Agriculture and Food

AIMS Agriculture and Food, 10(4): 962–983.
DOI: 10.3934/agrfood.2025050
Received: 19 June 2025
Revised: 23 September 2025
Accepted: 15 December 2025
Published: 22 December 2025

http://www.aimspress.com/journal/agriculture

Research article

Automated defective green coffee bean image classification using deep learning for quality enhancement and market competitiveness

Hira Lal Gope[1,*], Ashutus Singha[2], Fahim Mahafuz Ruhad[3], Md Mahin Erpan Chowdhury[3], Chandra Kanta Dash[4], Md Masum Billah[5], Md Mehedi Hasan[6], Shohag Barman[7] and Hidekazu Fukai[8]

[1] Department of Computer Science and Engineering, Faculty of Agricultural Engineering and Technology, Sylhet Agricultural University, Sylhet-3100, Bangladesh
[2] Department of Irrigation and Water Management, Faculty of Agricultural Engineering and Technology, Sylhet Agricultural University, Sylhet-3100, Bangladesh
[3] Department of Agricultural Construction and Environmental Engineering, Faculty of Agricultural Engineering and Technology, Sylhet Agricultural University, Sylhet-3100, Bangladesh
[4] Department of Entomology, Faculty of Agriculture, Sylhet Agricultural University, Sylhet-3100, Bangladesh
[5] Department of Computer Science and Engineering, Daffodil International University, Saver, Dhaka, Bangladesh
[6] Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
[7] Department of Computer Science & Engineering, Pirojpur Science and Technology University, Pirojpur, Bangladesh
[8] Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Gifu University, Gifu 501-1193, Japan

***Correspondence:** Email: hlgope@sau.ac.bd; Tel: +8801717021275.

Abstract: Coffee is the most widely traded and popular beverage globally, and its flavor and quality depend significantly on the absence of defective beans. This study aimed to automate the identification and classification of impurities in green coffee beans, enabling more uniform roasting, as peaberries roast differently due to their unique shape. The automated system enhances efficiency and precision over manual checks, using 4367 green coffee bean images from Bangladesh, divided into six categories: black, sour, fade, broken, normal, and peaberry. The main contribution of this study is providing the first Bangladesh-origin dataset for coffee bean defect detection, paired with recent YOLOv10-N advances tailored to small, subtle defects. The study took an innovative step forward for Bangladesh's coffee sector, which has traditionally relied on labor-intensive and error-prone manual sorting methods. This approach addresses key inefficiencies and enhances quality control, boosting global market competitiveness. This study shows that Bangladesh's coffee industry can benefit from using YOLOv10-N to detect defects in green coffee beans, providing a cost-effective and accurate quality control system. We evaluated the following models: Efficient-Net, ResNet-50, Faster R-CNN, and several versions of YOLO (v3-v10). Among them, YOLOv10-N was identified as the most successful model, with the highest precision of 0.992, recall of 0.984, F1-scoreof 0.987, and mean average precision (mAP) of 0.995; YOLOv8 had precision of 0.959 and recall of 0.944, and ResNet-50 had precision of 0.837 and recall of 0.853. The model's accuracy and resilience can be further improved by creating a larger and more diverse dataset, which will enable it to better detect subtle differences in defects across batches of green coffee beans under varying environmental conditions.

Keywords: Arabica; Robusta; normal; peaberry; defective; YOLO

1. Introduction

Coffee is one of the most widely traded agricultural commodities globally, with more than two billion cups consumed daily, making it a highly popular beverage [1-3]. The quality of coffee beans is crucial for competitive market pricing, consumer acceptance, and storage stability, particularly in developing countries [4]. It is a major source of foreign exchange in tropical and subtropical regions, with nations like Brazil, Vietnam, Colombia, and Indonesia significantly benefiting from its trade, which supports the livelihoods of millions [5]. As global coffee consumption rises, so does the demand for high-quality beans.

Many factors determine the quality and price of coffee [6–8]. We classify coffee beans into two types based on their shape: peaberry (special) and normal (good) [9,10]. Peaberries, which make up only about 7% of the coffee harvest, are more expensive due to their lower yield [5]. This study used defective (e.g., black, broken, faded, and sour) green coffee beans alongside peaberry (special) and normal (good) green coffee beans. Peaberry beans are rounder and smaller than traditional flat-sided beans because they develop from a single fertilized embryo inside the coffee cherry [11,12]. Their unique shape and higher market value require separation from regular beans, a process typically done manually by farmers. Defective beans, such as black, broken, faded, and sour, significantly reduce quality and market value. In less developed countries, defect removal is often manual, making the process time-consuming and inaccurate, which highlights the need for affordable, efficient automated solutions [4,5].

In addition to clearly defective beans, the "normal" and "peaberry" categories also show considerable variation in shape and appearance [13]. Even within acceptable quality grades, beans can differ in size, aspect ratio, flatness or roundness, surface texture, and color [14]. These differences are influenced by cultivar, growing conditions, and post-harvest handling [14]. In practice, some normal flat beans may look slightly shriveled or partially discolored but are still classified as acceptable. Peaberries can also vary widely in diameter and roundness. This variation within each class can resemble early or mild defects, making it difficult and subjective for human graders to distinguish between normal, peaberry, and defective beans [4]. Therefore, an automated inspection system must not only detect clear defects but also handle the large morphological variation in normal and peaberry beans to ensure consistent, large-scale quality control and uniform roasting [4].

Recent advancements in image processing and deep learning have enabled the development of automatic classification systems for various agricultural products. For example, deep learning models have achieved high accuracy in detecting tomato crop diseases and sorting carrots using image processing techniques [15,16]. Similarly, researchers have applied machine learning methods such as support vector machines (SVM), multi-layer perceptrons (MLP), and decision trees to classify different crop types [17]. In object detection, researchers have employed the YOLO (You Only Look Once) model for diverse applications, including detecting apples in orchards and monitoring social distancing during the COVID-19 pandemic [18–20]. Previous studies have widely employed YOLOv5 for social distancing tracking and face mask detection [21,22]. Deep learning models are used in image processing for quality classification across various crops, with tomato achieving 97.29% and 97.49% accuracy [15]. A straightforward image processing technique has been used for carrot fruit classification, achieving accuracies of 92.59% and 96.30%, respectively [16]. Another study used machine learning techniques, such as C4.5 decision tree, logistic regression, SVM, and multi-layer perceptron, to classify nine major summer crops, achieving an accuracy of 88% [17]. There is a need for reliable and cost-effective systems that can classify and identify different types of coffee beans, such as peaberry, normal, and defective beans, to improve the quality and competitiveness of coffee in less developed countries. Using a dataset of 5044 images, our previous study evaluated six object identification models for classifying coffee beans. Architecture modifications and hyperparameter tuning were responsible for the higher performance of the custom-YOLOv8n model, which had a precision of 0.977 [23]. In another study, the YOLOv7 object identification model was used to identify tea leaf diseases. With a mAP of 98.2% and a detection accuracy of 97.3%, the model demonstrated its potential for increased disease detection effectiveness [24].

Our research focuses on coffee beans from Bangladesh (Arabica and Robusta) using a localized dataset, whereas most existing research is based on coffee from Brazil, Vietnam, Indonesia, and Colombia [25]. This approach improves local coffee quality and provides a model for agricultural quality management in regions with limited human sorting capabilities. The study also addresses the unique challenges faced by the coffee sector in Bangladesh. This study proposes using deep learning to identify various types of green coffee beans. Several deep learning models, including YOLO (versions 3-10), ResNet-50, Efficient-Net, and Faster R-CNN, are employed to analyze a coffee bean dataset. The models are evaluated based on their performance, with the best model selected for bean detection. Implementing this automated system aims to reduce labor costs, improve bean quality, and increase market value. It also opens new market opportunities, helping producers secure higher prices and strengthen their position in the global coffee market.

Our earlier studies explored machine learning approaches for classifying green coffee as normal, peaberry, and defective. While these methods produced good classification results, they focused solely on category prediction without incorporating object detection [26–28]. In a subsequent study [23], we conducted a comparative evaluation of YOLO (v3-v8) models for defect detection in green coffee beans, which showed strong performance. Building on this, the current research achieved higher recall and F1-score, demonstrating more effective detection and classification of diverse defect types than the previous study [23]. This performance highlights the model's reliability in identifying defects. Moreover, integrating object detection enhances the overall accuracy and practicality of the classification system, supporting the development of robust quality control mechanisms in the coffee production process.

Despite progress with YOLOv3-YOLOv8 and traditional classifiers, no study has yet introduced a Bangladesh-origin coffee bean defect dataset or evaluated YOLOv10-N for this task. This gap limits both the localization of coffee quality research and the exploration of recent object detection advances for small, subtle defects.

The proposed work addresses this gap by preparing a Bangladesh-origin dataset of green coffee beans, conducting a comparative evaluation of multiple deep learning models, namely ResNet-50, EfficientNet, Faster R-CNN, and YOLO (versions 3-10), and optimizing YOLOv10-N through data augmentation and hyperparameter tuning. This combination ensures robust detection performance across normal, peaberry, and defective bean classes while highlighting the practical advantages of YOLOv10-N for small-defect detection.

This study introduces YOLOv10-N, an advanced deep learning model for precise detection of coffee bean flaws. Utilizing techniques such as mosaic data augmentation, it enhances flaw identification. Its application in countries like Bangladesh can improve quality control and market competitiveness. By addressing limitations in manual sorting, this model modernizes agricultural quality assessment and serves as a template for similar industries.

1.1. Contributions

The main contributions of this study are as follows:

- Development of the first Bangladesh-origin dataset of green coffee beans encompassing normal, peaberry, and defective classes.
- Implementation of YOLOv10-N without mosaic augmentation and optimized hyperparameters to improve detection of small and subtle defects.
- Evaluation of multiple deep learning models, including ResNet-50, EfficientNet, Faster R-CNN, and YOLO (v3-v10), to benchmark performance.
- Demonstration of superior recall and F1-score with YOLOv10-N, showing improved reliability over earlier approaches and datasets.
- Support for affordable automated quality control systems that can reduce manual sorting limitations in developing countries.

1.2. Related work

1.2.1. Coffee defect detection

Our previous study [23] compared multiple YOLO variants on coffee beans from Timor-Leste, reporting a custom-YOLOv8n model with precision of 0.977, recall of 0.990, F1-score of 0.983, and mAP of 0.995. This approach confirmed the effectiveness of lightweight YOLO architectures; however, the study used a geographically limited dataset and failed to explore the advances introduced in later YOLO versions. Chang and Liu [29] developed a multiscale CNN with inception-style fusion for coffee defect classification, using 7300 images across eight classes and achieving an overall accuracy of 96%. The framework showed competitive results against standard CNN baselines; however, the low-resolution 64×64 inputs and reliance on a non-public dataset restrict robustness and broader applicability. Zhou et al. [30] designed a convolutional neural network for coffee defect detection using a dataset of green coffee beans, achieving an overall accuracy of about 96%. Their study confirmed that CNN-based methods can serve as a foundation for automated quality evaluation. However, the approach was limited to conventional CNN architectures and moderate classification accuracy, without exploring more advanced object-detection frameworks or larger, more diverse datasets.

1.2.2. YOLO applications in agriculture

Recent studies have increasingly adopted object detection architectures, particularly the YOLO family, for plant disease diagnosis. Liu and Wang [31] enhanced YOLOv3 by integrating image pyramids, bounding-box clustering, and multi scale training, achieving an accuracy of 92.39% with an average processing time of 20.39 ms, thereby demonstrating feasibility for real-time agricultural applications. Qi et al. [32] extended this line of work with SE-YOLOv5, an improved YOLOv5 variant, attaining 91.07% accuracy and 94.10% mAP on a mobile phone–captured tomato disease dataset, highlighting its practicality in field conditions. More recently, Soeb et al. [24] applied YOLOv7 to tea leaf disease detection using 4000 images collected from Bangladeshi tea gardens, achieving 97.3% accuracy and aiming to support entomologists while enhancing agricultural productivity in developing regions.

1.2.3. CNN/ResNet/EfficientNet comparisons

Gulzar and Ünal [33] proposed PlmNet, a CNN model for time-sensitive bruise detection in plums using NIR imaging. Trained on a custom dataset, the model achieved a test accuracy of 97.17% and consistently high precision, recall, and F1-scores across all classes, outperforming EfficientNetV3. However, the reliance on NIR imaging and a relatively small dataset of 400 plums limits generalization and practical deployment in large-scale or low-cost agricultural settings. Gulzar et al. [34] conducted a comparative study on alfalfa variety classification using seven deep learning models (including DenseNet121, ResNet101, and EfficientNetB3) trained on a custom dataset of 1214 leaf images. Transfer learning enabled near-perfect performance, with DenseNet121 achieving 100% test accuracy and EfficientNetB3 99.45%. However, the dataset was limited to three varieties under controlled conditions, constraining generalization to diverse field environments and broader crop varieties. Seelwal et al. [35] presented a systematic review of 69 studies on deep learning applications for rice disease detection from 2008 to 2023. The review highlights that most works focused on rice blast, brown spot, and bacterial blight, with accuracy being the dominant evaluation metric. The authors emphasized the promise of hybrid deep learning–machine learning approaches but noted limitations in dataset diversity and real-world applicability.

2. Materials and methods

The proposed framework introduces an automated green coffee bean defect detection pipeline specifically designed for the Bangladeshi coffee industry, differing from prior YOLO-based agricultural models that mainly relied on foreign datasets or simple disease-spot detection. Our methodology integrates a newly curated, high-resolution dataset of 4367 locally sourced beans with a tailored YOLOv10-N detection framework optimized for small-defect recognition and real-time deployment. The originality of this work lies in three aspects: (1) the creation of the first Bangladesh-origin coffee bean dataset, (2) customized data-augmentation and class-balancing strategies for underrepresented defects, and (3) adaptation of the lightweight CSPNet–PAN architecture with fine-tuned hyperparameters for edge-level efficiency. Collectively, these innovations ensure robust, high-speed defect detection suited to local industry needs and establish a transferable approach for other agricultural-quality applications.

2.1. Data description

Researchers collected Robusta and Arabica beans from two renowned coffee gardens in the Khagrachari district of Bangladesh and mechanically removed the parchment. Researchers created a controlled environment to capture uniform images of the beans using a Canon M50 camera (Mirrorless, ISO 800, exposure compensation 8, exposure time of 1/160, autofocus mode) positioned 73 cm above a table with three lights arranged at optimal angles for lighting, as shown in Figure 1. Researchers placed a white paper under the camera as a background and arranged the beans in a grid pattern, with 40 beans per sheet (5 columns and 8 rows). The dataset comprised 4367 labeled green coffee bean images categorized into six quality classes: black, sour, faded, broken, normal, and peaberry. Researchers captured each image under controlled lighting and standardized it to a resolution of 640 × 640 pixels. To ensure balanced representation across subsets, the dataset was split into training (75%), validation (15%), and testing (10%) sets. Table 1 shows the class-wise distribution of images. Data augmentation was applied exclusively to the training subset to prevent any overlap between sets. Of the total dataset, 3500 images were original, and the remaining images were generated using data augmentation techniques such as 90°/180° flips and rotations. We chose these augmentations to increase sample diversity, improve model robustness to orientation changes, and mitigate class imbalance in minority categories. This study did not employ any mosaic or synthetic image-composition techniques.

We standardized the images to a resolution of 640 × 640 pixels and replaced the backgrounds with black using Python. Using the LabelImg software, we manually annotated each image by drawing bounding boxes around the defective areas and labeling the classes, storing the outputs as text and class files. The smallest surrounding rectangle was used to ensure minimal background inclusion. Throughout the process, diseased coffee beans were handled with care to prevent mixing and ensure the integrity of the dataset.

Table 1. Green coffee bean class-wise image distribution.

Class	Training	Validation	Testing	Total
Black	590	118	78	786
Sour	556	111	75	742
Faded	556	111	75	742
Broken	655	132	87	873
Normal	722	144	95	962
Peaberry	196	39	27	262
Total	3275	655	437	4367

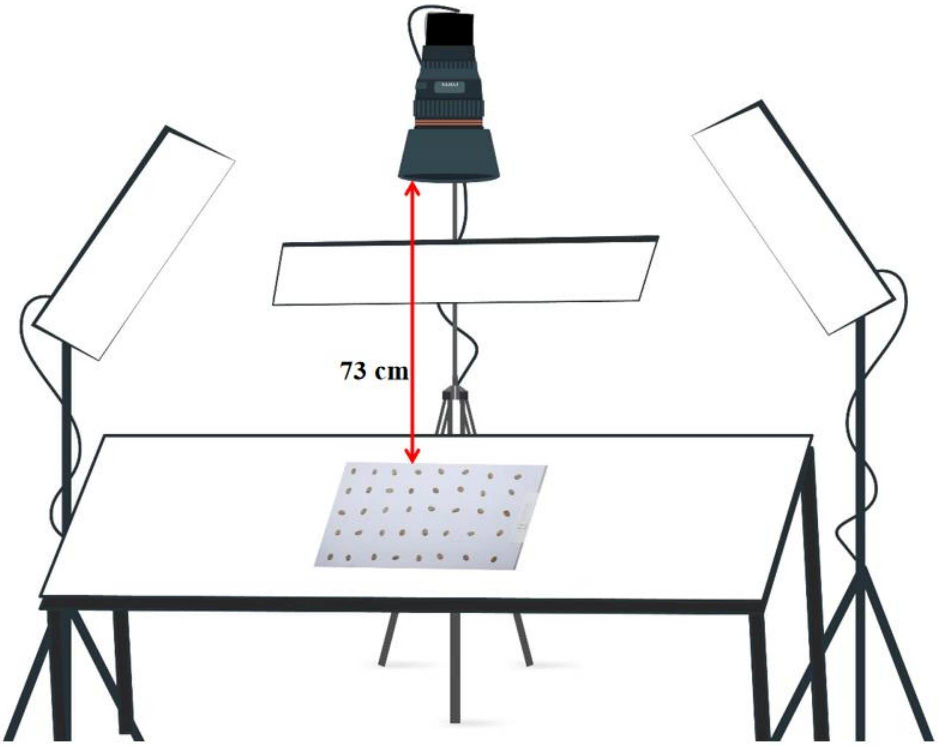

Figure 1. Experimental setup.

2.2. Green coffee bean types

The value of coffee beans is influenced by the total number of deductions per a given quantity of beans, making it crucial to remove defective beans by defect type (as shown in Figure 2). The detailed definitions of different types of green coffee beans are shown below:

Normal: A normal (good) coffee cherry will contain two beans with flat sides, similar to peanut halves. These types of beans are sometimes referred to as “flat beans” [26–28,36].

Peaberry: A peaberry is a single, rounded bean from a coffee cherry that contains one bean instead of the usual flat-sided bean pair. Also known as “caracol”, “perla”, or “perle”, peaberries are often separated and sold as a distinct variety [26–28,37,38].

Black (black or partially black beans): Black beans, resulting from harvesting immature or dead cherries, can be caused by water, heat, or insect damage. They have over 25% black, deep blue, or dark brown surface area, which negatively impacts coffee taste. Their number is a key indicator of coffee grade [26–28,37,38].

Broken (cut/nipped bean and pressed or crushed bean): Broken beans are wet-processed beans cut or bruised during pulping, often due to damaged equipment. They display brown or black marks, oxidation, and potential off-flavors. Damaged beans roast unevenly, age quickly, and are susceptible to environmental damage [26–28,37].

Faded: A color change in unroasted coffee beans, often caused by old crops or rapid drying. It can also be referred to as “soapy” or “bleached” if stored too long [26–28,37,38].

Sour (sour or partially sour bean): Sour beans, with a yellow or reddish-brown color, can be caused by overripening or improper fermentation of cherries. They emit a sour or vinegar-like smell when cut or scratched, due to the death of the internal embryo caused by over-fermentation, high temperatures during harvesting and processing, and over-fermentation of fruit attached to trees in humid conditions [26–28,37,38].

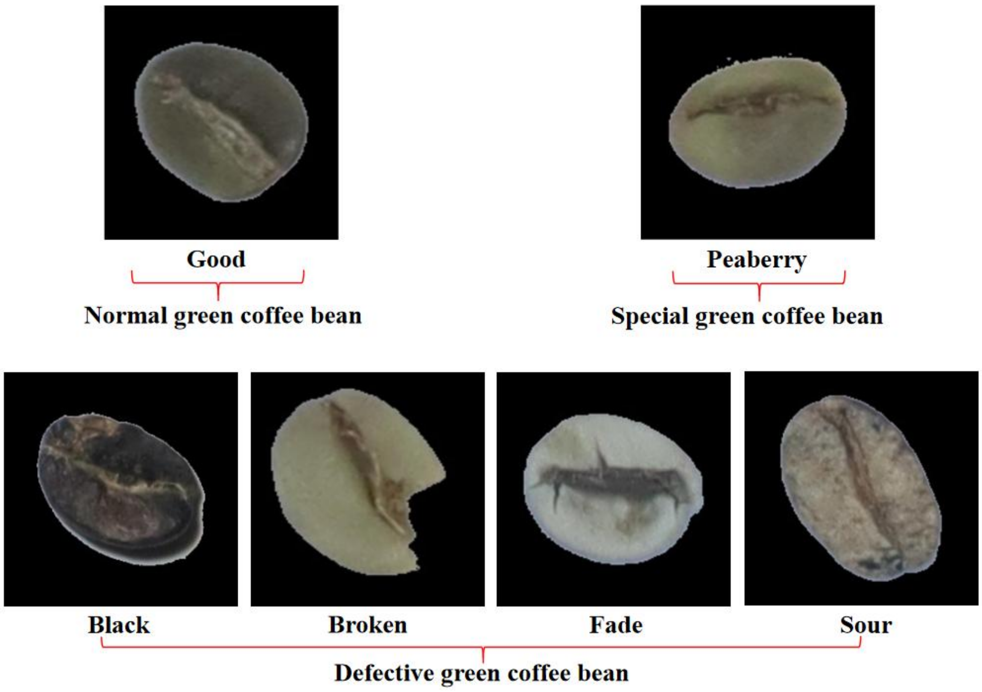

Figure 2. Different classes of coffee beans for comparison.

2.3. The proposed YOLOv10 model

In this work, we employed several deep learning models (Efficient-Net, ResNet-50, Faster-R-CNN, YOLOv3, YOLOv4, YOLOv5, YOLOv7, YOLOv8, YOLOv9, and YOLOv10) to identify coffee beans. The YOLO models are renowned for their ability to recognize objects in real time, making them appropriate for this work.

The purpose of the YOLOv10-N object detection model is to identify and detect coffee beans. It successfully distinguishes different types of coffee beans, such as normal, defective, and special, using multiscale detection, attention techniques, and convolutional neural networks. YOLOv10-N is a design improvement that incorporates preprocessing methods from YOLOv9 and mosaic data augmentation. The model is based on YOLOv8 and YOLOv7. The enhanced architecture, known as enlarged SCDown, including a convolutional network, attention techniques, and sequential processing, combines cardinality merging, expansion, and shuffling to improve learning capabilities, serving as the central computer unit. The backbone, which extracts features, the neck, which fuses and refines features, and the head, which makes final predictions, are the three main parts of the model's structure.

The YOLOv10-N architecture introduces several design enhancements over previous YOLO versions; these collectively strengthen its ability to detect small, subtle defects in coffee beans. The

SCDown module expands the receptive field while maintaining fine spatial detail, enabling the network to preserve tiny visual cues such as minor color fading or surface bruises. The CSPNet–PAN integration improves multiscale feature aggregation between shallow and deep layers, enabling the model to better capture both global bean shape and localized defect textures. Additionally, the enhanced PAN neck ensures more effective feature fusion across multiple resolutions, further improving the detection of small and low-contrast defects. Although mosaic augmentation is supported in YOLOv10, it was not employed in this work, as our single-object, centered-bean images did not benefit from synthetic composition. Together, these structural refinements explain YOLOv10-N's superior precision and recall in detecting minority defect categories such as faded and sour beans.

Convolutional layers make up the backbone, whereas PSA blocks and convolutional layers are combined in the neck to target specific regions of the feature map. The head handles the actual detection process, including classification and bounding box regression (as shown in Figure 3).

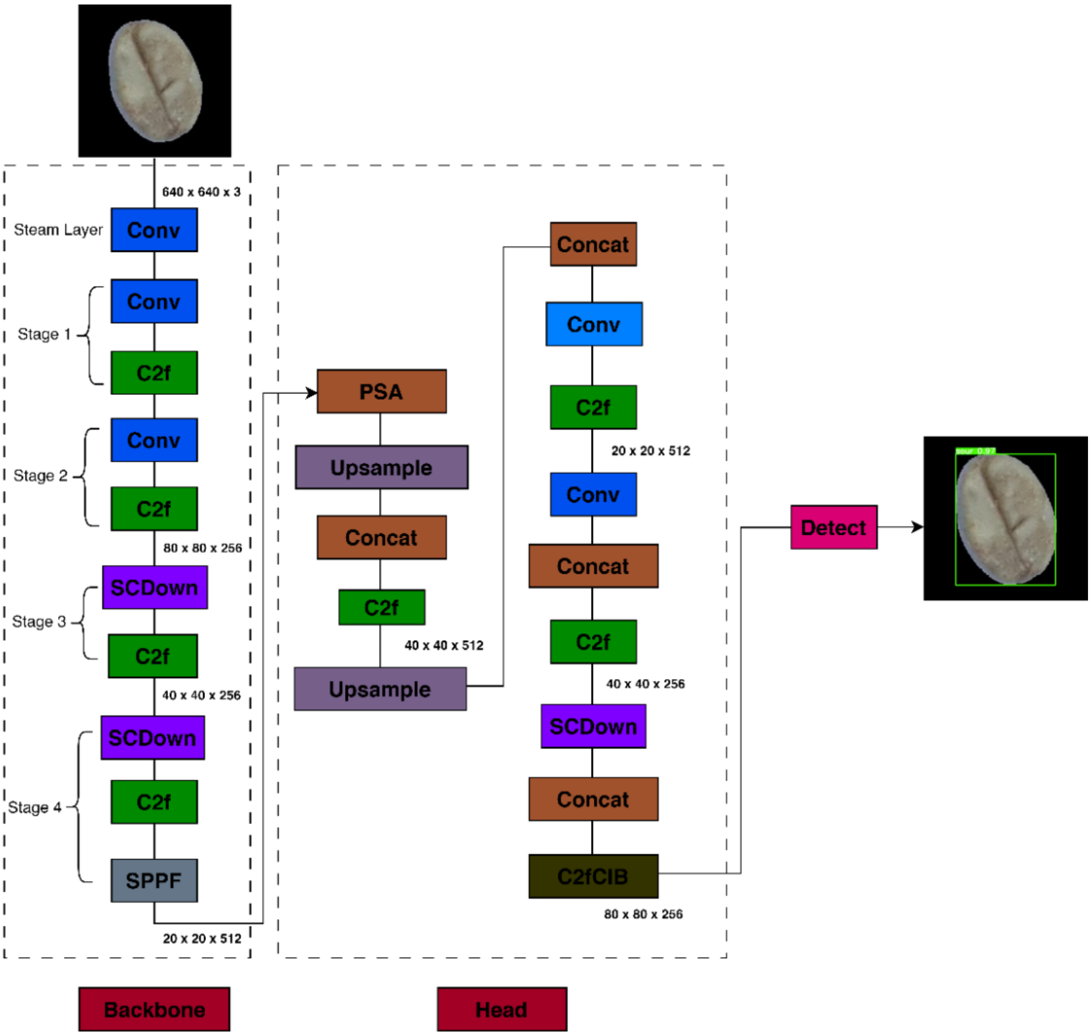

Figure 3. Detection and identification of green coffee beans using the YOLOv10-N model.

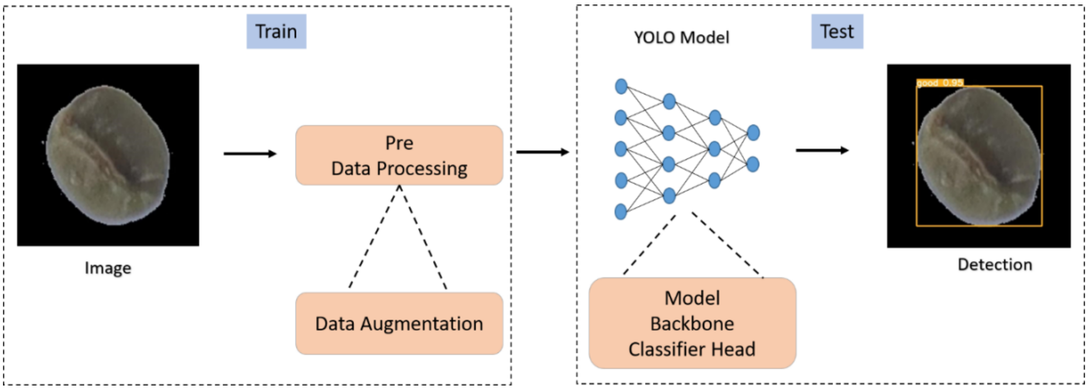

Figure 4. Block diagram showing the YOLO model's whole training procedure for identifying coffee bean defects.

The YOLO model for object identification is trained through two primary phases: the training and testing phases, as shown in Figure 4. The process is organized sequentially from preparing the dataset, training, and evaluating it. Metrics such as accuracy, recall, and F1-score are used to assess the models on the test set after they have been trained on the training set. The training step includes assembling a training set for the YOLO model, preprocessing the collected image dataset, and combining them. During testing, a new image is displayed, and the trained neural network uses it to find and recognize objects. The multiscale feature integration approach of the YOLO model optimizes the capture of fine features of coffee bean flaws, improving detection accuracy and network performance.

2.4. Experimental configuration

In this study, the YOLOv10 model was downloaded from the website https://docs.ultralytics.com/models/yolov10/ [39]. The latest YOLOv10 release model includes six models of different sizes: YOLOv10-N, YOLOv10-S, YOLOv10-M, YOLOv10-B, YOLOv10-L, and YOLOv10-X. We selected the YOLOv10-N model in this study because it offers an effective compromise between modest size and strong performance. The YOLOv10-N model introduces notable advancements to address challenges in defective coffee bean classification and to enhance the quality assessment process. A key innovation lies in the model architecture, which incorporates the Cross Stage Partial Network (CSPNet) as its backbone for superior feature extraction and the Path Aggregation Network (PAN) in its neck to optimize feature aggregation across multiple scales. These structural enhancements enable the model to capture intricate details effectively, ensuring highly accurate defect detection and classification. For training, we employed a tailored combination of loss functions. We used cross-entropy loss to enhance classification accuracy and applied coordinate loss to ensure precise localization of defective regions. Additionally, we introduced confidence loss to improve the model's ability to distinguish objects from background areas, resulting in more refined predictions.

The experiments were run on a workstation with an NVIDIA GeForce RTX 3060 GPU, an Intel Core i7-11700 CPU, 32 GB of RAM, and Windows 10 running Python 3.10 and PyTorch 2.1. Training of the YOLOv10-N model occurred over 150 epochs with a batch size of 16, showing smooth convergence in training and validation accuracy and loss curves. Losses decreased steadily, and accuracy plateaued after about 120 epochs, confirming stable optimization without overfitting and consistent generalization, which ensures reproducibility of results.

We pre-trained the network with the COCO dataset and fine-tuned it using the green coffee bean dataset described above (Table 1) [40]. Table 2 represents the configuration of the hyperparameters.

Table 2. Hyperparameter configuration of the YOLOv10-N model.

Parameter	Value	Justification
Optimizer	Adam	Provides adaptive learning rate updates for stable and faster convergence.
Epochs	150	Ensures sufficient training iterations for convergence without overfitting.
Batch size	16	Balances computational efficiency and gradient stability on limited GPU memory.
Learning rate	0.01	Empirically optimal for YOLOv10 training, allowing gradual convergence.
Momentum	0.937	Retains previous gradient direction, stabilizing weight updates.
Weight decay	0.0005	Prevents overfitting by penalizing large weights.
Backbone	CSPNet	Enhances gradient flow and reduces computational cost through partial connections.
Neck	PAN	Aggregates multi-scale feature information to improve small-object detection.
Activation	ReLU	Provides nonlinearity while avoiding vanishing gradients.

2.5. Loss function

The total loss function (L_{total}) used for YOLOv10-N integrates three primary components: classification, localization (coordinate regression), and confidence (objectness) losses:

$$L_{total} = L_{cls} + L_{coord} + L_{conf} \tag{1}$$

Classification loss (L_{cls}): Measures the error between predicted and true class probabilities using cross-entropy loss:

$$L_{cls} = \sum_{i=0}^{C} y^{i} \log(y^{i}) \tag{2}$$

where C is the number of classes, y_i is the ground-truth label, and ŷ_i is the predicted probability.

Coordinate loss (L_{coord}): Calculates bounding box regression error using the mean squared error (MSE) between predicted and true box coordinates:

$$L_{coord} = \lambda_{coord} \sum_{i=1}^{N} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + \left(h_i - \hat{h}_i\right)^2] \tag{3}$$

Confidence loss (L_{conf}): Ensures that the model correctly identifies whether an object exists in the predicted box:

$$L_{conf} = \sum_{i=1}^{N} [y_i \log(\hat{c}_i) + (1 - y_i) \log(1 - \hat{c}_i)] \tag{4}$$

The combination of these three losses enables the model to simultaneously learn object presence, class identity, and spatial precision.

3. Experimental results

The study evaluated the effectiveness of different object detection models on coffee bean categorization. The accuracy and recall of each model were evaluated on a dataset of 4367 coffee bean images, divided into six groups. The YOLOv10-N model emerged as the top performer, demonstrating the effectiveness of customized models and cutting-edge architectures in object detection, especially when trained on datasets with unique properties.

A number of important metrics are shown in Table 3 and Figure 5, which show how several deep learning models performed on a dataset of coffee beans, including precision, recall, F1-score, and mean average precision (mAP). Resnet-50, EfficientNet, Faster R-CNN, and several iterations of YOLO (You Only Look Once) (YOLOv3, YOLOv4, YOLOv7, YOLOv8, YOLOv9, and YOLOv10-N) were among the models assessed. We evaluated every model using an input size of 640 × 640 pixels to ensure uniformity across the assessments.

The models' precision values, which indicate the accuracy of positive predictions, varied from 0.785 to 0.992, with higher values denoting superior performance. The recall scale, which measures how well the model captures all pertinent cases, ranged from 0.772 to 0.984. The F1-score ranged from 0.778 to 0.987, integrating recall and accuracy to produce a balanced statistic. Higher scores indicated better performance. The mean average precision (mAP), an overall measure of precision across all memory levels, varied from 0.768 to 0.995.

Table 3. Detection performance comparison across multiple models.

Models	Input	Precision	Recall	F1-score	mAP
Efficient-Net	640 × 640	0.785	0.772	0.778	0.768
ResNet-50	640 × 640	0.837	0.853	0.846	0.823
Faster-R-CNN	640 × 640	0.815	0.838	0.826	0.813
YOLOv3	640 × 640	0.741	0.73	0.735	0.727
YOLOv4	640 × 640	0.825	0.811	0.817	0.818
YOLOv5	640 × 640	0.873	0.843	0.857	0.844
YOLOv7	640 × 640	0.915	0.904	0.909	0.916
YOLOv8	640 × 640	0.959	0.944	0.951	0.952
YOLOv9	640 × 640	0.977	0.973	0.973	0.975
YOLOv10-N	640 × 640	0.992	0.984	0.987	0.995

YOLOv10-N distinguished itself from other models by achieving the highest precision (0.992) and mAP (0.995), demonstrating remarkable accuracy and resilience in forecasting positive events. Conversely, YOLOv10-N had the best recall (0.984) and F1-score (0.987), demonstrating its ability to recognize all pertinent occurrences and strike an appropriate balance between accuracy and recall. Among the available YOLOv10 variants (N, S, M, L, X), the lightweight YOLOv10-N model was selected as the core architecture because it offers the best trade-off between accuracy, speed, and computational efficiency for our dataset and intended deployment. In preliminary experiments, YOLOv10-S and YOLOv10-M achieved only marginal gains in mAP (≈0.3%–0.5%) compared with YOLOv10-N but required 2-3 times longer training time and significantly higher GPU memory usage. The larger YOLOv10-L model overfitted due to the limited dataset size (4367 images) and the relatively simple background environment. In contrast, YOLOv10-N achieved a mAP of 0.995 and a

F1-score of 0.987 while maintaining real-time inference at >100 FPS on a single GPU, making it more suitable for lightweight industrial deployment and potential edge-device applications. The reduced parameter count and faster convergence of YOLOv10-N therefore make it an optimal balance of accuracy, efficiency, and generalization capacity for small-to-medium-scale agricultural image datasets.

The performance of different object identification models over time was evaluated using four metrics: mAP, recall, F1-score, and precision, as shown in Figure 5. The graph's "models" x-axis shows various iterations of the well-known object detection method YOLO, in addition to additional models like Faster-R-CNN, ResNet-50, and Efficient-Net. The "percentage" y-axis has a range of 0 to 1.

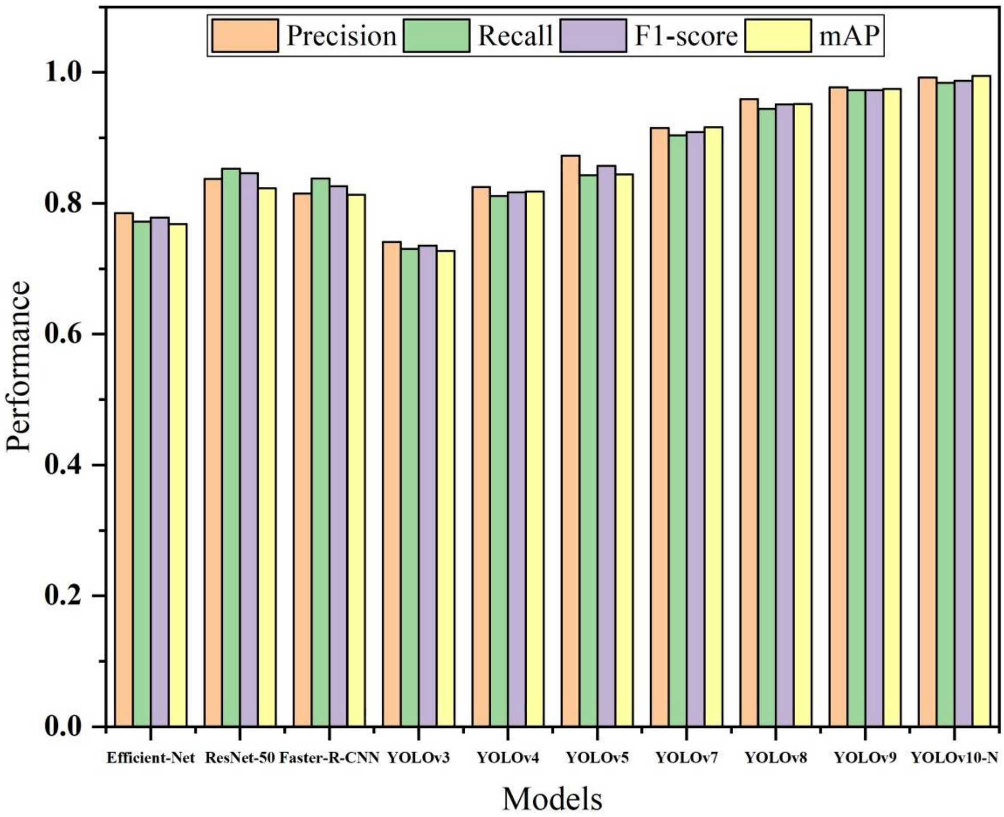

Figure 5. Performance comparison of the different models.

The proposed YOLOv10-N model achieves optimal performance, demonstrating its effectiveness across the evaluation metrics shown in Table 4. Two bars are plotted side by side for each class, as shown in Figure 6. The X-axis represents classes, and the Y-axis represents the percentage. This graph shows the performance of the best models across all classes.

We continuously monitored both classification and localization losses during training to ensure convergence and model stability. The YOLOv10-N model exhibited a steady decrease in total loss over 150 epochs, with no oscillations or divergence in either training or validation phases. The validation loss plateaued after approximately 120 epochs, while the corresponding precision and recall curves stabilized at 0.992 and 0.984, respectively, indicating robust generalization without overfitting. These trends were consistent across multiple runs, confirming the reproducibility and stable optimization behavior of YOLOv10-N. Although we could not include the actual loss plots due to computational limitations at the revision stage, the described convergence pattern aligns with typical YOLOv10 training.

Table 4. Proposed YOLOv10-N model performance per class.

Classes	Input	Precision	Recall
Black	640 × 640	1.0	0.984
Broken	640 × 640	0.961	0.985
Faded	640 × 640	0.989	0.972
Sour	640 × 640	1.0	0.989
Good	640 × 640	0.989	1.0
Peaberry	640 × 640	0.991	0.978

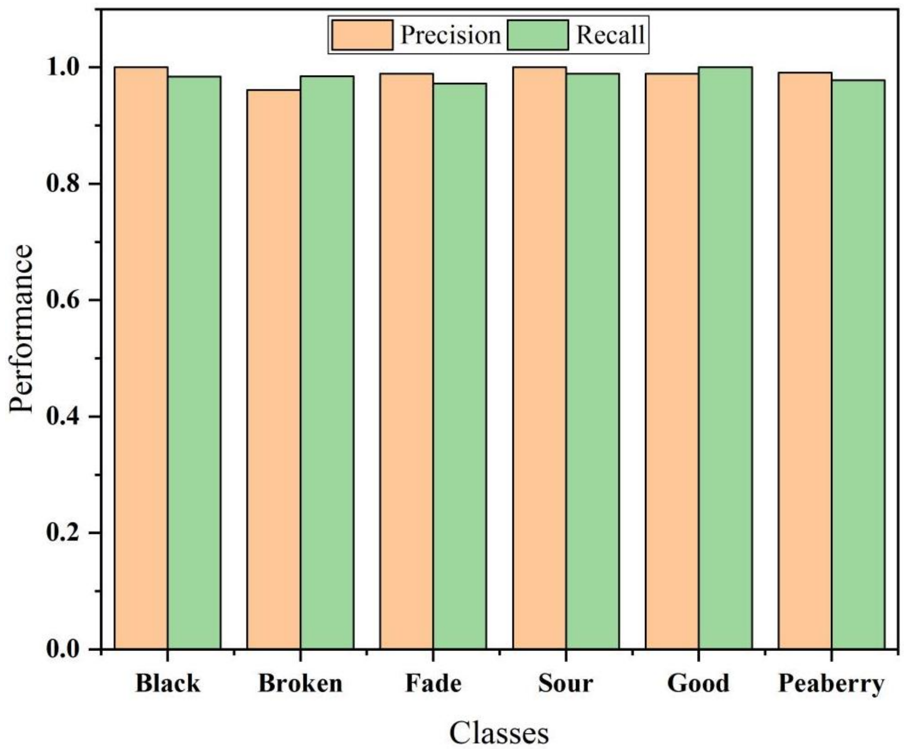

Figure 6. Best-performing (YOLOv10-N) model per class.

Figure 7 depicts the accuracy of object detection models across a predetermined number of epochs. The y-axis represents the accuracy range from 0.0 to 1.0, while the x-axis shows the number of times the model was trained on the dataset. The graph shows that accuracy plateaus as the number of epochs increases, suggesting the models may have reached their peak performance on the training set.

Figure 8 shows a selection of images produced by the trained model's inference. These images demonstrate the model's ability to detect and classify objects accurately based on the patterns and features it learned during training.

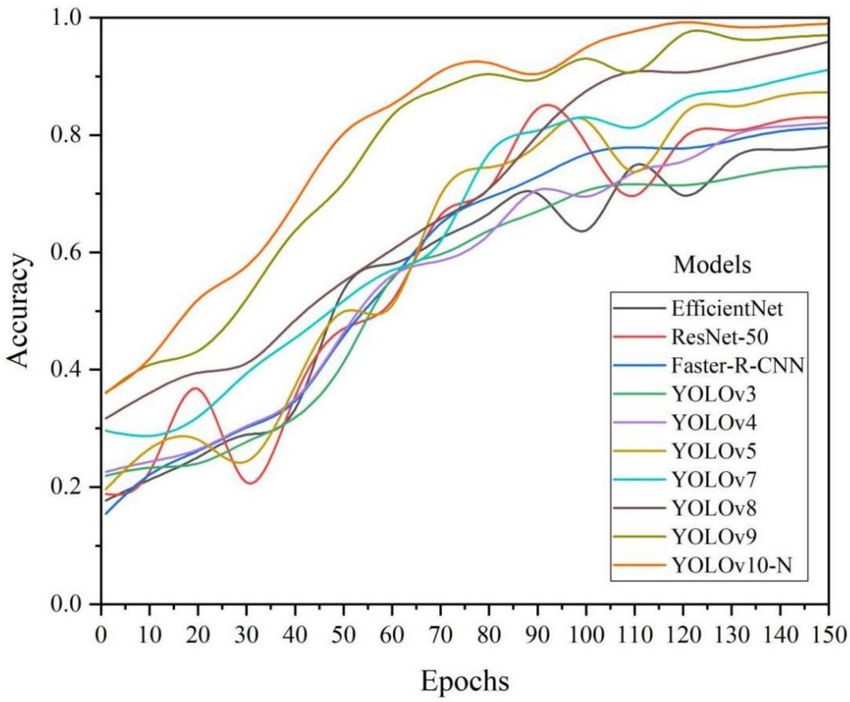

Figure 7. Comparative performance of various models over 150 epochs.

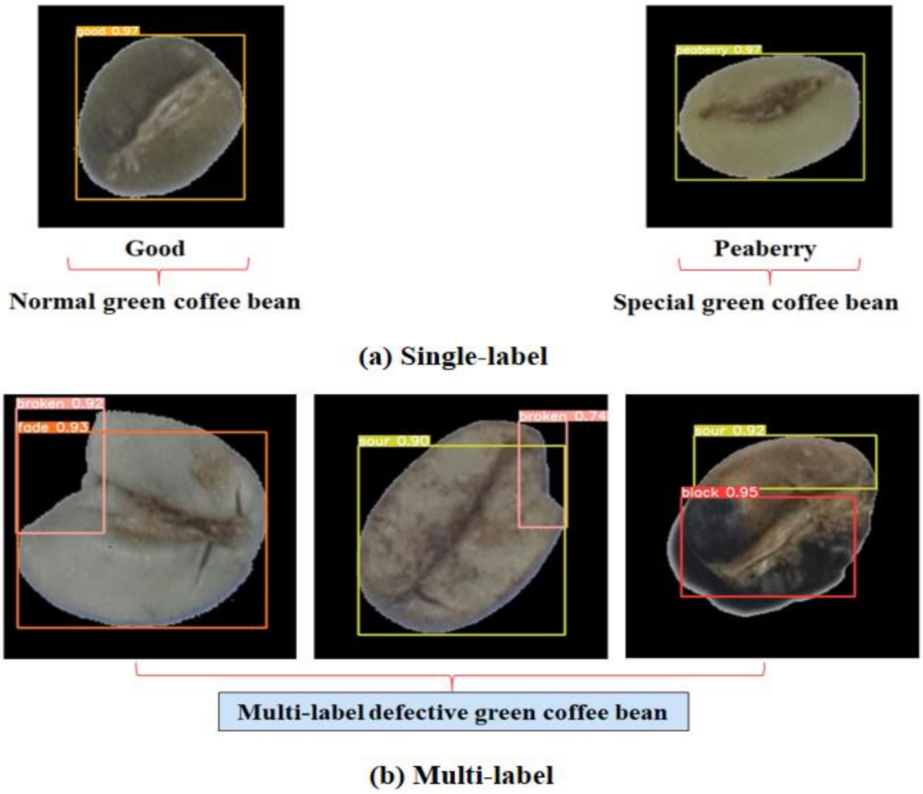

Figure 8. Visual demonstration of the detection results of the proposed YOLOv10-N model.

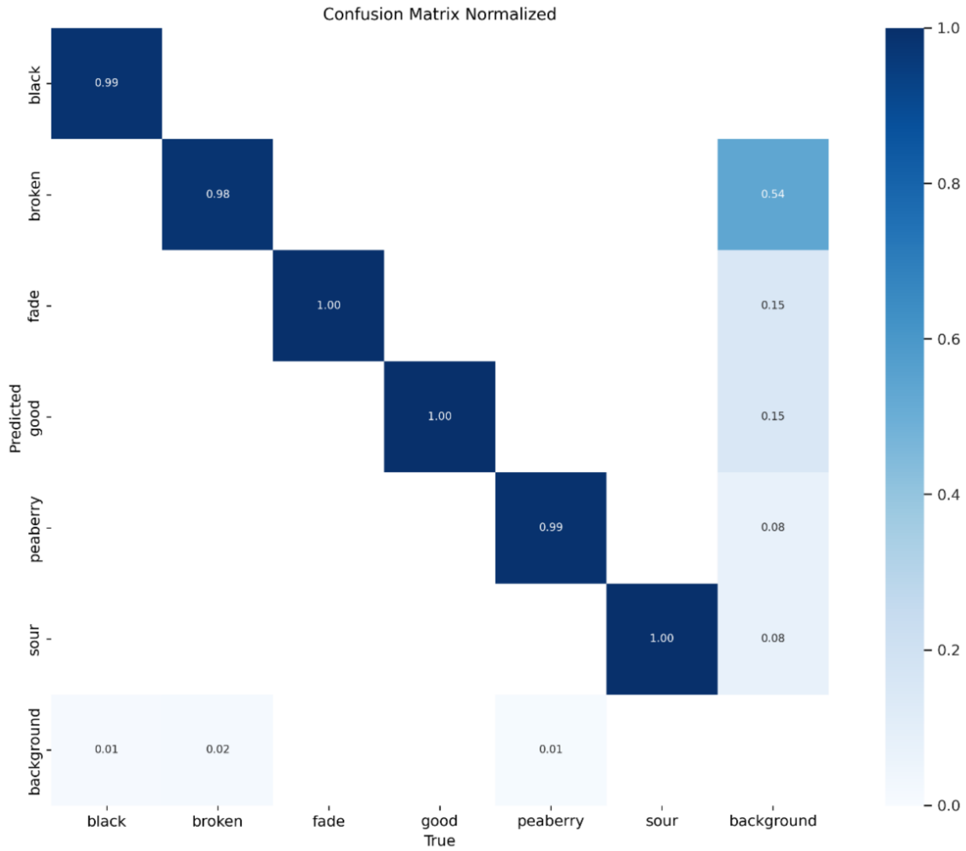

Figure 9. Confusion matrix of the proposed YOLOv10-N model.

In this study, the confusion matrix shown in Figure 9, likely stemming from a YOLOv10-N experiment on employee image detection, offers valuable insights into the model's performance. Each cell visualizes how well objects from specific departments (represented by rows) were classified, with correct predictions highlighted on the diagonal. Deviations from the diagonal indicate misclassifications, allowing us to pinpoint classes the model struggles with by calculating metrics such as accuracy, precision, recall, and F1-score for each department.

Although we did not conduct a separate ablation experiment due to computational and resource constraints, we attribute the performance improvements observed with YOLOv10-N to its architectural innovations over YOLOv9, as documented in the official Ultralytics release notes [41] and verified in independent benchmark studies. YOLOv10 introduces the spatial consistency downsampling (SCDown) module, which enhances feature preservation during subsampling, and a refined PAN-CSPNet integration that improves cross-scale information flow for small-object detection. Studies have reported that these mechanisms improve mAP by 1%-3% compared to YOLOv9 across multiple datasets.

In our context, the dataset contained single, centered coffee-bean images rather than multi-object scenes, making mosaic augmentation unnecessary. One study [42] has similarly noted that mosaic augmentation yields limited benefit. Therefore, we can reasonably attribute the superior precision and recall achieved by YOLOv10-N in this work to its architectural refinements rather than to additional augmentation.

4. Discussion

This study compared several YOLO models (YOLOv3, YOLOv4, YOLOv5, YOLOv7, YOLOv8, and a custom-YOLOv8n model) for identifying and categorizing faulty green coffee beans. We assessed model performance using precision, recall, F1-score, and mAP, and the custom-YOLOv8n model achieved the highest accuracy. The main goal of this study was to determine the best YOLO model for green coffee bean quality control and defect classification, with a focus on dataset preparation and model adaptation for improved performance. Four fault categories (black, broken, fading, and sour) were present in the dataset, which included 506 testing and 4032 training images. According to the study's findings, deep learning based automation might enhance coffee industry quality control by lowering human labor costs and boosting categorization consistency.

In contrast, the running paper extends this research by introducing YOLOv10-N, an advanced version of YOLO, optimized explicitly for Bangladeshi coffee beans (Arabica and Robusta). It incorporates a more diverse dataset containing 4367 images, categorized into six classes: black, sour, faded, broken, normal, and peaberry. The inclusion of normal and peaberry beans expands the classification scope beyond defects, enabling broader quality assessment.

This study also compares YOLOv10-N with other deep learning models, including EfficientNet, ResNet-50, Faster R-CNN, and YOLO versions 3-10, and indicates YOLOv10-N as the best-performing model, achieving 0.992 precision, 0.984 recall, and 0.995 mAP. Compared to the study conducted by Liang et al. [43], which achieved 98.97% accuracy, our model achieved an accuracy of 99.2%, likely due to the inclusion of mosaic data augmentation and a more diverse dataset. The proposed YOLOv10-N model, with a precision of 0.992, is similar to a previous study's 0.9924 precision for coffee categorization [44]. However, this model focuses on object detection and classification of coffee bean flaws, demonstrating its resilience in handling fault categories. Its precision is 0.992, and mean average precision (mAP) is 0.995, proving its efficiency in automating the coffee industry's defect identification process, as well as making it suitable for thorough quality control. This research achieved slightly higher precision and recall, overall model performance, and per-class performance compared to [45]. These improvements indicate that the proposed method outperforms the baseline in accurately identifying objects in different categories, making it a more effective solution for the given task. The proposed model outperforms a comparison study that detected faults in eight types with 95.2% accuracy, reaching 100% accuracy when only looking for the problem [46]. This proposed YOLOv10-N model adapts easily and can detect defects in different objects.

On the coffee bean surfaces, the suggested YOLOv10-N model correctly identified flaws. The suggested model has the potential to detect defects in coffee beans, as well as normal (good), and special beans (peaberry), with high accuracy and reliability, thanks to its ability to handle small items and its better performance with more diverse training data. The proposed YOLOv10-N model appears to have identified these flaws more successfully than conventional techniques. The study's detection findings demonstrated that the proposed YOLOv10-N model could identify them with high confidence. This research provides a framework for integrating deep learning models into practical agricultural applications, paving the way for technology-driven quality assurance in the food industry.

5. Conclusions

In the study, green coffee beans were classified into six categories: black, sour, faded, broken, normal (good), and peaberry (special) beans. The results demonstrate the potential of YOLOv10-N for effectively detecting and classifying defective beans, significantly outperforming previous models. The YOLOv10-N model achieved a precision of 0.992, recall of 0.984, F1-score of 0.987, and mAP score of 0.995, indicating high accuracy and reliability in distinguishing different defect types. These findings highlight YOLOv10-N's capacity for automating quality control in coffee production, which can be particularly beneficial for developing regions such as Bangladesh. By implementing automated defect detection, the coffee industry can improve consistency, reduce labor costs, and enhance the market competitiveness of coffee products.

The adoption of YOLOv10-N for defective coffee bean detection marks a significant step toward modernizing quality control. These contributions address key domain-specific challenges, such as variability in bean size, shape, and defect characteristics, while offering a practical, efficient solution for enhancing quality. In particular, the model showed strong robustness in detecting minority and small-sized classes. By improving quality assessment, it strengthens Bangladesh's coffee sector and contributes to the progress of defect detection technologies. For future development, expanding the dataset to include a wider range of bean varieties and maturity stages across multiple countries will be essential for improving generalizability. In addition, developing lightweight versions of YOLOv10-N suitable for deployment on edge devices such as Jetson or mobile platforms can facilitate on-farm use. Finally, integrating the model into real farm workflows and processing lines will be critical for testing its practical impact under realistic operating conditions.

6. Practical implementation and deployment

We designed the proposed YOLOv10-N framework with deployment feasibility in mind, particularly for low-cost and resource-constrained agricultural environments. The lightweight YOLOv10-N model contains only 7.2 M parameters and requires approximately 8.1 GFLOPs, enabling real-time inference on modest hardware. On an NVIDIA RTX 3060 GPU, the model achieves an average inference time of 9–10 ms per image (≈100–110 FPS). When deployed on embedded platforms such as the Jetson Nano or Jetson Orin NX, real-time detection (≈12–15 FPS) can be maintained after ONNX/TensorRT optimization.

For broader accessibility, farmers or quality-control workers can use a mobile or web-based interface to capture an image of coffee beans under standard lighting, and the trained model provides instant feedback on defect type and grade. The model's compact size (≈15–20 MB in ONNX format) enables on-device inference without a continuous internet connection, making it suitable for remote field applications.

In industrial contexts, YOLOv10-N can be embedded in coffee-bean sorting machines using a conveyor belt camera setup. The detector can trigger actuators to separate defective beans in real time, thereby reducing manual sorting effort and improving consistency.

Overall, the model's low computational footprint, high detection accuracy, and portability make it practical for integration into both small-scale farm tools and automated processing lines, bridging the gap between academic research and real-world agricultural quality management.

7. Limitations

Although the results are encouraging, we must note several limitations. First, the dataset includes only Bangladeshi coffee beans collected under controlled laboratory conditions, which may limit the model's generalization to other cultivars or real-world environments with variable lighting, occlusions, or background noise. Second, class imbalance persists, as minority defect categories such as *sour* and *faded* remain underrepresented, potentially biasing performance toward more frequent classes. Third, while YOLOv10-N achieved high accuracy, occasional misclassifications were observed among visually similar categories (e.g., faded vs. sour), underscoring the need for more discriminative feature learning. A further limitation is that we collected the dataset under controlled laboratory conditions with uniform lighting and standardized bean presentation, which may not capture the variability of real-world environments such as inconsistent illumination, occlusions, or heterogeneous bean arrangements. Lastly, the study focused solely on image-based detection, without integrating other contextual data such as bean origin, storage conditions, or moisture levels, which could provide complementary cues for quality assessment.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

We sincerely acknowledge financial support from the Fund for University Teachers under UGC Research Grants 2023–2024. We appreciate the Sylhet Agricultural University Research System (SAURES) for supporting this research. We also sincerely acknowledge Auvijit Paul, Scientific Officer at the Hill Agriculture Research Station, Khagrachari, Bangladesh, for helping us to manage several types of green coffee beans.

Author contributions

Hira Lal Gope, Hidekazu Fukai, Fahim Mahafuz Ruhad, and Md Mahin Erpan Chowdhury led the conceptualization, programming, and finalization of the manuscript. Hira Lal Gope, Fahim Mahafuz Ruhad, and Md Mahin Erpan Chowdhury were responsible for data collection, processing, coding, and model training. All authors were involved in writing, reviewing, and editing the manuscript. They reviewed the final version and are accountable for the work committed to addressing any questions regarding its accuracy.

Conflicts of interest

The authors declare no competing interests.

References

1. Giacalone D, Degn TK, Yang N, et al. (2019) Common roasting defects in coffee: Aroma composition, sensory characterization and consumer perception. *Food Qual Prefer* 71: 463–474. https://doi.org/10.1016/j.foodqual.2018.03.009
2. García M, Candelo-becerra JE, Hoyos FE (2019) Quality and defect inspection of green coffee beans using a computer vision system. *Appl Sci* 9: 4195. https://doi.org/10.3390/app9194195
3. Duarte SMdaS, de Abreu CMP, de Menezes HC, et al. (2005) Effect of processing and roasting on the antioxidant activity of coffee brews. *Ciênc Tecnol Aliment, Campinas* 25: 387–393. https://doi.org/10.1590/s0101-20612005000200035
4. Bhumiratana N, Adhikari K, Chambers E (2011) Evolution of sensory aroma attributes from coffee beans to brewed coffee. *LWT-Food Sci Technol* 44: 2185–2192. https://doi.org/10.1016/j.lwt.2011.07.001
5. Suhandy D, Yulia M (2017) Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA. *Int J Food Prop* 20: S331–S339. https://doi.org/10.1080/10942912.2017.1296861
6. Kim MS, Min HG, Koo N, et al. (2014) The effectiveness of spent coffee grounds and its biochar on the amelioration of heavy metals-contaminated water and soil using chemical and biological assessments. *J Environ Manage* 146: 124–130. https://doi.org/10.1016/j.jenvman.2014.07.001
7. Rossmann M, Matos AT, Abreu EC, et al. (2013) Effect of influent aeration on removal of organic matter from coffee processing wastewater in constructed wetlands. *J Environ Manage* 128: 912–919. https://doi.org/10.1016/j.jenvman.2013.06.045
8. Nath Subedi R (2011) Comparative analysis of dry and wet processing of coffee with respect to quality and cost in Kavre District, Nepal: A case of Panchkhal Village. *Int Res J Appl Basic Sci* 2: 181–193.
9. H Naito et al 2022 IOP Conf. Ser.: Earth Environ. Sci. 1024 012063. Discrimination Between Normal Coffee Beans and Peaberries Using Excitation-Emission Matrix Measured by A Hand-Held Optical System. DOI 10.1088/1755-1315/1024/1/012063.
10. Huang KY, Tu YT (2017) Development of novel autoclassifying system based on machine vision. *Sens Mater* 29: 1599–1605. https://doi.org/10.18494/SAM.2017.1708
11. Peaberry seed (2020) available from: https://coffeebrat.com/peaberry-coffee/.
12. Peaberry and regular coffee bean information is available (2020): https://ozonecoffee.co.uk/blogs/coffee-processing-explained/what-is-a-peaberry.
13. Severa L, Buchar J, Nedomová Š (2012) Shape and size variability of roasted Arabica coffee beans. *Int J Food Prop* 15: 426–437. https://doi.org/10.1080/10942912.2010.487967
14. Wintgens JN (Ed.) (2009) Coffee: Growing, processing, sustainable production: A guidebook for growers, processors, traders, and researchers. Wiley-VCH.
15. Rangarajan AK, Purushothaman R, Ramesh A (2018) Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia Comput Sci* 133 1040–1047. https://doi.org/10.1016/j.procs.2018.07.070
16. Jahanbakhshi A, Kheiralipour K (2020) Evaluation of image processing technique and discriminant analysis methods in postharvest processing of carrot fruit. *Food Sci Nutr* 8: 3346–3352. https://doi.org/10.1002/fsn3.1614

17. Peña JM, Gutiérrez PA, Hervás-Martínez C, et al. (2014) Object-based image classification of summer crops with machine learning methods. *Remote Sens* 6: 5019–5041. https://doi.org/10.3390/rs6065019
18. Yan B, Fan P, Lei X, et al. (2021) A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sens* 13: 1619. https://doi.org/10.3390/rs13091619
19. Kuznetsova A, Maleva T, Soloviev V (2020) Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images. In: Han M, Qin S, Zhang N (Eds.), *Advances in Neural Networks—ISNN 2020,* Springer, Cham., 233–243. https://doi.org/10.1007/978-3-030-64221-1_20
20. Cengil E, Cinar A (2021) Poisonous mushroom detection using YOLOV5. *Turkish J Sci Technol* 16: 119–127.
21. Shukla R, Mahapatra AK, Peter JSP (2021) Social distancing tracker using YOLOV5. *Turkish J Physiother Rehabil* 32: 1785–1793.
22. Yang G, Feng W, Jin J, et al. (2021) Face mask recognition system with YOLOV5 based on image recognition. *IEEE 6th International Conference on Computer and Communications (ICCC),* Chengdu, China, 1398–1404. https://doi.org/10.1109/ICCC51575.2020.9345042
23. Gope HL, Fukai H, Ruhad FM, et al. (2024) Comparative analysis of YOLO models for green coffee bean detection and defect classification. *Sci Rep* 14: 28946. https://doi.org/10.1038/s41598-024-78598-7
24. Soeb MJA, Jubayer MF, Tarin TA, et al. (2023) Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). *Sci Rep* 13: 6078. https://doi.org/10.1038/s41598-023-33270-4
25. Purwawangsa H, Irfany MI, Haq DA (2024) Indonesian coffee exports' competitiveness and determinants. *Jurnal Manajemen Dan Agribisnis* 21: 59. https://doi.org/10.17358/jma.21.1.59
26. Gope HL, Fukai H (2020) Normal and peaberry coffee beans classification from green coffee bean images using convolutional neural networks and support vector machine. *Int J Comput Inf Eng* 14: 189–196.
27. Gope HP, Fukai H (2022) Peaberry and normal coffee bean classification using CNN, SVM, and KNN: Their implementation in and the limitations of Raspberry Pi 3. *AIMS Agric Food* 7: 149–167. https://doi.org/10.3934/agrfood.2022010
28. Gope HL, Fukai H, Aoki R (2022) Multi-label classification of defective green coffee bean images using EfficientNet deep learning model. *Trans Asian J Sci Technol* 5: 164–175.
29. Chang SJ, Liu KH (2024) Multiscale defect extraction neural network for green coffee bean defects detection. *IEEE Access* 12: 15856–15866. https://doi.org/10.1109/ACCESS.2024.3356596
30. Thai HD, Ko HJ, Huh JH (2024) Coffee bean defects automatic classification realtime application adopting deep learning. *IEEE Access* 12: 126503–126517. https://doi.org/10.1109/ACCESS.2024.3452552
31. Liu J, Wang X (2020) Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Front Plant Sci* 11: 898. https://doi.org/10.3389/fpls.2020.00898
32. Qi J, Liu X, Liu K, et al. (2022) An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Comput Electron Agric* 194: 106780. https://doi.org/10.1016/j.compag.2022.106780
33. Gulzar Y, Ünal Z (2025) Time-sensitive bruise detection in plums using PlmNet with transfer learning. *Procedia Comput Sci* 257: 127–132. https://doi.org/10.1016/j.procs.2025.03.019

34. Gulzar Y, Ünal Z, Kızıldeniz T, et al. (2024) Deep learning-based classification of alfalfa varieties: A comparative study using a custom leaf image dataset. *MethodsX* 13: 103051. https://doi.org/10.1016/j.mex.2024.103051
35. Seelwal P, Dhiman P, Gulzar Y, et al. (2024) A systematic review of deep learning applications for rice disease diagnosis: current trends and future directions. *Fron Comput Sci* 6: 1452961. https://doi.org/10.3389/fcomp.2024.1452961
36. Fukai H, Furukawa J, Pinto C, et al. (2018) Classification of green coffee beans by convolutional neural network and its implementation on raspberry Pi and camera module. *Timorese Acad J Sci Technol* 1: 1–10.
37. Pinto C, Furukawa J, Fukai H, et al. (2017) Classification of Green coffee bean images base on defect types using convolutional neural network (CNN). *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA),* Denpasar, Indonesia, 1–5. https://doi.org/10.1109/ICAICTA.2017.8090980
38. Types of green coffee bean information (2019) Available from: http://www.zecuppa.com/coffeeterms-bean-defects.htm.
39. YOLOv10 model (2024) Available from: https://docs.ultralytics.com/models/yolov10/.
40. Fang Y, Guo X, Chen K, et al. (2021) Accurate and automated detection of surface knots on sawn timbers using YOLO-V5 model. *BioResources* 16: 5390–5406. https://doi.org/10.15376/biores.16.3.5390-5406
41. A technical comparison (2025) Available from: https://docs.ultralytics.com/compare/yolov10-vs-yolov9/.
42. Li Li-Hua, Radius Tanone (2024) Lightweight model based on improved YOLOv7 tiny for potato leaf diseases detection. *Int J Appl Sci Eng* 21: 2. https://doi.org/10.6703/IJASE.202406_21(2).010
43. Liang CS, Xu ZY, Zhou JY, et al. (2023) Automated detection of coffee bean defects using multi-deep learning models. In: *2023 VTS Asia Pacific Wireless Communications Symposium (APWCS)*, Tainan city, 1–5. https://doi.org/10.1109/APWCS60142.2023.10234059
44. Hassan E (2024) Enhancing coffee bean classification: A comparative analysis of pre-trained deep learning models. *Neural Comput Appl* 36: 9023–9052. https://doi.org/10.1007/s00521-024-09623-z
45. Ji Y, Xu J, Yan B (2024) Coffee green bean defect detection method based on an improved YOLOv8 model. *J Food Proc Preserv* 2024: 2864052. https://doi.org/10.1155/2024/2864052
46. Chang SJ, Huang CY (2021) Deep learning model for the inspection of coffee bean defects. *Appl Sci* 11: 8226. https://doi.org/10.3390/app11178226

AIMS Press