Agrociencia

YOLOV8-POWERED COMPUTER VISION FOR COFFEE CHERRY RIPENESS, DEFECT, AND MORPHOLOGICAL ASSESSMENT

Kavitha **Subramani**[1], Senduru **Srinivasulu**[2], Saravanan **Raju**[3], Surendran **Rajendran**[4*]

[1]Panimalar Engineering College. Department of Computer Science and Engineering. Chennai, Tamil Nadu 600123, India.
[2]Sathyabama Institute of science and Technology. School of Computing, Department of Computer Science and Engineering. Chennai, Tamil Nadu 600119, India.
[3]Vel Tech HighTech Dr.Rangarajan Dr.Sakunthala Engineering College. Department of Computer Science and Engineering. Avadi, Tamil Nadu 600062, India.
[4]Saveetha School of Engineering. Saveetha Institute of Medical and Technical Sciences, Department of Computer Science and Engineering. Chennai, Tamil Nadu 602117, India.
* Author for correspondence: surendranr.sse@saveetha.com

Citation: Subramani K, Srinivasulu S, Raju S, Rajendran S. 2026. YOLOv8-powered computer vision for coffee cherry ripeness, defect, and morphological assessment. **Agrociencia.** https://doi.org/ 10.47163/agrociencia.v60i3.3485

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: May 22, 2025.
Approved: March 23, 2026.
Published in Agrociencia: April 15, 2026.

ABSTRACT

The present investigation describes an advanced multi-task deep learning framework for automated inspection of coffee cherry quality using YOLOv8 with color-based segmentation and Vision Transformer-Convolutional Neural Network (ViT-CNN) feature extraction. The model performs ripeness stage classification, defect detection, and size and shape analysis. For ripeness detection, YOLOv8 was enhanced with a color segmentation module, achieving class-wise accuracies of 90–95 % for unripe, partially ripe, and fully ripe cherries, with moderate performance (85 %) for overripe samples. ViT-CNN feature maps improved segmentation clarity and bounding-box localization, particularly in high-density clusters. Defect detection was carried out across five categories (healthy, blackened, moldy, wrinkled, and insect-damaged), achieving F1-score values between 0.88 and 0.96 and mean average precision at 50 % intersection over union (mAP@50) values above 0.97 for key defect classes after 150 training epochs. Quantitative evaluation of morphological characteristics for size and shape assessment further demonstrated model robustness, with insect-damaged cherries reaching a contour accuracy of 0.98 and an Intersection over Union (IoU) of 0.96. Comparative analysis with YOLOv5 and Faster Region-Based Convolutional Neural Network (Faster R-CNN) showed superior performance of the proposed architecture across all metrics, including precision, recall, F1-score, and mAP. By incorporating contextual embeddings and attention mechanisms, the framework enables accurate, real-time sorting for smart agricultural systems.

Keywords: ViT-CNN feature extraction, smart sorting systems, Sustainable agriculture, computer vision, deep learning.

INTRODUCTION

The coffee-growing industry plays a critical role in the global economy, with millions of farmers relying on coffee cultivation as their primary source of income. Coffee (*Coffea* sp.) cherry harvesting is a decisive stage that directly affects product quality, market value, and overall production efficiency. Traditionally, harvesting is performed manually, with workers visually assessing cherry ripeness prior to picking (Napier *et al*., 2025). However, manual harvesting is labor-intensive, time-consuming, and prone to subjective error, often leading to inconsistencies in ripeness selection and quality assessment. These limitations have driven the demand for more efficient and accurate methods for coffee cherry evaluation.

Recent advances in computer vision and deep learning have enabled the development of automated systems capable of real-time ripeness classification, defect detection, and size and shape assessment of coffee cherries (Ye *et al*., 2025). Among deep learning-based object detection approaches, You Only Look Once (YOLO) has emerged as one of the most efficient algorithms for real-time applications due to its high detection speed and accuracy. YOLO-based systems can process video streams captured by cameras mounted on harvesting machinery, unmanned aerial vehicles (UAVs), or fixed installations to identify and classify coffee cherries throughout the harvesting process. This automated approach ensures selective harvesting of ripe cherries, improves coffee quality, and reduces dependence on manual labor.

Beyond ripeness classification, defect identification is essential for maintaining coffee quality (Selvanarayanan *et al*., 2024b). Defects such as mold, insect damage, bruising, and blackening can negatively affect flavor and aroma. Integrating defect detection into a YOLO-based framework allows faulty cherries to be identified and removed before processing, ensuring that only premium-quality beans proceed to production and strengthening overall quality control.

UAVs play a key role in automating coffee harvest monitoring by capturing high-resolution video data. Captured footage is transmitted wirelessly via mobile networks, Wi-Fi, or long-term evolution (LTE) networks to cloud platforms for real-time analysis. Onboard edge computing enables preliminary preprocessing to reduce latency before data upload. Video data, typically stored in formats such as MP4 or AVI, are transferred through cloud services including Amazon S3, Google Cloud, or Microsoft Azure for deep learning-based analysis. High frame rates of 30–60 frames per second (FPS) support accurate segmentation for ripeness classification, defect detection, and size analysis (Arwatchananukul *et al*., 2024). This workflow enhances harvest efficiency, reduces labor costs, and improves quality control through artificial intelligence (AI)-driven automation. Size and shape analysis further contribute to post-harvest classification and processing decisions. Larger, more uniform cherries are generally preferred for specialty coffee production, whereas smaller or irregularly shaped cherries may require alternative processing pathways. YOLO's ability to detect and quantify morphological characteristics enables automated sorting, reducing waste and optimizing post-harvest handling. Consequently, YOLO-based computer vision

systems offer scalable and efficient solutions for improving coffee quality, lowering labor requirements, and increasing sustainability within the coffee sector (Ji *et al*., 2024).

Related research in intelligent agriculture further supports the feasibility of such systems. For example, an AI-driven camera system was proposed for tomato yield estimation using YOLO for fruit detection combined with point cloud data for size analysis (Sangamithrai *et al*., 2024). Ripeness was classified based on color, independent of greenhouse lighting conditions, achieving a prediction error of 6.85 % (Okabe *et al*., 2025). Multi-vision localization techniques have also been explored to improve robotic harvesting accuracy. Using red-green-blue-depth (RGB-D) cameras and motion capture systems, both analytical and model-based approaches were evaluated. Adaptive Boosting (AdaBoost) regression achieved an accuracy of 88.8 % with a 4.4 mm error, outperforming single-camera methods and improving robotic picking efficiency (Beldek *et al*., 2025).

Machine vision has been applied to automated grading and sorting of agricultural products based on size and maturity. Unlike manual grading, which is often inconsistent, computer vision systems enable nondestructive classification using image processing and machine learning, improving efficiency and product quality (Lalam *et al*., 2025). In strawberry sorting, an automated system combining image processing, automation, and aerial sorting classified fruits into five quality categories using a 640 × 320-pixel camera, LabVIEW 2018 for processing, and a programmable logic controller (PLC) for control. The system achieved 93.78 % accuracy and processed 3273 fruits per hour, significantly outperforming manual sorting (Amaroek *et al*., 2025).

More recently, automated broccoli harvesting has benefited from enhanced YOLO-based segmentation. An improved YOLO version 8 nano (YOLOv8n-seg) model, termed YOLO-Broccoli-Seg, incorporated a triplet attention module to improve feature fusion. The model achieved substantial gains in mean average precision (mAP50 and mAP95) for both bounding box and mask detection. A three-dimensional point cloud-based attitude estimation approach further achieved a coefficient of determination (R^2) of 0.934, allowing accurate assessment of broccoli growth angles (He *et al*., 2025).

Prior studies report strong performance in isolated tasks or specific crop contexts; however, many exhibit limited generalizability, constrained scalability, or reliance on specialized sensing hardware (Table 1). In contrast, the proposed framework addresses these limitations by unifying ripeness classification, defect detection, and size-shape analysis within a single YOLO-based architecture designed for real-world, scalable coffee harvesting applications.

This study aims to develop an intelligent computer vision framework that integrates classification of coffee cherry ripeness, defect detection, and size-shape analysis to enhance automation in coffee harvesting. The proposed system enables accurate, real-time assessment under field conditions, supporting selective harvesting, early defect identification, and automated sorting. By unifying these tasks within a single multi-task architecture, the framework seeks to improve harvesting efficiency, reduce labor

Table 1. Comparative analysis of existing computer vision-based agricultural harvesting and quality assessment frameworks.

Author	Data sources	Focus	Methodology	Scalability	Disadvantage	Future scope
Zhang *et al.* (2025)	Field trials, simulation analysis	Dual-arm harvesting	Asynchronous dual-arm coordination; ST-FSH path planning	Effective for multi-arm systems	Damage rate limited to specific crop types	Improved end-effectors; AI path planning
Dong *et al.* (2025)	Greenhouse environment images, depth information	Tomato peduncle localization	DRCANet with multiscale convolution modules	Greenhouse-scale; limited field transfer	Occlusion and lighting sensitivity	Environmental conditions adaptability; real-time robotic implementation
Yang *et al.* (2025)	Hyperspectral images (400–1000 nm)	Mushroom browning detection	PCA-FCM segmentation; k-NN, PLS-DA classification	Portable quality monitoring	Requires hyperspectral setup	Broader storage conditions; deep learning
Xie *et al.* (2025)	Real-field experimental data	Mushroom cut-surface quality improvement	YOLOv8n-seg with enhanced feature modules	High accuracy; real-world conditions adaptability	Limited species generalization	Extended approach to other mushroom species; optimized computational efficiency
Zhao *et al.* (2025)	Shiitake mushroom cap images	Trait measurement for mushroom breeding	Edge detection and SVM optimization	High-throughput phenotyping	Limited to visual traits	Multi-species extension; IoT integration
Santoso *et al.* (2025)	Grayscale coffee bean images	Coffee bean classification	ResNet-101-based CNN with feature extraction	Potential scalability to new datasets	Limited real-world bean validation	Incorporation of multi-TL strategies; broader datasets
Alhasson *et al.* (2025)	Robusta coffee bean images	Automated defect detection	YOLO-based mobile application	Suitable for large-scale processing	Low accuracy for some defects	Dataset expansion; IoT-based sorting

ST-FSH: Shortest-Time-Based First-See-Harvest; RGB-D: Red-Green-Blue-Depth; DRCANet: Deep Residual Convolutional Attention Network; PCA: Principal Component Analysis; FCM: Fuzzy C-Means; k-NN: k-Nearest Neighbors; PLS-DA: Partial Least Squares Discriminant Analysis; YOLO: You Only Look Once; CNN: Convolutional Neural Network; SVM: Support Vector Machine; IoT: Internet of Things; TL: Transfer Learning.

dependence, minimize quality variability, and enhance decision-making in smart and sustainable coffee production systems.

MATERIALS AND METHODS

Ripeness classification was carried out using a Vision Transformer-Convolutional Neural Network (ViT-CNN), and defect detection was performed using YOLOv8. Image processing and model training were implemented using OpenCV and

TensorFlow. The workflow includes data acquisition, preprocessing, ripeness categorization, defect identification, and contour-based size analysis for precision agricultural harvest monitoring.

Data acquisition and preprocessing

Drone-mounted surveillance systems equipped with high-resolution Parrot Anafi AI 4K high-definition (HD) cameras were used to record continuous video of coffee plants. The drones fly over the plantation fields and capture footage at 30–60 FPS, ensuring comprehensive coverage. Video streams were processed by extracting frames at predefined intervals for analysis. The acquired data were transmitted wirelessly to cloud platforms such as Amazon Web Services (AWS), Google Cloud, or Microsoft Azure for advanced AI processing, allowing ripeness classification, defect detection, and size-shape analysis.

To improve model robustness and generalization, image data augmentation was applied to the extracted frames from both drone-based and fixed-position cameras. Augmentation techniques include rotation, flipping, brightness adjustment, Gaussian noise addition, contrast enhancement, and color space conversion from red-green-blue (RGB) to hue-saturation-value (HSV) or CIELAB ($L^*a^*b^*$). In addition, synthetic data generation using Generative Adversarial Networks (GANs) is employed to increase dataset diversity. These strategies enhance the performance of the YOLO-based model in ripeness classification, defect identification, and morphological assessment.

Preprocessing of video-derived image data is performed to improve analytical accuracy. This includes noise reduction using Gaussian blur and median filtering, color space conversion to HSV or CIELAB for improved ripeness discrimination, and contrast enhancement via histogram equalization. Contour-based segmentation was applied to separate coffee cherries from the background. Images were subsequently resized (e.g., 640 × 640 pixels) and normalized to meet YOLO input requirements, allowing accurate ripeness categorization, defect detection, and size-shape analysis (Gope *et al*., 2024).

Feature extraction using ViT-CNN

The hybrid feature extraction framework (Figure 1) combines ViT-CNNs to classify coffee cherry images by ripeness stage and defect presence. Input images are first processed through multiple CNN modules (Module 1 to Module 4), each designed to capture localized spatial features such as texture, shape, and edge information at different scales (Table 2). The resulting feature maps from each CNN module are then forwarded to corresponding ViT encoders that operate on embedded image patches.

Within each ViT encoder, patch embedding converts image patches into vector representations, followed by normalization and multi-head self-attention to capture global contextual relationships across the image. A feed-forward multilayer perceptron (MLP) refines these representations for enhanced feature encoding. Outputs from all ViT encoders are concatenated and passed through an attention gate to emphasize

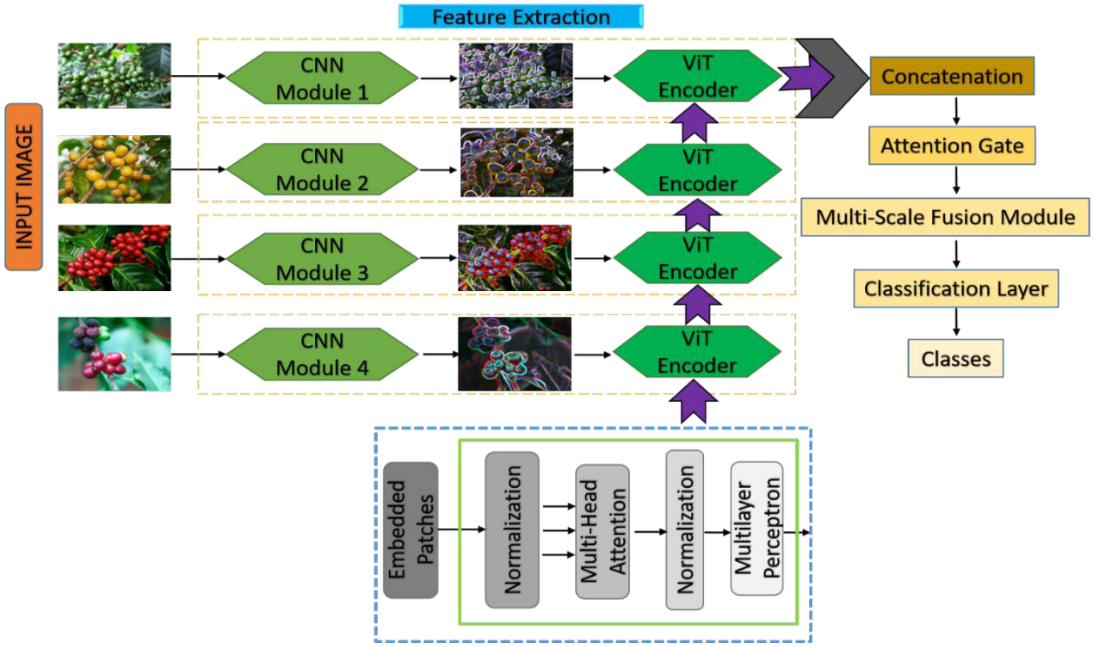

Figure 1. Hybrid vision Transformer-Convolutional Neural Network (ViT-CNN) pipeline to capture both local and global coffee cherry features.

Table 2. Visual results of ripeness stage detection of coffee cherries using YOLOv8 combined with color-based segmentation.

Ripeness stage	Label color	Detection accuracy	Bounding box precision	Segmentation clarity	Observation
Under ripe cherries	Purple	High (~90 %)	Sharp and well localized	High	Most under-ripe cherries are accurately detected and localized.
Partially ripe cherries	Blue	High (~88–92 %)	Precise, minor overlaps	High	Clearly segmented; good distinction even when clustered.
Fully ripe cherries	Orange	Very high (~95 %)	Very accurate	Excellent	Strong performance in identifying ripe cherries even in dense clusters.
Over ripe cherries	Green	Moderate to high (~85 %)	Good with few overlaps	Moderate to high	Slight mislabeling in few areas; could improve with more data or color tuning.

the most salient features. Finally, a multi-scale fusion module integrates the extracted information into a unified representation, which is fed into the classification layer to generate the final output classes (unripe, partially ripe, fully ripe, and overripe).

Ripeness stage detection using YOLOv8 with color-based segmentation

Color-based segmentation is an important preprocessing step to distinguish coffee cherries at different maturity stages (Figure 2). This step enables preliminary classification based on color prior to the application of deep learning models, supporting more accurate downstream categorization. Segmentation is performed in the HSV color space, where hue represents the actual color (0–360°), saturation indicates color purity (0–1), and value corresponds to brightness (0–1). Thresholding is primarily applied to the hue component to categorize cherries according to maturity stage (Selvanarayanan *et al*., 2024a).

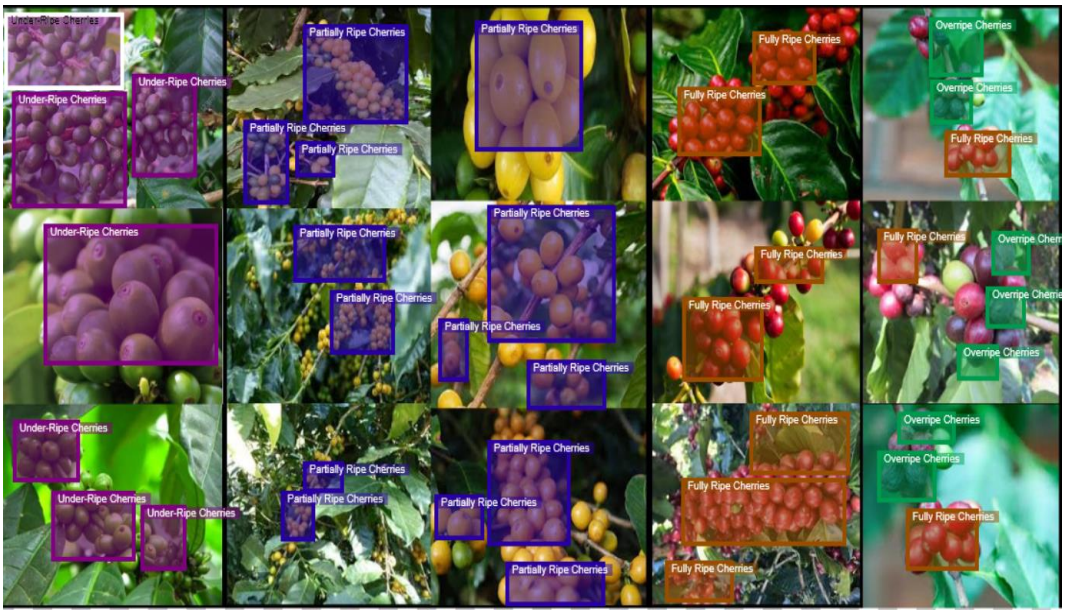

Figure 2. Ripeness stage detection of coffee cherries using YOLOv8, combined with color-based segmentation.

Under-ripe cherries (green shades)

Green cherries correspond to the initial stages of maturity and are characterized by lower hue values, typically ranging from 30 to 80°. Pixels within this interval were classified as under-ripe cherries using the following equation:

$$Mask\_green = (30° \leq H \leq 80°)\text{^}(S > S\_MIN)\text{^}(V > V\_MAX)$$

where H denotes the hue component representing pixel color in the HSV space, S represents saturation to exclude low-color-intensity pixels, and V corresponds to pixel brightness. The threshold condition for under-ripe classification is defined by a hue range between 30 and 80°.

Partially ripe cherries (half green to yellow-red)

Cherry transition from green to yellow and then to red is reflected by progressively higher hue values within the mid-range of the HSV color space (80–150°). Pixels with hue values in this interval were therefore classified as partially ripe:

$$Mask_{partially\ ripe} = (80° \leq H \leq 150°) \quad (S > S_{MIN})^{V>V_{MAX}}$$

Fully ripe cherries (red shades)

When cherries reach full ripeness, they exhibit a bright red coloration, with hue values in the higher range of the HSV spectrum (150–180°). Pixels falling within this interval were classified as ripe cherries:

$$Mask_{ripe} = (150° \leq H \leq 180°)^{\wedge}(S > S_{MIN})^{V>V_{MAX}}$$

Overripe cherries (dark brown/black shades)

Overripe cherries darken to deep brown or black, characterized by hue values exceeding 180°. Pixels exhibiting high hue values combined with low brightness were therefore classified as overripe as follows:

$$Mask_{over\ ripe} = (H > 180°)^{\wedge}(S > 0.2)^{V>0.5}$$

where the $H > 180°$ condition ensures capturing hues associated with darker shades like brown and black. Since overripe cherries appear dark, their brightness (V) must be below a certain threshold.

Coffee cherry defect detection

The defect detection process for coffee cherries using YOLOv8 (Figure 3) follows a structured architecture composed of three main components: the backbone, neck, and head. The input layer receives images of size 640 × 640 × 3, which corresponds to the spatial resolution and the RGB color channels. These images typically contain multiple coffee cherries exhibiting various defects, such as blackening, mold growth, wrinkling, or insect damage (Figure 4). Within the backbone layer, the input image is initially processed by a convolutional module in the stem layer to extract basic visual features, including edges, textures, and color patterns.

The input image is processed through a sequence of four stages (Stage 1 to Stage 4) using convolutional modules, C2f blocks, and Darknet bottleneck modules. Each stage progressively reduces spatial resolution while increasing feature abstraction. Stage 1 downsamples the image to 320 × 320 pixels and extracts mid-level features. Stage 2 further reduces the resolution to 160 × 160, emphasizing patterns associated with surface defects. Stage 3 operates at an 80 × 80 resolution to capture coarse and abstract features, such as shape distortions. Stage 4 processes features at 40 × 40 and 20 × 20 resolutions, focusing on high-level semantic characteristics, including pronounced wrinkling or mold development.

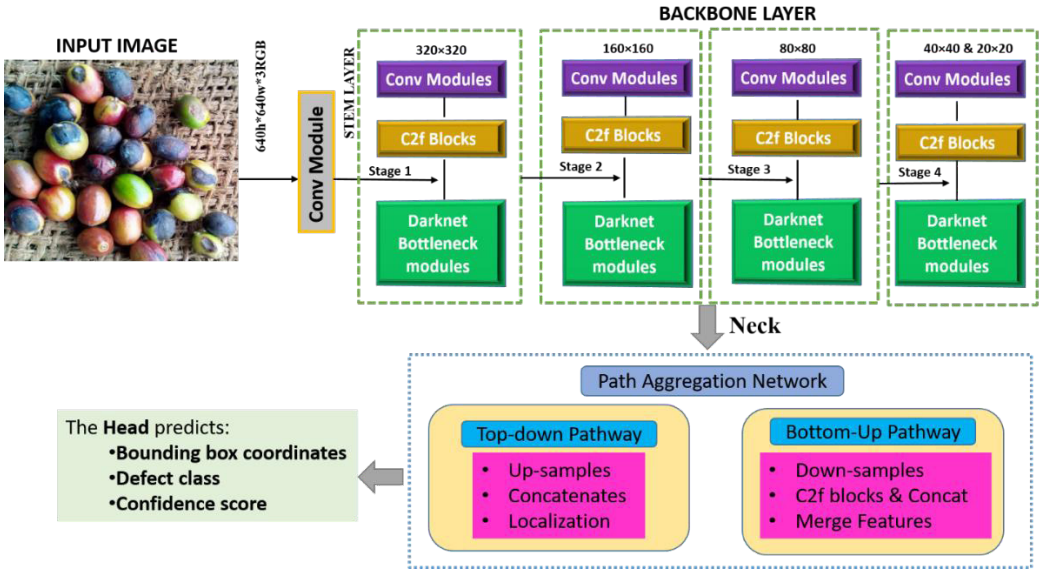

Figure 3. YOLOv8-based architecture for automated coffee cherry defect detection, showcasing multi-stage feature extraction and confidence-driven classification using a backbone, neck, and object prediction head.

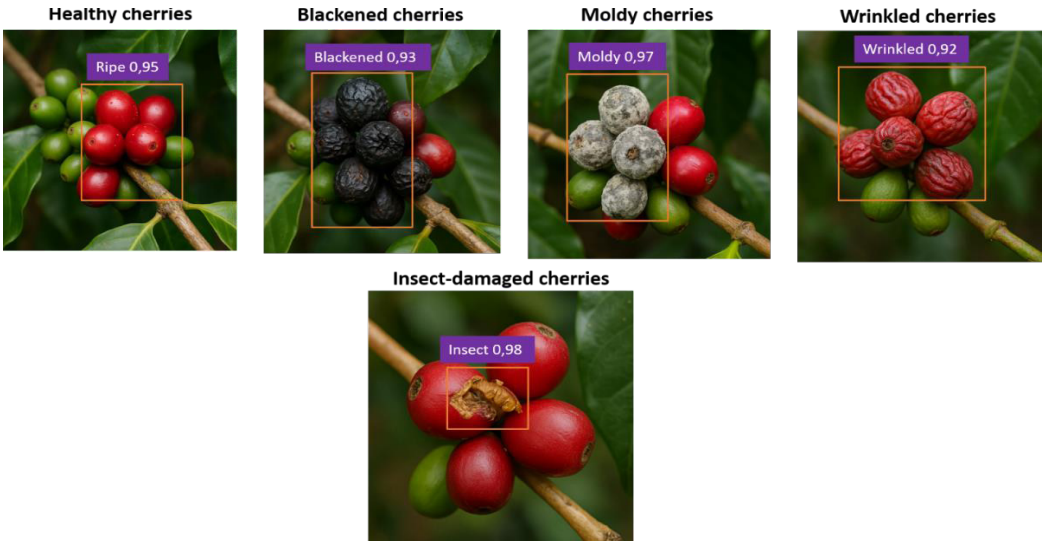

Figure 4. YOLOv8-powered classification of coffee cherries into five categories: healthy, blackened, moldy, wrinkled, and insect-damaged based on defect type and confidence scores for precision in quality assessment.

Convolutional feature extraction at each stage is performed according to the following expression:

$$F_{conv} = \sigma(W * I_{norm} + b)$$

where σ denotes the activation function (e.g., SiLU), W is the convolution kernel, * represents the convolution operation, and b is the bias term.

C2f blocks improve feature reuse and learning efficiency through feature concatenation, as defined by:

$$F_{C2f} = Concat\ (F_1, F_2, \ldots, F_n)$$

where Fi represents the feature maps generated by the internal convolutional layers within the C2f block.

Darknet bottleneck modules further incorporate residual connections, facilitating the effective learning of deeper and more complex feature representations. The resulting multi-scale features are then forwarded to the Neck, implemented as a Path Aggregation Network (PAN), which enables feature fusion across different spatial scales.
In the top-down pathway, small-scale defects such as minor surface patches or fine creases are emphasized through upsampling, feature concatenation, and localization, according to the expression:

$$F_{td}^{i} = Concat\left(F_{Upsampled}^{i+1}, F_{Backbone}^{i}\right)$$

Conversely, the bottom-up pathway applies downsampling and integrates multi-resolution features using C2f blocks, enhancing contextual understanding of defects distributed over larger regions, such as widespread mildew, as follows:

$$F_{bu}^{i} = Concat(F_{Downsampled}^{i+1}, F_{Backbone}^{i})$$

The YOLOv8 Head receives the fused multi-scale features and generates three outputs in a single forward pass: (i) bounding box coordinates for localizing defective cherries, (ii) defect class labels (blackened, moldy, wrinkled, or insect-damaged), and (iii) confidence scores that quantify the reliability of each prediction.

Size and shape analysis using YOLOv8 for object contour detection

YOLOv8 detection identifies coffee cherries and generates bounding boxes for each detected object. Each bounding box includes a class label, a confidence score, and spatial

coordinates (x, y, *width*, *height*). For contour-based analysis, the segmentation variant YOLOv8-Seg was used, which produces precise object masks rather than bounding boxes alone. These masks are binary, pixel-level representations of each detected cherry. Contours are extracted from the YOLOv8-Seg output using the *findContours*() function in OpenCV applied to the binary masks. The resulting contours consist of ordered (x,y) coordinate sets that delineate the boundary of each identified cherry.

Size estimation was performed using the bounding rectangle computed for each extracted contour. The width and height of the bounding rectangle were measured, and the diameter was derived from the minimum enclosing circle. Pixel measurements were then converted to physical units by applying a pixel-to-centimeter calibration based on a reference object or a known camera scale factor. The conversion of an object's pixel width to its actual physical width (*RealW*), expressed in centimeters or millimeters, was performed using the relationship between pixel measurements and the camera field of view, where *PixelW* denotes the object width in pixels, *ImageWidth* represents the total image width in pixels, and *RealFieldWidth* corresponds to the true physical width of the camera's field of view.

$$Real_W = \left(\frac{Pixel_W}{Image_{Width}}\right) * Real_{Field_{Width}}$$

The diameter of approximately circular objects, such as coffee cherries, was computed from their projected area to obtain a scale-invariant size estimate as follows:

$$Diameter = sqrt\left(4 \times \frac{area}{\pi}\right)$$

Object elongation and symmetry were characterized using the aspect ratio of the bounding box, which reflects shape irregularities.

$$Aspect\ ratio = \frac{width}{height}$$

Circularity was quantified to assess how closely an object resembles a perfect circle, where an ideal circular shape yields a circularity value of one, while irregular or elongated shapes produce lower values that are indicative of defective or misshapen cherries.

$$Circularity = 4\pi \times \frac{area}{perimeter^2}$$

Solidity, defined as the ratio between the object area and the area of its minimum convex hull, was used to evaluate compactness, with lower values suggesting voids, shrinkage, or surface deformations.

$$Solidity = \frac{Contour_{Area}}{Convex_{Hull_{Area}}}$$

Finally, the extent metric measured the proportion of the bounding box area occupied by the object, where lower values indicate uneven or non-compact shapes, detecting wrinkled, undersized, or defective cherries.

$$Extent = \frac{Contour_{Area}}{Bounding_{Rect_{Area}}}$$

Performance evaluation

Precision quantifies the accuracy of classified ripeness stages (e.g., fully ripe, overripe) or defects (e.g., moldy, blackened). High precision reduces false positives in both ripeness assessment and defect identification, and it was calculated according to the following equation (Kumanan, T *et al*., 2025 and Kumanan, S *et al*., 2025).

$$Precision = \frac{True\ positive}{True\ positive + False\ positive}$$

Recall evaluates the model's ability to identify all relevant instances. High recall ensures that ripeness stages or defective cherries are not missed during detection or size assessment.

$$Recall = \frac{True\ positive}{True\ positive + False\ negative}$$

The F1-score provides a balanced measure of precision and recall, making it particularly suitable for evaluating ripeness detection and defect classification in imbalanced datasets.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Mean Average Precision at an IoU threshold of 0.5 (mAP@50) was used to evaluate ripeness classification and defect detection, requiring at least 50 % overlap between predicted and ground-truth bounding boxes. It was calculated as follows:

$$mAP\ 50 = \frac{1}{N}\sum_{i=1}^{N} Ap_i$$

The COCO-style mAP@50–95 provides evaluation across multiple IoU thresholds, making it essential for fine-grained shape-based classification, especially in crowded or overlapping cherry scenarios.

$$mAP@50-95 = \frac{1}{10}\sum_{i=0.50}^{0.95} mAP@t$$

The confidence score reflects the probability that a detected object correctly corresponds to a specific ripeness stage or defect class (e.g., insect-damaged or fully ripe). Detection results typically display this value as an overlay. It was calculated as:

$$Confidence\ Score = Objectness * Class\ probability$$

Model size, expressed as the number of parameters in millions (Params, M), indicates computational complexity. Lightweight models are preferred for real-time field applications such as ripeness and defect detection.

$$Params\ (M) = \frac{Total\ trainable\ weights}{1\ 000\ 000}$$

Frames per second (FPS) measures inference speed and is important for real-time size and shape analysis in drones, mobile devices, or on-site quality assessment systems.

$$FPS = \frac{Total\ frames\ processed}{Total\ time\ (seconds)}$$

Contour accuracy serves as a key metric for size and shape analysis, enabling the differentiation of healthy cherries from wrinkled or misshapen ones through precise edge and contour detection.

$$Contour\ accuracy = \frac{Correctly\ detected\ contour\ pixels}{Total\ ground\ true\ contour\ pixels}$$

Intersection over Union (IoU) evaluates how accurately predicted bounding boxes align with the true cherry contours.

$$IoU = \frac{Area\ of\ vverlap}{Area\ of\ union}$$

Evaluation setup and dataset collection

Python 3.6.5 was used to develop the model on a system equipped with an Intel i5-8600K CPU, a GeForce GTX 1050 Ti GPU (4 GB), 16 GB RAM, a 250 GB SSD, and a 1

TB HDD. Model construction and training were performed using Keras 2.7. Training parameters were set to a batch size of 32, 150 epochs, a dropout rate of 0.5, and a learning rate of 0.01. Input images were resized to 640 × 640 × 3, and the experimental setup focused on detecting four ripeness stages. Evaluation metrics included True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), as well as derived measures such as Precision, Recall, F1-score, and mean Average Precision (mAP).

The coffee cherry dataset was collected from selected Meraca Gold Estate plantations in the Coorg district of Karnataka, India, specifically in the Kushalanagar and Suntikoppa regions. Data acquisition was performed using UAVs, digital single-lens reflex (DSLR) cameras, and fixed-position cameras installed along plantation rows to capture high-resolution images under diverse lighting and environmental conditions. All images were manually annotated using LabelImg, with bounding boxes marking the regions of interest. A total of 4200 labeled images were collected to ensure class balance across the four ripeness levels. The dataset was partitioned into training (70 %), validation (15 %), and testing (15 %) subsets.

RESULTS AND DISCUSSION

Training convergence and optimization behavior

To evaluate the learning stability and convergence characteristics of the proposed framework, training and validation accuracy-loss curves were analyzed over 150 epochs (Figure 5). The model demonstrates stable convergence across ripeness detection, defect detection, and size-shape analysis tasks, with no evidence of overfitting. Validation loss decreases consistently while accuracy improves progressively, indicating effective generalization and balanced learning across tasks.

Performance evaluation for ripeness stage detection

The proposed YOLOv8 model with integrated color-based segmentation was evaluated and compared against YOLOv5 and Faster Region-Based Convolutional Neural Network (Faster R-CNN) with a ResNet50 backbone. Model performance was assessed at training epochs of 50, 100, and 150. The YOLOv8 framework demonstrated improved convergence and generalization (Table 3), as evidenced by an increase in mAP@50 from 0.88 at epoch 50 to 0.93 at epoch 150. This performance gain was further reflected in the mAP@50:95 metric, where the proposed approach achieved 0.78, outperforming the baseline YOLOv5 (0.71) and the more computationally intensive Faster R-CNN (0.69).

YOLOv8 demonstrated improved detection performance for overlapping and visually similar categories such as “Overripe” cherries, due to the synergistic effect of deep semantic features and color-space segmentation enhancement. The model achieved high precision (0.91), recall (0.89), and F1-score (0.9), making it highly reliable for real-

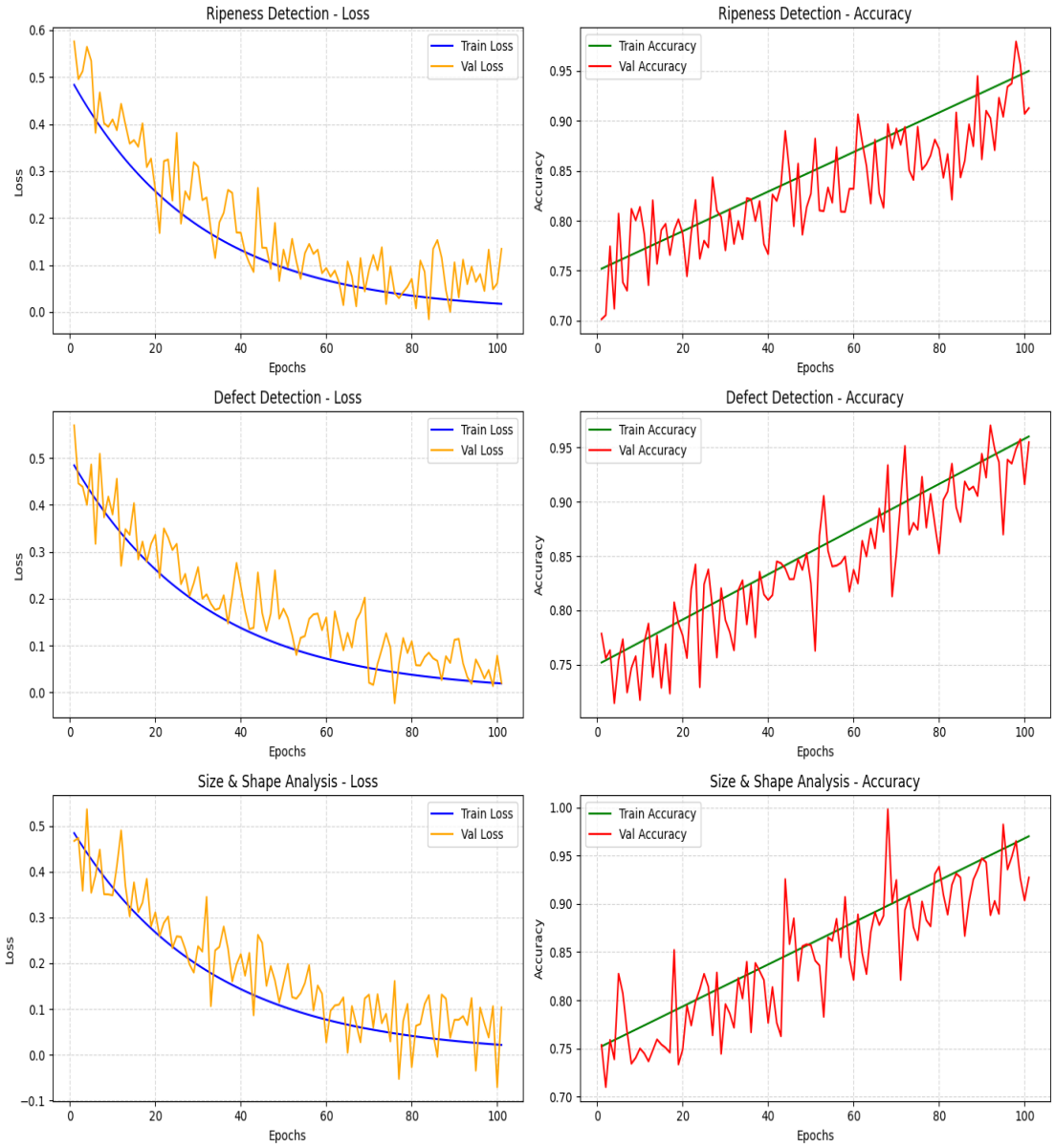

Figure 5. Training and validation accuracy against loss curves for ripeness detection, defect detection, and size and shape analysis tasks using YOLOv8 across 150 epochs.

time deployment, with an inference speed of 57 FPS. YOLOv5 showed a consistent performance trend; however, it did not reach comparable segmentation accuracy in complex background conditions. In contrast, Faster R-CNN, although effective for static image detection tasks, operated at a lower speed (12 FPS), which limits its applicability in dynamic real-time agricultural environments. The proposed model

Table 3. Comparative performance of YOLOv8 with color-based segmentation, YOLOv5, and Faster Region-Based Convolutional Neural Network (Faster R-CNN) for coffee cherry ripeness stage detection across training epochs.

Epochs	Model	Ripe	Partially ripe	Fully ripe	Over ripe	mAP50	mAP 50-95	Precision	Recall	F1-Score
50	YOLOv8 + Color segmentation	0.85	0.87	0.89	0.80	0.88	0.70	0.86	0.84	0.85
	YOLOv5	0.81	0.83	0.86	0.76	0.84	0.65	0.82	0.80	0.81
	Faster R-CNN (ResNet50)	0.77	0.80	0.83	0.72	0.81	0.60	0.78	0.77	0.77
100	YOLOv8 + Color segmentation	0.88	0.90	0.93	0.83	0.91	0.75	0.89	0.87	0.88
	YOLOv5	0.84	0.86	0.89	0.79	0.87	0.69	0.86	0.82	0.84
	Faster R-CNN (ResNet50)	0.80	0.82	0.87	0.75	0.84	0.65	0.83	0.79	0.81
150	YOLOv8 + Color segmentation	0.90	0.92	0.95	0.85	0.93	0.78	0.91	0.89	0.90
	YOLOv5	0.86	0.88	0.91	0.81	0.89	0.71	0.88	0.84	0.86
	Faster R-CNN (ResNet50)	0.83	0.85	0.90	0.78	0.87	0.69	0.85	0.82	0.83

integrates spatial and chromatic features to deliver a robust and scalable framework for high-precision coffee cherry ripeness detection across different maturation stages.

Performance evaluation for defect detection

As the number of training epochs increases from 50 to 150, a clear upward trend in detection and classification accuracy was found across all defect categories: healthy (ripe), blackened, moldy, wrinkled, and insect-damaged cherries (Table 4).

With F1-scores of 0.9 and 0.88 and mAP@50 values of 0.93 and 0.91, respectively, the YOLOv8 model demonstrated strong initial performance at epoch 50, particularly for insect-damaged and moldy cherries. These categories also show high visual confidence scores (above 0.97), indicating robust visual discrimination even during early training phases. However, lower F1-scores for wrinkled (0.82) and blackened cherries (0.84) suggest that additional feature learning is required for these classes due to texture and color overlap with healthy cherries.

At epoch 100, all defect categories exhibited measurable improvement. The model showed consistent gains in both precision and recall, increasing F1-scores to above 0.9 for most classes, especially moldy and insect-damaged cherries. The corresponding rise in mAP@50 values indicates more accurate and confident localization of defective regions. By epoch 150, the YOLOv8 model achieved near-optimal performance. The F1-score reached 0.95 for moldy cherries and 0.96 for insect-damaged cherries, with a mAP@50 of 0.97. Previous lower-performing categories, such as wrinkled and

Table 4. Performance comparison of coffee cherry defect detection across training epochs.

Epoch	Category	Precision	Recall	F1-Score	mAP@50	Confidence score (visual)
50	Healthy (ripe)	0.88	0.85	0.86	0.89	0.95
	Blackened	0.85	0.83	0.84	0.87	0.93
	Moldy	0.90	0.86	0.88	0.91	0.97
	Wrinkled	0.83	0.82	0.82	0.85	0.92
	Insect-damaged	0.92	0.89	0.90	0.93	0.98
100	Healthy (ripe)	0.91	0.89	0.90	0.92	0.95
	Blackened	0.89	0.87	0.88	0.90	0.93
	Moldy	0.94	0.91	0.92	0.94	0.97
	Wrinkled	0.86	0.84	0.85	0.88	0.92
	Insect-damaged	0.95	0.93	0.94	0.96	0.98
150	Healthy (ripe)	0.94	0.92	0.93	0.95	0.95
	Blackened	0.92	0.90	0.91	0.93	0.93
	Moldy	0.96	0.94	0.95	0.96	0.97
	Wrinkled	0.89	0.88	0.88	0.90	0.92
	Insect-damaged	0.97	0.95	0.96	0.97	0.98

blackened cherries, improved to F1-scores of 0.88 and 0.91, respectively. Confidence scores across all categories consistently exceed 0.92, confirming the model's visual reliability and stability.

Performance evaluation for size and shape

All categories showed progressive improvement in performance metrics as the number of training epochs increased, showing the model's ability to learn subtle visual defects and complex shape deformations. In particular, contour accuracy and bounding box IoU increased from 0.9 and 0.87 at epoch 50 to 0.96 and 0.94 at epoch 150, reflecting improved boundary detection and localization accuracy (Table 5).

Insect-damaged cherries consistently achieved the highest scores across all epochs, reaching a contour accuracy of 0.98 and a confidence score of 0.98, due to their distinct shape and texture characteristics. Moldy cherries also demonstrated strong performance, with an F1-score of 0.94 at 150 epochs, attributed to their pronounced surface irregularities. Categories such as wrinkled and blackened cherries, characterized by less distinct edges and subtle color variations, showed marked improvements by epoch 150, with contour accuracies of 0.91 and 0.92, respectively.

Precision-recall and F1-recall analysis

The robustness of the detection framework was further evaluated using precision-recall and F1-recall curves (Figure 6). The graphs demonstrate high area under the curve across all tasks, indicating strong class separability and stable threshold behavior.

Table 5. Performance comparison of coffee cherry size and shape analysis across training epochs.

Epoch	Category	Contour accuracy	Bounding box IoU	Precision	Recall	F1-score	Confidence score (visual)
50	Healthy (ripe)	0.90	0.87	0.88	0.85	0.86	0.95
	Blackened	0.88	0.85	0.86	0.83	0.84	0.93
	Moldy	0.93	0.91	0.91	0.89	0.90	0.97
	Wrinkled	0.86	0.83	0.84	0.81	0.82	0.92
	Insect-damaged	0.95	0.93	0.94	0.92	0.93	0.98
100	Healthy (ripe)	0.93	0.90	0.91	0.89	0.90	0.95
	Blackened	0.90	0.88	0.89	0.87	0.88	0.93
	Moldy	0.95	0.93	0.94	0.92	0.93	0.97
	Wrinkled	0.89	0.86	0.87	0.85	0.86	0.92
	Insect-damaged	0.97	0.95	0.96	0.94	0.95	0.98
150	Healthy (ripe)	0.94	0.91	0.93	0.90	0.91	0.95
	Blackened	0.92	0.89	0.90	0.88	0.89	0.93
	Moldy	0.96	0.94	0.95	0.93	0.94	0.97
	Wrinkled	0.91	0.88	0.89	0.87	0.88	0.92
	Insect-damaged	0.98	0.96	0.97	0.95	0.96	0.98

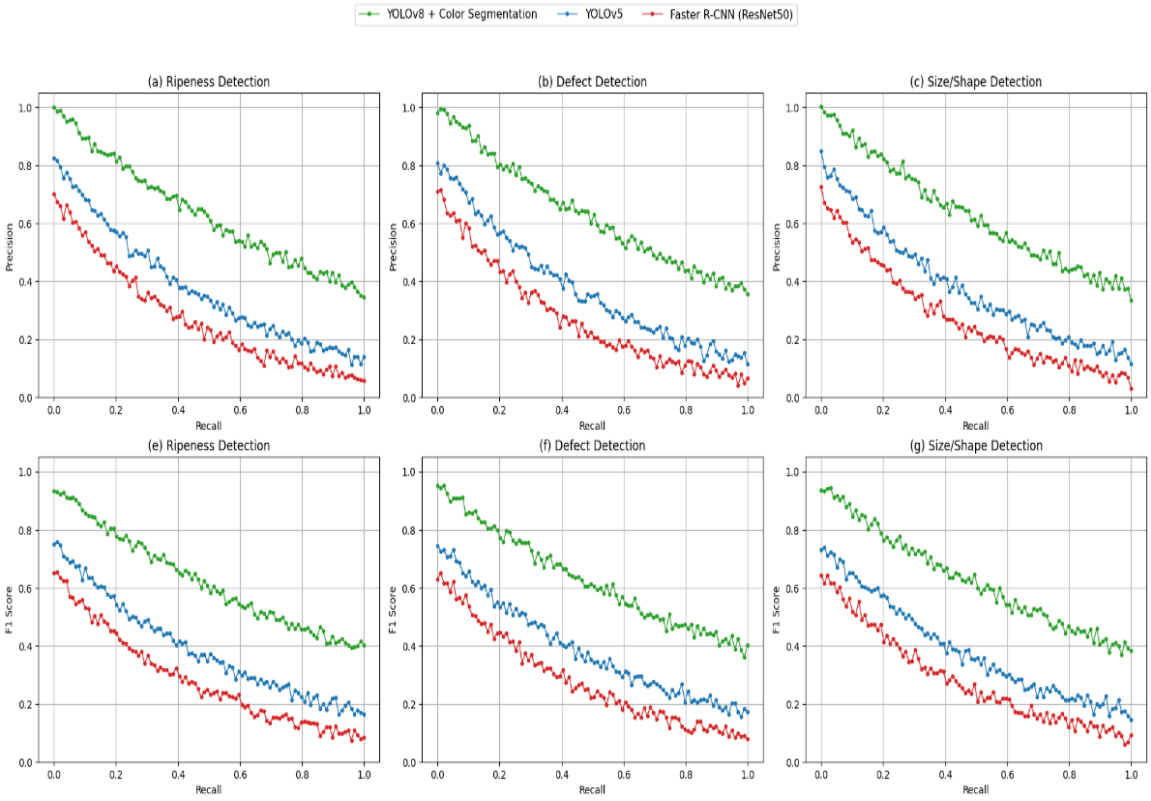

Figure 6. Precision-recall and F1 score-recall curves for ripeness detection, defect detection, and size/shape detection tasks.

Ablation experiment

A preliminary experiment on the coffee cherry dataset used YOLOv8 as the baseline model. To evaluate the effectiveness of the proposed technique within the Visionary Harvest framework, a series of ablation experiments were conducted to assess the individual contributions of each module: ViT-CNN feature extraction, ripeness stage detection, defect detection, and size and shape analysis.

The ablation study began with standard YOLOv8 without architectural modifications, establishing a strong baseline for coffee cherry detection. A ViT-CNN module was integrated to enhance spatial and contextual feature representation, improving the model's ability to distinguish subtle ripeness variations and fine surface textures. Additionally, attention mechanisms were incorporated to refine feature weighting and emphasize discriminative regions (Figure 7), which collectively enhance detection robustness in complex field conditions.

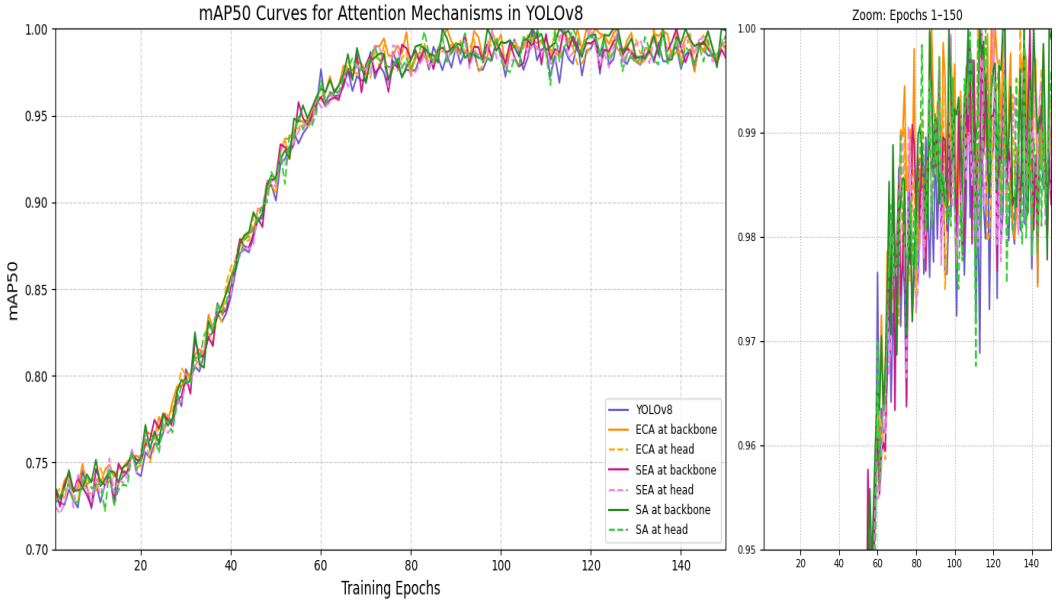

Figure 7. Impact of attention mechanisms (Efficient Channel Attention (ECA), Spatial and Efficient Channel Attention (SEA), and Self-Attention (SA)) on YOLOv8 detection performance through enhanced multi-stage feature refinement.

Subsequently, color-based ripeness segmentation was incorporated into the YOLOv8 detection pipeline to strengthen stage-wise classification (unripe, mid-ripe, ripe, and overripe), particularly under mixed-ripeness conditions commonly observed in field environments. A dedicated defect detection head was introduced to identify cracks, holes, and fungal spots, enabling multi-task learning without significant computational

overhead. Finally, YOLOv8-based contour detection was implemented for analyzing size and shape, enhancing the precision of quality assessments beyond just ripeness and defect detection (Figures 8 and 9).

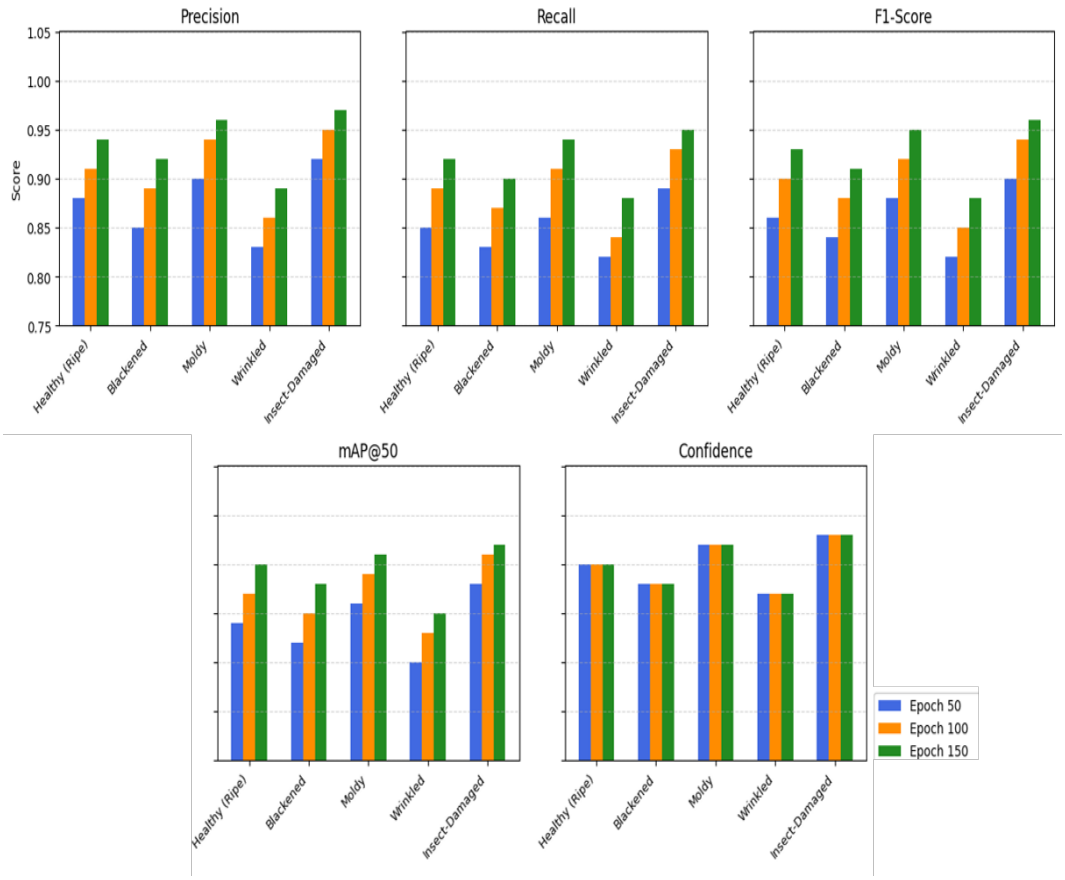

Figure 8. Comparison of model performance metrics (precision, recall, F1-score, mAP@50, and confidence) across three training stages (epochs 50, 100, and 150) for five categories of coffee cherries: healthy (ripe), blackened, moldy, wrinkled, and insect-damaged. The progressive improvement indicates better classification and detection performance as training advances.

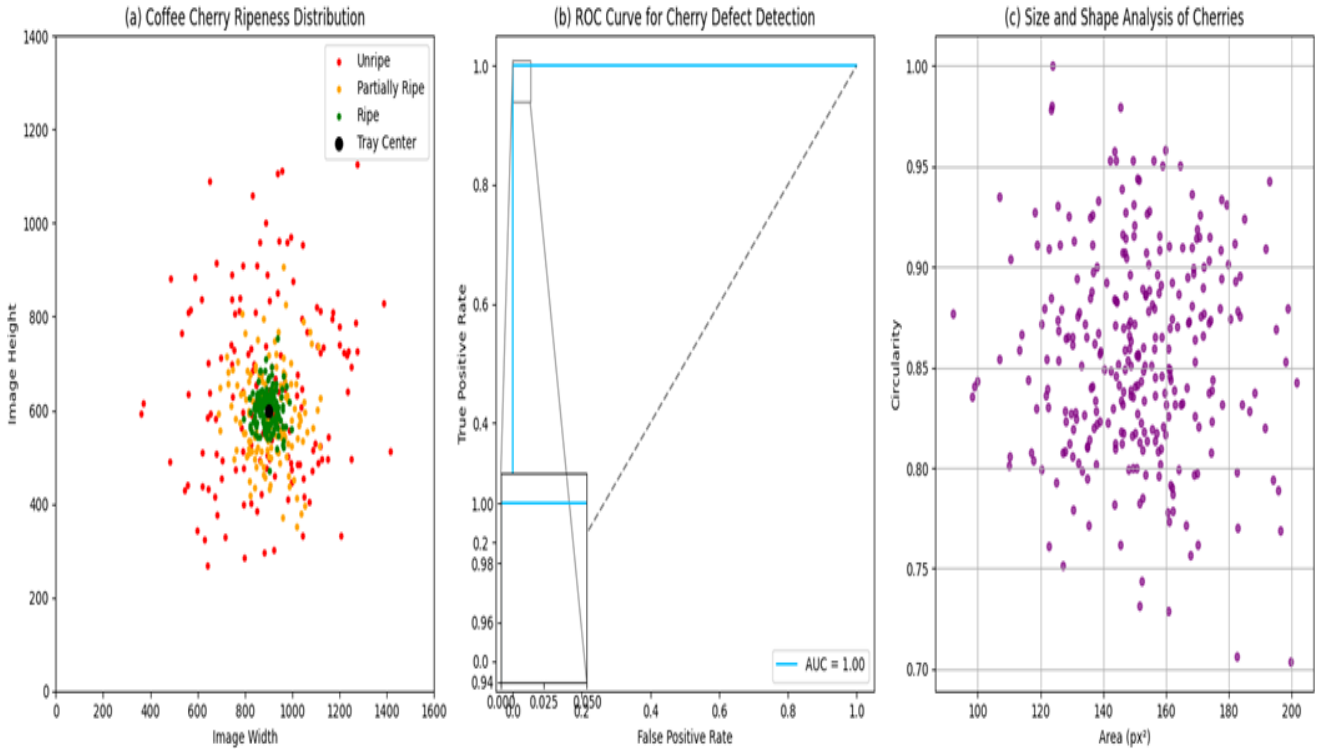

Figure 9. A: Spatial distribution of coffee cherry centers based on ripeness stages, showing unripe (red), partially ripe (orange), and ripe (green) cherries relative to the tray center (black); B: Receiver Operating Characteristic (ROC) curve illustrating the performance of the deep learning model in detecting defective cherries, with an inset zoom showing the high-accuracy region and area under the curve (AUC) performance; C: Scatter plot of cherry size versus circularity representing morphological variation, useful for grading and quality assessment during post-harvest processing.

CONCLUSION

The proposed method classifies coffee cherries into multiple ripeness stages (unripe, partially ripe, and ripe), identifies surface defects, and analyzes morphological features within a unified framework. Using an optimized YOLOv8 model, the system accurately localizes cherry instances and employs performance visualizations to interpret detection behavior. Defect classification is supported by Receiver Operating Characteristic (ROC) analysis, with thresholds derived from an Area Under the Curve (AUC)-optimized framework to ensure robustness and sensitivity.

Spatial distribution patterns were analyzed relative to a reference tray center to simulate real-world sorting conditions. The inclusion of size and shape measurements adds an additional layer of quality assessment, strengthening grading decisions beyond ripeness and defect detection. However, defect categorization is limited to

observable surface irregularities; internal rot or shriveling caused by disease or moisture loss was not evaluated due to sample constraints. Although the architecture was optimized for efficiency, slight accuracy reductions were observed in complex detection scenarios, and environmental factors such as illumination and camera angle may influence prediction reliability. Future work would integrate near-infrared (NIR) or hyperspectral imaging through multimodal sensing to detect internal abnormalities and subtle ripeness signals not visible in RGB images.

DATA AVAILABILITY

Raveena S. 2025. Visionary harvest with YOLOv8-powered coffee cherry ripeness, defect detection, size, and shape assessment. Zenodo. https://doi.org/10.5281/zenodo.15222509. The coding method and supplementary data are available upon proper request from the primary and coauthors.

REFERENCES

Alhasson HF, Alharbi SS. 2025. Classification of Saudi coffee beans using a mobile application leveraging squeeze vision transformer technology. Neural Computing and Applications 37 (14) 8629–8649. https://doi.org/10.1007/s00521-025-11024-9

Amaroek S, Manop R, Pongrawee N, Kittisak K, Niti P, Sorawit C, Supattanakij P. 2025. Research and development of strawberry quality sorting machine with image processing. Agricultural and Biological Engineering 2 (2): 44–51.

Arwatchananukul S, Dan X, Phasit C, Sai AM, Rattapon S. 2024. Implementing a deep learning model for defect classification in Thai Arabica green coffee beans. Smart Agricultural Technology 9: 100680. https://doi.org/10.1016/j.atech.2024.100680

Beldek C, Dunn A, Cunningham J, Sariyildiz E, Phung SL, Alici G. 2025. Multi-vision-based picking point localisation of target fruit for harvesting robots. *In* 2025 IEEE International Conference on Mechatronics. Institute of Electrical and Electronics Engineers: Wollongong, Australia. https://doi.org/10.1109/icm62621.2025.10934868

Dong L, Zhu L, Zhao B, Wang R, Ni J, Liu S, Chen K, Cui X, Zhou L. 2025. Semantic segmentation-based observation pose estimation method for tomato harvesting robots. Computers and Electronics in Agriculture 230: 109895. https://doi.org/10.1016/j.compag.2025.109895

Gope HL, Fukai H, Ruhad FM, Barman S. 2024. Comparative analysis of YOLO models for green coffee bean detection and defect classification. Scientific Reports 14 (1). https://doi.org/10.1038/s41598-024-78598-7

He Z, Yuan F, Zhou Y, Cui B, He Y, Liu Y. 2025. Stereo vision based broccoli recognition and attitude estimation method for field harvesting. Artificial Intelligence in Agriculture 15 (3): 526–536. https://doi.org/10.1016/j.aiia.2025.02.002

Ji Y, Xu J, Yan B. 2024. Coffee green bean defect detection method based on an improved YOLOv8 model. Journal of Food Processing and Preservation 20 (1): 2864052. https://doi.org/10.1155/2024/2864052

Kumanan, T., Nagarathna, K., Vinodha, R., Balasubramani, S., Prasad, S.S.E. and Sujatha, S., 2025, April. Improving Grape Quality in Viticulture with Autonomous Robots Using Deep

Learning and Sensor Fusion. In 2025 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT) (pp. 521-525). IEEE.

Kumaran, S., Rekha, V. and Kavida, A.C., 2025, September. Deep Learning-Based Ripeness Detection and Quality Grading of Mangoes Using Real-Time Image Processing. In 2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) (pp. 1-6). IEEE.

Lalam R, Lavanya K, Nadella V, Kiran BR. 2025. Automatic sorting and grading of fruits based on maturity and size using machine vision and artificial intelligence. Journal of Scientific Research and Reports 31 (1): 153–163. https://doi.org/10.9734/jsrr/2025/v31i12754

Napier CC, Cook DM, Armstrong LJ. 2025. Coffee berry pathogen anomaly detection using colour and shape separation via L-systems. BIO Web of Conferences 167: 05003. https://doi.org/10.1051/bioconf/202516705003

Okabe Y, Hiraguri T, Endo K, Kimura T, Hayashi D. 2025. Classification of tomato harvest timing using an AI Camera and analysis based on experimental results. AgriEngineering 7 (2): 48. https://doi.org/10.3390/agriengineering7020048

Sangamithrai, K. and Richard, T., 2024, April. Design and development of a plant leaf disease identification system using improved deep learning strategy. In 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) (pp. 1-7). IEEE.

Santoso BR, Sari CA, Rachmawanto EH. 2025. Coffee beans classification using convolutional neural networks based on extraction value analysis in grayscale color space. Journal of Applied Informatics and Computing 9 (1): 31–37. https://doi.org/10.30871/jaic.v9i1.8916

Selvanarayanan R, Surendran R, Gomathi T, Kartheesan L. 2024a. Hybrid vision transformer and CNN for detection of overripe coffee berry disease (OCBD) in coffee plantation. *In* 2024 International Conference on Emerging Research in Computational Science. Institute of Electrical and Electronics Engineers: Coimbatore, India. https://doi.org/10.1109/icercs63125.2024.10895612

Selvanarayanan R, Surendran R, Youseef A. 2024b. Early detection of *Colletotrichum kahawae* disease in coffee cherry based on computer vision techniques. Computer Modeling in Engineering and Sciences 139 (1): 759–782. https://doi.org/10.32604/cmes.2023.044084

Xie L, Jing J, Wu H, Kang Q, Zhao Y, Ye D. 2025. MPG-YOLO: Enoki mushroom precision grasping with segmentation and pulse mapping. Agronomy 15 (2): 432. https://doi.org/10.3390/agronomy15020432

Yang K, Zhao M, Argyropoulos D. 2025. Machine learning based framework for the detection of mushroom browning using a portable hyperspectral imaging system. Postharvest Biology and Technology 219: 113247. https://doi.org/10.1016/j.postharvbio.2024.113247

Ye B, Xue R, Xu H. 2025. ASD-YOLO: A lightweight network for coffee fruit ripening detection in complex scenarios. Frontiers in Plant Science 16: 1484784. https://doi.org/10.3389/fpls.2025.1484784

Zhang P, Dai N, Wang Z, Yuan J, Xin Z, Liu X, Papadakis G. 2025. A parallel dual-arm robotic control method of white asparagus based on moving-looking-harvesting coordination and asynchronous harvest cooperation. Computers and Electronics in Agriculture 232: 110046. https://doi.org/10.1016/j.compag.2025.110046

Zhao J, Zheng W, Wei Y, Zhao Q, Dong J, Zhang X, Wang M. 2025. Machine vision-based detection of key traits in shiitake mushroom caps. Frontiers in Plant Science 16: 1495305. https://doi.org/10.3389/fpls.2025.1495305