Colegio de Postgraduados

# DeepVision for Precision Farming: Automated Ripeness Grading of Coffee Cherries Using Deep Learning Algorithms

**Cuellar-Cortés, Enrique Alejandro[1]; Martínez-Miranda, Carlos Antonio[1]; Sandoval-Gonzalez, Oscar Osvaldo[1*], Rodríguez-Jarquín, José Pastor[1]; Landeta-Escamilla, Ofelia[1]; Portillo-Rodríguez, Otniel[2]**

[1] Tecnológico Nacional de México, División de Estudios de Posgrado e Investigación, Orizaba, Veracruz, México. 94320.
[2] Universidad Autónoma del Estado de México, Facultad de Ingeniería, Toluca, Estado de México, México, 50130.
* Correspondence: oscar.sg@orizaba.tecnm.mx

**Citation**: Cuellar-Cortés, E. A., Martínez-Miranda, C. A., Sandoval-Gonzalez, O. O., Rodríguez-Jarquín, J. P., Landeta-Escamilla, O., & Portillo-Rodríguez, O. (2025). DeepVision for Precision Farming: Automated Ripeness Grading of Coffee Cherries Using Deep Learning Algorithms. *Agro Productividad*. https://doi.org/10.32854/rxr1f675

**Academic Editor**: Jorge Cadena Iñiguez
**Associate Editor**: Dra. Lucero del Mar Ruiz Posadas
**Guest Editor**: Daniel Alejandro Cadena Zamudio

**Received**: July 18, 2025.
**Accepted**: November 10, 2025.
**Published on-line**: December 11, 2025.

*Agro Productividad, 18*(11). November. 2025. pp: 171-187.



## ABSTRACT

**Objective**: To develop an automated computer vision system using the YOLO framework for classifying coffee cherry ripeness stages, specifically addressing selective harvesting challenges faced by small-scale producers through accessible and cost-effective technology.

**Design/methodology/approach**: YOLOv8, YOLOv9 and YOLOv11 models were implemented and evaluated using transfer learning on a dataset of 7,778 images in four maturity classes. The methodology included comprehensive preprocessing, data augmentation, and rigorous validation with a 70/20/10 train/validation/test split. Statistical validation employed paired t-tests and Wilcoxon signed-rank tests to compare model performance.

**Results**: YOLOv11m demonstrated superior performance with overall mAP@50 of 0.864, precision of 0.806, and recall of 0.813. Class-wise analysis revealed highest performance for unripe cherries (AP: 0.905) and most challenging detection for overripe class (AP: 0.817). The model showed consistent 7.5% improvement in mAP@50 over YOLOv8m, with statistical significance confirmed through rigorous testing.

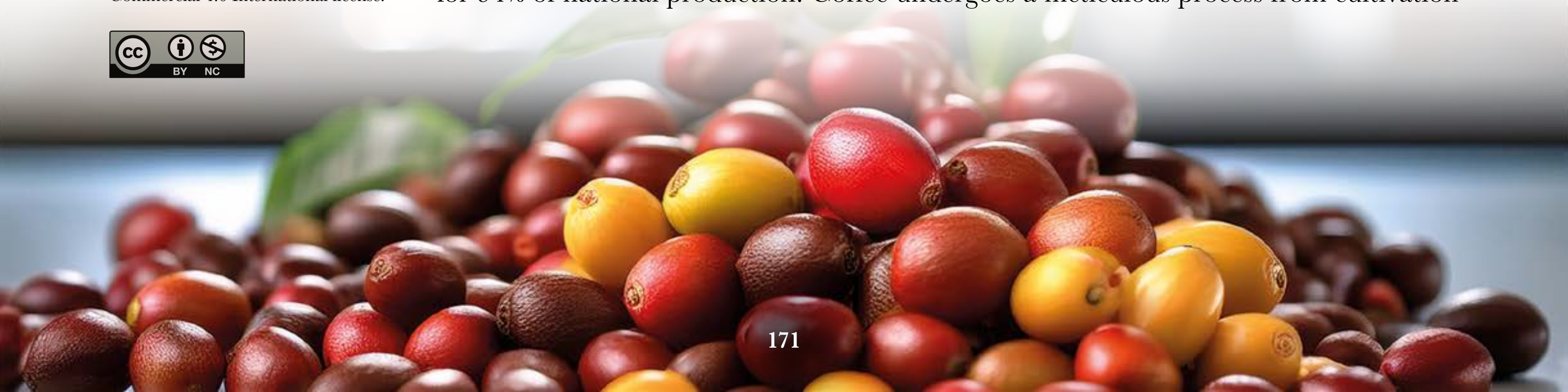
**Study limitations/implications**: The natural imbalance of classes, particularly in overripe, represents the main limitation. The results imply the feasibility of implementation in affordable mobile applications.

**Findings/conclusions**: The effectiveness of YOLOv11 for maturity classification in real conditions was demonstrated. The system represents an accessible solution that can optimize harvesting and improve grain quality.

**Keywords**: Coffee, Maturity, Automatic Classification, Computer Vision, YOLOv11, Precision Agriculture.

## INTRODUCTION

Coffee is one of the most widely consumed beverages globally, and Mexico ranks among the world's leading producers, with around 600,000 hectares of crops distributed across states such as Chiapas, Veracruz, Puebla, Oaxaca, and Guerrero, which account for 94% of national production. Coffee undergoes a meticulous process from cultivation



to consumption, which is why harvesting the coffee cherry at the right stage of ripeness is so important, as it directly influences the quality of the coffee. Currently, this determination depends largely on human experience, which leads to subjectivity and variability. In addition, existing fruit sorting machinery is expensive and designed for large volumes, making it inaccessible to small coffee growers. Given this problem, there is a need to develop accessible and automated systems that support decision-making in the field.

This study is driven by the hypothesis that the latest iterations of the YOLO framework, specifically YOLOv11, can be effectively leveraged through transfer learning to create a highly accurate and computationally efficient model for the automated ripeness grading of coffee cherries directly in the field. It is expected that this model will outperform previous architectures and achieve an accuracy level suitable for practical implementation in a cost-effective mobile application, thereby addressing the critical gap in technology access for small-scale producers.

The primary objective of this work is to develop and evaluate an automated computer vision system based on the YOLO framework for classifying coffee cherries into four distinct ripeness stages: green, semi-ripe, ripe, and overripe. This involves a comparative analysis of YOLOv8, YOLOv9, and YOLOv11 models to identify the optimal architecture that balances high performance with operational efficiency for real-world deployment.

An example of this is the application of computer vision in agriculture, which has gained relevance in recent years, as it processes and analyzes images or videos to extract information, recognize patterns, and make decisions based on that data.

Several studies have addressed this technique for coffee fruit classification, such as the article reported by Anita & Albarda (2020), which shows limited accuracy of 72.12% with KNN *vs.* 24.41% with ANN in coffee cherry classification, highlighting the need for more sophisticated approaches. In contrast, convolutional neural networks (CNN) have demonstrated superior performance, as in the case of the research by Tamayo-Monsalve *et al.* (2022), in which they applied Transfer Learning with pre-trained architectures such as VGG16, VGG19, and Inception-V3, achieving accuracies of up to 98% in maturity classification. Similarly, in the research by Huang *et al.* (2020), they implemented a CNN system for real-time classification of green coffee beans, achieving a recognition rate of 93%. In addition, Bazame *et al.* (2023) developed a computer vision system based on YOLOv4 and YOLOv4-tiny to detect and classify coffee fruits directly on the branches, achieving 81% mean accuracy (mAP) in real field conditions. The model demonstrated superiority over YOLOv3, especially in complex scenarios with high fruit density and detection of green fruits similar to leaves. Another related study is that of Fuentes *et al.* (2020), in which they implemented a computer vision system with neural networks that classifies coffee fruits as "ripe" *vs.* "unripe" with 97.6% accuracy using only 196 images, demonstrating the potential of these technologies to reduce labor costs, save time, and improve the quality of the final product in the coffee production chain.

For the implementation of neural networks, the edge computing environment has been explored, as in the case of Vilcamiza *et al.* (2022), where they deployed a roast quality classifier on an NVIDIA Jetson Nano, achieving 91.33% accuracy, while Pei *et al.* (2019) developed a real-time system for color classification in cherries using TensorFlow and Python. Muhlisin *et al.* (2022) extended these applications by developing an Android app for fruit classification in general using ResNet50 and VGG16. Similarly, Muhlisin *et al.* (2024) developed an Android application with ResNet50 to classify the ripeness of roasted coffee beans, reporting 100% accuracy in controlled tests, but 83.3% in real-world scenarios.

In addition to coffee fruit ripeness, this technology has been applied to other coffee cultivation issues, such as the research by Santana *et al.* (2023), who automated the counting of coffee plants using YOLOv3, with 96.8% accuracy after six months. On the other hand, Abreu Júnior *et al.* (2022) used neural networks and multispectral images from Sentinel-2 to estimate coffee yields with a high coefficient of determination ($R^2$=0.82). Lewis & Espineli (2020) also created a CNN system to detect nutritional deficiencies in leaves, while Eron *et al.* (2024) developed CoffeApp, a mobile application with YOLOv7 for comprehensive crop monitoring. Low-power smart sensors for edge processing have even been explored (Hsu *et al.*, 2023).

The above relates to coffee fruits, but computer vision is applied to other crops, as shown by Yang *et al.* (2024) in classifying banana ripeness with 99.2% accuracy using ResNet101, while Ko *et al.* (2021) applied stochastic fusion to tomatoes with 96% accuracy. Xu *et al.* (2022) achieved 99.7% accuracy in classifying corn seeds with P-ResNet, and Khan *et al.* (2024) recognized bean varieties with 97.5% accuracy. Studies such as those by Liu *et al.* (2020) and Lottes *et al.* (2018) compared color-based methods versus CNN for plant and weed detection, finding greater robustness in the latter. Another study was conducted by Rivera & Reyes-Duke (2024), who applied convolutional neural networks (CNN) with Roboflow to classify the maturity of cocoa fruits, achieving 92.3% mean accuracy (mAP) with images captured at different distances, demonstrating that training with spatial variability improves the robustness of the model. On the other hand, Nisal Ratnayake *et al.* (2021) developed the Polytrack algorithm for automated pollination monitoring, achieving high accuracy with a value of 0.975 and recall with 0.972 in tracking insects in complex agricultural environments. The system includes flower identification and low-resolution processing. In addition, research such as that by Patrício & Rieder (2018) analyzed 25 studies on computer vision in the five main grain crops, identifying the potential of CNNs and deep belief networks (DBNs) for pest detection. Complementarily, the article by Darwin *et al.* (2021) focused specifically on yield and flowering estimation, achieving an average accuracy of 92.51% in precision agriculture. Madhavi *et al.* (2023) provided a sustainability perspective, analyzing how segmentation and maturity detection in tomatoes using CNN optimizes resource use.

This research demonstrates the feasibility of using CNN and image processing techniques for fruit classification. Although there are articles on the detection of coffee cherry ripeness in controlled environments, most have focused on post-harvest stages such as the classification of coffee beans or other fruits such as apples and tomatoes. This project stands out for specifically addressing the classification of ripeness in the field, directly on the plant, with a focus on accessibility for small producers.

## MATERIALS AND METHODS

The comprehensive methodology for developing the coffee cherry ripeness classifier is shown in Figure 1. It describes the constitution of the dataset from public sources. The preprocessing pipeline and data augmentation techniques to improve robustness are explained. The stratified division into training, validation, and test sets is specified. This is followed by the development of the model, including hardware configuration, the training strategy with YOLOv11, and the comparative evaluation of its variants (n, m, x). The process concludes with the analysis of performance metrics (mAP, precision, recall) that inform the selection of the optimal model and its final evaluation with unseen data, ensuring the validity of the results.

### Data Preparation and Architecture Selection

#### Model Selection

To address the task of detecting and classifying the degree of maturity of coffee cherries in images, the YOLOv11 model from Ultralytics (Ultralytics, 2024) was selected. This architecture represents a significant improvement over its predecessors, with advances in network design and training methodologies, making it ideal for applications that require a balance between accuracy and efficiency. For this work, the transfer learning strategy was adopted. The decision is based on the main advantages of this approach: i) a significant reduction in training time, and ii) the ability to achieve high performance with a limited data set, which is particularly advantageous in domains with restricted data availability (Géron, 2023).

#### Dataset Construction

An extensive dataset of coffee cherry images was compiled, covering the four defined stages of ripeness: green, semi-ripe, ripe, and overripe. To ensure the robustness and functionality of the model under real growing conditions, the collection prioritized variability in backgrounds and lighting conditions, simulating the scenarios that the system would encounter in the field. Figure 2 shows representative examples of each maturity category.

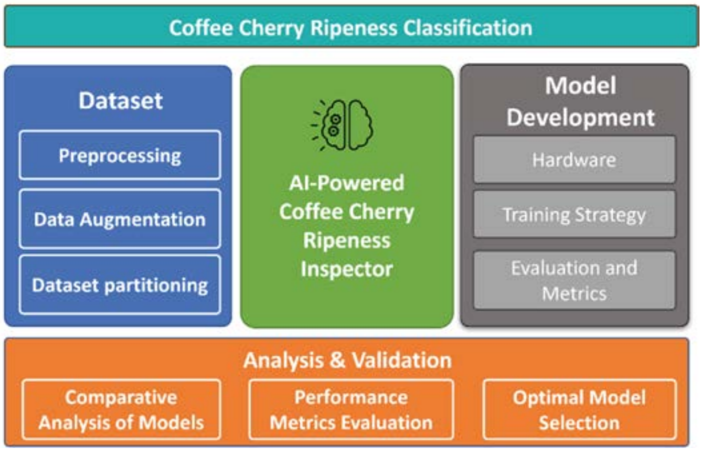



**Figure 1**. Methodological Framework Diagram.


**Figure 2**. Examples of coffee cherries according to their state of ripeness. A) Green, B) Semi-ripe, C) Ripe, D) Overripe.

**Data Sources and Annotation Process**

The dataset was compiled from multiple public sources, including the Kaggle repository (Josué Duff, 2023) and Roboflow Universe (Chess, 2024; Putra, 2024; Park, 2024), totaling 7,778 images of coffee cherries in different stages of ripeness, using the Roboflow platform (Roboflow, 2024) for unification and standardization. Subsequently, a meticulous annotation process was carried out in which each coffee cherry in the images was delimited using bounding boxes. Each of these annotations was assigned to the corresponding class label. This step is essential, as it establishes the supervised relationship between the input (the image) and the expected output (the location and class of the object), allowing the model to learn the discriminative patterns of each stage of ripeness (Ultralytics, 2024).

**Preprocessing and Data Augmentation**

A preprocessing and data augmentation pipeline was applied to the compiled dataset to ensure the quality and consistency of the model inputs. This process was performed using the Roboflow platform and included the following steps:

a. Auto-Orientation: Automatic image orientation correction was applied to ensure consistent visualization and processing, regardless of the original rotation metadata.
b. Resizing: All images were resized to 640×640 pixels using the "Stretch to Fit" method. This standardization is a fundamental requirement for the YOLO architecture and ensures that all inputs have the same dimensions during training.

**Division of the Dataset**

After preprocessing, the final dataset, consisting of 7,778 images, was divided into three stratified subsets for robust model development (Roboflow, 2024):

a. Training (70%-5,446 images): Used for direct adjustment of model weights.
b. Validation (20%-1,556 images): Used for hyperparameter tuning and evaluation during training to prevent overfitting.
c. Test (10%-776 images): Reserved exclusively for the final evaluation of the model, providing an unbiased estimate of its generalization error (Goodfellow *et al.*, 2016).

This division ensures a rigorous and reliable evaluation of the model's performance on unseen data. Figure 3 shows examples of photographs belonging to the training, validation, and test data subsets.

To improve the robustness and generalization ability of the model, a data augmentation process was applied exclusively to the training subset. This technique expands the dataset through random transformations, simulating variations that the model might encounter in real conditions and preventing overfitting (Roboflow, 2024). The transformations applied, illustrated in Figure 4, were as follows: Rotation: Random angular variation within the range of $\pm 15°$ (Figure 4A), brightness: Random adjustment of the brightness level within a range of $\pm 20\%$ (Figure 4B), and blur: Application of a Gaussian blur filter with a kernel of up to 2.5 pixels (Figure 4C).

Subsequently, augmented images were generated by randomly and combinatorially applying the previously defined transformations. This process produced three synthetic variants for each original image in the training subset (Roboflow, 2024), quadrupling its size from 5,446 to 16,337 images.

After this process, the final dataset consisted of 18,669 images, with the following distribution: training (16,337 images, 88%), validation (1,556 images, 8%), and testing (776 images, 4%) this size was maintained to ensure statistical validity while preserving data authenticity. This sample size provides sufficient statistical power for reliable performance


**Figure 3**. Distribution of the dataset into training, validation, and test subsets. A) Training (70%), B) Validation (20%), and C) Test (10%).

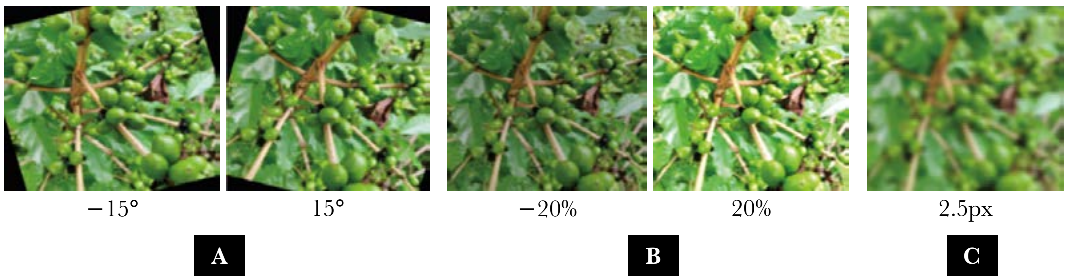



**Figure 4**. Examples of data augmentation techniques applied to an image from the training set. A) Rotation (±15°), B) Brightness (±20%), C) Blur (up to 2.5 px).

estimation, with a margin of error of approximately ±3.5% at 95% confidence level for the reported metrics. Furthermore, maintaining the original test distribution prevents potential biases introduced by synthetic data in evaluation. Figure 5 shows the final configuration in the Roboflow interface, where the distribution of the subsets and the complete preprocessing and augmentation pipeline applied can be seen.

## Model Development

### Hardware Configuration and Software Environment

Model training was implemented in Python 3.10.18, selected for its robust ecosystem for machine learning and the flexibility of the Ultralytics framework. Local training was chosen on a workstation equipped with an NVIDIA GeForce RTX 3080 Ti GPU (12 GB VRAM), 24-core CPU, and 128 GB of RAM, balancing computational capacity and operating costs. The software environment used PyTorch 2.5.1+cu121 and the Ultralytics 8.3.152 library.

### Training Strategy

The model was trained using the backpropagation algorithm. The hyperparameter configuration is detailed below:

a. Epochs: 60 complete cycles on the training dataset.
b. Image size: 640×640 pixels.

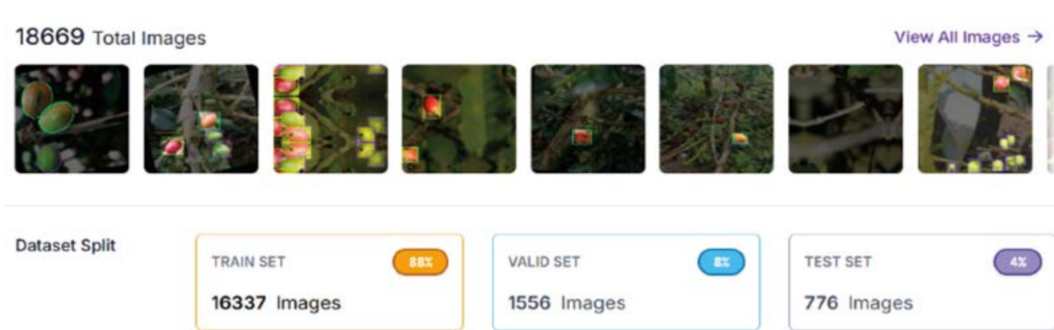



**Figure 5**. Roboflow platform interface with the final dataset configuration. The distribution of images among the training (88%), validation (8%), and test (4%) subsets is displayed, along with the preprocessing (Auto-Orient, Resize) and data augmentation (Rotation, Brightness, Blur) stages implemented.

c. Batch size: 12 samples per parameter update.
d. Optimizer: AdamW for stable and efficient convergence.
e. Accuracy: Training in simple accuracy (AMP disabled).

To mitigate overfitting, the following were implemented:

a. Patience: 10 epochs for early stopping.
b. Initial learning rate (lr0): 0.001.
c. Weight decay: 0.0005 as L2 regularization.

**Evaluation and Metrics**

The evaluation was performed using the validation subset (1,556 images) with the following metrics:

a. mAP@0.5: Average accuracy with IoU threshold at 50%.
b. mAP@0.5:0.95: Strict metric that averages multiple IoU thresholds (0.5 to 0.95).
c. Precision: Proportion of correct positive detections.
d. Recall: Proportion of actual positive instances detected.

**Comparative Statistical Analysis**

To statistically validate the superiority of the YOLOv11m model over previous architectures, a rigorous statistical analysis was performed that included both parametric and non-parametric tests. The comparative analysis was designed considering the five evaluation categories (Overall, Overripe, Ripe, Semi-ripe, Unripe) as paired observations between models. For each performance metric (Precision, Recall, mAP@50, mAP@50-95), values obtained by YOLOv11m were compared against YOLOv8m and YOLOv9m.

Two complementary statistical approaches were employed using parametric tests with paired T-test and non-parametric tests using the Wilcoxon Signed-Rank Test. A significance level of $\alpha=0.05$ was established for all tests. However, given the exploratory nature of the study and the limited sample size ($n=5$ observations per model), statistical trends ($p<0.10$) were also considered in the interpretation of results, following recommendations for research with small samples.

## RESULTS AND DISCUSSION

### Comparative Analysis of YOLO Architectures

A comprehensive comparative analysis was conducted among YOLOv8m, YOLOv9m, and YOLOv11m architectures to identify the most suitable model for coffee cherry maturity classification. All models were trained and evaluated using an identical dataset of 7,778 images with a 70/20/10 split for training, validation, and testing respectively, ensuring consistent evaluation conditions. The performance metrics, including Precision (P), Recall (R), mAP@50, and mAP@50-95 across all maturity classes, are systematically presented in Tables 1, 2, and 3.

The results demonstrate a clear evolutionary improvement across YOLO versions, with YOLOv11m emerging as the superior architecture for this specific application. The overall performance metrics reveal consistent enhancements in both detection accuracy and classification precision through successive model generations.

## Performance Metrics Analysis

YOLOv11m achieved the highest overall performance across all key metrics, with precision of 0.806, recall of 0.813, mAP@50 of 0.864, and mAP@50-95 of 0.727. This represents a significant improvement over YOLOv9m (mAP@50: 0.840) and YOLOv8m (mAP@50: 0.804), indicating approximately 2.9% and 7.5% performance gains respectively in the primary detection metric.

The evaluation by maturity class revealed consistent patterns across all models:

a) **Immature Class**: All models demonstrated excellent performance, with YOLOv11m achieving the highest precision (0.856) and mAP@50 (0.905). The distinctive green coloration of immature cherries appears to facilitate reliable identification.
b) **Semi-mature Class**: YOLOv11m maintained strong performance with precision of 0.798 and mAP@50 of 0.861, showing balanced capabilities in detecting intermediate maturity stages.

**Table 1**. Details the performance of the YOLOv8m model, establishing a performance baseline.

| Class | Images | Instances | Precision (P) | Recall (R) | mAP@50 | mAP@50-95 |
|---|---|---|---|---|---|---|
| Overall | 776 | 8117 | 0.750 | 0.761 | 0.804 | 0.691 |
| Overripe | 229 | 845 | 0.707 | 0.696 | 0.760 | 0.694 |
| Ripe | 424 | 2896 | 0.754 | 0.790 | 0.811 | 0.691 |
| Semi-ripe | 538 | 1830 | 0.742 | 0.769 | 0.801 | 0.710 |
| Unripe | 441 | 2546 | 0.796 | 0.789 | 0.842 | 0.668 |

**Table 2**. Captures the results obtained by the YOLOv9m model, showing improvements in the evaluation metrics.

| Class | Images | Instances | Precision (P) | Recall (R) | mAP@50 | mAP@50-95 |
|---|---|---|---|---|---|---|
| Overall | 776 | 8117 | 0.790 | 0.793 | 0.840 | 0.714 |
| Overripe | 229 | 845 | 0.745 | 0.725 | 0.794 | 0.718 |
| Ripe | 424 | 2896 | 0.795 | 0.823 | 0.848 | 0.714 |
| Semi-ripe | 538 | 1830 | 0.782 | 0.801 | 0.837 | 0.734 |
| Unripe | 441 | 2546 | 0.839 | 0.822 | 0.880 | 0.690 |

**Table 3**. Performance of the YOLOv11m model, completing the comparative overview of the most recent architectures.

| Class | Images | Instances | Precision (P) | Recall (R) | mAP@50 | mAP@50-95 |
|---|---|---|---|---|---|---|
| Overall | 776 | 8117 | 0.806 | 0.813 | 0.864 | 0.727 |
| Overripe | 229 | 845 | 0.760 | 0.744 | 0.817 | 0.731 |
| Ripe | 424 | 2896 | 0.811 | 0.844 | 0.872 | 0.727 |
| Semi-ripe | 538 | 1830 | 0.798 | 0.822 | 0.861 | 0.747 |
| Unripe | 441 | 2546 | 0.856 | 0.843 | 0.905 | 0.703 |

c) **Mature Class**: YOLOv11m achieved superior results with precision of 0.811 and mAP@50 of 0.872, indicating enhanced capability to identify optimal harvest maturity.

d) **Over-mature Class**: This category remained the most challenging for all architectures, though YOLOv11m showed notable improvement with mAP@50 of 0.817 compared to YOLOv9m (0.794) and YOLOv8m (0.760). The visual similarity between mature and over-mature stages presents inherent classification difficulties.

To evaluate the performance of the selected YOLO11m model in detail, Figure 6 shows its normalized confusion matrix. The confusion matrix evaluates the YOLOv11 model for classifying coffee maturity. The recall rate is high for Unripe (0.89) and Ripe (0.84), demonstrating the model's robustness in these classes. Performance is lower for Overripe (0.73).

The Precision-Recall curve shown in Figure 7 for the YOLO11m model confirms its robustness, achieving an average precision (mAP@50) of 0.864 across all classes. The analysis by class reveals exceptionally consistent performance: the unripe class obtained the highest value (0.904), followed by ripe (0.872), semi-ripe (0.861), and overripe (0.817). This result indicates the model's uniform ability to correctly identify all stages of ripeness, with the overripe class being the most challenging due to its phenotypic variability and visual similarity to ripe fruit.

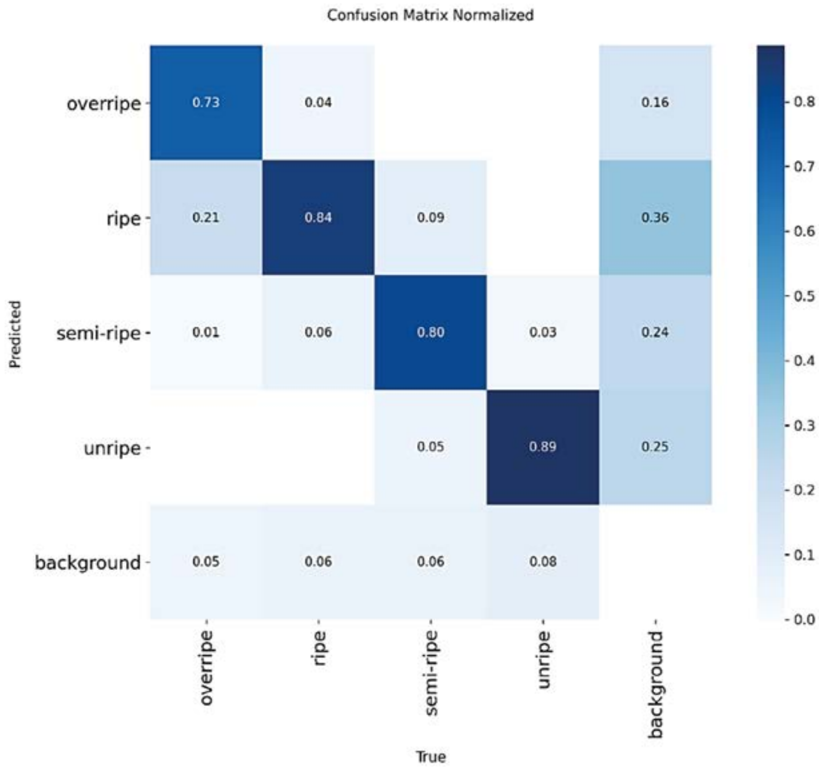



**Figure 6**. Confusion matrix of the YOLO11m model.

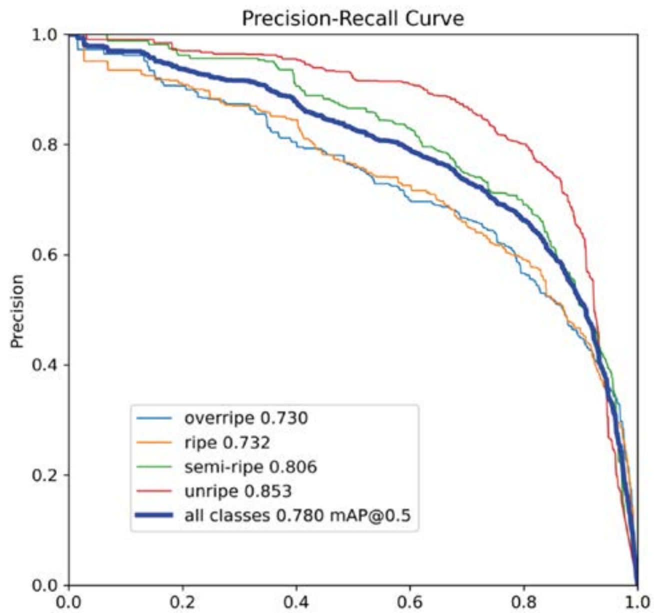



**Figure 7**. Precision-Recall Curve of YOLO11m model.

The Precision-Confidence Curve in Figure 8 is observed that the unripe class (red line) consistently presents the highest precision across all confidence thresholds. This suggests that the model is more robust and reliable when identifying cherries that are clearly immature.

On the other hand, the overripe class (light blue line) shows the lowest precision, especially at medium confidence thresholds, which could indicate greater model confusion when differentiating these cherries from other categories due to visual similarities or variability in appearance.

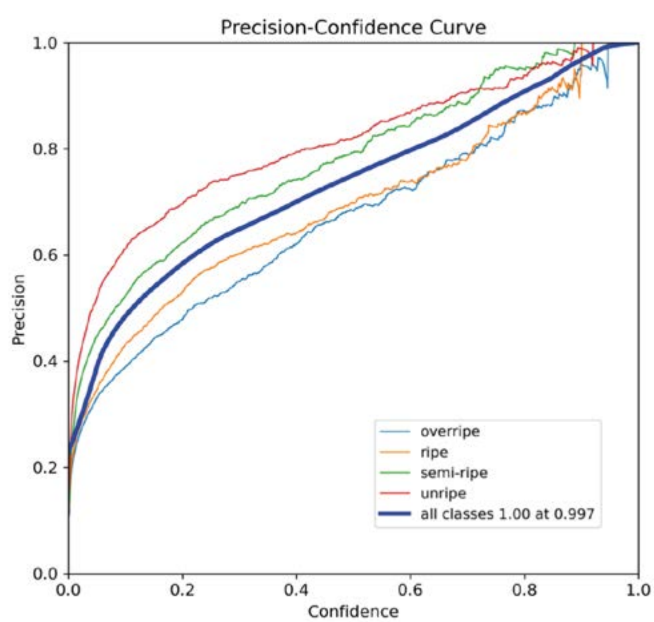



**Figure 8**. Precision-Confidence Curve generated by the YOLOv11 object detection model for classifying the maturity of coffee cherries into four categories: overripe, ripe, semi-ripe, and unripe.

Figure 9 shows the F1-Confidence Curve of the YOLOv11 model for coffee cherry maturity detection. The F1-score, the harmonic mean of precision and recall, evaluates the model's balance.

The overall F1-score (thick blue line) reaches its maximum value of 0.72 at an optimal confidence threshold of 0.403.

By class, the unripe class (red line) achieves the highest F1 (above 0.8), demonstrating the greatest accuracy. The semi-ripe and ripe classes obtain an F1 of around 0.7. The overripe class exhibits the lowest F1, suggesting greater challenges. The analysis confirms robust performance, especially in the early maturity stages.

**Final Evaluation on Test Set**

The results show consistent improvements of YOLOv11m over YOLOv8m across all evaluated metrics (7.5% in mAP@50, 7.5% in precision, 6.9% in recall). Although the paired t-test indicates statistically significant differences ($p<0.001$), the Wilcoxon test does not reach the conventional significance threshold ($p=0.0625$). However, the consistency of the improvements observed across all maturity classes suggests that YOLOv11m represents a substantial improvement over the previous architecture. The comparative analysis of average metrics by model, depicted in Figure 10, reveals that YOLOv11m consistently outperforms the other architectures in all measured parameters. The visualization demonstrates the consistent superiority of YOLOv11m across all evaluation metrics, including precision, recall, mAP@50, and mAP@50-95.

**Qualitative Validation Using Predictive Mode**

To complement the quantitative evaluation, the model's prediction mode was implemented using the OpenCV library in a customized Python environment. This

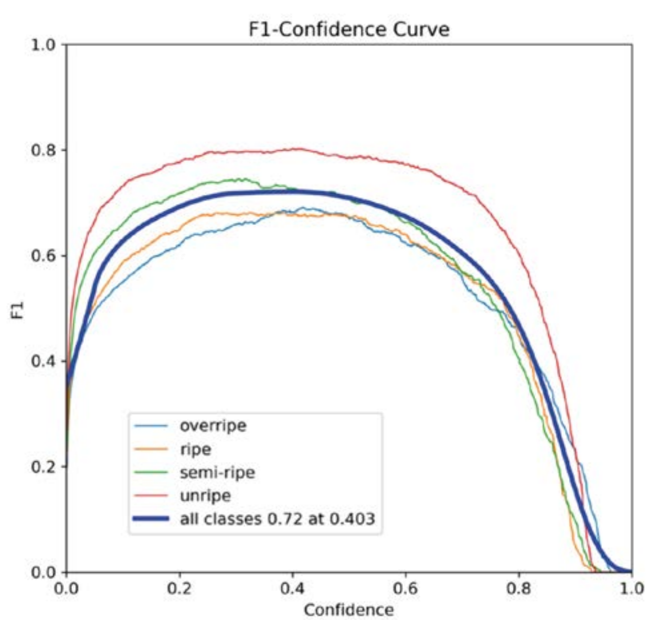



**Figure 9**. F1-Confidence Curve for the YOLOv11 model across the four ripeness classes of coffee cherries.

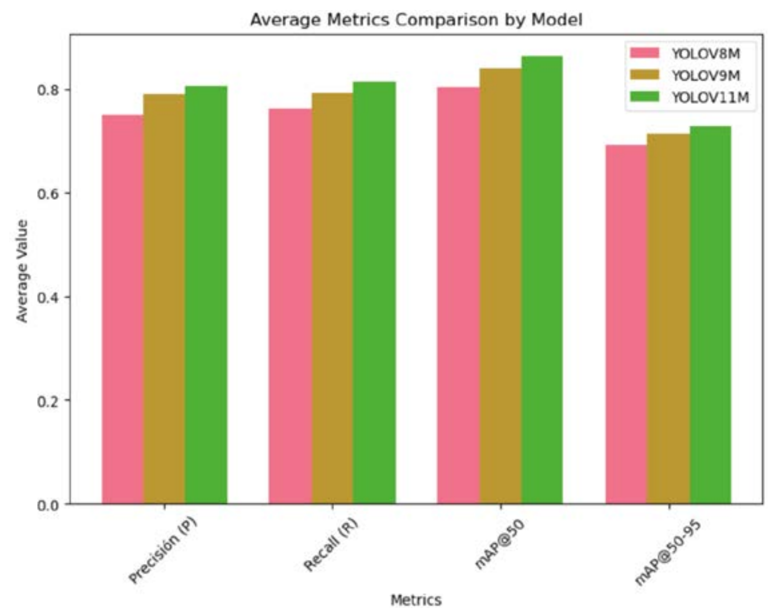



**Figure 10**. Comparative evaluation of average performance metrics across YOLO model architectures.

qualitative validation allowed for visual verification of the YOLOv11m model's performance in practical scenarios with real-world data (Ultralytics, 2025). The validation included tests with images representative of each maturity category: immature (Figure 11A), semi-mature (Figure 11B), mature (Figure 11C), and over-mature (Figure 11D).

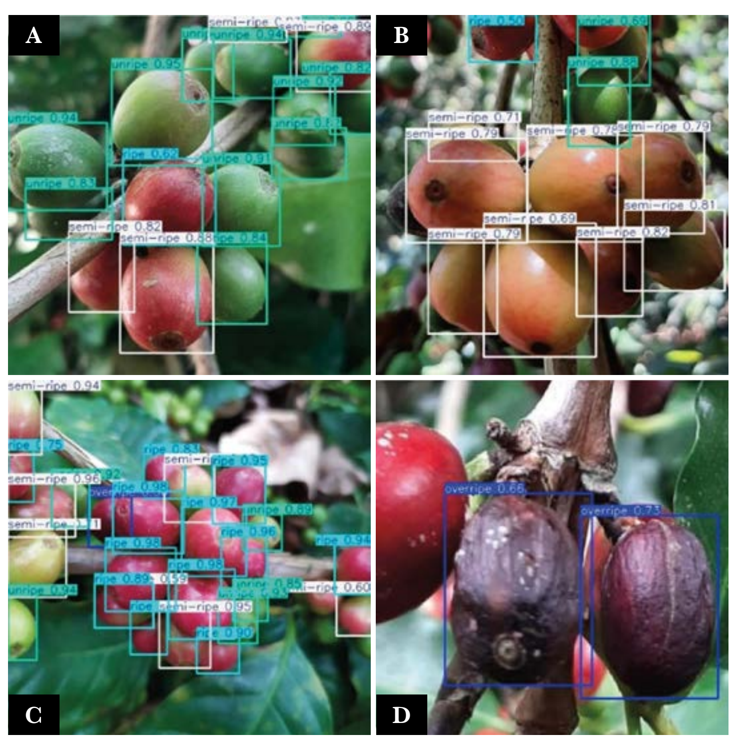



**Figure 11**. Images of qualitative validation using predictive mode. A) immature, B) semi-mature, C) mature, and D) overripe.

Although this modality does not generate numerical performance metrics, it allowed us to visually verify: the correct generation of bounding boxes around coffee cherries, the accuracy of classification labels assigned according to maturity status, the robustness of the model in the face of variations in lighting, occlusion, and complex backgrounds, and the quasi-real-time processing capacity during video inference.

This methodological approach demonstrated the practical effectiveness of the model for its potential implementation in harvest assistance systems.

This research validates the feasibility of using YOLOv11 for the automatic classification of coffee cherry ripeness under field conditions. The selected YOLOv11m model achieved a mAP@50 of 0.864, surpassing the performance reported in the similar study by Bazame *et al*. (2023), who achieved 81% mAP using YOLOv4 for in-branch coffee fruit detection. This improvement can be attributed to the evolutionary advancements in the YOLO architecture, which offer better feature extraction and detection capabilities. Furthermore, our results are comparable to the high accuracy (97.6%) reported by Fuentes *et al*. (2020), despite the fact that our system addresses a more complex classification task with four maturity classes instead of a binary "ripe" *vs*. "unripe" classification. This demonstrates the robustness of our approach in handling finer-grained maturity distinctions.

The class-wise performance analysis revealed a consistent pattern across all models, where the immature class was consistently the easiest to identify, a finding that aligns with Tamayo-Monsalve *et al*. (2022), who also noted that distinct color features facilitate high classification accuracy. Conversely, the overripe class presented the greatest challenge, a phenomenon also observed in other crops like bananas by Yang *et al*. (2024), who attributed similar difficulties to the phenotypic variability and visual similarity between adjacent maturity stages. This underscores a common challenge in agricultural computer vision that may require targeted data strategies or post-processing logic in future work.

The selection of YOLOv11m represented the optimal balance between accuracy and computational efficiency, a critical consideration for practical deployment. This focus on efficiency for edge deployment mirrors the approach of Vilcamiza *et al*. (2022) and Muhlisin *et al*. (2024), who successfully implemented CNN models on NVIDIA Jetson Nano and Android platforms, respectively. Although the larger YOLOv11x variant showed marginally better localization ability, its higher computational cost and increased tendency to overfit did not justify its use for a potential mobile application, which is a key objective of this study.

The model's excellent generalization ability, confirmed through rigorous testing on the hold-out test set and qualitative validation, positions this technology as a practical solution. The successful application of computer vision for direct, on-plant classification, as demonstrated here, extends the findings of Rivera & Reyes-Duke (2024) in cocoa maturity detection and Santana *et al*. (2023) in coffee plant counting, confirming the versatility of deep learning for in-field agricultural tasks. These results are particularly relevant in the Mexican context, where the computational efficiency of YOLOv11m facilitates its potential integration into affordable mobile applications. This addresses a critical gap highlighted in the introduction, offering an accessible tool for small producers, in contrast to the expensive, large-volume machinery that is currently inaccessible to them. By optimizing the harvesting process, this system has the direct implication of improving the final quality of the coffee bean, as harvesting at the correct ripeness stage is paramount for cup quality. All datasets used in this study were obtained from publicly available sources (Kaggle, Roboflow Universe) and were utilized in strict compliance with their respective licenses and terms of use. In accordance with principles of research transparency and reproducibility, the complete implementation code, trained model weights, and detailed documentation have been made publicly available under an open-source license at: https://github.com/oosg/Automated_Ripeness_Grading_Coffee_YOLOV11m

## CONCLUSIONS

This study successfully demonstrated the development of an automatic system for classifying the degree of maturity of coffee cherries using the YOLOv11 framework. The YOLOv11m model emerged as the optimal architecture, achieving an ideal balance between performance and computational efficiency, with an mAP@50 metric of 0.864 in the test set. The results confirm the feasibility of using computer vision to address the challenge of maturity classification directly in the field. Class-wise analysis revealed consistently high performance across all categories, with AP values exceeding 0.81, with the "unripe" class being the best identified (0.904) and "overripe" being the most challenging (0.817), probably due to its greater phenotypic variability. The research validates that transfer learning with YOLOv11 provides a sufficiently robust basis for this task, without requiring aggressive fine-tuning. The model's excellent generalization ability, confirmed in both quantitative and qualitative evaluation, positions this technology as a practical and accessible solution for small producers. The main contribution of this work lies in demonstrating that it is possible to bridge the gap between high performance in the laboratory and practical applicability in the field, offering a tool that can be integrated into low-cost mobile applications to optimize the harvesting process and improve the final quality of Mexican coffee.

## ACKNOWLEDGMENTS

To the National Technological Institute of Mexico (TecNM) and the Secretariat of Education, Science, Technology, and Innovation (SECIHTI) for their institutional support and the resources provided for the development of this research project.

## REFERENCES

Anita, S., & Albarda. (2020). Classification Cherry's Coffee using k-Nearest Neighbor (KNN) and Artificial Neural Network (ANN). 2020 International Conference on Information Technology Systems and Innovation (ICITSI), 117-122. https://doi.org/10.1109/ICITSI50517.2020.9264927.

Tamayo-Monsalve, M. A., Ruiz, E. M., Pulgarin, J. P. V., Ortiz, M. A. B., Arteaga, H. B. A., Rubio, A. M., Alzate-Grisales, J. A., Garzon, D. A., Cano, V. R., Arias, S. O., Osorio, G., & Soto, R. T. (2022). Coffee Maturity Classification Using Convolutional Neural Networks and Transfer Learning. *IEEE Access*, 10, 42971-42982. https://doi.org/10.1109/ACCESS.2022.3166515.

Huang, N., Chou, D., Lee, C., Wu, F., Chuang, A., Chen, Y., & Tsai, Y. (2020). Smart agriculture: realtime classification of green coffee beans by using a convolutional neural network. *IET Smart Cities, 2*(4), 167-172. https://doi.org/10.1049/iet-smc.2020.0068.

Bazame, H. C., Molin, J. P., Althoff, D., & Martello, M. (2023). Detection of coffee fruits on tree branches using computer vision. *Scientia Agricola*, 80. https://doi.org/10.1590/1678-992x-2022-0064.

Fuentes, M. S., Zelaya, N. A. L., & Avila, J. L. O. (2020, September). Coffee Fruit Recognition Using Artificial Vision and neural NETWORKS. 2020 5th International Conference on Control and Robotics Engineering (ICCRE). https://doi.org/10.1109/iccre49379.2020.9096441.

Vilcamiza, G., Trelles, N., Vinces, L., & Oliden, J. (2022). A coffee bean classifier system by roast quality using convolutional neural networks and computer vision implemented in an NVIDIA Jetson Nano. 2022 Congreso Internacional de Innovacion y Tendencias En Ingenieria, CONIITI 2022 - Conference Proceedings. https://doi.org/10.1109/CONIITI57704.2022.9953636.

Pei, Y., Lian, M., Jiang, Y., Ye, J., Han, X., & Gu, Y. (2019). Real-time Cherry Color Grading Based on Machine Vision. 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), 1-6. https://doi.org/10.1109/ICSIDP47821.2019.9173073.

Mimma, N.-E.-A., Ahmed, S., Rahman, T., & Khan, R. (2022). Fruits Classification and Detection Application Using Deep Learning. *Scientific Programming*, 2022, 1-16. https://doi.org/10.1155/2022/4194874.

Muhlisin, E. I., Nurmalasari, R. R., Kamelia, L., & Sururie, R. W. (2024). Implementation Of Convolutional Neural Network (CNN) in The Android-Based Application for Detecting Coffee Bean Maturity. 2024 10th International Conference on Wireless and Telematics (ICWT), 1-5. https://doi.org/10.1109/icwt62080.2024.10674676.

Santana, L. S., Ferraz, G. A. e S., Santos, G. H. R. dos, Bento, N. L., & Faria, R. de O. (2023). Identification and Counting of Coffee Trees Based on Convolutional Neural Network Applied to RGB Images Obtained by RPA. *Sustainability, 15*(1), 820. https://doi.org/10.3390/su15010820.

Abreu Júnior, C. A. M. de, Martins, G. D., Xavier, L. C. M., Vieira, B. S., Gallis, R. B. de A., Fraga Junior, E. F., Martins, R. S., Paes, A. P. B., Mendonça, R. C. P., & Lima, J. V. do N. (2022). Estimating Coffee Plant Yield Based on Multispectral Images and Machine Learning Models. *Agronomy, 12*(12), 3195. https://doi.org/10.3390/agronomy12123195.

Lewis, K. P., & Espineli, J. D. (2020). Classification and detection of nutritional deficiencies in coffee plants using image processing and convolutional neural network (Cnn). *International Journal of Scientific and Technology Research, 9*(4).

Hsu, T.-H., Chen, G.-C., Chen, Y.-R., Liu, R.-S., Lo, C.-C., Tang, K.-T., Chang, M.-F., & Hsieh, C.-C. (2023). A 0.8 V Intelligent Vision Sensor With Tiny Convolutional Neural Network and Programmable Weights Using Mixed-Mode Processing-in-Sensor Technique for Image Classification. *IEEE Journal of Solid-State Circuits, 58*(11), 3266-3274. https://doi.org/10.1109/jssc.2023.3285734.

Yang, L., Cui, B., Wu, J., Xiao, X., Luo, Y., Peng, Q., & Zhang, Y. (2024). Automatic Detection of Banana Maturity—Application of Image Recognition in Agricultural Production. *Processes, 12*(4), 799. https://doi.org/10.3390/pr12040799.

Ko, K., Jang, I., Choi, J. H., Lim, J. H., & Lee, D. U. (2021). Stochastic Decision Fusion of Convolutional Neural Networks for Tomato Ripeness Detection in Agricultural Sorting Systems. *Sensors, 21*(3), 917. https://doi.org/10.3390/s21030917.

Xu, P., Tan, Q., Zhang, Y., Zha, X., Yang, S., & Yang, R. (2022). Research on Maize Seed Classification and Recognition Based on Machine Vision and Deep Learning. *Agriculture, 12*(2), 232. https://doi.org/10.3390/agriculture12020232.

Khan, Md. A. A., Rahman, Md. A., Hossain, M. L., & Habib, Md. T. (2024). Machine Vision Based Local Hyacinth Bean Breed Recognition Using Convolutional Neural Network. 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (ICACCESS), 1-6. https://doi.org/10.1109/icaccess61735.2024.10499455

Liu, H., Sun, H., Li, M., & Iida, M. (2020). Application of Color Featuring and Deep Learning in Maize Plant Detection. *Remote Sensing, 12*(14), 2229. https://doi.org/10.3390/rs12142229

Lottes, P., Behley, J., Milioto, A., & Stachniss, C. (2018). Fully Convolutional Networks With Sequential Information for Robust Crop and Weed Detection in Precision Farming. *IEEE Robotics and Automation Letters, 3*(4), 2870-2877. https://doi.org/10.1109/lra.2018.2846289

Rivera, J. A. M., & Reyes-Duke, A. M. (2024). Convolutional Neural Network for the Detection of Cocoa Maturity with an Approach for the Analysis of Images Captured at Different Distances. *Engineering Headway*, 12, 3-8. https://doi.org/10.4028/p-jm3tgd

Nisal Ratnayake, M., Dyer, A. G., & Dorin, A. (2021). Towards Computer Vision and Deep Learning Facilitated Pollination Monitoring for Agriculture. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2915-2924. https://doi.org/10.1109/cvprw53098.2021.00327

Patrício, D. I., & Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, 153, 69-81. https://doi.org/10.1016/j.compag.2018.08.001

Darwin, B., Dharmaraj, P., Prince, S., Popescu, D. E., & Hemanth, D. J. (2021). Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture: A Review. *Agronomy, 11*(4), 646. https://doi.org/10.3390/agronomy11040646.

Madhavi, K., Babu, Y. S., Ramesh, G., Dua, D., & Reddy, V. B. (2023). Review on Tomato Ripe Detection and Segmentation Using Deep learning Models for Sustainable Agricultural Development. E3S Web of Conferences, 430, 1058. https://doi.org/10.1051/e3sconf/202343001058

Géron, A. (2023). Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow. O'REALLY.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. Cambridge, MA: MIT Press.

Chollet, F. (2018). Deep Learning with Python. Shelter Island, NY: Manning.

Fang, W., Wang, L., & Ren, P. (2020). Tinier-YOLO: A real-time object detection method for constrained environments. *IEEE Access*, 8, 1935-1944. doi:10.1109/ACCESS.2019.2961955

Chess. (2024). Coffee Cherry Beans Detection Dataset [Dataset]. Roboflow Universe. https://universe.roboflow.com/chess-m1x7u/coffee-cherry-beans-detection

Putra, A. D. (2024). Coffee Dataset 2 Dataset [Dataset]. Roboflow Universe. https://universe.roboflow.com/alfito-dwi-putra/coffee-dataset-2

Park, S. (2024). Coffee fruit dataset [Dataset]. Roboflow Universe. https://universe.roboflow.com/seongchan-park/coffee-fruit

Josué Duff. (2023). Coffee Fruit Maturity [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DS/3679843

Ultralytics. (2024). YOLO11 Model Training Guide. https://docs.ultralytics.com/models/yolo11/train/

Roboflow. (2024). Roboflow: Computer Vision Platform [Software]. https://roboflow.com/