attention modules that provide fast multi-scale pooling while selectively emphasizing informative spatial regions. YOLOv12 further extends this line with an attention-centric design that introduces lightweight attention-based feature aggregation, trading a modest increase in complexity for additional accuracy.

In this study, YOLOv11n and YOLOv12n were used as lightweight, off-the-shelf YOLO variants, following a strategy similar to a previous study (Dai *et al*., 2024) that reuses YOLO detector backbones and necks as feature extractors and replaces the detection head with a task-specific, attention-based classification module. Our main novelty, therefore, lies not in proposing a new detector architecture but in how these compact YOLO variants are combined, adapted with SE blocks for the harvest-readiness task, and integrated into a mobile decision-support system for banana bunch management.

A previous work (Baglat *et al*., 2025b) on YOLO generations applied to banana bunch identification concluded that YOLOv12n demonstrated the most balanced performance, achieving an AP^{50test} of 93%, and an average inference latency of 5.1 ms per image. The same detection framework is employed in this work without modification. However, a subsequent layer of filter was added so that, after the model detects all the banana bunches in the images, the filter will crop the image based on the detected bounding box around the bunch that is focused on the image. If no bunch is detected, then the filter will trigger a "No Bunch Detected" flag, and this image will not be further analysed.

The cropped banana bunch image is then passed to the harvest-readiness classification model, designed to distinguish between "Cut" and "Keep" categories. This classifier is based on a modified YOLOv11n-cls architecture enhanced with SE blocks**,** which were integrated into both the backbone and the head to improve channel-wise feature recalibration. These SE modules were used to improve the network's ability to focus on discriminative features relevant for harvest decisions. The rationale was that an enhanced dynamic channel-wise feature recalibration (Hu *et al*., 2018) can improve class-specific attention of YOLO. For this purpose, first, the squeeze operation applies global average pooling across each feature map to generate a compact channel descriptor. This is defined as:

$$z_c = \frac{1}{HW}\sum_{(i,j)\in\Omega} u_c(i,j) \qquad \text{(Eq. 1)}$$

where $u_c(i,j)$ is the activation at position (i,j) for channel c, and Ω denotes the spatial domain over height H and width W. This helps in summarizing the overall activation strength of each channel by reducing each feature map to a singular scalar, which then is used in the excitation block to generate channel-wise weight. This step also reduces the risk of overfitting, as a global average pooling was used instead of flattening the feature map. In the excitation step, this descriptor is passed through a two-layer fully connected bottleneck network with a non-linearity based on rectified linear unit, and then a sigmoid activation is used to generate channel-wise weights. These weights are then employed in the recalibration step, where each channel is rescaled to prioritize important features while suppressing less informative ones.

Mobile application design and implementation

The implemented architecture of the mobile application is pre-

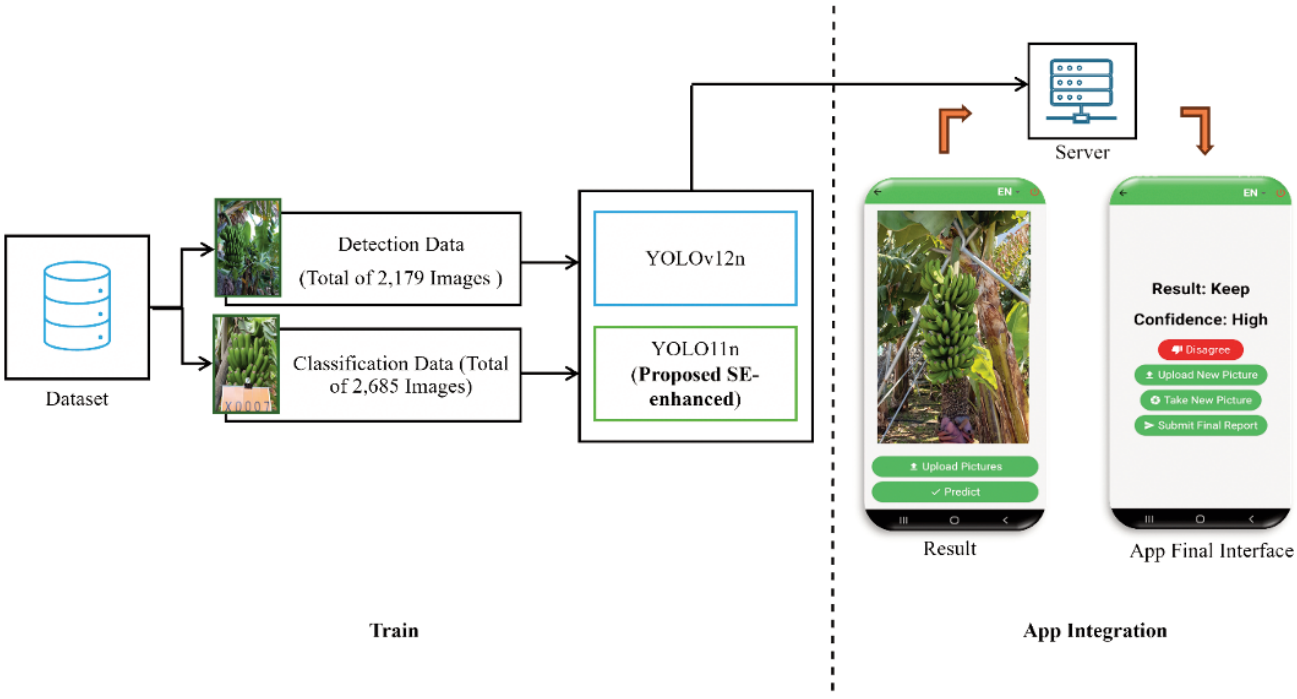

Figure 2. Proposed framework for banana bunch harvesting decision support.

sented in Figure 3. A web-based client-server solution was adopted to address performance and scalability constraints identified during the initial testing. When evaluated with local deployment, the mobile application exhibited substantially slower inference times due to the computational demands of deep learning models on resource-constrained mobile hardware. Additionally, performance varied substantially across different mobile devices based on their processing capabilities, memory availability, and chipset architectures. Thus, to ensure consistent performance and broader device compatibility, the YOLO models were deployed on a centralized backend server developed using a Flask framework. The mobile application communicates with these models through Application Programming Interfaces (APIs), enabling real-time inference by using the dedicated server hardware. This architecture also facilitates model updates and maintenance without requiring application redistribution to end users.

Communication between the mobile client and server is established using the Hypertext Transfer Protocol Secure (HTTPS), which provides a standardized method for data exchange over the Internet. The mobile application utilizes both GET requests to retrieve information from the server and POST requests to submit image data for inference processing. The server responds with structured data in JavaScript Object Notation (JSON) format, containing the requested information, and, if it is for a classification, it returns the model forecast. The application does not require a constant internet connection during the whole field session. Users can take photos even when there is no network coverage, and these images are stored locally on the device. When a connection becomes available, the user can upload the stored images through the application and obtain predictions. In this way, only the moment of sending the images and receiving the predictions depends on network connectivity.

Regarding the user flow, after secure login, the users must select a field from which the data they will analyse is collected (Figure 4). The idea is for the application management to know from which field the samples are collected. This can be done using a quick response code scan or manual entry (if the users do not intend to specify the field, they can select a "default field" option that allows them to use the application without sharing the field identification), which takes them to the next screen for selecting the mode of inputting the image. This screen has 3 options, specifically, to open a camera, to upload a picture from the gallery, and to return to the previous screen to change the field.

The user can either use the camera to take the photos or upload multiple photos for a decision on them. When the user clicks on the camera option, the camera opens, and the user can take multiple photos of the same bunch and touch the predict button at the bottom of the screen to send them for prediction. The prediction score for each image, rounded to 2 decimal digits for each one, is averaged out to provide the final prediction score for banana bunch readiness. This comprises a kind of ensemble approach where the user can provide multiple viewpoints of the same bunch.

This way, this multi-image ensemble methodology addresses the inherent variability in banana ripening patterns within a single bunch, where part of the bunch may exhibit different maturation levels due to factors such as position within the bunch, sunlight exposure, and microenvironmental conditions. The hypothesis is that by capturing multiple perspectives and computing the averaged prediction score, the system can reduce the impact of outlier predictions from individual images and provide a more reliable maturity assessment that better represents the bunch's overall readiness for harvest. This approach also helps mitigate potential imaging artifacts, lighting variations, and partial occlusions that might affect single-image predictions, potentially improving the performance and reliability of the maturity classification system.

The result is then displayed on the "Result" screen, where the

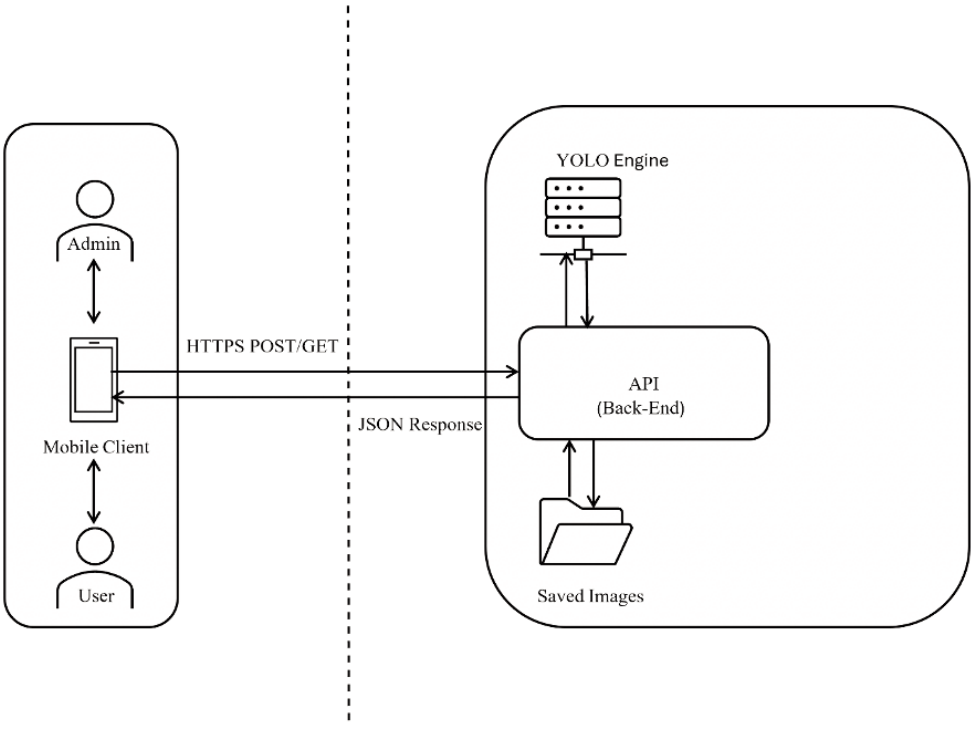

Figure 3. System architecture of the mobile application with YOLO-based image analysis backend.

user can see if the bunch is marked as “Keep” or “Cut”. Along with this, the confidence of the model is also displayed to the user, showing how confident the model is about its decision. It is important to note that all images indicated as “No Bunch Detected” are not considered in the average, which will only be performed if at least one image with a detected bunch is provided. Otherwise, the user will receive a message indicating “No Bunch Detected”. Confidence is categorized based on the average prediction score. It is marked as «High» when the score is equal to or greater than 80%, labelled as «Moderate» when the score falls between 65% and 79% (both values inclusive), and considered «Low» when the score is lower than 65%. This screen also allows the users to disagree with the result if they feel the model’s prediction is different from their on-field opinion. If the user presses the disagree button, the image is stored in the central server at a location and is marked properly for further analysis and future refinement of the model. From the same screen, users can now again upload or take more photos and keep collecting the results, and at the end, can submit the final report to produce a log of the session summary. Figure 4 shows the indicated procedures for the mobile application user interface for banana bunch harvesting decision support. The application is also multilingual and currently supports English and Portuguese. The administrative flow, presented in Figure 5, indicates how the administrators can access the application, which is

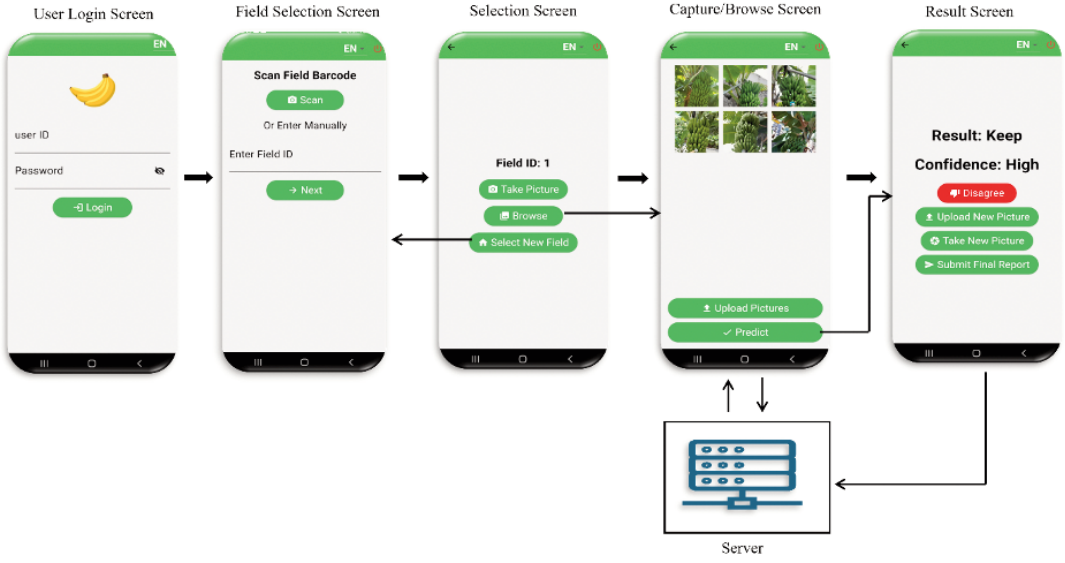

Figure 4. User flow interfaces of the proposed mobile application.

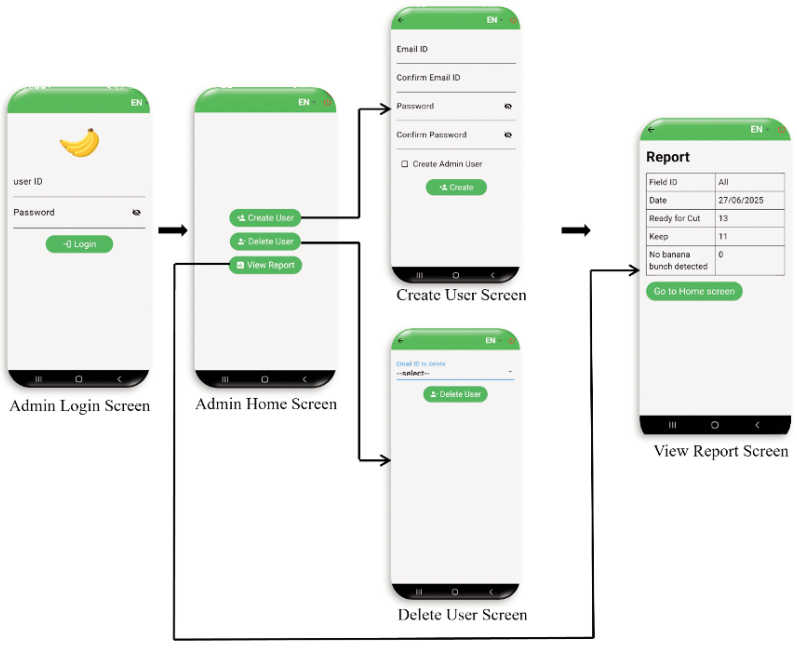

Figure 5. Administrative flow interfaces of the proposed mobile application.

controlled by the administrative screen. This screen allows for creating and managing users and accessing reports of the banana bunches. Administrators need to first log in to access this area, and they can create or delete user accounts using API endpoints designated for user management. The administrative flow also allows viewing per-field harvest reports that aggregate counts of "Cut," "Keep," and "No Bunch Detected" decisions for specified dates and fields, allowing for oversight of harvesting operations and model performance over time.

Each task within the mobile application, such as classifying banana bunches, managing user accounts, or submitting feedback, is mapped to a distinct API endpoint. These endpoints handle core backend logic, log request activity for monitoring, and return responses in the JSON format, which is widely adopted for lightweight, structured data exchange in mobile ecosystems.

Field evaluation and usability feedback

The field validation and test consisted of two main parts. The first one was a quantitative performance evaluation under real-world conditions. The second one was structured usability feedback that was collected from end users, including field harvesters and administrative staff. For the feedback acquisition, an initial clarification session was performed to inform the users on how to use the application and what features they can explore. A blank version of the 35-item questionnaire used to collect user feedback is provided in the *Supplementary File*.

The test version of the mobile application was downloaded by the harvesters in the field, using a file provided, and it was installed on various versions of Android and models of phones, including Xiaomi, Samsung, and Oppo. It was chosen to test the app when the field had ready-to-harvest bunches to include all test cases where every stage of banana bunch maturity could be covered. The testing happened in two phases, morning and afternoon, where the morning session happened on the field Lugar de Baixo (32.679952° N 17.086920° W) and the afternoon session was conducted at Ponta do Sol (32.680026° N and 17.086764° W). The users worked with the application in the banana field in various lightning conditions and with banana bunches at various maturity stages. Both test locations had mobile data coverage, and, during these field tests, the internet connection was stable enough to support real-time communication with the server for most predictions. The users were divided into different groups, who tested the application on different parts of the field, where fields had different altitudes and exposure, to cover all possible scenarios. The overall time taken during the usability testing was around 8 hours by 18 participants.

The feedback process was in two steps, with one online feedback step, and the other was a more intensive offline feedback step. The online feedback was recorded directly via looking at the log on the server and was more towards the correctness of model to their on-field judgement. Harvesters could provide feedback *via* the built-in "Disagree" option whenever the model's prediction did not match their judgment. The second feedback was usability feedback and was done offline. Overall, the server-based deployment effectively supported real-time harvesting decisions in the field while enabling ongoing feedback collection from users. However, the current design still requires at least occasional internet access, which can be a limitation in regions with no connectivity.

Results

To evaluate the effectiveness of harvest-readiness classification, a previously collected dataset (Hayat *et al*., 2023) of 2,685 banana bunch images at a resolution of 640×640 pixels was used to benchmark various models. This analysis evaluated multiple YOLOv11n-based architectures and a reimplementation of DenseNet121 for fair comparison with a previous study by Hayat *et al*. (2024), which used transfer learning on 256×256 images and achieved 83.2% accuracy, 85.4% precision, 82.0% recall, and an F1-score of 82.5%. For this study, all models were trained for 100 epochs (batch size 16), with early stopping (patience of 20 epochs) monitoring validation top 1 accuracy. For each model, the checkpoint with the highest validation top 1 accuracy was selected. Table 1 shows performance held on the test set, which was not used during training. The model did not apply additional training optimizations such as augmentations, label smoothing, or staged fine-tuning. This simplified setup was chosen to isolate model architecture performance.

Table 1. Comparative performance of YOLOv11n-based models and DenseNet121 for banana bunch harvest-readiness classification.

Models	Image size	Accuracy (%)	Precision (%)	Recall (%)	AUC (%)	Inference time (ms/image)	Params (M)	GFLOPs
YOLOv12n-cls (built from YAML, trained from scratch)	640	60.00	61.50	79.61	55.57	3.9	1.7	3.5
YOLOv11n-cls (built from Pretrained ImageNet-only)	640	92.00	93.51	93.20	97.66	2.6	1.5	3.2
YOLOv11n-cls (built from YAML, trained from scratch)	640	92.00	93.71	91.59	96.38	2.5	1.5	3.2
YOLOv11n-cls-SE (built from YAML, trained from scratch)	640	89.00	91.12	89.64	95.68	2.8	1.2	2.9
YOLOv11n-cls (built from YAML and pretrained weights)	640	92.00	93.51	93.20	97.66	2.0	1.5	3.2
YOLOv11n-cls-SE (built from YAML-SE pretrained weights	640	94.00	95.10	94.17	98.03	2.8	1.2	2.9
DenseNet121	640	78.29	73.41	97.41	88.43	12.76	7.0	46.5
DenseNet121 (Hayat *et al*., 2024)	256	83.20	85.40	82.00	-	-	-	-

Model performance

Building on state-of-the-art work's findings, this work employs a YOLO-based model with SE blocks, trained for harvesting readiness classification. The evaluation focused on accuracy, precision, recall, and Area Under the Receiver Operating Characteristic Curve (AUC), inference time, and model complexity. Table 1 presents the attained results. Among the models tested, YOLOv11n with SE blocks and pretrained weights (on ImageNet) performed best. Training stopped automatically at epoch 37 (due to early stopping), and this model reached 94% accuracy, 98% AUC, and an inference time of 2.8 ms per image. This is a considerable improvement over the standard YOLOv11n (pretrained on ImageNet) model (92% accuracy without SE), and during training, the best result was observed at epochs 16, highlighting the benefits of combining channel-wise attention with strong initial feature representations from pretraining.

When the YOLOv11n model was trained from a cold start, the accuracy was 92%, with the best result observed at epoch 52. However, with the SE-enhanced version, it dropped to 89%, indicating that, in this case, the standard YOLOv11n provides some benefit even without using pretraining. These results suggest that, for this relatively small binary classification task ("Cut" *vs* "Keep"), pretrained YOLOv11n backbones provide a good balance between model capacity and data size, and that SE blocks are most effective when combined with pretrained weights rather than when training from scratch.

Comparatively, YOLO12n, despite having more parameters and Floating-Point Operations Per Second (FLOPs), performed worse, achieving 60% accuracy with a higher inference time of 3.9 ms. Training stopped early at epoch 17, suggesting that model size alone does not guarantee better performance, especially when training from scratch. In this setting, YOLOv12n-cls showed unstable training behaviour and signs of overfitting, which is likely related to the limited dataset size and the binary nature of the task, in alignment with their official documentation. Thus, although YOLOv12n is a newer architecture and was effective for bunch detection in previous work, it did not transfer well as a classifier for harvest readiness. For this reason, YOLOv12n is retained for the detection stage (where it achieved the best trade-off between AP^{50test} and speed), whereas YOLOv11n-cls with SE blocks is used for the classification stage, where it clearly outperforms YOLOv12n on this dataset.

DenseNet121 achieved 78% accuracy with high recall but poor precision, leading to a high number of false positives. The model was also trained for up to 100 epochs with the same early stopping condition, and training stopped at epoch 66. Moreover, it was considerably slower (12.8 ms per image) and more resource-intensive, with 7.0 million parameters and 46.5 GFLOPs. Nevertheless, the original DenseNet121 model evaluated at a lower resolution (256×256) in the earlier study by Hayat *et al*. (2023) showed better accuracy (83%), suggesting the model might be better suited for lower-resolution inputs.

Feedback analysis

A structured usability assessment was conducted with 18 participants, with 8 certified banana harvesters and 10 administrative members with experience in banana harvest. The questionnaire was holistic and covered a wide range of areas, including user interface and experience overall usability, harvesting decision correctness, and personal details. The questionnaire was a combination of closed and open-ended questions.

Field testing of the harvesting process

Among the eight certified harvesters, 2 (25%) reported being very comfortable, 2 (25%) somewhat comfortable, 2 (25%) neutral, 1 (12.5%) somewhat uncomfortable, and 1 (12.5%) very uncomfortable with new technologies and mobile phone applications. Mobile app use also varied: 4 (50%) use apps often, 3 (37.5%) sometimes, and 1 (12.5%) rarely. Overall, 50% of the participants reported feeling comfortable, 25% were neutral, and 25% reported discomfort. Accordingly, harvesters with lower general familiarity were more likely to report difficulty using our app, whereas frequent users generally reported little or no difficulty. This pattern suggests most difficulties came from overall technology or app familiarity rather than problems of the app itself (n=8). All reported using Android smartphones, and the majority (62.5%, n=5) were aged 35-44 years, with the remainder split between 45-54 years (12.5%, n=1) and 55–64 years (25%, n=2). All participants were male.

Mobile application familiarity varied, with 50% (n=4) frequently using mobile apps, 37.5% (n=3) using them occasionally, and 12.5% (n=1) rarely engaging with mobile applications. Respondents reported using a mix of application types, most commonly social media, productivity, entertainment, and utility tools. None had prior experience with agricultural or harvesting-related applications. Primary motivations for using the application were evenly split between learning the optimal harvest time (50%, n=4) and improving harvesting efficiency and productivity (50%, n=4).

Average inference times were benchmarked at 16ms during field testing, and the system exhibited stable performance under

Table 2. Summary of user feedback and future system requirements.

Feedback category	User group	Feedback / future requirement
Font size / visibility	Harvesters and admin	Increase font size for outdoor readability
Report export functionality	Admin	One-click export options (PDF, CSV)
Offline functionality	Admin and Harvesters	Offline image queue for poor network coverage
Additional features	Admin and Harvesters	Pest and disease detection, harvest estimation
Image upload improvements	Admin and Harvesters	Allow multiple photos and gallery selection
Documentation automation	Admin	Auto-generate reports from app predictions
Harvest estimation	Admin	Forecast expected bunch numbers for logistical planning
Model decision transparency	Harvesters	Clearer explanations of predictions
User interface preferences	All	Maintain Portuguese as the primary language
Branding / design	Admin	Enhanced logo and larger fonts for outdoor use

operational conditions, with predictions consistently returned to the user interface within an acceptable response time. The total number of images taken during the testing was 186, out of which our model predicted 138 as "keep" and 48 as "cut" (random tests were also performed without any banana bunch in the photo, and the model correctly identified all these cases). The keep predictions were 100% accurate and no disagreement was logged, whereas on 14 instances, the harvester logged disagreement, which, according to them, were "cut", but the model predicted them as "keep". However, 1 image out of 14 disagreed and was blurred, hence it was not considered in the accuracy calculation. Therefore, the model accuracy was 92.9%.

Evaluation of usability and flow

Interface design (only of the user flow) received positive feedback as all respondents approved of button colours (100%), and colour contrast between background and text was rated excellent by 37.5% (n=3), good by 25% (n=2), and fair by 37.5% (n=3). The application logo was well received by 87.5% (n=7), with one participant suggesting changes. Font size was suitable for 75% (n=6), while 25% (n=2), primarily older participants over 45 years, reported difficulty reading text outdoors and recommended larger default text or a user-selectable font size. Portuguese was unanimously preferred as the interface language. Suggested improvements included providing clearer explanations of how the model arrives at its decisions and introducing a consolidated daily summary screen for all bunch evaluations. Overall, harvester feedback demonstrated that the current design is generally accessible and well-received, though minor adjustments in font size and decision explanation could further improve usability, particularly for older workers and those operating in outdoor field environments.

Afterwards, a total of 11 participants evaluated all interfaces of the application, providing more structured feedback across 34 usability and feature-related questions. The majority (63.6%, n=7) reported prior banana harvesting experience, while 36.4% (n=4) did not. Most respondents (72.7%, n=8) accessed the application on Android devices, with 18.2% (n=2) using iOS, and one participant (9.1%) not disclosing their operating system. The age distribution was skewed toward older users, with 45.5% (n=5) aged 45-54, followed by equal representation (18.2%, n=2 each) from the 18-24 and 25-34 groups, one participant (9.1%) aged 35-44, and one (9.1%) aged 55-64. Males constituted 72.7% (n=8) of the sample, with females representing 27.3% (n=3).

Regarding mobile application usage, 90.9% (n=10) were frequent users of mobile apps, and 9.1% (n=1) reported occasional use. All respondents reported using multiple categories of applications, spanning social, productivity, entertainment, and utility tools. Only 27.3% (n=3) had prior experience with agricultural or harvesting-related applications, mostly for plant identification or pest detection. Motivations for using the application were split between learning the optimal harvest time (36.4%, n=4), improving harvesting efficiency and productivity (36.4%, n=4), and managing harvesting activities (18.2%, n=2), with one participant (9.1%) citing other reasons. Comfort with technology was high, with 63.6% (n=7) describing themselves as very comfortable, 27.3% (n=3) somewhat comfortable, and one (9.1%) neutral.

Feature requests were dominated by the need for pest and disease detection, with some also requesting bunch size and quantity estimation capabilities. Participants suggested capturing multiple images from different angles rather than relying solely on a single full-bunch image to improve decision accuracy (this recommendation is implemented in the current version of the application by using the previously described average method). Several users proposed adding quality differentiation criteria, such as finger length, thickness, and bruise area, to aid in categorization and improve production profitability. Others recommended estimating the number of bunches to be harvested from approved images, allowing for a possible improved scheduling of deliveries to processing facilities, optimizing time, and avoiding product loss or delays.

Technical improvements included enabling uploads from the device gallery, allowing users to capture images in the field without internet access, and uploading them later when connected. This improvement is now implemented in the application. Several participants emphasized the importance of a simple, intuitive, and user-friendly interface, particularly for field workers with possible limited formal education.

Interface design feedback was largely positive, as all respondents approved of button colours, and 45.5% (n=5) rated colour contrast as excellent, while 54.5% (n=6) rated it as good. The application logo was well received (81.8% approval), with only one negative response. Font size was deemed suitable by 90.9% (n=10), while one participant preferred smaller text. Portuguese was the preferred language for most respondents (90.9%).

Navigation and layout were universally rated as easy to understand, with 100% agreement on the ease of finding relevant fields and satisfaction with field placement. No participants indicated issues entering information. Most reported that error messages (90.9%), button intuitiveness (90.9%), screen clarity (90.9%), and icon labelling (90.9%) were satisfactory. Application functions generally met expectations (81.8%), though one participant noted uncertainty regarding bunch maturity in prediction results. Viewing and managing images was effortless for all respondents, and most (90.9%) found success messages clear and next steps easy to understand.

Overall satisfaction with the interface was high, with 36.4% (n=4) being very satisfied, 54.5% (n=6) being satisfied, and no participant expressing dissatisfaction. Feature gap identification was limited, with 36.4% (n=4) suggesting additional functionality, again centring on pest and disease detection, while 63.6% (n=7) felt the app met their needs. Overall, the feedback reflects a strong desire for features that enhance operational efficiency, provide actionable insights, and maintain ease of use in real-world agricultural conditions. Table 2 summarizes the main feedback points and suggested improvements gathered during the evaluation phase.

Conclusions

This work presents the development of an end-to-end mobile-based solution that uses YOLO-based architectures to automate both banana bunch detection and harvest readiness classification. The developed solution can operate in real-time and is suitable for in-field banana harvesting support. The main contribution of this study is the development of a decision-support pipeline, its deployment, and evaluation in real banana orchards, supported by YOLO-based models adapted with task-specific SE enhancements for classification**.** The SE-enhanced YOLOv11n classifier is implemented as a lightweight, task-specific refinement on an existing YOLO backbone and achieved 94% accuracy with 2.8 ms inference times. This work also extends prior research on YOLOv12n detection models (Baglat *et al*., 2025b), now providing a complete decision-support pipeline for harvest readiness.

User feedback highlighted that the system can likely support harvesting decisions. Furthermore, the ability to collect structured feedback for future model improvements establishes a closed-loop

learning system for digital agriculture. Expert harvesters and administrative staff found the mobile application easy to use and suitable for their daily harvesting tasks. The mobile application is also publicly available on the Google Play Store to support broader adoption and future usability studies. This work acknowledges limitations regarding the examined population in the usability studies, which was restricted to Madeira Island, and the images used for model development were also exclusively from Madeira Island. The reported findings may not generalize to other banana-growing regions with different cultivars, environmental conditions, or agricultural practices. The geographical constraint limits the diversity of banana varieties, maturation patterns, and growing conditions represented in the dataset. Different regions may exhibit variations in banana morphology, coloration patterns during ripening, and local cultivation methods that could affect model performance. Future work should include validation studies with datasets from multiple geographic regions and diverse banana cultivars to assess model transferability and robustness across different agricultural environments. It would also be beneficial to extend this work to other crops and the possible integration of field records or harvest logs that farmers typically maintain to track crop development. This integration would enable the mobile application to contribute to farm management systems by automatically logging maturity assessments alongside traditional record-keeping practices. Such integration could also facilitate the development of predictive models for optimal harvest scheduling and allow quality control documentation (required for agricultural certifications).

References

Altaf S, Ahmad S, Zaindin M, Soomro MW, 2020. Xbee-based WSN architecture for monitoring of banana ripening process using knowledge-level artificial intelligent technique. Sensors (Basel) 20:4033.

Bac CW, Van Henten EJ, Hemming J, Edan Y, 2014. Harvesting robots for high-value crops: state-of-the-art review and challenges ahead. J Field Robot 31: 888-911.

Baglat P, Hayat A, Mostafa SS, Mendonça F, Morgado Dias F, 2025a. Banana bunch dataset: multi-field acquisition with various environmental conditions. Available from: https://zenodo.org/records/15642838

Baglat P, Hayat A, Mostafa SS, Mendonça F, Morgado-Dias F, 2025b. Comparative analysis and evaluation of YOLO generations for banana bunch detection. Smart Agr Technol 12:101100.

Chuquimarca L, Vintimilla B, Velastin S, 2023. Banana ripeness level classification using a simple CNN model trained with real and synthetic datasets. Proc. 18th Int. Conf. Computer Vision Theory and Applications (VISAPP). Lisbon; pp. 536-543.

Dadzie BK, Orchard JE, 1997. Routine post-harvest screening of banana/plantain hybrids: criteria and methods. Accessed: 12 October 2025. Available from: https://cgspace.cgiar.org/bitstreams/295b5ef7-3a0f-4f6c-bec3-661330300f40/download

Dai G, Tian Z, Fan J, Sunil CK, Dewi C, 2024. DFN-PSAN: Multi-level deep information feature fusion extraction network for interpretable plant disease classification. Comput Electron Agric 216:108481.

Dewi C, Mahmudy WF, Arisoesilaningsih E, Solimun S, 2021. Review of non-destructive banana ripeness identification using imagery data. Proc. 6th Int. Conf. Sustainable Information Engineering and Technology, Malang; pp. 348-354.

Ergün E, 2025. High precision banana variety identification using vision transformer based feature extraction and support vector machine. Sci Rep 15:10366.

FAO, WHO, 2022. Standard for bananas (CXS 205-1997): 2022 amendment. Accessed: 22 July 2025. Available from: https://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/en/

Ferdaus MH, Prito RH, Rasel AAS, Ahmed M, Saykot MJH, Shanta SS, et al., 2025. BananaImageBD: A comprehensive banana image dataset for classification of banana varieties and detection of ripeness stages in Bangladesh. Data Brief 58:111239.

Fu L, Duan J, Zou X, Lin G, Song S, Ji B, Yang Z, 2019. Banana detection based on color and texture features in the natural environment. Comput. Electron. Agric. 167:105057.

Fu L, Duan J, Zou X, Lin J, Zhao L, Li J, Yang Z, 2020. Fast and accurate detection of banana fruits in complex background orchards. IEEE Access 8:196835-196846.

Fu L, Wu F, Zou X, Jiang Y, Lin J, Yang Z, Duan J, 2022a. Fast detection of banana bunches and stalks in the natural environment based on deep learning. Comput Electron Agric 194:106800.

Fu L, Yang Z, Wu F, Zou X, Lin J, Cao Y, Duan J, 2022b. YOLO-Banana: a lightweight neural network for rapid detection of banana bunches and stalks in the natural environment. Agronomy 12:391.

Hayat A, Baglat P, Mendonça F, Mostafa SS, Dias FM, Garces H, 2023. Banana bunch harvesting dataset. Mendeley Data, V1. Available from: https://data.mendeley.com/datasets/kjrsb7ztr9/1

Hayat A, Baglat P, Mendonça F, Mostafa SS, Morgado-Dias F, 2024. Machine learning system for commercial banana harvesting. Eng Res Express 6:035202.

Hu J, Shen L, Albanie S, Sun G, Wu E, 2018. Squeeze-and-excitation networks. IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Salt Lake City; pp. 7132-7141.

Knott M, Perez-Cruz F, Defraeye T, 2023. Facilitated machine learning for image-based fruit quality assessment. J Food Eng 345:111401.

Ni J, Gao J, Deng L, Han Z, 2020. Monitoring the change process of banana freshness by GoogLeNet. IEEE Access 8:228369-228376.

Piedad E, Larada JI, Pojas GJ, Ferrer LVV, 2018. Postharvest classification of banana (Musa acuminata) using tier-based machine learning. Postharvest Biol Technol 145:93-100.

Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C, 2016. Deepfruits: A fruit detection system using deep neural networks. Sensors (Basel) 16:1222.

Wang G, Gao Y, Xu F, Sang W, Han Y, Liu Q, 2025. A banana ripeness detection model based on improved YOLOv9c multi-factor complex scenarios. Symmetry 17:231.

Wu F, Duan J, Chen S, Ye Y, Ai P, Yang Z, 2021. Multi-target recognition of bananas and automatic positioning for the inflorescence axis cutting point. Front Plant Sci 12:705021.

Zhang R, Li X, Zhu L, Zhong M, Gao Y, 2021. Target detection of banana string and fruit stalk based on YOLOv3 deep learning network. Proc. IEEE 2nd Int. Conf. on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang; pp. 346-349.

Zhang Y, Lian J, Fan M, Zheng Y, 2018. Deep indicator for fine-grained classification of banana's ripening stages. EURASIP J Image Video Process 2018:46.

Received: 12 October 2025; Accepted: 18 December 2025.

Contributions: Preety Baglat, investigation, writing-original draft preparation and incorporation of revisions, methodology, and app development; Preety Baglat, Sidharth Gupta, review/validation, editing, and app development; Francisco Silva, Helena Garcês, Ruben Sousa, Diana Côrte, review, industry collaboration, field management support, and operational feedback; Fábio Mendonça, Sheikh Shanawaz Mostafa, Fernando Morgado-Dias, supervision, review/validation, and editing, All authors read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflicts of interest: the authors declare no conflict of interest.

Funding: this research was funded by Bolsa de Investigação (BI) within Project BASE: Banana Sensing (PRODERAM20-16.2.2-FEADER-1810); Bolsa de Investigação (BI) within Project PRR (TD-C16-i03-SIH); Instituto de Desenvolvimento Empresarial da Região Autónoma da Madeira; and ARDITI - Agência Regional para o Desenvolvimento da Investigação, Tecnologia e Inovação under the scope of project M1420-09-5369-FSE-000002-Post-Doctoral Fellowship, co-financed by the Madeira 14-20 Program-European Social Fund. Acknowledgement is also given to ITI/LARSyS, funded by FCT (Fundação para a Ciência e a Tecnologia) through projects 10.54499/LA/P/0083/2020 and UID/50009/2025.

Availability of data and materials: the dataset used in this study will be made publicly available on Mendeley for banana bunch harvesting (Hayat *et al.*, 2023) and Zenodo for bunch detection (Baglat *et al*., 2025a). The questionnaire instrument is included in the Supplementary File S1; raw questionnaire responses are held by GESBA and may be shared upon reasonable request and with their permission. No personal or sensitive information is present in the image data.

Acknowledgments: the authors thank GESBA (https://www.gesba.pt/) https://www.gesba.pt/), for administrative and field support. Questionnaire data were collected under the required administrative and ethical approvals, and informed consent was obtained from all participants. Image data contain no personally identifiable information. The study was approved by the Ethics Committee (ID: 10084399) and the Data Protection Office (ID: 10993715), in compliance with GDPR and applicable regulations.