OPEN ACCESS 29

ISSN 2356-5462
http://socj.telkomuniversity.ac.id/ijoict/

Intl. Journal on ICT
Vol. 12, No. 1, Jun 2026. pp. 29-42
doi:10.21108/ijoict.v12i1.10584

Real-Time Coffee Bean Defect Detection Based on SNI 01-2907-2008 Standards Using Lightweight YOLOv5s Architecture

Nanda Aptana Irsyadul Bahy[1], Achmad Pratama Rifai[2,*]

[1,2] *Department Mechanical and Industrial Engineering, Universitas Gadjah Mada*
Bulaksumur, Caturtunggal, Depok, Sleman Regency, Special Region of Yogyakarta, Indonesia, 55281

[*] achmad.p.rifai@ugm.ac.id

Abstract

Physical quality control of coffee beans in Indonesia relies heavily on manual sorting, which is limited in consistency, time efficiency, and objectivity. This study develops an automated real-time detection system for 20 physical defect categories defined in SNI 01-2907-2008, using the lightweight YOLOv5s object detection architecture trained via transfer learning with two-phase GridSearch hyperparameter optimization and data augmentation on a dataset of 107 images containing 13,863 annotations. The optimized model achieves a global mAP@50 of 0.867 and mAP@0.5-0.95 of 0.601, with an average inference time of 14 ms per image, thereby confirming its real-time suitability. Per-class analysis reveals high detection accuracy for morphologically distinctive classes, such as pod beans and small shell skins with mAP@50 scores exceeding 0.98, but reduced performance for visually ambiguous classes like slight insect damage with an mAP@50 of 0.626, attributed to texture bias and contrast ambiguity. Compared with the only prior SNI 01-2907-2008-aligned study, which achieved 53.35% accuracy using image classification across 18 simplified classes, this object detection approach offers superior accuracy and spatial localization of defects. The model's low computational footprint enables deployment on low-cost edge devices, providing a practical and standardized quality-inspection solution for Indonesian coffee SMEs.

Keywords: Green Bean Coffee, Object Detection, SNI 01-2907-2008, Defects, YOLOv5s.

I. INTRODUCTION

Coffee is a very popular agricultural commodity in Indonesia. Indonesia is one of the world's largest coffee producers. With high demand and production, coffee has become one of Indonesia's primary sources of income, both through imports and exports [1]. Coffee production in Indonesia is dominated by smallholder plantations, which contribute 99.56% of total production and the majority of coffee products produced and marketed consist of green coffee beans, with Robusta beans representing the largest share, followed by Arabica and Liberica varieties [2]. The micro, small, and medium enterprise sector in Indonesia is experiencing rapid growth in coffee production. The substantial scale of coffee production and the diverse range of coffee products in Indonesia necessitate the implementation of regulations to ensure the consistent quality of these products. In this regard, the Indonesian government has introduced SNI 01-2907-2008, a regulatory framework designed to govern the coffee industry effectively. This regulation specifies quality standards, testing methodologies, eligibility criteria for passing tests, and classifications of different coffee types [3]. According to SNI, the coffee quality control process is defined by the defect value of coffee beans, which are categorized into several types of defects. The standard for determining coffee quality is differentiated between Arabica and Robusta, which are divided into classes 1 to 6. In a competitive global market for coffee, it is crucial for stakeholders to adapt

Received on 19 Jan 2026. Revised on 31 Mar 2026. Accepted and Published on 5 Apr 2026.

and evolve their business processes. Presently, the quality control and defect identification processes for coffee beans are conducted manually, relying significantly on the precision and experience of individual inspectors [4]. Considering the situation in Indonesia, where almost all coffee production comes from MSMEs with different company standards and business processes, it is necessary to develop a model for equalizing classification results and improving process efficiency, especially in quality control.

On the other hand, the development of implementable technology is a determining factor in improving performance and efficiency, such as the application of artificial intelligence (AI) and the Internet of Things (IoT), particularly in agriculture, which can perform monitoring, controlling, and analysing in real time, thereby increasing productivity through more prudent resource management [5]. In AI applications, machine learning and deep learning are used to detect and classify products. In the coffee bean industry, the process of classifying bean quality is essential because the quality of the coffee beans represents the resulting flavour [6]. The utilization of a deep learning model based on Convolutional Neural Networks (CNN) demonstrates significant capabilities in the analysis and detection of defective images. Using CNNs can improve accuracy and reduce the detection process time, especially in handling the sorting and quality control. The sorting and quality control stages are important roles that are currently still carried out traditionally using human labour. Utilizing CNNs requires developing a robust model that efficiently processes coffee bean image data, enabling automated classification that meets industry standards. Several studies have used CNN in agricultural practices, especially in the coffee commodity sector.

The classification of defects in coffee beans provides a foundation for implementing interventions and enhancing the coffee farming process. This encompasses a wide range of practices, including land preparation, planting, and post-harvest processing. The information derived from this classification system also plays a crucial role in determining the selling price of coffee beans and in ensuring compliance with existing standards and regulations. In this study, the detection model will focus on classification based on SNI-10-2907-2008, considering that this regulation is the standard for coffee beans under Indonesian regulations. The model will use the YOLOv5 object detection model to detect physical defects. YOLOv5s was selected as the primary model for this study based on a systematic evaluation of the accuracy and efficiency trade-off across the YOLOv5 family [7]. YOLOv5s is selected as the optimal model for MSMEs because its 9.1 million parameters provide sufficient depth to accurately detect subtle coffee bean defects, unlike the heavily pruned YOLOv5n with its 2.6 million parameters. Furthermore, larger models like the 25.1 million parameter YOLOv5m, as well as the equivalent small variants of newer architectures such as YOLOv8s through YOLOv11s, demand excessive computational power, causing severe inference latency on standard low-cost hardware. Therefore, YOLOv5s achieves the most practical balance of detection accuracy, real-time processing speed, and stable edge deployment for real world operational environments.

To date, prior deep learning studies aligned with SNI 01-2907-2008 have relied exclusively on image classification paradigms with simplified defect taxonomies. Kesiman et al. [4] employed an image classification approach that achieved only 53.35% accuracy across a simplified 17-class scheme and was unable to spatially localize defects. More recently, Nugroho et al. [8] developed a YOLO based mobile application to detect defects in Robusta coffee beans but severely reduced the taxonomy to only three classes. While their application achieved 95.3% accuracy for the black bean class, the performance dropped significantly to 62.2% for the moldy/bleached bean class. Critically, these prior SNI-based works share fundamental limitations regarding spatial localization and taxonomic completeness. Traditional image classification inherently lacks the ability to spatially localize individual defect instances, which is an indispensable capability for automated physical sorting on a conveyor system. Furthermore, none of the existing literature has addressed the complete, unmodified 20-class defect taxonomy defined in SNI 01-2907-2008. This study addresses both gaps by applying object detection via the YOLOv5s architecture across all 20 SNI-defined defect classes, providing simultaneous spatial localization and defect classification in a lightweight model suitable for deployment on low-cost edge hardware in Indonesian coffee MSMEs.

II. LITERATURE REVIEW

According to [9] Convolutional Neural Network (CNN) has been implemented to obtain multispectral image data or images capable of processing large-scale data and producing better, more accurate, and more feasible techniques for coffee bean classification. The advantages of CNN support the exploration of other CNN models

in classifying types of defects in green coffee beans. There are two types of CNN applications, namely image classification and object detection, with different models in their implementation.

Rivalto et al [10] and Murinto et al [11] used a CNN to classify coffee varieties in Indonesia with reasonable accuracy, while [12] and [13] using a CNN to classify coffee beans into good and no good quality achieved 93% and 99% accuracy after the pretrained stage. In another study using a different classification model, namely the Deep Convolutional Neural Network (DCNN) and Light Deep Convolutional Neural Network (LDCNN), an accuracy of nearly 98% was achieved with a small number of parameters [14],[15]. This low number of parameters improved computational efficiency and resource utilization. The use of CNN is the result of image processing and machine learning that can detect the physical condition of coffee beans and can be optimized not only to distinguish good and bad beans. Several studies have used various types of defects, including broken beans, insect-infected beans, fungus-infected beans, sour beans, black beans, withered beans, and immature beans [16], [17], [18]. The results of several studies using image classification show a percentage of around 96% using a CNN model modified with Multiscale Defect Extraction.

Developments related to coffee bean defect detection based on the continuously evolving number of classifications depend on established regulations and the development of detection models, making research related to coffee bean defect detection increasingly advanced. In [19] study, Using SCAA regulations that classify 17 types of coffee bean defects into two categories, physical defect detection was performed using various kinds of models, namely MobileNetV2, EfficientNetV2, InceptionNetV2, ResNetV2, and MobileNetV3. The MobileNetV3 model achieved the best accuracy, reaching 95.85%. In a different study using defect classification based on SNI 01-2907-2008, which contains 20 types of defect classifications that were then modified into 17 types of defects, and training using image classification through the MobileNet and InceptionResNetV2 models, the accuracy reached 53.35% with InceptionResNetV2 [4]. In addition to using image classification, some models use another method called object detection. Object detection is a part of computer vision that detects spatial objects through semantic inference and spatial localization, which cannot be performed by image classification [20].

Huang [21] used object detection to classify good and bad defects using YOLOv3, achieving an accuracy of 94.63%. Another study used another object detection model, Faster R-CNN, which was compared with image classification, namely the VGG-16 model, to detect four types of classification, namely peaberry, longberry, defective, and premium beans, achieving an accuracy of 93% and 86% for each model [22]. In a comparison of other models between object detection and image classification for detecting and classifying coffee beans into four types of defects, black, ruptured, insect-infected, and fade, as well as one classification for good beans. [23] YOLOv5 was used for object detection and combined with other image classification models, such as Slim-CNN and VGG-16, achieving accuracies of 98.52% for YOLOv5, 93.63% for Slim-CNN, and 96.89% for VGG-16. In other cases, [24] YOLOv5 was used to classify bean types between Ethiopian Sidamo Arabica, Guatemalan Shb Arabica, Tanzanian Superior Robusta, and Indonesian Flores Robusta, achieving mAP@50 of 99% and mAP@50-95 of 94%.

III. RESEARCH METHOD

In this study, the transfer learning method was used in the object detection model to recognize the location and classification of physical defects in green coffee beans in accordance with the SNI 01-2907-2008 regulation on coffee beans. This standard is used in Indonesia to classify coffee bean defects and to grade coffee beans. In SNI 01-2907-2008, physical defect classification is divided into 20 types, namely black bean, partial black bean, broken black bean, pod bean, brown bean, large shell skin, medium shell skin, small shell skin, parchment bean, large husk, medium husk, small husk, broken bean, immature bean, bean with slight insect damage, bean with severe insect damage, bean with fungus damage, large foreign matter, medium foreign matter, and small foreign matter. The YOLOv5 model was selected for comparison due to its established reliability in object detection, particularly for coffee beans [23], [24]. The research stages are illustrated in Fig. 1.

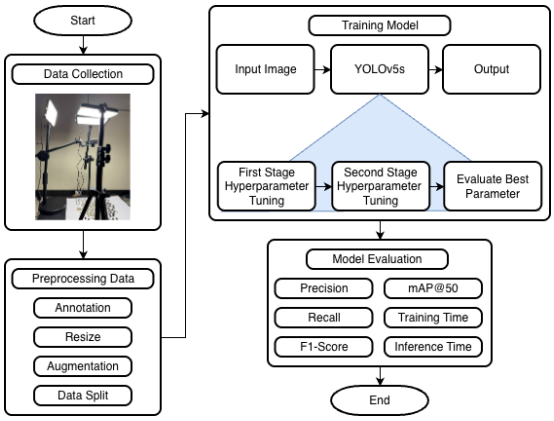

Fig. 1. Research Stages

A. *Data Gathering*

This study focuses on implementing SNI 01-2907-2008 using computer vision to detect physical defects in coffee beans. The data collection process was meticulously executed in a controlled environment, with a data collection height of 40 centimeters and a white background utilized as a medium for arranging the coffee beans. A single image was used for data collection, consisting of 50 grams of coffee beans. This data was obtained from the coffee production results of the Gandiva Processing Unit in Malang using a random combination of Arabica and Robusta green beans. This dataset consists of 107 images with 20 defect classifications, captured with an iPhone 11 Pro smartphone with a 12 MP camera.

B. *Preprocessing Data*

Object detection is a methodology for training data models that involves the annotation process described in Fig. 2. This stage consists of labelling objects according to their classification using bounding boxes to establish spatial references and precise class labels for model training. The annotation stage provides the ground truth required by the object detection model. Based on annotations across the entire dataset, 13,863 annotations were obtained. The annotated dataset was subsequently partitioned into a training and validation set with proportions of 70% and 30%, respectively. This dataset was then resized to 640 x 640, which is the sweet spot for models, especially YOLO, as the default input size.

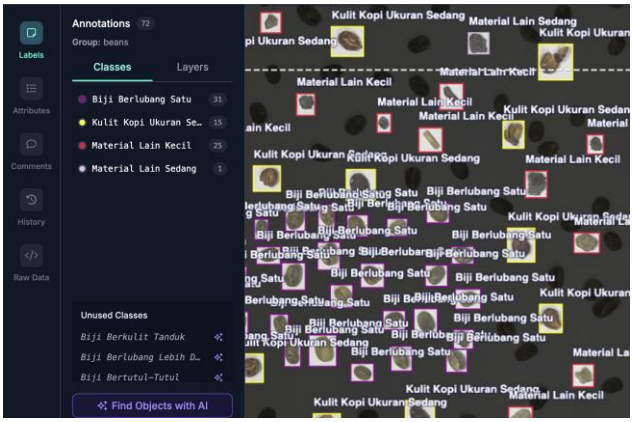

Fig. 2. Example of Image Annotations

Due to the limited dataset, the dataset was developed through an augmentation process. Augmented data can overcome the problem of overfitting in CNNs due to limited samples and build deep learning models that generalize to variations in orientation, position, and lighting that may occur in real conditions, thereby increasing the model's robustness [25]. The augmentation process was carried out by adjusting parameters such as flip, rotate, crop, rotation, shear, blur, and noise. After augmentation, the number of training images was 222. Table I describes the configuration and settings of each augmentation.

TABLE I
AUGMENTATION PARAMETERS

Augmentation Function	Parameter
Flip	Horizontal, Vertical
90º Rotate	Clockwise, Counter-Clockwise, Upside Down
Crop	0% Minimum Zoom, 29% Maximum Zoom
Rotation	Between -15° and +15°
Shear	±10° Horizontal, ±10° Vertical
Blur	Up to 2px
Noise	Up to 1.96% of pixels

C. *Model Architecture*

The dataset has been pre-processed and is ready for training with the CNN model. The selected YOLOv5 model is a model with a one-stage detection algorithm (one-stage detector). The YOLOv5 model predicts bounding boxes and class probabilities simultaneously during a single forward pass. The YOLOv5 architecture is shown in Fig. 3.

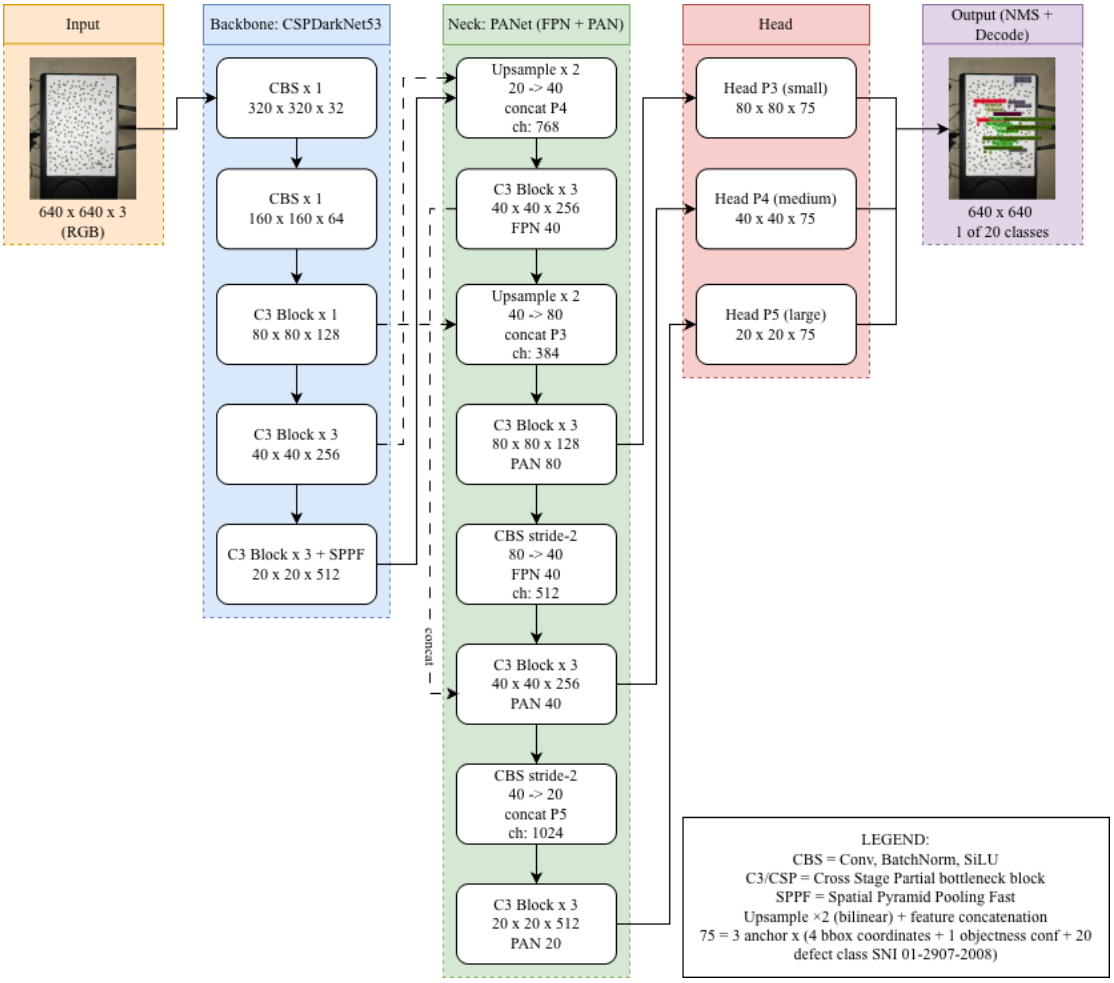

Fig. 3. YOLOv5 Architecture.

The initial stage of the YOLOv5 model utilizes a backbone network based on CSPDarkNet for hierarchical feature extraction. This process entails reducing spatial resolution (down sampling) while increasing spatial depth through two stages to minimize gradient redundancy. The neck section integrates FPN and PAN structures to facilitate two-way feature fusion (top-down and bottom-up), thereby enriching semantic information and

strengthening hierarchical spatial details through a concatenation mechanism. The final stage of the process, namely the head, is responsible for processing the aggregated features into three grid scales (80 x 80, 40 x 40, and 20 x 20) as outputs in the form of bounding boxes, object classes, and confidence levels.

D. Model Training

The model will be trained on Google Colab using Python on a Tesla 4 GPU, employing the YOLOv5 model version small(s). The selection of this model is predicated on considerations of efficiency and the model's accuracy performance across light parameters ranging from 7.2 million. The model will be employed in real-time applications that require accelerated image inference without any loss of accuracy. The model will undergo hyperparameter tuning using the values listed in Table II.

TABLE II
TUNING HYPERPARAMETER CONFIGURATION

Hyperparameter	Configuration
Learning rate (lr)	0.001 and 0.01
Batch size	8 and 16
Optimizer	SGD and Adam
Epochs	25 and 50
Image Size	640 x 640

Model training will be conducted by tuning hyperparameters using GridSearch to determine the optimal combination of learning rate, batch size, optimizer, and epochs for a dataset with a constant image size. The hyperparameter tuning stage has been shown to improve model performance, reduce overfitting, and enhance model robustness [26]. The tuning stage employs a two-phase approach using GridSearch. In the initial phase, tuning is conducted using the learning rate, batch size, and optimizer parameters with five epochs. Subsequently, the optimal results from this tuning are further tuned using epoch parameters of 25 and 50.

E. Evaluation Matrix

The model's performance was assessed using various metrics, including precision, recall, F1-score, and mean average precision. A thorough examination of the confusion matrix results was also conducted to gain insights into the model's accuracy and bias. In the YOLOv5 architecture, detection validation is determined by the Intersection over Union (IoU) score, which indicates the degree of overlap between the predicted and ground truth boxes [27]. The list of evaluation matrices to be performed is presented in Table III. The components that constitute the evaluation matrix are as follows:

- True Positive (TP): the model has successfully detected the target object correctly.
- False Positive (FP): the model detects objects that do not align with the established ground truth.
- True Negative (TN): the model correctly detects false data.
- False Negative (FN): the model incorrectly classifies a correct object as negative.

TABLE III
EVALUATION MATRIX [28]

	Predicted Positive	**Predicted Negative**	
Ground Truth Positive	True Positive (TP) Correct detection	False Negative (FN) [Type II Error]	Recall = TP / (TP + FN)
Ground Truth Negative	False Positive (FP) [Type I Error]	True Negative (TN) Correct rejection	Specificity = TN / (TN + FP)
	Precision = TP / (TP + FP)	NPV = TN / (TN + FN)	Accuracy = (TP+TN) / (TP+TN+FP+FN)

The following are equations for calculating several evaluation metrics such as F1-score and Mean Average Precision (mAP), which are calculated using equations (1), (2), and (3).

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FN + FP} \quad (1)$$

$$Average\ Precision = \int_0^1 p(r)dr \quad (2)$$

$$Mean\ Average\ Precision = \frac{1}{N}\sum_{i}^{i=N} AP_i \quad (3)$$

In equations (1), TP, FP, and FN denote the total number of True Positives, False Positives, and False Negatives, respectively. Meanwhile in equation (2), p(r) represents precision at recall level r. Furthermore, for equation (3), N indicates the total number of classes, specifically the 20 SNI-defined categories and AP_i is Average Precision for class i, computed as the area under the Precision-Recall curve. mAP@50 uses a fixed IoU threshold of 0.5 and mAP@0.5:0.95 averages over IoU thresholds {0.50, 0.55, …, 0.95}. Summary of the research methodology is presented in Table IV.

TABLE IV
SUMMARY OF RESEARCH METHODOLOGY

Stage	Activity	Tool / Setting
Data Collection	Image capture (50 g beans, 40 cm height)	iPhone 11 Pro, 12 MP, white background
Preprocessing	Annotation, resize, augmentation	Bounding-box labeling; 640×640 resize; flip, rotate, crop, shear, blur, noise
Model Training	Transfer learning + 2-phase GridSearch hyperparameter tuning	YOLOv5s, Google Colab Tesla T4 GPU; lr, batch, optimizer, epoch search
Evaluation	Precision, Recall, F1, mAP@50, mAP@50-95, Inference time	IoU threshold 0.5 / 0.5–0.95; Confusion matrix per-class
Model Comparison	Comparative benchmarking across YOLOv5 variants	YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l

IV. RESULTS AND DISCUSSION

Training and validation of the model were performed in Google Colab using a Tesla 4 GPU and Python 3. The libraries used were Ultralytics for the model and Scikit-learn for GridSearch. During model training, a loss matrix was added to visualize the model's performance. The tuning stage was also performed using 5-fold cross-validation to achieve an optimal balance between bias and variance. The results of hyperparameter tuning are shown in Table V. It shows that the best results from the first hyperparameter tuning are achieved with a learning rate of 0.01, a batch size of 8, and the SGD optimizer, with mAP@50 of 0.61, F1-Score of 0.6, and training time of 356.4 seconds. Based on [29], AP can represent the model's ability to allocate and classify objects with spatial tolerance. These tuning results were then used to observe model tuning during training at epochs 25 and 50. This was done to improve computational efficiency and to observe the effect of each parameter on model training, ensuring each contributed positively to model performance.

The results of the second hyperparameter tuning show a comparison between models trained with 25 epochs and 50 epochs. The model with 25 epochs has a faster training time of 900 seconds (15 minutes), while the model with a longer epoch has a longer training time of around 30 minutes and 25 seconds. With twice the number of epochs, the training time will also increase by twice as much. The results with a lower epoch also show a lower inference time for the best epoch, indicating that a model with an epoch of 50 provides the best performance. The model performance shows mAP@50 of 0.867, mAP@50-95 of 0.601, and inference time of approximately 10.8-16.7 ms. This is shown in Table VI and Fig. 4, which present the performance of each hyperparameter and the loss graph for the best model.

TABLE V
FIRST STAGE OF HYPERPARAMETER TUNING

No	Tuning Hyperparameter	Precision	Recall	mAP@50	F1-Score	Training Time (second)
1	lr 0,001; batch 8, Opt Adam	0.468	0.473	0.474	0.470	367.2
2	lr 0,01; batch 8, Opt Adam	0.253	0.435	0.261	0.320	392.4
3	lr 0,01; batch 8, Opt SGD	0.647	0.560	0.610	0.600	356.4
4	lr 0,001; batch 8, Opt SGD	0.177	0.368	0.223	0.239	342.0
5	lr 0,01; batch 16, Opt Adam	0.184	0.338	0.207	0.238	392.4
6	lr 0,001; batch 16, Opt Adam	0.521	0.558	0.566	0.539	360.0
7	lr 0,001; batch 16, Opt SGD	0.209	0.293	0.197	0.244	320.4
8	lr 0,01; batch 16, Opt SGD	0.490	0.552	0.530	0.519	324.0

TABLE VI
SECOND STAGE OF HYPERPARAMETER TUNING

Tuning Hyperparameter	Precision	Recall	mAP@50	mAP@50-95	F1-Score	Training Time	Inference Time (ms)
Epoch 25	0.752	0.660	0.767	0.516	0.703	900.0 s	14 – 20.8
Epoch 50	0.817	0.816	0.867	0.601	0.816	1825.2 s	10.8 – 16.7

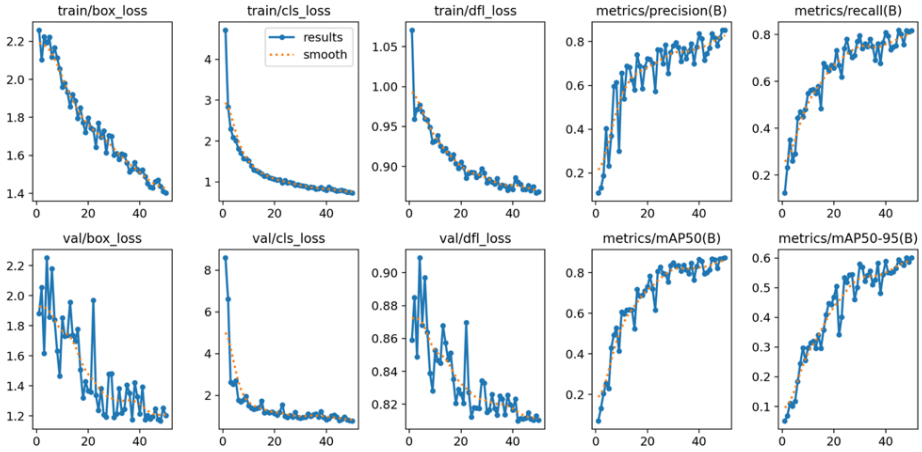

Fig. 4. YOLOv5s Loss Curve at 50 Epochs

Based on Fig. 4, the Box Loss curves in training and validation show a consistent downward trend, indicating that the bounding box regression in the model is becoming increasingly accurate at estimating the center coordinates, width, and height of objects. Meanwhile, the decrease in Class Loss indicates that the model has converged, suggesting it can learn distinguishing features between classes. Furthermore, there is a significant decrease in Distribution Focal Loss, indicating that the model can handle uncertainty in box predictions at the appropriate confidence level. The downward trend in the loss graph suggests a reversal in model performance, namely an increase in its ability to predict objects and provide bounding boxes. This increase is reflected in the performance metrics, such as precision, recall, mAP@50, and mAP@50-95, which increase consistently as the number of epochs increases.

The precision matrix shows that the model achieves a relatively high accuracy of 0.817 in predicting the presence of an object. In addition, the recall matrix shows similar performance with a value of 0.816, indicating that the model can find most objects in the image. The balance between recall and precision shows that the model is robust. A robust model can also be indicated by comparing the loss curves for the training and validation data, also called model fit. Based on a comparison of the train and val curves, there is no indication of underfitting or overfitting, so the model can be considered a good fit or robust. The mAP value indicates the final model performance and serves as the threshold for IoU calculation. The curve shows that at a low level of

strictness, mAP@50-95 reaches 0.617. This value is a good indicator of the model's ability to accurately extract complex features and predict green coffee bean objects in general, based on their classification. The results of the second hyperparameter tuning then show how the model performs in each trained class. Table VII and Fig. 5 show the performance of each classification of coffee bean defect prediction using SNI 01-2907-2008 regulations.

TABLE VII
RESULT FROM YOLOV5S MODEL FOR EACH CLASSIFICATION CLASS

Class	Precision	Recall	mAP@50	mAP@50-95	F1-Score
Black Bean	0.844	0.749	0.818	0.410	0.794
Partial Black Bean	0.618	0.918	0.829	0.576	0.739
Broken Black Bean	0.858	0.927	0.909	0.412	0.891
Pod Bean	0.772	0.995	0.986	0.766	0.869
Brown Bean	0.664	0.926	0.825	0.558	0.773
Large Shell Skin	0.888	0.822	0.939	0.708	0.854
Medium Shell Skin	0.766	0.892	0.910	0.704	0.824
Small Shell Skin	0.955	0.939	0.985	0.679	0.947
Parchment Skin	0.980	0.862	0.958	0.711	0.917
Large Husk	0.816	0.649	0.734	0.573	0.723
Medium Husk	0.820	0.971	0.970	0.723	0.889
Small Husk	0.935	0.803	0.917	0.590	0.864
Broken Bean	0.938	0.797	0.879	0.503	0.862
Immature Bean	0.974	0.877	0.976	0.621	0.923
Bean with Slight Insect Damage	0.576	0.624	0.626	0.462	0.599
Bean with Severe Insect Damage	0.732	0.792	0.840	0.649	0.761
Bean with Fungus Damage	0.840	0.524	0.740	0.587	0.645
Large Foreign Matter	0.758	0.948	0.961	0.727	0.842
Medium Foreign Matter	0.619	0.714	0.688	0.511	0.663
Small Foreign Matter	0.989	0.597	0.843	0.550	0.745

Table VI shows that the robust architecture's generalization capacity, measured by mAP@50 reaches 0.867 at epoch 50. However, Table VII shows that the matrix decomposition per class reveals crucial performance heterogeneity. There is morphological variability that can trigger significant accuracy gaps, such as pod bean and immature bean, with mAP@50 of more than 0.97, compared to beans with slight insect damage, with mAP@50 of 0.626. In addition, a significant decrease in value between mAP@50 and mAP@50-95, such as black beans, which dropped from 0.818 to 0.41, indicates that even though the model is very good at recognizing object types, the accuracy of bounding box predictions at strict thresholds still needs optimization to ensure adequate precision standards. The validation-stage prediction results show the spatial locations and classifications of each green coffee bean with physical surface defects, as shown in Fig. 5.

Fig. 5. Prediction Result of YOLOv5s

The provided image illustrates a batch of YOLOv5s prediction results for coffee bean defect detection as shown in Fig. 5, demonstrating the model's capacity to handle high-density object localization and multi-class classification within a single frame. The mosaic grid reveals successful identification of various defect categories such as broken bean, black bean, and bean with insect damage with confidence scores consistently maintained above 0.25 to capture subtle variations. Despite the visual complexity and label congestion caused by the sheer volume of beans in each 50 g sample, the YOLOv5s architecture effectively distinguishes between distinct defect types and foreign materials using specific color-coded bounding boxes. This automated approach shows high sensitivity to small-scale features, suggesting a robust performance in replicating manual grading standards through deep learning. Furthermore, to provide a more comprehensive evaluation of these visual detections, the quantitative metrics and a detailed comparison against other architectures are presented in Table VIII, which outlines the comparative model performance.

TABLE VIII
COMPARATIVE MODEL PERFORMANCE

Model	Precision	Recall	mAP@ 50	mAP@50-95	Training Time	Inference Time (ms)	Params (M)
YOLOv5n	0.687	0.639	0.708	0.495	5 minutes 17 seconds	16.5	2.5
YOLOv5s	**0.817**	**0.816**	**0.867**	**0.601**	**5 minutes 53 seconds**	**14.4**	**9.1**
YOLOv5m	0.908	0.866	0.910	0.654	10 minutes 16 seconds	15,2	25.1
YOLOv5l	0.904	0.876	0.913	0.656	14 minutes 53 seconds	24.1	53.1

To contextualize the performance of the selected YOLOv5s architecture, comparative experiments were conducted using YOLOv5n (nano), YOLOv5m (medium), and YOLOv5l (large) under identical training conditions, specifically learning rate of 0.01, batch size of 8, SGD optimizer, and 50 epochs. The results presented in Table VIII show that the YOLOv5n model achieved a mAP@50 of 0.708 with a precision of 0.687 and recall of 0.639. Although YOLOv5n is the most computationally efficient variant with only 2.5 million parameters and the shortest training time of 5 minutes and 17 seconds, its detection performance is substantially lower than YOLOv5s across all metrics. The 15.9-point gap in mAP@50 (0.708 versus 0.867) reflects the model's insufficient representational capacity for a 20-class fine-grained defect taxonomy. This aggressive parameter reduction in YOLOv5n limits the feature discriminability required to distinguish visually similar categories, such as slight versus severe insect damage or the three size variants of husk and shell skin defects. Consequently, while YOLOv5n might suffice for binary detection tasks on ultra-constrained hardware, it proves inadequate for a complete classification scheme.

In contrast, YOLOv5m achieved a mAP@50 of 0.910 with precision of 0.908 and recall of 0.866, demonstrating higher detection performance than YOLOv5s. However, this marginal accuracy gain of 4.3 points in mAP@50 comes with a substantially higher computational cost. YOLOv5m requires 25.1 million parameters, nearly three times the parameter count of YOLOv5s, and its training time of 10 minutes and 16 seconds is approximately 1.75 times longer. Furthermore, the inference time of 15.2 ms is slightly higher than YOLOv5s. Similarly, YOLOv5l achieved a mAP@50 of 0.913 with the highest precision of 0.904 and recall of 0.876, representing the best absolute accuracy among all variants. However, this comes at a significantly greater cost, requiring 53.1 million parameters, approximately 5.8 times more than YOLOv5s, and the longest inference time of 24.1 ms, which is 67% slower than YOLOv5s. The training time of 14 minutes and 53 seconds is 2.5 times longer than YOLOv5s.

These computational demands make YOLOv5l and YOLOv5m impractical for deployment on low-cost edge devices such as Raspberry Pi or NVIDIA Jetson Nano, which are the target hardware for Indonesian coffee MSMEs. The slight improvement in mAP@50 between YOLOv5s and YOLOv5m (0.867 versus 0.910) does not justify the 2.9-fold increase in parameters and the associated latency and energy consumption in resource-constrained environments. YOLOv5s therefore represents the optimal balance between detection performance with a mAP@50 = 0.867 and an inference time of 14.4 ms, validating it as the preferred architecture for real-time coffee bean quality inspection in compliance with SNI 01-2907-2008. he comparison across the four YOLOv5 variants confirms that YOLOv5s provides the most effective balance of accuracy and efficiency for 20-class defect detection, making it as the most viable architecture for implementation on low-cost hardware in SME environments. To further analyze the classification accuracy across all 20 defect categories and identify any specific inter-class misclassifications, the confusion matrix for the YOLOv5s model is presented in Fig. 6.

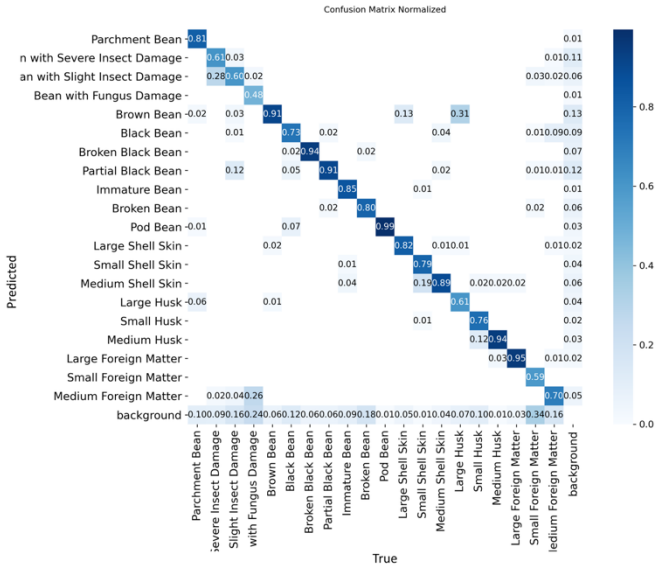

Fig. 6. Confusion Matrix YOLOv5s

Based on the results of the YOLOv5s model training, Fig. 6 shows the X-axis as the actual and the Y-axis as the predicted. The results of the confusion matrix show that all classifications have been successfully predicted with a score of at least 0.5, indicating the model's performance in distinguishing true positives and true negatives. However, there is still a massive misclassification of 0.28 between beans with severe insect damage that are predicted to beans with slight insect damage. Visually and computationally, there may be failures in distinguishing between colour bits and holes. Dark pixels in holes have identical features caused by spatial resolution or lighting that is not distinctive enough to differentiate between surface and depth. In addition, the low success rate for beans with fungus damage, which were only 0.48 and 0.26 incorrectly predicted as medium foreign materials and 0,24 lost to background, indicates that the model experiences negative texture bias. White and blue spots on beans with fungus damage are misinterpreted as rough textures from foreign materials or considered as background noise, resulting in detection failure.

The prediction for large husk beans shows that the prediction is off by 0.31 to brown beans. This is due to the actual condition of large husk beans, which often change colour (degradation) due to oxidation or contamination, making them visually similar to brown beans. In addition, morphological similarities due to the size of large husk beans, which may still contain coffee beans inside, result in geometric and colour similarities. The confusion matrix also shows that other materials, especially medium and small ones, exhibit false negatives of 0.34 and 0.16, respectively. This indicates the presence of visual camouflage between dark colours and textures with low contrast (contrast ambiguity) between the object and the background, resulting in the loss of spatial feature representation, as well as limitations in feature resolution on small objects due to the failure to produce a confidence score above the validation threshold of 0.25, causing the object to be considered background noise.

This study shows that number of epoch is positively correlated with improved coffee bean defect detection performance, even with longer training times. The trade-off between training time and performance in the YOLOv5s model provides clues for adjusting model usage to requirements. However, this model shows that the length of the training duration provides a relatively faster image prediction time of two to three ms. The speed of accurate prediction can encourage the implementation of coffee bean defect detection based on SNI 01-2907-2008 in real time. Based on other research in the field of agricultural defect detection using YOLOv5, it has been shown that this model is the optimal solution for precise physical surface defect inspection with low computational requirements, thus ensuring real-time implementation on low-power devices [30], [31], [32].

In detecting physical surface defects in green coffee beans under the SNI 01-2907-2008 classification, previous research has achieved an accuracy of up to 53% by simplifying the classification into 17 types using an image classification method [4]. This simplification combines other materials with three sizes into a single size and classifies small husk that is not present in the training dataset. Although a direct quantitative

comparison is constrained by the use of different proprietary datasets, the transition from whole-image classification to our YOLO-based object detection framework demonstrates a fundamental enhancement in both detection capability and practical compliance with the SNI standard. In this study, the current architecture has high reliability in general classification with the strictest threshold at mAP@50-95, producing a success rate of 60.1%. The accuracy indicates that a model using an object detection method performs better with additional spatial location information for green coffee beans with defects. Additionally, the prediction speed, averaging 14.36 ms (0.01436 seconds) per image, demonstrates the model's suitability for use on low-power devices. Therefore, SMEs may utilize this model to detect physical surface defects in green coffee beans.

V. CONCLUSION

Based on this study using transfer learning, it has been proven that implementing object detection for classifying the physical quality of green coffee beans can be an effective automation alternative for meeting the SNI 01-2907-2008 standard. The result indicated a mAP@50 of 0.867, which represents the model's capacity to replicate human visual perception of major physical defects. However, findings on texture bias in certain beans and limitations in bounding-box precision indicate that the system still requires targeted feature optimization to achieve prediction accuracy within strict industry tolerances. Overall, this model provides a strong technical foundation for transforming the subjective manual sorting process into a standardized, fast, and efficient digital inspection system. Future research will prioritize improving detection for underperforming defect categories and validating the model's real-world robustness. To address current limitations, expanding the dataset for underrepresented defects, applying focal loss, and exploring near-infrared imaging are recommended to detect sub-surface damage invisible to standard RGB cameras. Furthermore, to transition from laboratory evaluation to industrial application, the model requires validation within an IoT edge-computing pipeline. This involves deploying the system on edge devices (e.g., NVIDIA Jetson Nano) integrated with standardized conveyor illumination, automated sorting mechanisms, and cloud dashboards to evaluate its real-time throughput and practical viability in operational coffee processing environments.

DATA AND COMPUTER PROGRAM AVAILABILITY

All the data used in this study can be obtained from the corresponding author upon a reasonable request. Note: If anyone needs to request the data, feel free to reach out to me (achmad.p.rifai@ugm.ac.id).

ACKNOWLEDGMENT

The authors would like to thank to Dhira Anantawijaya and Much. Andrianto Wicaksono, the coffee farmer and coffee processor in Gandiva Coffee Processing Unit (Malang, Indonesia), for providing free green coffee bean samples.

FUNDING

This study is funded by Badan Riset dan Inovasi Nasional (BRIN) and Lembaga Pengelola Dana Pendidikan (LPDP) Republic of Indonesia, Riset dan Inovasi untuk Indonesia Maju (RIIM) Grant No.172/IV/KS/11/2023 and 6815/UN1/DITLIT/Dit-Lit/KP.01.03/2023 year 2.

REFERENCES

[1] K. Khusnul, Suratno, N. I. Asyiah, and S. Hariyadi, “Analysis of the Effect of Several Types of Shade on the Productivity of Robusta Coffee”, in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2021, doi: 10.1088/1742-6596/1751/1/012060.

[2] Badan Pusat Statistik Indonesia, "Statistik Kopi Indonesia 2023 [Online]", Vol. 8, 2024, Accessed: Apr. 27, 2025. [Online]. Available: https://www.bps.go.id/id/publication/2024/11/29/d748d9bf594118fe112fc51e/statistik-kopi-indonesia-2023.html

[3] SNI 01-2907-2008, "Biji Kopi", 2008.

[4] M. W. A. Kesiman, I. Sulaiman, I. M. D. Maysanjaya, and K. T. Dermawan, "Benchmarking A New Dataset for Coffee Bean Defects Classification Based on SNI 01-2907-2008", in *2023 International Conference on Information Technology Research and Innovation, ICITRI 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 75–80, doi: 10.1109/ICITRI59340.2023.10249345.

[5] S. O. Araújo, R. S. Peres, J. C. Ramalho, F. Lidon, and J. Barata, "Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives", Dec. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI),* doi: 10.3390/agronomy13122976.

[6] T. A. Heryanto and I. G. B. B. Nugraha, "Classification of Coffee Beans Defect Using Mask Region-based Convolutional Neural Network", in *2022 International Conference on Information Technology Systems and Innovation, ICITSI 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 333–339, doi: 10.1109/ICITSI56531.2022.9970890.

[7] G. Jocher *et al.*, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation", Nov. 2022, *Zenodo,* doi: 10.5281/zenodo.3908559.

[8] E. D. Nugroho *et al.*, "Development of YOLO-Based Mobile Application for Detection of Defect Types in Robusta Coffee Beans", 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC.

[9] K. Muchtar *et al.*, "Edge AI-Based Detection for Defective Coffee Beans using Deep Learning and Streamlit Framework", *IEEE Access*, pp. 67977–67992, Apr. 2025, doi: 10.1109/ACCESS.2025.3561189.

[10] A. Rivalto, Pranowo, and A. J. Santoso, "Classification of Indonesian coffee types with deep learning", in *AIP Conference Proceedings*, American Institute of Physics Inc., Apr. 2020, doi: 10.1063/5.0000678.

[11] M. Murinto, M. Rosyda, and M. Melany, "Klasifikasi Jenis Biji Kopi Menggunakan Convolutional Neural Network dan Transfer Learning pada Model VGG16 dan MobileNetV2", *JRST (Jurnal Riset Sains dan Teknologi)*, Vol. 7, No. 2, p. 183, Sep. 2023, doi: 10.30595/jrst.v7i2.16788.

[12] N.-F. Huang, D.-L. Chou, and C.-A. Lee, "Real-Time Classification of Green Coffee Beans by Using a Convolutional Neural Network", *3rd International Conference on Imaging, Signal Processing and Communication*, p. 171, 2019, doi: 10.1109/ICISPC.2019.8935644.

[13] A. Pratondo, T. Zani, A. Novianty, and B. Pudjoatmodjo, "Raw Coffee Bean Classification for Roasting Suitability Assessment Using Transfer Learning", in *2023 IEEE 11th Conference on Systems, Process and Control, ICSPC 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 390–395, doi: 10.1109/ICSPC59664.2023.10419990.

[14] L. Y. Ke, E. Chen, and C. H. Hsia, "Green Coffee Bean Defect Detection Using Shift-Invariant Features and Non-Local Block", in *Proceedings of the 2023 IEEE 6th International Conference on Knowledge Innovation and Invention, ICKII 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 430–431, doi: 10.1109/ICKII58656.2023.10332580.

[15] C. H. Hsia, Y. H. Lee, and C. F. Lai, "An Explainable and Lightweight Deep Convolutional Neural Network for Quality Detection of Green Coffee Beans", *Applied Sciences (Switzerland)*, Vol. 12, No. 21, Nov. 2022, doi: 10.3390/app122110966.

[16] S. J. Chang and K. H. Liu, "Multiscale Defect Extraction Neural Network for Green Coffee Bean Defects Detection", *IEEE Access*, Vol. 12, pp. 15856–15866, 2024, doi: 10.1109/ACCESS.2024.3356596.

[17] C. S. Liang, Z. Y. Xu, J. Y. Zhou, C. M. Yang, and J. Y. Chen, "Automated Detection of Coffee Bean Defects using Multi-Deep Learning Models", in *Proceedings - 2023 VTS Asia Pacific Wireless Communications Symposium, APWCS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, doi: 10.1109/APWCS60142.2023.10234059.

[18] P. Wang, H. W. Tseng, T. C. Chen, and C. H. Hsia, "Deep convolutional neural network for coffee bean inspection", *Sensors and Materials*, Vol. 33, No. 7, pp. 2299–2310, Jul. 2021, doi: 10.18494/SAM.2021.3277.

[19] S. Arwatchananukul, D. Xu, P. Charoenkwan, S. Aung Moon, and R. Saengrayap, "Implementing a deep learning model for defect classification in Thai Arabica green coffee beans", *Smart Agricultural Technology*, Vol. 9, Dec. 2024, doi: 10.1016/j.atech.2024.100680.

[20] P. C. Manojkumar, L. S. Kumar, and B. Jayanthi, "Performance Comparison of Real Time Object Detection Techniques with YOLOv4", in *Proceedings of 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication, IConSCEPT 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, doi: 10.1109/IConSCEPT57958.2023.10169970.

[21] N. F. Huang *et al.*, "Smart agriculture: real-time classification of green coffee beans by using a convolutional neural network", *IET Smart Cities*, Vol. 2, No. 4, pp. 167–172, Dec. 2020, doi: 10.1049/iet-smc.2020.0068.

[22] G. A. Pratama, E. Y. Puspaningrum, and H. Maulana, "Convolutional Neural Network dan Faster Region Convolutional Neural Netowrk untuk Klasifikasi Kualitas Biji Kopi Arabika", *Jurnal Informatika dan Teknik Elektro Terapan*, Vol. 12, No. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4887.

[23] A. J. Manansala and E. C. C. Paglinawan, "Classification of Coffea Liberica Quality Using Convolution Neural Networks (Slim-CNN, YOLOv5, and VGG-16)", in *2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, doi: 10.1109/ICCCNT61001.2024.10723931.

[24] D. Buonocore, M. Carratù, and M. Lamberti, "Classification of coffee bean varieties based on a deep learning approach", in *18th IMEKO TC10 Conference: Measurement for Diagnostics, Optimisation and Control to Support Sustainability and Resilience,* Warsaw, Sep. 2022, p. 14. doi: 10.21014/tc10-2022.002.

[25] D. Tsalsabila Rhamadiyanti, "Analisa Performa Convolutional Neural Network dalam Klasifikasi Citra Apel dengan Data Augmentasi", *Media Online)*, Vol. 5, No. 1, pp. 154–162, 2024, doi: 10.30865/klik.v5i1.2023.

[26] F. G. Lalamentik, O. A. Lantang, and F. D. Kambey, "Implementation of Parameter Tuning for Optimizing CNN Model Performance", *Jurnal Teknik Informatika*, Vol. 20, No. 1, pp. 77–86, Feb. 2025, [Online]. Available: https://ejournal.unsrat.ac.id/index.php/informatika.

[27] G. B. Mohan, N. A. Prasad, V. Amirthavarshini, K. Ananya, N. Hrishikeasan, and V. Viswanathan, "Segmentation of Instances in an Image with Custom Neural Networks", in *2024 3rd International Conference on Artificial Intelligence for Internet of Things, AIIoT 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, doi: 10.1109/AIIoT58432.2024.10574618.

[28] J. H. Cabot and E. G. Ross, "Evaluating prediction model performance", *Surgery (United States)*, Vol. 174, No. 3, pp. 723–726, Sep. 2023, doi: 10.1016/j.surg.2023.05.023.

[29] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal Visual Object Classes (VOC) Challenge", *Int. J. Comput. Vis.*, Vol. 88, No. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.

[30] H. Wu, R. Zhu, H. Wang, X. Wang, J. Huang, and S. Liu, "Flaw-YOLOv5s: A Lightweight Potato Surface Defect Detection Algorithm Based on Multi-Scale Feature Fusion", *Agronomy*, Vol. 15, No. 4, Apr. 2025, doi: 10.3390/agronomy15040875.

[31] J. Yao, J. Qi, J. Zhang, H. Shao, J. Yang, and X. Li, "A real-time detection algorithm for kiwifruit defects based on yolov5", *Electronics (Switzerland)*, Vol. 10, No. 14, Jul. 2021, doi: 10.3390/electronics10141711.

[32] Y. Zhou, Z. Li, S. Xue, M. Wu, T. Zhu, and C. Ni, "Lightweight SCD-YOLOv5s: The Detection of Small Defects on Passion Fruit with Improved YOLOv5s", *Agriculture (Switzerland)*, Vol. 15, No. 10, May 2025, doi: 10.3390/agriculture15101111.