

Application of MobileNetV2 for Coffee Bean Roast Level Classification to Support Coffee Quality Control

Angga Bayu Santoso^{1*}, Okma Arnilia², Sahrial Ihsani Ishak³, I Gusti Nyoman Agung Bisma Tatwa⁴, Riki Rikardo⁵

^{1,5}Information System, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Lampung, Indonesia

²Informatics, Faculty of Education and Teacher Training, UIN Siber Syekh Nurjati Cirebon, Jawa Barat, Indonesia

³Informatics Engineering, Faculty of Engineering and Informatics, Universitas Dian Nusantara, Jawa Barat, Indonesia

⁴Computer Science, Faculty of Mathematics and Computer Science, Institut Pertanian Bogor, Jawa Barat, Indonesia

Received: 14-08-2026
Accepted: 02-09-2026

Keywords:

Coffee Bean Classification;
Deep Learning;
Roast Level Classification;
MobileNetV2;
Transfer Learning.

Correspondent Email:

anggabayu@teknokrat.ac.id

Abstract. *This study addresses the problem of manual classification of coffee bean roast level, which is often inconsistent and inefficient, potentially affecting product quality and market value. To overcome this issue, the study proposes an automated classification approach using a deep learning model based on MobileNetV2. The objective of this study is to develop an accurate and efficient model for classifying coffee beans into four roast levels: Green, Light, Medium, and Dark, while ensuring its applicability in real-world conditions with limited computational resources. The dataset consists of 1,200 images of roasted and unroasted coffee beans, processed through resizing, normalization, and data augmentation to improve generalization. The model was trained using transfer learning and evaluated using accuracy, precision, recall, and F1-score metrics. The results show that the model achieves an accuracy of 97.78% on the testing dataset, with a precision of 0.98, recall of 0.98, and F1-score of 0.98. When implemented in a web-based application, it achieves 97.8% accuracy with an average response time of 1.8 seconds and is capable of handling up to 100 concurrent users with stable performance. These findings indicate that the proposed model is effective and reliable for supporting automated coffee bean roast level classification and enhancing efficiency in coffee quality control.*

1. INTRODUCTION

Indonesia is one of the largest coffee-producing countries in the world, contributing significantly to the global coffee supply through diverse coffee varieties originating from regions such as Sumatra, Java, Bali, and Sulawesi [18]. Coffee is an important agricultural commodity with significant economic value, whose quality is influenced by various production and post-harvest processes, including the classification of coffee bean roast level [2]. The roast level of coffee beans can

affect the physical and sensory characteristics of coffee, including flavor, aroma, and overall product quality [2]. Therefore, accurate roast level classification is important to support consistent quality control and maintain the standardization and market value of coffee products [18].

Traditionally, coffee bean roast level classification is performed manually through visual inspection of characteristics such as color, texture, shape, and physical appearance [6]. However, this process is subjective because

it depends on the experience and perception of individual workers and can be affected by environmental conditions, particularly lighting [1]. Differences in human judgment may result in inconsistent classification outcomes and affect product standardization [18]. Manual sorting also requires considerable time and labor, making it less efficient for large-scale production [2]. These limitations highlight the need for an automated system that can classify coffee bean roast level more objectively, consistently, and efficiently [5].

The development of artificial intelligence, particularly computer vision and deep learning, provides opportunities to automate image-based classification in the coffee industry. Traditional machine learning methods, such as Naïve Bayes and Support Vector Machine (SVM), generally require manually designed features, which can limit their ability to capture complex visual patterns [22], [5]. In contrast, Convolutional Neural Networks (CNNs) can automatically learn relevant features directly from image data and have demonstrated strong performance in various coffee-related classification tasks [11].

Several studies have applied deep learning to coffee-related image analysis, including coffee quality, roasting levels, varieties, and maturity classification [4], [6]. EfficientNet-based models have demonstrated high classification performance, while ResNet architectures have shown strong feature extraction capabilities [2], [6], [7]. MobileNet-based architectures have also been explored because of their lower computational requirements and potential for mobile deployment [3]. Meanwhile, object detection models such as YOLO have been applied to related coffee classification tasks, such as maturity detection, but generally require greater computational resources than image classification models [5], [25]. These studies demonstrate the potential of deep learning while highlighting the importance of selecting an architecture according to the specific classification objective and deployment environment.

Despite these developments, research on lightweight deep learning for coffee bean roast level classification remains limited. Previous studies mainly address coffee variety, quality assessment, or maturity detection [4], [6], while

high-performance models often require substantial computational resources [7]. Thus, further research is needed to evaluate lightweight CNNs that balance classification accuracy and computational efficiency for practical deployment [8].

To address this gap, this study applies MobileNetV2 for coffee bean roast level classification. MobileNetV2 is a lightweight CNN architecture that utilizes depthwise separable convolution, inverted residual blocks, and linear bottlenecks to reduce computational complexity while maintaining effective feature extraction [1]. Its lightweight design makes it suitable for image classification applications and potential deployment on mobile or embedded devices with limited computational resources [17].

The application of MobileNetV2 can support automated coffee sorting by providing more consistent and objective roast level classification than manual inspection. Its computational efficiency also provides opportunities for implementation in production or processing environments where high-performance hardware may not be readily available. This approach is aligned with efforts to improve coffee quality control through the use of digital technologies that enhance consistency, efficiency, and decision-making in coffee processing [18].

The research gap addressed in this study is the limited investigation of lightweight deep learning models specifically for coffee bean roast level classification while considering computational efficiency and practical implementation. The novelty of this study lies in applying MobileNetV2 specifically to coffee bean roast level classification and evaluating the balance between classification performance and computational efficiency. Therefore, this study aims to develop a MobileNetV2-based model, evaluate its performance using accuracy, precision, recall, and F1-score, and assess its potential suitability for efficient real-time implementation to support coffee quality control.

In addition to model development, this study also implements the trained model into a web-based application to evaluate its practical performance in terms of response time and scalability under concurrent usage, providing a

more comprehensive assessment of its readiness for real-world coffee quality control.

2. LITERATURE REVIEW

2.1. Coffee Bean Roast Level Classification

Coffee bean roast level classification plays an important role in assessing coffee quality, as different roast levels affect the color, flavor, aroma, and sensory characteristics of coffee beans [2]. Conventional classification based on visual appearance is subjective and can be inconsistent due to differences in human perception and environmental conditions [6]. Therefore, an automated image-based classification approach is needed to achieve more consistent and efficient roast level identification [5].

2.2. Image Classification

Image classification is a fundamental task in computer vision that aims to categorize images into predefined classes based on visual features [20]. In image processing, data is represented as pixel values, which are typically normalized to improve model performance and training stability [20]. The normalization process can be expressed as:

$$x' = \frac{x}{255} \quad (1)$$

where x is the original pixel value and x' is the normalized value.

2.3. Machine Learning and Deep Learning

Traditional machine learning methods such as Support Vector Machine (SVM) have been applied to image classification tasks [22]. The decision function of SVM can be formulated as:

$$f(x) = w \cdot x + b \quad (2)$$

where w represents weights, x is the input feature vector, and b is the bias. However, these methods rely on manual feature extraction, which limits their ability to capture complex patterns in image data [5].

To overcome these limitations, deep learning approaches have been introduced. Deep learning models are capable of automatically learning hierarchical features

from raw data, leading to improved performance in complex classification tasks [4].

2.4. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a deep learning architecture specifically designed for image processing tasks [6]. CNN extracts features through convolution operations, which can be mathematically expressed as:

$$y(i,j) = \sum_m \sum_n x(i+m, j+n) \cdot w(m,n) \quad (3)$$

where x is the input image, w is the convolution kernel, and y is the output feature map.

To introduce non-linearity, activation functions such as Rectified Linear Unit (ReLU) are applied:

$$f(x) = \max(0, x) \quad (4)$$

CNN models have demonstrated superior performance in various image classification tasks, including agricultural applications [19].

2.5. MobileNetV2 Architecture

MobileNetV2 is a lightweight CNN architecture designed for efficient computation while maintaining high accuracy [1]. It utilizes depthwise separable convolution to reduce the number of parameters and computational cost compared to standard convolution [8]. Additionally, MobileNetV2 incorporates inverted residual blocks and linear bottlenecks to improve feature representation while preserving efficiency. These characteristics make MobileNetV2 highly suitable for real-time applications and deployment in resource-constrained environments such as mobile and web-based systems [17].

2.6. Evaluation Metrics

Model evaluation in this study is conducted using several metrics, namely accuracy, precision, recall, and F1-score, to comprehensively assess classification performance [14]. These metrics are important because they provide both an overall evaluation and a detailed understanding of the model's performance for each class [15].

All evaluation metrics are derived from the confusion matrix, which consists of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). A general formulation of classification evaluation can be expressed as:

$$Metric = \frac{Correct\ Predictions}{Total\ Predictions} \quad (5)$$

Accuracy measures overall correctness, precision evaluates the accuracy of positive predictions, recall measures the ability to identify actual positives, and F1-score balances precision and recall. Together, these metrics provide a comprehensive evaluation of model performance both overall and per class [14][15].

3. RESEARCH METHODS

The research methodology includes dataset preparation, data augmentation, MobileNetV2 development using transfer learning, model training, evaluation, and web implementation. The model is trained using Adam and categorical crossentropy, evaluated with accuracy, precision, recall, and F1-score, and deployed in a Flask-based application for real-time classification. The methodology is presented in Figure 1.

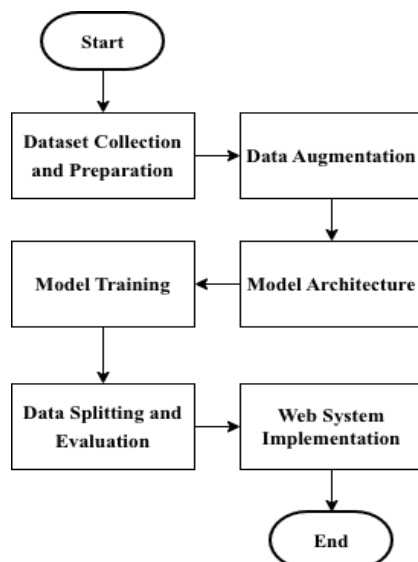


Figure 1. Research Methodology

3.1. Dataset Collection and Preparation

This study utilizes a custom dataset consisting of 1,200 coffee bean images categorized into four roast levels: *Green*, *Light*,

Medium, and *Dark*. The dataset was collected from coffee plantations in Indonesia to represent variations encountered in real-world conditions. During data acquisition, variations in lighting conditions, camera angles, and background settings were considered to improve the model's ability to generalize to diverse environments.

Each roast level contains 300 images, resulting in a balanced dataset with an equal number of samples across all four classes. This balanced distribution helps minimize class imbalance and provides an equitable basis for model training and evaluation.

The images were captured using a smartphone camera at an approximate distance of 15–20 cm from the coffee beans under natural indoor lighting. A plain background was used to minimize visual interference. Each image contains a single layer of coffee beans representing one roast level, with samples collected from multiple batches to reduce potential batch-specific bias. The class labels were assigned based on the recorded roasting level and visually verified according to differences in bean color and surface characteristics. Samples with ambiguous visual characteristics were excluded to maintain labeling consistency.

Representative images of the four roast levels are presented in Table 2. The preprocessing stage includes resizing all images to 224×224 pixels to match the input dimensions of MobileNetV2 and normalizing pixel values to the range of [0,1] by dividing each pixel value by 255. Data augmentation is applied only to the training set to increase data diversity, improve generalization, and reduce the risk of overfitting. The distribution of the dataset is presented in Table 1.

Table 1. Dataset Distribution

Class	Count	Percent	Train	Val	Test
Dark	300	25%	210	45	45
Green	300	25%	210	45	45
Light	300	25%	210	45	45
Medium	300	25%	210	45	45
Total	1200	100%	840	180	180

To provide a clear visual reference for each roasting category used in this study, Table 2 presents sample images of coffee beans

classified into four roast levels: Green, Light, Medium, and Dark.

Table 2. Classification of Coffee Bean Samples Based on Roasting Level

Class	Roast Level	Sample
Green	Un-roasted	
Light	Light roast	
Medium	Medium roast	
Dark	Dark roast	

Table 2 presents the classification of coffee beans into four classes based on roasting level, namely Green, Light, Medium, and Dark, each of which reflects a distinct stage of thermal processing that directly affects the beans' color, surface texture, and physical characteristics. The Green class represents un-roasted beans with a yellowish-green color and a dry, matte surface; the Light class shows beans roasted at a light level, appearing light golden brown with a surface that is not yet oily; the Medium class represents a moderate roasting level, characterized by a darker brown color and a slightly glossy surface; and the Dark class represents the highest roasting level, marked by a dark brown to blackish color with a shiny, oily surface caused by the release of natural bean oils during intensive roasting. These visual differences form an important basis for automated classification using computer vision or deep learning methods, as each roasting level

exhibits relatively consistent and visually distinguishable color and texture patterns.

3.2. Data Augmentation

Data augmentation plays an important role in improving the robustness and generalization of deep learning models, particularly when the available dataset is limited. In this study, data augmentation is applied in real-time during the training process using TensorFlow's *ImageDataGenerator*. This approach generates variations of coffee bean images dynamically in each training epoch without increasing the size of the original dataset. By exposing the MobileNetV2 model to diverse image variations, the augmentation process helps the model adapt to variations that may occur during real-world coffee quality control.

The augmentation techniques used in this study include rotation, zooming, width shifting, height shifting, and horizontal flipping. Rotation simulates variations in the orientation of coffee beans, while zooming represents differences in the distance between the camera and the coffee beans. Width and height shifting simulate changes in the position of coffee beans within the image, whereas horizontal flipping increases visual diversity. These transformations are intended to reduce overfitting and improve the model's ability to recognize roast-level characteristics across different image conditions. The augmentation parameters used in the MobileNetV2 training process are presented in Table 3.

Table 3. Data Augmentation Parameters

Technique	Parameter	Value	Purpose
Rotation Range	Rotation angle	20°	Orientation variation
Zoom Range	Zoom factor	0.2	Distance variation
Width Shift Range	Horizontal shift	0.1	Position variation
Height Shift Range	Vertical shift	0.1	Position variation
Horizontal Flip	Flip horizontally	TRUE	Perspective variation
Rescale	Normalization	1/255	Value standardization

3.3. Model Architecture

The model used in this study is MobileNetV2, modified for four-class classification. It is selected due to its efficiency and lightweight design, making it suitable for resource-constrained environments. The architecture uses depthwise separable convolution to reduce parameters while maintaining high accuracy.

MobileNetV2 also employs inverted residual blocks and linear bottlenecks for efficient feature extraction. Transfer learning is applied using a pre-trained ImageNet model to improve performance and reduce training time. The final layers include Global Average Pooling, a Dense layer, Dropout for regularization, and an output layer with four neurons. Dropout helps prevent overfitting by promoting more generalized feature learning. The model architecture is presented in Table 4.

Table 4. Model Architecture

Layer	Shape	Params	Desc
Input	(224,224,3)	0	RGB image input
MobileNet V2 Base	(7,7,1280)	2,257,984	Pre-trained model
Global Avg Pooling	(1280)	0	Feature reduction
Dense	128	163,968	Fully connected layer
Dropout (0.5)	128	0	Regularization
Output Layer	4	516	Softmax output (4 classes)

3.4. Model Training

The model training process is conducted using Google Colab with Tesla T4 GPU acceleration to improve computational efficiency. The Adam optimizer is employed due to its adaptive learning capability, while categorical crossentropy is used as the loss function for multi-class coffee bean roast level classification. A fixed random seed of 42 is applied to ensure reproducibility. The model is trained with an initial learning rate of 1×10^{-4} , a batch size of 32, and a maximum of 30 epochs.

MobileNetV2 is initialized with ImageNet pre-trained weights and trained using a fine-tuning strategy. In the initial stage, the MobileNetV2 backbone is frozen while the classification head, consisting of Global Average Pooling, Dense, Dropout, and output layers, is trained. Subsequently, selected deeper layers are unfrozen and fine-tuned using a lower learning rate to adapt the model to the visual characteristics of different coffee bean roast levels. The resulting model contains 2,421,952 parameters, with approximately 285,000 trainable parameters during the initial training stage.

To improve training efficiency and reduce overfitting, EarlyStopping and ReduceLROnPlateau callbacks are applied. EarlyStopping stops training when validation performance no longer improves, while ReduceLROnPlateau reduces the learning rate when validation loss reaches a plateau. This configuration supports the development of a robust MobileNetV2 model for automated coffee roast level classification and quality control.

3.5. Data Splitting and Evaluation

The dataset is divided into three subsets: training, validation, and testing, using a ratio of 70:15:15. Stratified sampling is applied to maintain a consistent distribution of the four coffee bean roast levels across all subsets. The training set is used to train the MobileNetV2 model, the validation set is used to monitor and optimize model performance during training, and the testing set is reserved for final performance evaluation [21].

Model performance is evaluated using accuracy, precision, recall, and F1-score. Accuracy measures the overall proportion of correctly classified images, while precision and recall assess the classification performance for each roast level. The F1-score provides a balanced evaluation by combining precision and recall [14]. In addition, a confusion matrix is used to identify and analyze misclassification patterns among the different coffee bean roast levels [15].

3.6. Web System Implementation

The trained model is deployed as a web-based application to support practical coffee bean roast level classification. The system uses

Flask as the backend framework, with HTML, CSS, and JavaScript for the frontend interface [24]. The trained model is converted into TensorFlow.js format to enable browser-based inference without requiring high-performance server-side computation [24]. The application consists of three main modules: image upload, image preprocessing, and roast level prediction, providing a practical solution for automated coffee quality control and sorting [26].

The application is deployed on an Ubuntu Server 22.04 LTS cloud environment with 2 vCPUs and 4 GB of RAM. Flask 3.0 is served using Gunicorn with Nginx as a reverse proxy, while TensorFlow.js 4.x is used for in-browser inference. Functional testing is conducted using Google Chrome and Mozilla Firefox on desktop and mobile devices to verify the functionality and accessibility of the deployed application.

4. RESULTS AND DISCUSSION

The Results and Discussion section evaluates the performance of *MobileNetV2* for coffee bean roast level classification using evaluation metrics, confusion matrix, training analysis, web-based implementation, and comparison with previous studies.

4.1. Model Performance on Testing Data

The trained *MobileNetV2* model demonstrates strong performance on the test dataset consisting of 180 images, achieving an accuracy of 97.78%, with 176 correctly classified and 4 misclassified images. The results indicate that the combination of *MobileNetV2*, data augmentation, and the adopted training strategy provides effective classification performance across the four coffee bean roast levels.

The test dataset contains an equal number of samples for each class, with 45 images per class, minimizing the influence of class imbalance on the evaluation results. As shown in Table 5, the model achieves high precision, recall, and F1-score across all roast levels. The Dark class achieves balanced performance, while the Green class shows slightly lower recall due to a small number of misclassifications. The Light class achieves the highest recall, whereas the Medium class obtains the highest F1-score. Overall, the results demonstrate that *MobileNetV2* can effectively

distinguish different coffee bean roast levels under the experimental conditions.

However, the high test accuracy should be interpreted with caution because the 1,200-image dataset was collected from a limited range of sources and acquisition conditions. Therefore, the results may not fully represent performance on images from different regions, cameras, backgrounds, or lighting conditions. Evaluation using an independent external dataset is recommended to further assess the model's generalization capability.

Table 5. Model Evaluation Results per Class

Class	Precision	Recall	F1-Score	Support
Dark	1.00	0.98	0.99	45
Green	1.00	0.96	0.98	45
Light	0.92	1.00	0.96	45
Medium	1.00	0.98	0.99	45
Macro Average	0.98	0.98	0.98	180

Table 5 shows strong performance across all four coffee bean roast levels, with macro-average precision, recall, and F1-score of approximately 0.98. Dark and Medium achieve the highest F1-score (0.99), while Light records perfect recall (1.00) but lower precision (0.92). Overall, the model demonstrates consistent classification performance across all classes.

4.2. Confusion Matrix Analysis

The confusion matrix provides a detailed evaluation of the classification performance across the four coffee bean roast levels. Most predictions are concentrated along the main diagonal, indicating a high number of correctly classified samples. The misclassifications mainly involve the Light class, particularly samples from the Dark, Green, and Medium classes that are predicted as Light. These errors may be associated with similarities in bean appearance and variations in lighting conditions. The confusion matrix is presented in Table 6.

Table 6 shows that 176 of 180 test images were correctly classified, resulting in an accuracy of 97.78%. The Dark, Light, and Medium classes achieve 44–45 correct predictions, while the Green class records 43 correct predictions. The errors are

primarily associated with predictions toward the Light class, suggesting that visual similarities between roast levels may contribute to misclassification.

Table 6. Confusion Matrix

Actual \ Predicted	Dark	Green	Light	Medium
Dark	44	0	1	0
Green	0	43	2	0
Light	0	0	45	0
Medium	0	0	1	44

Although the model demonstrates strong performance, the results should be interpreted with caution. The classification performance may be influenced by the dataset characteristics, including similarities in background, lighting, and image acquisition conditions. Therefore, the confusion matrix demonstrates strong performance on the test set but does not necessarily guarantee generalization to unseen real-world conditions. Further evaluation using more diverse and independently collected coffee bean images is recommended to assess the model's robustness.

4.3. Learning Curve and Convergence

The learning curve shows a stable convergence of the *MobileNetV2* model during training. Training accuracy increases from 0.942 to 0.998, while validation accuracy improves from 0.928 to 0.983. At the same time, training loss decreases from 0.165 to 0.008, and validation loss decreases from 0.198 to 0.063, indicating consistent optimization throughout the training process. The learning curve and convergence results are presented in Table 7.

Table 7. Learning Curve and Model Convergence

Epoch	Train Acc.	Val. Acc.	Train Loss	Val. Loss
5	0.942	0.928	0.165	0.198
10	0.976	0.961	0.078	0.112
15	0.988	0.972	0.042	0.089
20	0.993	0.978	0.025	0.076
25	0.996	0.981	0.015	0.068
30	0.998	0.983	0.008	0.063

Table 7 indicates that both training and validation performance improve consistently as the number of epochs increases. The relatively small gap between training and validation accuracy at epoch 30 suggests no strong indication of overfitting during training. The gradual decrease in both training and validation loss further indicates stable model convergence. However, the results are based on the same dataset distribution, so further evaluation using independent data is needed to confirm generalization under different real-world conditions.

4.4. Web Application Results

The trained *MobileNetV2* model is implemented in a web-based application for coffee bean roast level classification. The system allows users to upload images and obtain prediction results through a simple web interface. As shown in Table 8, the application achieves a production accuracy of 97.8% and an average response time of 1.8 seconds, indicating efficient prediction performance. The upload success rate reaches 99.5%, demonstrating reliable image processing.

The application also supports up to 100 concurrent users with stable performance, while browser compatibility reaches 95% under the tested conditions. Overall, the results demonstrate that the integration of *MobileNetV2* into a web-based application provides an efficient and accessible solution for automated coffee bean roast level classification and quality control.

Table 8. Web Application Results

Metric	Result	Description
Production Accuracy	97.8%	Close to testing accuracy
Response Time	1.8 sec	Real-time performance
Upload Success	99.5%	High reliability
Concurrent Users	100	Stable performance
Browser Support	95%	High compatibility

Table 8 shows that the web application achieves 97.8% production accuracy, a 1.8-second response time, and a 99.5% upload success rate, indicating efficient and reliable operation. It supports up to 100 concurrent users

with 95% browser compatibility, demonstrating its practicality for automated coffee roast level classification and quality control.

4.5. Web Application Analysis

The developed web application consists of three main components: the frontend interface, backend processing module, and prediction engine. The frontend is built using HTML, CSS, and JavaScript, while Flask is used as the backend to manage image preprocessing and communication with the MobileNetV2 model. The application provides a simple and accessible interface for coffee bean roast level classification.

The system supports multiple users with an average response time of 1.8 seconds, enabling fast prediction results. Its web-based deployment allows users to access the classification system across different devices without requiring complex installation. Overall, the implementation demonstrates the practical integration of deep learning into a lightweight web application for automated coffee quality control. The dashboard and classification interfaces are presented in Figures 2 and 3.

The web application automatically classifies coffee bean roast levels from uploaded images using MobileNetV2 and displays the predicted class with its confidence score. The system supports four roast levels—*Dark*, *Green*, *Light*, and *Medium*—and provides fast and consistent classification results. In practical applications, it can support coffee bean sorting and quality control by reducing reliance on manual visual inspection. The application can also serve as a decision-support and educational tool for implementing deep learning in coffee processing.

4.6. Comparative Analysis

This section compares the proposed MobileNetV2 model with previous studies using ANN [27] and ResNet50 [28] for coffee bean image classification, as presented in Table 9. The comparison is intended to provide a general reference for evaluating the performance and practical advantages of the proposed approach. Since the studies used different datasets, preprocessing techniques, image acquisition conditions, and class definitions, the reported accuracy values should not be considered as direct benchmarks.

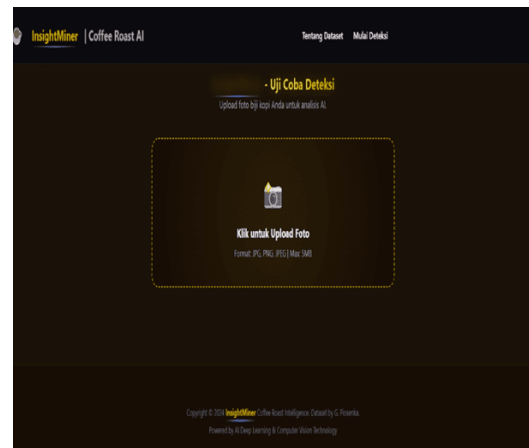


Figure 2. Dashboard Website

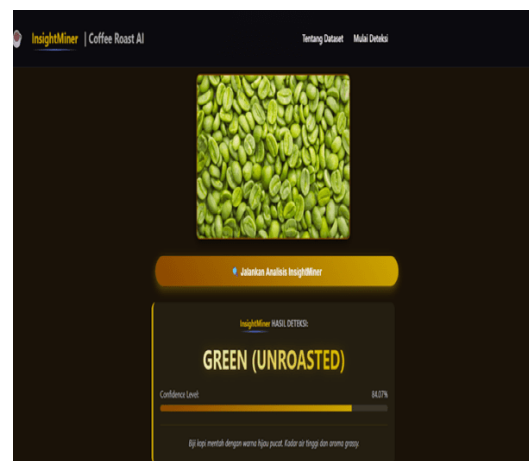


Figure 3. Classification Web Interface

Table 9. Research Comparison

Study	Method	Acc.	Adv.	Limit.
Previous	ANN [27]	97.22 %	Simple	Single variety only
Previous	ResNet50 [28]	94.50 %	Accurate	Heavy; bean-type task
This Study	MobileNetV2	97.78 %	Lightweight	Limited data

The ANN-based approach reported by Sarino et al. [27] achieved an accuracy of 97.22% on a single coffee variety, relying on manually extracted RGB features that may limit generalization to more diverse roast conditions. ResNet50, as reported by Aqsel and Rachmawanto [28] for coffee bean type identification, achieved 94.50% accuracy and provides effective feature extraction, but its deeper architecture requires greater computational resources. In comparison, the proposed MobileNetV2 achieved 97.78%

accuracy on four roast levels, offering a lightweight architecture with a good balance between classification performance and computational efficiency for web-based coffee roast level classification.

Although MobileNetV2 achieved the highest accuracy in this comparison, the results should not be considered definitive evidence of superiority because the compared studies used different datasets and experimental conditions. The limited diversity of the current dataset may also affect generalization. Future research should conduct controlled comparisons using the same dataset, preprocessing, data split, and evaluation metrics.

4.7. Error Analysis

Error analysis was conducted to examine the misclassified samples produced by the MobileNetV2 model. The analysis considers the predicted class, confidence score, and possible visual factors associated with classification errors. As shown in Table 10, the errors mainly involve the Light class, with confidence scores ranging from 51% to 55%.

Table 10. Classification Error Analysis

Original	Prediction	Confidence	Cause
Dark	Light	51%	Excessive dark lighting
Green	Light	55%	Green-light transition beans
Medium	Light	53%	Color variation within class

Table 10 shows that all analyzed errors are predicted as Light, indicating that this class has greater visual overlap with the other roast levels. The Dark-to-Light error may be influenced by lighting conditions, while the Green-to-Light error may result from similar visual characteristics. The Medium-to-Light error may be related to natural color variation within the class. The relatively low confidence scores also indicate that these samples represent borderline cases for the model.

These factors are considered possible causes based on visual inspection rather than definitive explanations. To improve robustness, future studies should use more diverse samples and acquisition conditions, together with additional augmentation such as brightness and

contrast variations. Independent external datasets are also recommended to further evaluate the model's generalization capability.

4.8. Ablation Study

The ablation study evaluates the contribution of data augmentation and dropout to the performance of the MobileNetV2 model for coffee bean roast level classification. The results are presented in Table 11.

Table 11. Ablation Study Results

Configuration	Test Accuracy	Improvement
Baseline	94.1%	-
+ Augmentation	98.0%	+3.9%
+ Dropout	98.3%	+0.3%
Final Model	98.3%	+4.2%

Table 11 shows that the baseline model achieves 94.1% accuracy. Adding data augmentation increases the accuracy to 98.0%, providing the largest improvement of 3.9%. This indicates that augmentation helps the model learn more robust visual features across different coffee bean roast levels. The addition of dropout further improves accuracy to 98.3%, with a 0.3% improvement, providing additional regularization during training.

Overall, the final configuration achieves a 4.2% improvement over the baseline. The results indicate that data augmentation contributes most substantially to model performance, while dropout provides an additional improvement. These findings support the use of both techniques in developing an effective MobileNetV2 model for coffee bean roast level classification.

5. CONCLUSION

This study applies MobileNetV2 for coffee bean roast level classification to support coffee quality control. The model achieves 97.78% accuracy on the testing dataset, with strong precision, recall, and F1-score across the four roast levels: *Dark*, *Green*, *Light*, and *Medium*. The confusion matrix confirms that most samples are correctly classified, with only a small number of misclassifications.

The learning curve shows stable training without strong indications of overfitting, while the ablation study demonstrates that data augmentation provides the largest performance

improvement, with dropout providing an additional gain. The developed web application achieves 97.8% production accuracy with an average response time of 1.8 seconds, demonstrating its practical potential for automated coffee roast level classification and quality control.

Overall, the proposed approach offers a promising balance of classification performance, computational efficiency, and practical usability. However, the limited diversity of the dataset may affect generalization under different real-world conditions. Future research should expand the dataset and evaluate the model using more diverse and independent coffee bean images to improve robustness and generalization.

REFERENCES

- [1] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.
- [2] I. M. Pakaya, R. Radi, and B. Purwantana, "Classification of roasting level of coffee beans using convolutional neural network with MobileNet architecture for Android implementation," *J. Tek. Pertanian Lampung*, vol. 13, no. 3, pp. 924–932, 2024, doi: 10.23960/jtep-l.v13i3.924-932.
- [3] R. G. R. Arsyian, A. Kurniawardhani, and I. V. Paputungan, "Classification of roasting maturity levels of coffee beans using CNN method based on MobileNetV2," *J. Res. Social Sci., Econ., Manage.*, vol. 5, no. 6, 2026, doi: 10.59141/jrssem.v5i6.1269.
- [4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.
- [7] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.
- [8] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
- [9] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1314–1324.
- [10] Z. Yang, R. O. Sinnott, J. Bailey, and Q. Ke, "A survey of automated data augmentation algorithms for deep learning-based image classification tasks," *Knowl. Inf. Syst.*, vol. 65, pp. 2805–2861, 2023, doi: 10.1007/s10115-023-01853-2.
- [11] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv preprint arXiv:1712.04621, 2017.
- [12] N. Srivastava et al., "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
- [13] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 1050–1059.
- [14] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.
- [15] A. Tharwat, "Classification assessment methods: A detailed review," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2021, doi: 10.1016/j.aci.2018.08.002.
- [16] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, 2nd ed. New York, NY, USA: Springer, 2021.
- [17] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. Cham, Switzerland: Springer, 2022.
- [18] International Coffee Organization, *Coffee Development Report 2023*. London, U.K.: International Coffee Organization, 2023.
- [19] Food and Agriculture Organization of the United Nations, *The State of Food and Agriculture 2022*. Rome, Italy: FAO, 2022.
- [20] F. Chollet, *Deep Learning with Python*, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2021.
- [21] S. Raschka, Y. Liu, and V. Mirjalili, *Machine Learning with PyTorch and Scikit-Learn*. Birmingham, U.K.: Packt Publishing, 2022.
- [22] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst.*, vol. 13, no. 4, pp. 18–28, 1998, doi: 10.1109/5254.708428.

- [23] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [24] TensorFlow, “TensorFlow.js: Machine learning for JavaScript,” TensorFlow Documentation. [Online]. Available: <https://www.tensorflow.org/js>. Accessed: Aug. 29, 2026.
- [25] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” arXiv preprint arXiv:1804.02767, 2018.
- [26] M. Abadi et al., “TensorFlow: A system for large-scale machine learning,” in *Proc. 12th USENIX Symp. Operating Systems Design and Implementation (OSDI)*, 2016, pp. 265–283.
- [27] J. N. C. Sarino, M. M. Bayas, E. R. Arboleda, E. C. Guevarra, and R. M. Dellosa, “Classification of coffee bean degree of roast using image processing and neural network,” *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 3231–3233, 2019.
- [28] A. M. Aqsel and E. H. Rachmawanto, “A roasted coffee bean identification using ResNet50 model,” *J. Appl. Informatics Comput.*, vol. 9, no. 6, pp. 3042–3056, 2025.