

Analysis of YOLO26 Model Performance with Transfer Learning in Detecting Coffee Bean Defects

Adrian Chen^{1*}, Eka Puji Widiyanto²

^{1,2}Multi Data Palembang University, Indonesia

^{1*}adrian.chen132@mhs.mdp.ac.id, ²ekapujiw2002@mdp.ac.id



*Corresponding Author

Article History:

Submitted: 20-05-2026

Accepted: 03-06-2026

Published: 08-06-2026

Keywords:

Coffee Defect; Deep Learning; Hyperparameter Tuning; Object Detection; Transfer Learning; YOLO26.

Brilliance: Research of

Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

Background: Indonesia is one of the world's largest coffee producers, yet post-roasting coffee bean defects remain a critical challenge that reduces product quality and market competitiveness. Manual sorting processes are inconsistent and prone to human visual limitations. Objective: This study aims to analyze the performance of the YOLO26 nano model with transfer learning for detecting and classifying post-roasting coffee bean defects, and to evaluate the effect of grid search-based hyperparameter tuning on model performance. Methods: A first-party dataset of 4,567 images covering five defect categories — insect damage, under roast, quaker, nugget, and shell — and one non-defect category was collected from three local Indonesian coffee roasting companies. After augmentation, the dataset expanded to 10,595 images with a 70:20:10 training-validation-testing split ratio. YOLO26, a deep learning-based object detection model released in January 2026, was applied using transfer learning and optimized through grid search hyperparameter tuning across optimizer, learning rate, epoch, classification loss, and weight decay configurations. Results: The model was evaluated using precision, recall, F1 score, mean Average Precision at IoU 50 (mAP50), and mean Average Precision at IoU 50-95 (mAP50-95) on the test dataset, with results demonstrating competitive multi-class detection performance across all defect categories. Conclusion: YOLO26 nano with transfer learning and hyperparameter tuning is a viable approach for automated post-roasting coffee bean defect detection, contributing to quality control advancements in the Indonesian coffee industry.

INTRODUCTION

Indonesia is one of the world's largest coffee producers, with coffee production and consumption increasing each year. According to data from the Central Statistics Agency (BPS), Indonesia's coffee exports in 2023 reached 276,300 tons with a value of USD 916.5 million, or approximately IDR 14.19 trillion (Badan Pusat Statistik Indonesia, 2024). In addition, a Snapcart survey conducted in September 2023 found that 79% of Indonesian people are coffee drinkers (Snapcart, 2023), reflecting the importance of coffee in daily life. This significant economic value highlights the need to maintain coffee bean quality as a primary factor influencing market prices internationally (Faojiah, Rahmah, Pramulya, Novita, & Sutrisno, 2025).

The quality of coffee beans is directly determined by the presence of defects, particularly in post-roasting stages. Defective coffee beans, caused by factors such as pests, plant diseases, or errors in post-harvest processing, can reduce cup quality and make products less competitive globally. Research has demonstrated that defects such as insect damage, under-roasted beans, quakers, nuggets, and shell beans significantly affect the physicochemical characteristics of coffee, including lower caffeine levels, higher pH values, darker color intensity, and the dominance of burnt aromas over the nutty aromas valued in specialty-grade coffee (Anggraini, Damat, & Manshur, 2024). Furthermore, the proportion of defective beans is one of the most critical factors in coffee grading, where beans failing sortation can only be marketed as commercial-grade coffee at substantially lower prices (Al-Mahish, Alfayadh, Al-Mahish, & Alfayadh, 2024; Hu, Quan, Dai, & Qiu, 2024).

Manual sortation, still widely employed in Indonesian coffee production, is prone to inconsistency and depends on human visual perception influenced by factors such as lighting conditions and individual visual acuity (Hafifah, Muchtar, Ahmadiar, & Esabella, 2022). This creates a critical need for automated, reliable defect detection systems.

Previous research has explored the use of computer vision and YOLO-based models for coffee bean defect detection. (Ahmed et al., 2025) compared four object detection models on Khawlani coffee beans and found YOLOv8 superior with mAP of 0.995. (Adiwijaya, Sarno, & Wijaya, 2024) combined YOLOv8 with the SAHI framework for real-time Robusta coffee defect detection, obtaining mAP of 0.979. More recently, Hebert & Alamsyah (2026) applied YOLOv12 for detecting 15 defect classes in green beans based on SCA standards, achieving precision of 87% and mAP50 of 84%. Chamorro-Pinchao, Puská-Chulde, Trejo-España, Caranqui-Sánchez, & García-Santillán (2025) used YOLOv11 nano and small variants for binary post-roasting defect classification, achieving accuracy up to 88.03%.



Despite these advances, no prior study has applied YOLO26 — the most recent YOLO architecture released in January 2026 — to the domain of post-roasting coffee bean defect detection. YOLO26 introduces key innovations including NMS-free inference, DFL removal, MuSGD optimizer, Progressive Loss (ProgLoss), and Small-Target-Aware Label Assignment (STAL), making it particularly well-suited for edge deployment scenarios (Chakrabarty, 2026; Hidayatullah & Tubagus, 2026). This study fills this research gap by applying YOLO26 nano with transfer learning and systematic grid search-based hyperparameter tuning on a dataset collected directly from local Indonesian coffee roasting industries.

LITERATURE REVIEW

Research on automated coffee bean quality assessment has evolved considerably over recent years. Traditional approaches using texture features and machine learning (Jumarlis et al., 2022) have progressively given way to deep learning methods. Murinto, Rosyda, & Melany (2023) applied CNN with transfer learning using VGG16 and MobileNetV2 architectures to classify coffee bean types (Arabica, Robusta, Liberica), achieving the highest accuracy of 96% with MobileNetV2. Michael, Rusman, & History (2023) addressed coffee defect classification using MobileNetV2 and DenseNet201 with grid search hyperparameter tuning, where DenseNet201 reached 93% accuracy across 11 defect categories.

In the domain of object detection for coffee defects, which builds on foundational work in real-time object detection, (Ahmed et al., 2025) demonstrated YOLOv8 outperforming YOLOv5, YOLOv9, and Faster R-CNN on Khawlani coffee beans with mAP50 of 99.5%. (Adiwijaya et al., 2024) showed that combining YOLOv8 with the SAHI framework improved detection of small-scale partial black defects, achieving mAP of 0.979 on Robusta coffee. Chamorro-Pinchao et al. (2025) evaluated YOLOv11 nano and small for binary defect classification in post-roasting beans, with the nano variant achieving 88.03% accuracy at 126.7 fps — surpassing human operator accuracy of 78.9%.

Transfer learning has proven effective in agricultural image classification by leveraging pretrained feature representations (Rochman & Junaedi, 2020). For green bean classification, (Arwatchananukul, Xu, Charoenkwan, Aung Moon, & Saengrayap, 2024) evaluated five CNN architectures on 17 defect categories of Thai Arabica green beans, with MobileNetV3 achieving the highest accuracy of 90.19% after augmentation. Hebert & Alamsyah (2026) applied YOLOv12 for 15-class SCA-standard defect detection, showing the importance of hyperparameter configuration (learning rate, batch size, epoch) on detection performance.

The consistent trend across these studies is that each successive YOLO version has demonstrated competitive performance in the coffee defect detection domain. Systematic hyperparameter tuning has been shown to be critical for maximizing detection performance (Isa, Rosli, Yusof, Maruzuki, & Sulaiman, 2022). However, a clear gap exists: no study has applied YOLO26 — the latest and most computationally efficient YOLO iteration — to post-roasting coffee bean defect detection. Furthermore, the combination of a first-party dataset from Indonesian local roasting industries with systematic grid search hyperparameter tuning represents a novel methodological contribution. Previous Indonesian studies have addressed coffee bean classification and defect detection for local industry contexts (Hanifah, Antoni, Ramadhani, & Yuliska, 2024; Rahmawati, Rianto, Riana, Studi Ilmu Komputer, & Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, 2021). This study aims to address both gaps simultaneously.

METHOD

This study followed a systematic series of stages to ensure rigor and reproducibility. The process began with a literature review, followed by dataset collection and annotation, data augmentation, model implementation, hyperparameter tuning, and performance evaluation.

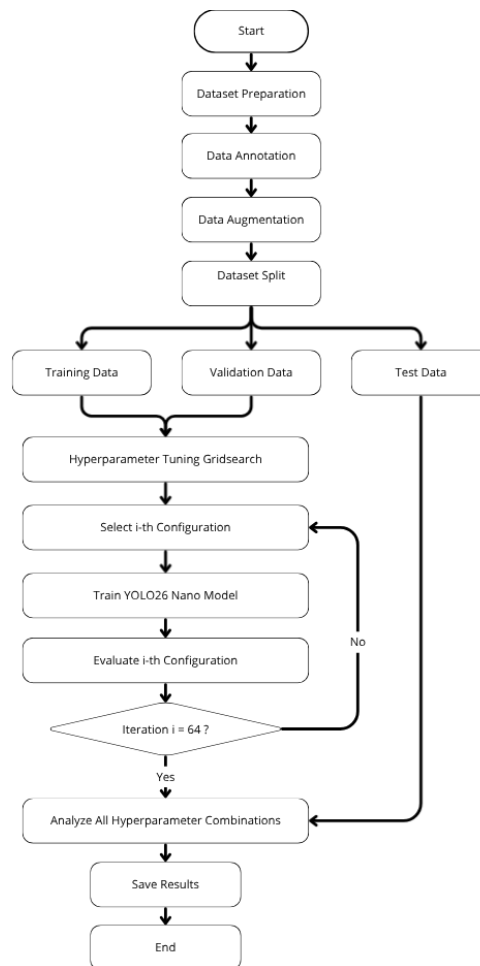


Figure 1. Research Methodology Flowchart

The dataset used in this study is a first-party dataset collected directly from three local coffee roasting companies in Palembang, Indonesia. A total of 4,567 images were collected covering six categories: insect damage, nugget, quaker, roasted-beans (non-defect), shell, and under roast. The distribution of the dataset is presented in Table 1.

Table 1. Details of the Dataset Used

No.	Category	Example Image	Count	Description
1	Insect Damage		752	Beans with visible insect-induced holes or entry points
2	Nugget		745	Hard, dark, compact beans with deformed shapes
3	Quaker		796	Under-developed beans appearing lighter after roasting
4	Roasted-Beans		718	Normal, well-roasted coffee beans (non-defect)
5	Shell		793	Hollow or thin shell-like beans due to incomplete development
6	Under Roast		763	Insufficiently roasted beans retaining pale or greenish tones
Total			4,567	

To expand the dataset and improve model generalization, data augmentation was applied including horizontal flip, vertical flip, random rotation (+-15 degrees), and brightness and contrast adjustment. After augmentation, the total dataset expanded to 10,595 images. The dataset was split with a ratio of 70% for training, 20% for validation, and 10% for testing.

Image annotation was performed using Roboflow with bounding box labeling for each defect instance. All images were resized to 640x640 pixels prior to training, consistent with the default input size of YOLO26 nano.

The detection model used in this study is YOLO26 nano with transfer learning, leveraging pretrained weights from the COCO dataset. Transfer learning reduces the need for large labeled datasets by reusing features from pretraining, which has been shown to improve performance in visual classification tasks (Purwono et al., 2022; Rochman & Junaedi, 2020). YOLO26 is the latest YOLO architecture released in January 2026, introducing key innovations including NMS-free inference, DFL removal, MuSGD optimizer, Progressive Loss (ProgLoss), and Small-Target-Aware Label Assignment (STAL) (Chakrabarty, 2026).

Hyperparameter optimization was conducted using grid search across five parameters: optimizer (SGD, AdamW), initial learning rate (0.001, 0.01), final learning rate (0.01, 0.1), epoch count (50, 100), classification loss weight (0.4, 0.5), and weight decay (0.0, 0.001). AdamW decouples weight decay from the gradient update to improve generalization, and the choice of optimizer has been shown to significantly affect YOLO detection performance (Isa et al., 2022; Moraes et al., 2025). The complete search space resulted in 64 unique training configurations.

Model performance was evaluated on the test set using Precision (P), Recall (R), F1-Score, mean Average Precision at IoU 0.50 (mAP50), and mean Average Precision at IoU 0.50 to 0.95 (mAP50-95), defined by Equations (1) through (5).

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \tag{1}$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \tag{2}$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \tag{3}$$

$$\text{mAP50} = \text{mean}(\text{AP}) \text{ at IoU} = 0.50 \tag{4}$$

$$\text{mAP50-95} = \text{mean}(\text{AP}) \text{ at IoU} = 0.50:0.05:0.95 \tag{5}$$

Where TP is True Positive, FP is False Positive, and FN is False Negative. Training was conducted on Kaggle Notebooks using a GPU T4 x2 accelerator.

RESULT

This section presents the experimental results of the YOLO26 nano model trained with transfer learning across 64 hyperparameter configurations. Prior to training, the dataset underwent augmentation and preprocessing, expanding from 4,567 to 10,595 images. The dataset was partitioned into training (70%), validation (20%), and testing (10%) subsets.

The complete results of all 64 grid search configurations, sorted by mAP50-95 in descending order, are presented in Table 2.

Table 2. Complete Grid Search Hyperparameter Tuning Results (Sorted by mAP50-95, Best Model in Bold)

No	Optimizer	LR Init	LR Final	Epochs	Cls Loss	W. Decay	Precision	Recall	F1-Score	mAP50	mAP50-95
0	SGD	0.01	0.1	50	0.4	0.0	0.9939	0.9933	0.9936	0.9948	0.9785
1	SGD	0.01	0.1	50	0.4	0.001	0.9936	0.9940	0.9938	0.9948	0.9806
2	SGD	0.01	0.1	50	0.5	0.0	0.9939	0.9940	0.9940	0.9949	0.9788
3	SGD	0.01	0.1	50	0.5	0.001	0.9951	0.9938	0.9945	0.9948	0.9804
4	AdamW	0.01	0.1	50	0.4	0.0	0.9961	0.9948	0.9955	0.9949	0.9815
5	AdamW	0.01	0.1	50	0.4	0.001	0.9962	0.9948	0.9955	0.9949	0.9817
6	AdamW	0.01	0.1	50	0.5	0.0	0.9961	0.9944	0.9953	0.9949	0.9807
7	AdamW	0.01	0.1	50	0.5	0.001	0.9933	0.9957	0.9945	0.9949	0.9772
8	SGD	0.01	0.1	100	0.4	0.0	0.9948	0.9939	0.9943	0.9949	0.9811
9	SGD	0.01	0.1	100	0.4	0.001	0.9941	0.9946	0.9943	0.9949	0.9828
10	SGD	0.01	0.1	100	0.5	0.0	0.9933	0.9956	0.9944	0.9949	0.9819
11	SGD	0.01	0.1	100	0.5	0.001	0.9956	0.9961	0.9959	0.9950	0.9830
12	AdamW	0.01	0.1	100	0.4	0.0	0.9889	0.9947	0.9918	0.9947	0.9641
13	AdamW	0.01	0.1	100	0.4	0.001	0.9959	0.9966	0.9963	0.9950	0.9826
14	AdamW	0.01	0.1	100	0.5	0.0	0.9964	0.9964	0.9964	0.9950	0.9826

15	AdamW	0.01	0.1	100	0.5	0.001	0.9961	0.9964	0.9963	0.9950	0.9830
16	SGD	0.01	0.01	50	0.4	0.0	0.9934	0.9937	0.9935	0.9948	0.9786
17	SGD	0.01	0.01	50	0.4	0.001	0.9941	0.9934	0.9937	0.9948	0.9804
18	SGD	0.01	0.01	50	0.5	0.0	0.9936	0.9943	0.9940	0.9949	0.9793
19	SGD	0.01	0.01	50	0.5	0.001	0.9942	0.9946	0.9944	0.9949	0.9801
20	AdamW	0.01	0.01	50	0.4	0.0	0.9962	0.9948	0.9955	0.9950	0.9813
21	AdamW	0.01	0.01	50	0.4	0.001	0.9962	0.9953	0.9957	0.9949	0.9816
22	AdamW	0.01	0.01	50	0.5	0.0	0.9967	0.9949	0.9958	0.9949	0.9808
23	AdamW	0.01	0.01	50	0.5	0.001	0.9963	0.9946	0.9954	0.9949	0.9808
24	SGD	0.01	0.01	100	0.4	0.0	0.9955	0.9942	0.9949	0.9948	0.9812
25	SGD	0.01	0.01	100	0.4	0.001	0.9940	0.9954	0.9947	0.9949	0.9830
26	SGD	0.01	0.01	100	0.5	0.0	0.9920	0.9953	0.9936	0.9949	0.9817
27	SGD	0.01	0.01	100	0.5	0.001	0.9947	0.9957	0.9952	0.9949	0.9830
28	AdamW	0.01	0.01	100	0.4	0.0	0.9955	0.9970	0.9962	0.9950	0.9833
29	AdamW	0.01	0.01	100	0.4	0.001	0.9962	0.9967	0.9964	0.9950	0.9834
30	AdamW	0.01	0.01	100	0.5	0.0	0.9971	0.9955	0.9963	0.9949	0.9828
31	AdamW	0.01	0.01	100	0.5	0.001	0.9966	0.9958	0.9962	0.9950	0.9827
32	SGD	0.001	0.1	50	0.4	0.0	0.9727	0.9765	0.9746	0.9926	0.9663
33	SGD	0.001	0.1	50	0.4	0.001	0.9739	0.9769	0.9754	0.9927	0.9667
34	SGD	0.001	0.1	50	0.5	0.0	0.9832	0.9832	0.9832	0.9939	0.9675
35	SGD	0.001	0.1	50	0.5	0.001	0.9780	0.9797	0.9788	0.9934	0.9673
36	AdamW	0.001	0.1	50	0.4	0.0	0.9906	0.9885	0.9895	0.9947	0.9745
37	AdamW	0.001	0.1	50	0.4	0.001	0.9923	0.9927	0.9925	0.9948	0.9775
38	AdamW	0.001	0.1	50	0.5	0.0	0.9953	0.9959	0.9956	0.9949	0.9821
39	AdamW	0.001	0.1	50	0.5	0.001	0.9957	0.9957	0.9957	0.9949	0.9825
40	SGD	0.001	0.1	100	0.4	0.0	0.9818	0.9840	0.9829	0.9941	0.9718
41	SGD	0.001	0.1	100	0.4	0.001	0.9845	0.9843	0.9844	0.9942	0.9732
42	SGD	0.001	0.1	100	0.5	0.0	0.9860	0.9903	0.9881	0.9945	0.9729
43	SGD	0.001	0.1	100	0.5	0.001	0.9885	0.9871	0.9878	0.9945	0.9734
44	AdamW	0.001	0.1	100	0.4	0.0	0.9960	0.9962	0.9961	0.9950	0.9842
45	AdamW	0.001	0.1	100	0.4	0.001	0.9932	0.9933	0.9933	0.9949	0.9789
46	AdamW	0.001	0.1	100	0.5	0.0	0.9895	0.9926	0.9911	0.9947	0.9764
47	AdamW	0.001	0.1	100	0.5	0.001	0.9960	0.9970	0.9965	0.9950	0.9844
48	SGD	0.001	0.01	50	0.4	0.0	0.9737	0.9749	0.9743	0.9925	0.9660
49	SGD	0.001	0.01	50	0.4	0.001	0.9738	0.9750	0.9744	0.9926	0.9661
50	SGD	0.001	0.01	50	0.5	0.0	0.9771	0.9800	0.9786	0.9934	0.9670
51	SGD	0.001	0.01	50	0.5	0.001	0.9793	0.9786	0.9789	0.9933	0.9670
52	AdamW	0.001	0.01	50	0.4	0.0	0.9948	0.9963	0.9955	0.9949	0.9827
53	AdamW	0.001	0.01	50	0.4	0.001	0.9938	0.9947	0.9943	0.9949	0.9825
54	AdamW	0.001	0.01	50	0.5	0.0	0.9951	0.9965	0.9958	0.9949	0.9833
55	AdamW	0.001	0.01	50	0.5	0.001	0.9962	0.9964	0.9963	0.9950	0.9829
56	SGD	0.001	0.01	100	0.4	0.0	0.9821	0.9838	0.9830	0.9941	0.9716
57	SGD	0.001	0.01	100	0.4	0.001	0.9847	0.9819	0.9833	0.9941	0.9723
58	SGD	0.001	0.01	100	0.5	0.0	0.9865	0.9880	0.9872	0.9944	0.9726
59	SGD	0.001	0.01	100	0.5	0.001	0.9856	0.9899	0.9878	0.9945	0.9730

60	AdamW	0.001	0.01	100	0.4	0.0	0.9962	0.9965	0.9963	0.9950	0.9842
61	AdamW	0.001	0.01	100	0.4	0.001	0.9908	0.9923	0.9916	0.9948	0.9809
62	AdamW	0.001	0.01	100	0.5	0.0	0.9964	0.9954	0.9959	0.9950	0.9835
63	AdamW	0.001	0.01	100	0.5	0.001	0.9959	0.9968	0.9963	0.9950	0.9843

As shown in Table 2, model 47 achieved the highest mAP50-95 of 0.9844, with configuration: AdamW optimizer, initial LR of 0.001, final LR of 0.1, 100 epochs, classification loss weight of 0.5, and weight decay of 0.001. This configuration is highlighted in bold.

Figure 2. presents the training curves of model 47 across 100 epochs, showing the progression of training and validation losses (box loss, classification loss, DFL loss) and detection metrics (Precision, Recall, mAP50, mAP50-95).

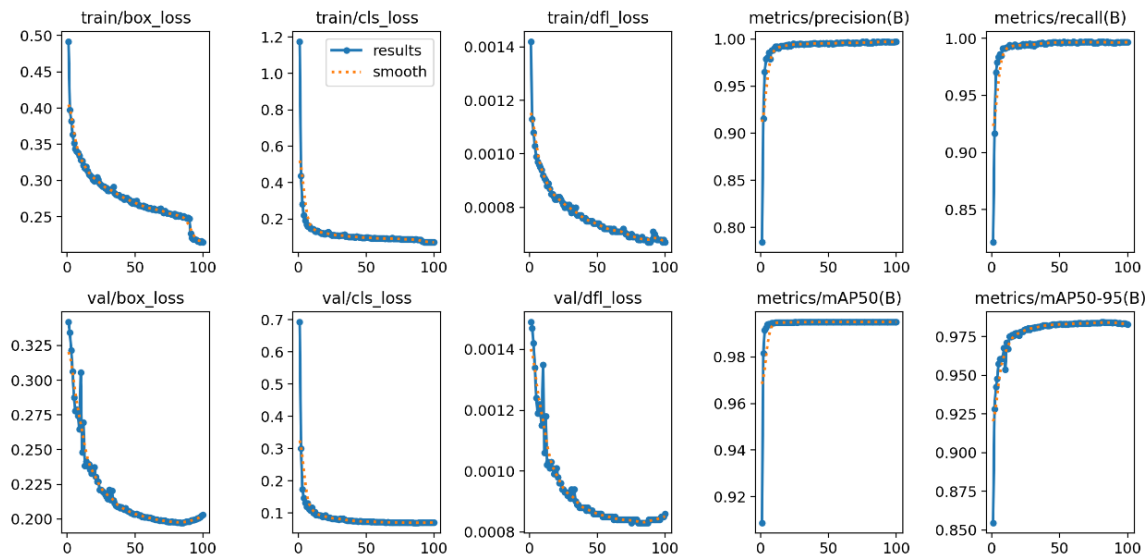


Figure 2. Training and Validation Curves of the Best Model (Model 47)

As shown in Figure 2, all training losses decreased consistently from epoch 1 to epoch 100 without signs of overfitting. Validation losses followed the same downward trend, converging in parallel with training losses. Precision and Recall both converged toward 1.00 from approximately epoch 20 onward, while mAP50 stabilized around 0.995 and mAP50-95 reached a plateau near 0.984.

Figure 3 presents the F1-Confidence Curve for model 47, illustrating the relationship between confidence threshold and F1-Score across all six classes.

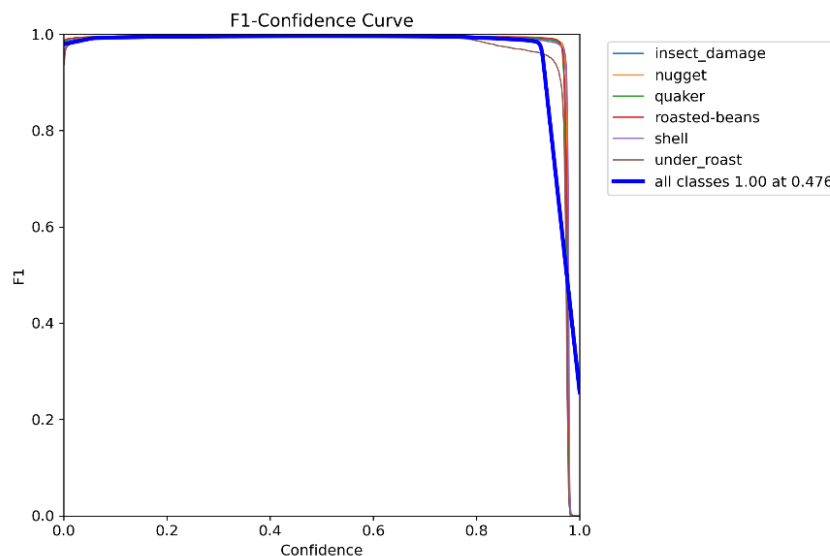


Figure 3. F1-Confidence Curve of the Best Model (Model 47)

As shown in Figure 3, all individual classes achieved an F1-Score approaching 1.00 across confidence thresholds from 0.0 to approximately 0.90. The peak F1-Score for all classes combined reached 1.00 at a confidence threshold of 0.476.

The confusion matrix of model 47 on the test set is presented in Figure 4, with its normalized version in Figure 5.

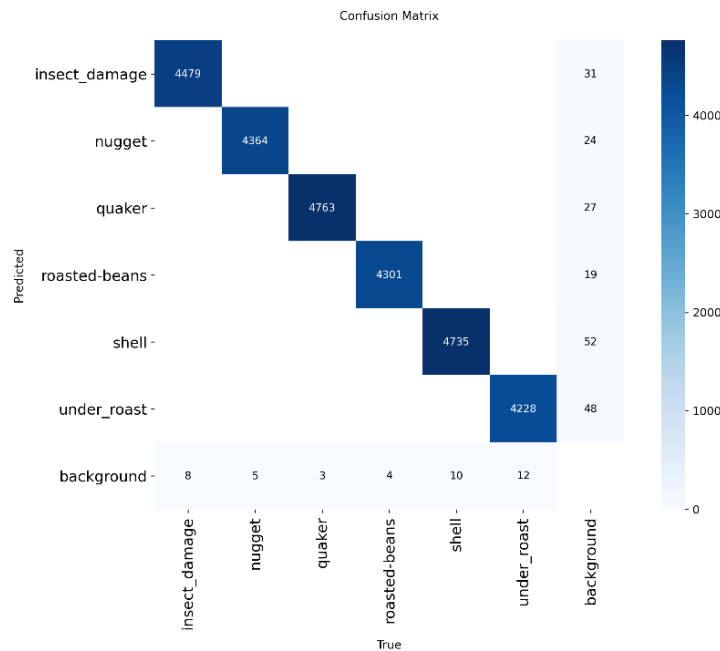


Figure 4. Confusion Matrix of the Best Model (Model 47) on Test Set

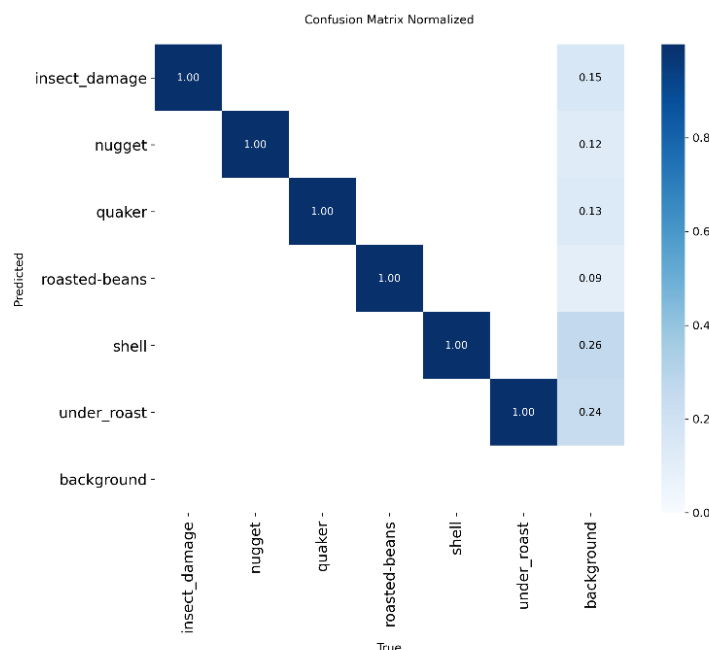


Figure 5. Normalized Confusion Matrix of the Best Model (Model 47) on Test Set

The confusion matrix confirms that the model correctly classified the vast majority of instances across all six categories. True positive counts were: insect_damage (4,479), nugget (4,364), quaker (4,763), roasted-beans (4,301), shell (4,735), and under_roast (4,228). The normalized matrix shows all classes at 1.00 on the diagonal, indicating near-perfect inter-class discrimination. The primary misclassification was false negatives toward the background class, with shell (52 instances) and under_roast (48 instances) showing the highest rates.

The per-class performance metrics of model 47 on the test set are presented in Table 3.

Table 3. Per-Class Performance Metrics of the Best Model (Model 47) on Test Set

No	Class	Images	Instances	Precision	Recall	F1-Score	mAP50	mAP50-95
1	Insect Damage	76	2,251	0.998	0.997	0.997	0.995	0.990
2	Nugget	78	2,069	0.999	0.995	0.997	0.995	0.989
3	Quaker	76	2,394	0.999	0.997	0.998	0.995	0.983
4	Roasted-Beans	77	2,152	0.996	0.995	0.995	0.995	0.986
5	Shell	77	2,380	0.996	0.997	0.996	0.995	0.992
6	Under Roast	73	2,089	0.996	0.998	0.997	0.995	0.973
Overall		457	13,335	0.997	0.996	0.996	0.995	0.985

As shown in Table 3, model 47 achieved consistently high performance across all six defect categories. The highest mAP50-95 was recorded for shell (0.992), followed by insect_damage (0.990), nugget (0.989), roasted-beans (0.986), quaker (0.983), and under_roast (0.973). The overall mAP50 of 0.995 and mAP50-95 of 0.985 confirm strong multi-class detection performance.

DISCUSSION

The experimental results provide several key insights regarding the effectiveness of YOLO26 nano with transfer learning and grid search hyperparameter tuning for post-roasting coffee bean defect detection.

The best-performing configuration (model 47: AdamW, LR init 0.001, LR final 0.1, 100 epochs, cls_loss 0.5, weight_decay 0.001) achieved an overall mAP50 of 0.995 and mAP50-95 of 0.985 on the test set. The consistent superiority of AdamW configurations in the top-ranking results can be attributed to AdamW's adaptive learning rate mechanism with decoupled weight decay, which facilitates stable convergence particularly when using a warm-up learning rate schedule (LR init 0.001 to final 0.1). This schedule allows the model to begin with cautious updates during early epochs and gradually scale up as training progresses, reducing the risk of overshooting local minima. AdamW consistently outperforms standard Adam in deep learning tasks with appropriate hyperparameter tuning (Michael et al., 2023; Moraes et al., 2025).

The superior performance of 100-epoch configurations over 50-epoch counterparts (mean mAP50-95: 0.979 vs. 0.977) suggests the model benefited from extended training to fully leverage the transfer learning initialization. This is consistent with transfer learning principles, wherein pretrained feature representations require sufficient fine-tuning iterations to adapt to domain-specific visual characteristics.

The per-class analysis (Table 3) reveals that the under_roast class achieved the lowest mAP50-95 (0.973) despite the highest recall (0.998). This indicates that while the model rarely misses under_roast instances, it occasionally produces lower-confidence detections that reduce average precision at stricter IoU thresholds. The pale coloration of under-roasted beans may be more challenging to localize precisely, especially when beans partially overlap. Similarly, the shell class showed higher background false negatives (52 instances), attributable to the morphologically thin and translucent appearance of shell beans.

The F1-Confidence Curve (Fig. 3) demonstrates that all classes achieve $F1 = 1.00$ at a confidence threshold of 0.476, confirming that the model is highly calibrated without class-imbalanced confidence distributions. This is noteworthy given the dataset was collected from three distinct companies, introducing real-world variability in lighting, camera distance, and bean presentation.

Comparing with prior studies in the literature, the proposed YOLO26 nano model demonstrates competitive or superior performance. The results are summarized in Table 4.

Table 4. Comparison of Results with Previous Studies

Study	Model	Domain	Classes	Precision	mAP50	mAP50-95
Ahmed et al. (2025)	YOLOv8	Green beans	Multi	—	0.995	—
Adiwijaya et al. (2024)	YOLOv8 + SAHI	Robusta	Multi	—	0.979	—
Chamorro-Pinchoa et al. (2025)	YOLOv11 nano	Post-roasting	2	—	—	—
Hebert & Alamsyah (2026)	YOLOv12	Green beans	15	0.870	0.840	—
This study (Model 47)	YOLO26 nano	Post-roasting	6	0.997	0.995	0.985

As shown in Table 4, the proposed YOLO26 nano model achieves competitive or superior performance compared to all referenced prior works, while addressing a more complex six-class post-roasting detection task using a first-party Indonesian dataset.

Several limitations should be acknowledged. First, the dataset was collected from only three coffee roasting companies in Palembang, which may limit generalizability to operations with different equipment or bean origins. Second, the grid search was conducted on the validation set and while the best model was evaluated on a held-out test set, the risk of implicit overfitting to the search space remains a consideration. Third, inference speed benchmarks under deployment conditions were not measured in this study.

CONCLUSION

This study successfully analyzed the performance of the YOLO26 nano model with transfer learning for detecting and classifying post-roasting coffee bean defects across six categories: insect damage, nugget, quaker, roasted-beans (non-defect), shell, and under roast. A first-party dataset of 4,567 images from three local Indonesian coffee roasting companies was augmented to 10,595 images.

The grid search hyperparameter tuning across 64 configurations identified the optimal setting as AdamW optimizer, initial learning rate of 0.001, final learning rate of 0.1, 100 training epochs, classification loss weight of 0.5, and weight decay of 0.001 (model 47). This configuration achieved Precision of 0.997, Recall of 0.996, F1-Score of 0.997, mAP50 of 0.995, and mAP50-95 of 0.985 on the test set.

The analysis reveals that AdamW consistently outperformed SGD across configurations, particularly when combined with a warm-up learning rate schedule (0.001 to 0.1). Extended training over 100 epochs provided consistent improvements over 50-epoch configurations. Per-class analysis showed all six categories achieved mAP50 of 0.995, with mAP50-95 ranging from 0.973 (under_roast) to 0.992 (shell).

Compared to prior studies applying YOLOv8, YOLOv11, and YOLOv12 to coffee bean defect detection, the proposed YOLO26 nano model demonstrates superior or competitive performance while addressing a more complex multi-class post-roasting detection task using a first-party Indonesian dataset.

For future research, it is recommended to (1) expand the dataset to include coffee roasting companies from diverse geographical regions across Indonesia, (2) evaluate the model under real-time deployment conditions on edge devices to assess inference speed and practical feasibility, and (3) investigate class-specific augmentation strategies to further improve detection accuracy for visually challenging categories such as under_roast and shell.

REFERENCES

- Adiwijaya, N. O., Sarno, R., & Wijaya, D. R. (2024). Real Time Detection of Coffee Bean Defects Using YOLO Method and SAHI (Slicing Aided Hyper Inference) Framework. *2024 Beyond Technology Summit on Informatics International Conference, BTS-I2C 2024*, 286–291. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/BTS-I2C63534.2024.10941784>
- Ahmed, M. I. B., Rahman, A., Ali, F., Youldash, M., Althubaiti, S., Alghamdi, A. S., ... Gollapalli, M. (2025). Coffee Bean Defect Detection with Deep Learning: Khawlani Coffee Case Study from Jizan, Saudi Arabia. *Traitement Du Signal*, 42(5), 2619–2630. <https://doi.org/10.18280/ts.420515>
- Al-Mahish, M., Alfayadh, R., Al-Mahish, M., & Alfayadh, R. (2024). The impact of coffee quality attributes and ratings on specialty coffee bean prices. *Beverage Plant Research 2024 1:E039*, 4(1). <https://doi.org/10.48130/BPR-0024-0029>
- Angraini, A. P., Damat, D., & Manshur, H. A. (2024). Studi Karakteristik Fisikokimia dan Organoleptik Bubuk Kopi dari Biji Cacat Hitam Kopi Robusta dengan Perbedaan Suhu Penyangraian. *Food Technology and Halal Science Journal*, 6(2), 117–132. <https://doi.org/10.22219/fths.v6i2.28307>
- Arwatchananukul, S., Xu, D., Charoenkwan, P., Aung Moon, S., & Saengrayap, R. (2024). Implementing a deep learning model for defect classification in Thai Arabica green coffee beans. *Smart Agricultural Technology*, 9. <https://doi.org/10.1016/j.atech.2024.100680>
- Badan Pusat Statistik Indonesia. (2024, June 23). Ekspor Kopi Menurut Negara Tujuan Utama. Retrieved May 23, 2025, from <https://www.bps.go.id/id/statistics-table/1/MTAxNCMx/ekspor-kopi-menurut-negara-tujuan-utama--2000-2023.html>
- Chakrabarty, S. (2026). *YOLO26: An Analysis of NMS-Free End to End Framework for Real-Time Object Detection*. Retrieved from <http://arxiv.org/abs/2601.12882>
- Chamorro-Pinchao, A., Pusedá-Chulde, M., Trejo-España, D., Caranqui-Sánchez, V., & García-Santillán, I. (2025). Binary classification of defects in multiple coffee beans using lightweight convolutional neural networks for embedded systems. *Data and Metadata*, 4, 840. <https://doi.org/10.56294/dm2025840>
- Faojiah, R. S., Rahmah, D. M., Pramulya, R., Novita, E., & Sutrisno, A. (2025). Model Dinamik Produksi Kopi Berdasarkan Pemenuhan Total Kebutuhan Kopi Indonesia. *Jurnal Teknotan*, 19(2), 81–92. <https://doi.org/10.24198/jt.vol19n2.2>

- Hafifah, Y., Muchtar, K., Ahmadiar, A., & Esabella, S. (2022). Perbandingan Kinerja Deep Learning Dalam Pendeteksian Kerusakan Biji Kopi. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 1928. <https://doi.org/10.30865/jurikom.v9i6.5151>
- Hanifah, P., Antoni, H. I., Ramadhani, S. R., & Yuliska, Y. (2024). PENGEMBANGAN APLIKASI MOBILE UNTUK DETEKSI CACAT BIJI KOPI ROBUSTA BERDASARKAN STANDAR NASIONAL INDONESIA. *Jurnal Komputer Terapan*, 10(2), 215–224. <https://doi.org/10.35143/jkt.v10i2.6472>
- Hebert, H., & Alamsyah, D. (2026). Detection of Coffee Bean Defects in Speciality Coffee Association Standards using YOLOv12. *INOVTEK Polbeng-Seri Informatika*, XI(1), 85–95.
- Hidayatullah, P., & Tubagus, R. (2026). *YOLO26: A Comprehensive Architecture Overview and Key Improvements A PREPRINT*.
- Hu, G. L., Quan, C. X., Dai, H. P., & Qiu, M. H. (2024). Characterization of defective coffee beans and blends differentiation based on 1H qNMR technique. *Current Research in Food Science*, 9, 100870. <https://doi.org/10.1016/J.CRFS.2024.100870>
- Isa, I. S., Rosli, M. S. A., Yusof, U. K., Maruzuki, M. I. F., & Sulaiman, S. N. (2022). Optimizing the Hyperparameter Tuning of YOLOv5 for Underwater Detection. *IEEE Access*, 10, 52818–52831. <https://doi.org/10.1109/ACCESS.2022.3174583>
- Michael, A., Rusman, J., & History, A. (2023). *Klasifikasi Cacat Biji Kopi Menggunakan Metode Transfer Learning dengan Hyperparameter Tuning Gridsearch*. 9(1), 37–45. Retrieved from <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- Moraes, A. M., Pugliese, L. F., Santos, R. F. dos, Vitor, G. B., Braga, R. A. da S., & Silva, F. R. da. (2025). Effectiveness of YOLO Architectures in Tree Detection: Impact of Hyperparameter Tuning and SGD, Adam, and AdamW Optimizers. *Standards*, 5(1), 9. <https://doi.org/10.3390/standards5010009>
- Murinto, M., Rosyda, M., & Melany, M. (2023). Klasifikasi Jenis Biji Kopi Menggunakan Convolutional Neural Network dan Transfer Learning pada Model VGG16 dan MobileNetV2. *JRST (Jurnal Riset Sains Dan Teknologi)*, 7(2), 183. <https://doi.org/10.30595/jrst.v7i2.16788>
- Purwono, Ma'arif, A., Rahmانيar, W., Fathurrahman, H. I. K., Frisky, A. Z. K., & Haq, Q. M. U. (2022). Understanding of Convolutional Neural Network (CNN): A Review. *International Journal of Robotics and Control Systems*, 2(4), 739–748. <https://doi.org/10.31763/ijrcs.v2i4.888>
- Rahmawati, A., Rianto, Y., Riana, D., Studi Ilmu Komputer, P., & Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, S. (2021). *Deteksi Defect Coffee Pada Citra Tunggal Green Beans Menggunakan Metode Ensemble Decision Tree Detection of Defect Coffee in Green Beans Single Image Using the Ensemble Decision Tree Method* (Vol. 20).
- Rochman, F., & Junaedi, H. (2020). IMPLEMENTASI TRANSFER LEARNING UNTUK IDENTIFIKASI ORDO TUMBUHAN MELALUI DAUN. *Jurnal Syntax Admiration*.
- Snapcart. (2023, October 2). Indonesia's Coffee Consumption Trends in 2023 – Snapcart. Retrieved from <https://snapcart.global/indonesias-coffee-consumption-trends-in-2023/>