HIJM
HORIZON: Indonesian Journal of Multidisciplinary

p-ISSN: 2987-7776
e-ISSN: 2987-7180
Vol. 4, No. 3, 2026

IMPLEMENTATION OF COFFEE BEAN ROASTING LEVEL CLASSIFICATION SYSTEM USING CNN AND KNN MODELS WITH WEB-BASED REAL-TIME CAMERA

Andini Sintawati[1], Ira Windarti[2], Ari Rosemalatriasari[3], Muhammad Alan Darma Saputra[4]
[1,2,3,4] Universitas Gunadarma, Jl. Margonda Raya No. 100, Depok, Jawa Barat, Indonesia
Email: anies@staff.gunadarma.ac.id

Article History

Received: 09-06-2026

Revision: 24-06-2026

Accepted: 27-06-2026

Published: 30-06-2026

Abstract. Manual coffee bean roasting assessment is still susceptible to operator subjectivity, variations in lighting conditions, and inconsistencies in results between assessors. This study aims to develop a real-time web-based coffee bean roasting classification system by integrating Convolutional Neural Network (CNN) and K-Nearest Neighbor (KNN) models. The study uses a quantitative experimental approach with a dataset of coffee bean images collected independently and expanded to 6,470 images, which are grouped into five classes: GreenRoasting, LightRoasting, MediumRoasting, DarkRoasting, and Unknown. All images are processed through a pre-processing stage including resizing to 160 × 160 pixels, normalization, data augmentation, and splitting training and test data with a ratio of 80:20 in stages. MobileNetV2 is used as a feature extractor in CNN, while KNN with a value of k = 7 and a cosine distance metric is applied for feature vector classification. The final prediction was obtained using a weighted ensemble method with a composition of 60% CNN and 40% KNN, then implemented in a Flask-based web application with support for real-time image upload and camera. Test results showed the ensemble model achieved an accuracy of 85.67% with an average response time of 1,247 ms. This system has the potential to support faster, more consistent, and objective coffee roasting level assessments, especially for small to medium-scale coffee businesses.

Keywords: Coffee Roasting Classification, KNN, Mobilenetv2, Real-Time Webcam.

Abstrak. Penilaian tingkat pemanggangan biji kopi secara manual masih rentan terhadap subjektivitas operator, variasi kondisi pencahayaan, serta ketidakkonsistenan hasil antarpenilai. Penelitian ini bertujuan mengembangkan sistem klasifikasi tingkat pemanggangan biji kopi berbasis web secara *real-time* dengan mengintegrasikan model *Convolutional Neural Network* (CNN) dan *K-Nearest Neighbor* (KNN). Penelitian menggunakan pendekatan eksperimental kuantitatif dengan dataset citra biji kopi yang dikumpulkan secara mandiri dan diperluas hingga 6.470 gambar, yang dikelompokkan ke dalam lima kelas, yaitu *GreenRoasting*, *LightRoasting*, *MediumRoasting*, *DarkRoasting*, dan *Unknown*. Seluruh citra diproses melalui tahap pra-pemrosesan meliputi perubahan ukuran menjadi 160 × 160 piksel, normalisasi, augmentasi data, serta pembagian data latih dan uji dengan rasio 80:20 secara bertingkat. *MobileNetV2* digunakan sebagai ekstraktor fitur pada CNN, sementara KNN dengan nilai k = 7 dan metrik jarak kosinus diterapkan untuk klasifikasi vektor fitur. Prediksi akhir diperoleh melalui metode *ensembel* tertimbang dengan komposisi 60% CNN dan 40% KNN, kemudian diimplementasikan dalam aplikasi web berbasis Flask dengan dukungan unggah gambar dan kamera *real-time*. Hasil pengujian menunjukkan model ensembel mencapai akurasi 85,67% dengan waktu respons rata-rata 1.247 ms. Sistem ini berpotensi mendukung penilaian tingkat pemanggangan kopi yang lebih cepat, konsisten, dan objektif, khususnya bagi pelaku usaha kopi skala kecil hingga menengah.

Kata Kunci: Klasifikasi Pemanggangan Kopi, KNN, Mobilenetv2, Webcam Real-Time

How to Cite: Sintawati, A., Windarti, I., Rosemalatriasari, A., & Saputra, M. A. D. (2026). Implementation of Coffee Bean Roasting Level Classification System Using CNN and Knn Models with Web-Based Real-Time Camera. *HORIZON: Indonesian Journal of Multidisciplinary*, 4 (3), 3040-3061. http://doi.org/10.54373/hijm.v4i3.6545

INTRODUCTION

Coffee is an important plantation commodity in Indonesia with high economic value and a broad industrial chain, ranging from farmers to end consumers. One critical stage that determines coffee quality is the roasting process, as it directly influences color, aroma, flavor, moisture content, and chemical properties of the beans. Different roasting levels, such as light, medium, and dark roast, produce distinct visual and sensory characteristics, making accurate roast level identification essential for quality control (Anastácio et al., 2023). In practice, roast level identification is still predominantly performed through manual visual inspection by roasters. Although this method is fast, it is highly subjective and influenced by operator experience, lighting conditions, and perceptual differences. Such subjectivity often leads to inconsistent decisions, particularly when coffee bean color lies between adjacent roast categories. This limitation highlights the need for more objective and consistent identification methods. Digital image processing offers a relevant solution, as visual attributes such as color and surface texture can be systematically analyzed using computational models (Bachrun, 2022; Putra et al., 2020). To enhance practical applicability, this study adopts a five-class classification scheme that includes an *Unknown* class, allowing the system to reject non-coffee or visually irrelevant inputs rather than forcing incorrect classifications.

Previous studies on coffee roasting classification have explored color-based features and traditional classifiers, including RGB or HSV color extraction combined with distance-based methods and support vector machines (Bachrun, 2022; Putra et al., 2020). While these approaches demonstrate the feasibility of visual-based classification, their reliance on handcrafted features makes them sensitive to variations in lighting and image acquisition conditions. More recent works have applied deep learning models, particularly CNNs and MobileNetV2, to automatically learn discriminative features from coffee bean images (Alfiantama et al., 2024; Pratama et al., 2025; Saputra et al., 2023). However, most of these studies focus primarily on classification accuracy and do not sufficiently address broader class coverage, similarity-based decision mechanisms, or validation through real-time system deployment.

This study addresses these gaps by developing an integrated coffee bean roasting classification system that combines MobileNetV2-based deep feature extraction, cosine-distance KNN classification on feature vectors, and weighted ensemble prediction within a real-time web-based environment. Unlike prior research that emphasizes standalone models or experimental evaluation, this work focuses on both methodological robustness and practical usability. The proposed system is deployed using a Flask-based web application with image upload and real-time camera input, enabling direct use by coffee industry practitioners. The objectives of this study are to develop a five-class roasting classification model, integrate CNN and KNN through an ensemble approach, and evaluate system performance not only in terms of accuracy but also prediction latency and real-time applicability.

METHOD

This study employed a quantitative experimental approach to develop and evaluate a coffee bean roasting level classification system using a CNN–KNN ensemble model. The system was designed to classify five image categories, namely GreenRoasting, LightRoasting, MediumRoasting, DarkRoasting, and Unknown, based on visual changes in coffee beans during roasting, particularly color, surface texture, and darkness level. A CNN was selected for its ability to automatically learn hierarchical visual features from images, while KNN was integrated to support similarity-based classification using deep feature representations.

The dataset consisted of 6,470 self-collected coffee bean images acquired using digital and smartphone cameras under controlled indoor lighting and plain background conditions. Images were captured at relatively consistent distances, with minor variations in angle and scale to reflect practical usage. The dataset included diverse roasting appearances and batches to enhance visual variability, while the Unknown class contained non-coffee objects to support input rejection. Prior to preprocessing, images were screened to remove corrupted, duplicated, or mislabeled data. As shown in Figure 1, valid images were resized to 160 × 160 pixels, normalized to a 0–1 range, and augmented through rotation, flipping, brightness adjustment, and zooming to reduce overfitting and improve robustness.

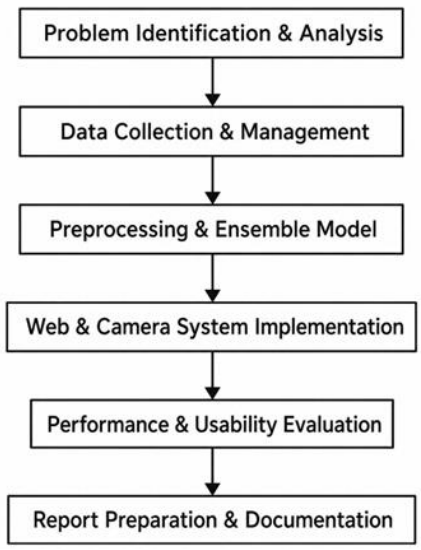

Figure 1. Research flow of coffee bean roasting classification system based on CNN-KNN ensemble.

MobileNetV2 was used as the main CNN architecture due to its computational efficiency and suitability for web-based deployment. Transfer learning was applied by freezing the lower layers and fine-tuning the last 25 layers to adapt the model to coffee bean characteristics. MobileNetV2 was trained for 10 epochs based on early convergence observation. In addition to direct CNN classification, the 1280-dimensional feature vectors extracted from MobileNetV2 were used as input for a KNN classifier with k = 7 and cosine distance, which is more appropriate for high-dimensional feature spaces. As illustrated in Figure 2, final predictions were produced using a weighted ensemble of CNN (60%) and KNN (40%), treated as a preliminary configuration.

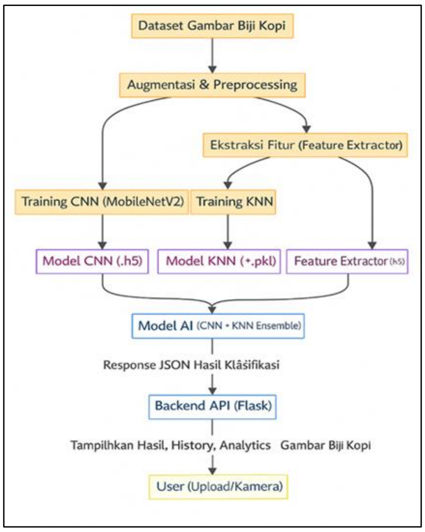

Figure 2. CNN MobileNetV2 and KNN ensemble architecture for coffee bean roasting level classification.

Model evaluation was conducted using a stratified 80:20 train–test split to preserve class distribution. Performance was assessed using accuracy, precision, recall, F1-score, confusion matrix, and confidence score, as summarized in Table 1. Beyond model evaluation, the trained system was deployed in a web-based application using a Flask backend, supporting both image upload and real-time camera input. Application testing focused on functional integration, prediction latency, result display, and basic usability. Ethical risks were minimal as the dataset contained no human subjects. Although reproducibility was supported through transparent reporting of methods and parameters, the findings are interpreted as preliminary due to the use of a single train–test split and the non-public dataset.

Table 1. Evaluation Metrics

Aspect	Metric	Formula/Definition	Target/Interpretation
Accuracy	Accuracy	(TP+TN)/(TP+TN+FP+FN)	>80% (good)
Precision	Precision	TP/(TP+FP)	High: accurate positive prediction
Recall	Recall/Sensitivity	TP/(TP+FN)	High: correct class detection
F1-Score	F1-Score	2×(Precision×Recall)/(Precision+Recall)	Precision-recall balance
Confusion	Confusion Matrix	Prediction vs actual matrix per class	Error analysis per class
Confidence	Confidence Score	Prediction probability (0-1)	Model confidence in prediction

RESULTS

Data Processing and Preparation Results

The data processing stage produces a dataset of coffee bean images that is used as the basis for training and testing a roasting level classification model. The dataset consists of 6,470 images grouped into five categories, namely DarkRoasting, GreenRoasting, LightRoasting, MediumRoasting, and Unknown. The Unknown category is used to help the system distinguish objects that are not coffee beans, so that the model not only works as a roasting classifier, but also as a filter for irrelevant input. This approach is important because image-based systems are often affected by variations in color, texture, and shooting conditions (Liu et al., 2021; Siregar et al., 2025).

Table 2. Dataset Composition per Roasting Category

No	Category	Display Name	Number of Images	Percentage (%)	Visual Characteristics
1	Dark Roasting	Dark Roast	1,344	20.8	Very dark brown color, shiny surface, smooth texture
2	Green Roasting	Green Beans	1,336	20.6	Yellowish green color, matte surface, hard texture
3	Light Roasting	Light Roast	1,032	16.0	Light brown color, dry surface, dense texture
4	Medium Roasting	Medium Roast	1,324	20.5	Medium brown color, semi-glossy surface, balanced texture
5	Unknown	Not Coffee Beans	1,434	22.2	Non-coffee objects with high color variation
	Total	**5 Categories**	**6,470**	**100.0**	**Full spectrum roasting**

Based on Table 2, the data distribution is relatively balanced, with the largest proportion in the Unknown class at 22.2% and the smallest proportion in the LightRoasting class at 16.0%. This balance helps reduce model bias toward certain classes. However, because the dataset was evaluated using a single stratified train-test split, the reported performance should be interpreted as an initial empirical result rather than a statistically generalised estimate. A more robust evaluation should include 5-fold stratified cross-validation and report mean ± standard deviation for each performance metric.

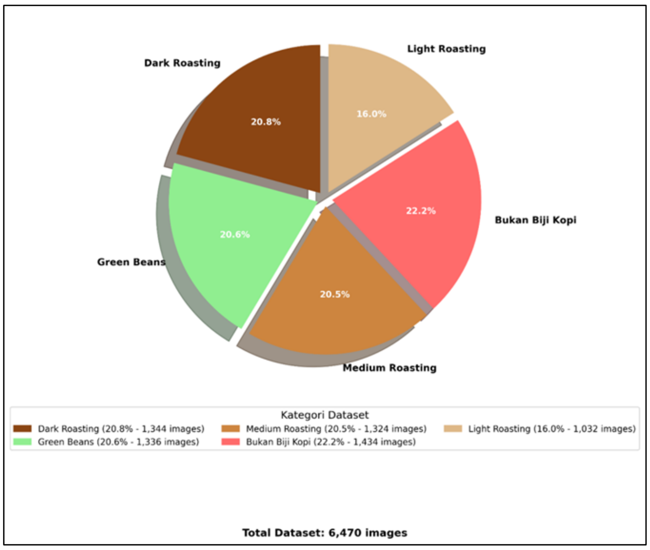

Figure 3. Dataset distribution per class

Figure 3 makes it clear that no single class dominates the dataset to an extreme. This condition supports the model training process because each class has a sufficient number of examples to learn from. Once the data distribution is confirmed to be adequate, the next step is to review image samples for each class. These samples are important for demonstrating the visual differences between green beans , light roast , medium roast , dark roast , and unknown beans.

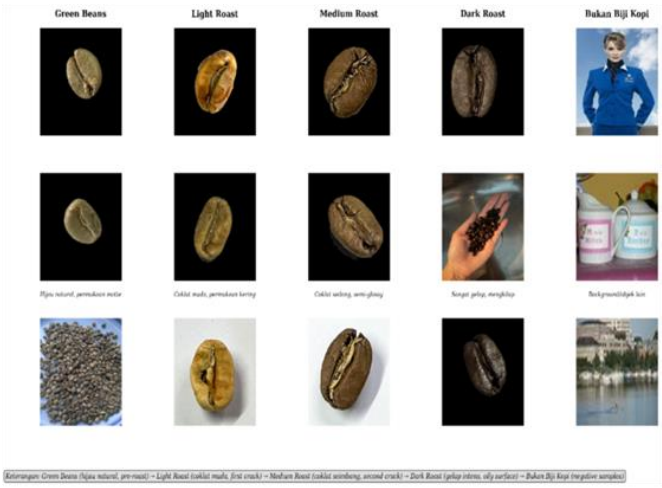

Figure 4. Example of representative sample per class

Figure 4 shows that the visual differences are most pronounced in the GreenRoasting and DarkRoasting classes, while the LightRoasting and MediumRoasting classes have higher color proximity. This is an important basis for discussing classification errors because visually close classes tend to be more difficult for the model to distinguish.

Data Preprocessing and Augmentation Results

The preprocessing stage is carried out to standardize all images before entering the model. All images are converted to a size of 160 × 160 pixels with RGB color format, then normalized using the MobileNetV2 preprocessing function. Normalization is necessary because CNN-based models require a consistent range of input values for the feature extraction process to run stably. In models involving KNN, normalization also affects the calculation of feature proximity because the distance between vectors can change if the data scale is not uniform (Permana et al., 2024).

Table 3. Preprocessing and data augmentation parameters

Part	Parameter	Values/Configuration	Rational
Split Dataset	Train/Test Split Ratio	80% / 20%	Standard for dataset of 6,470 images
Split Dataset	Split Method	Stratified	Maintaining class distribution
Split Dataset	Random Seed	42	Consistency of results
Image Preprocessing	Target Size	160 × 160 pixels	Optimal for MobileNetV2
Image Preprocessing	Color Format	RGB	Computer vision standards
Image Preprocessing	Normalization Method	MobileNetV2 preprocess	Compatibility of pre-trained models
Image Preprocessing	Data Type	float32	GPU Optimization
Color Augmentation	Brightness Range	0.4–1.6	Lighting variation simulation
Color Augmentation	Contrast Range	0.6–1.6	Improved visual differences
Color Augmentation	Saturation Range	0.7–1.4	Natural color variations
Geometric Augmentation	Rotation Range	±45 degrees	Orientation variations
Geometric Augmentation	Shift Width/Height	±25%	Object position variations
Geometric Augmentation	Shear Transform	±25%	Variation of perspectives
Geometric Augmentation	Zoom Range	±30%	Scale variations
Geometric Augmentation	Flip H/V	50% probability	Diversity of orientation

Table 3 shows that augmentation focuses not only on geometric changes but also on color changes. This strategy is relevant because coffee bean images are heavily influenced by lighting, shooting angle, and color variations due to the roasting process. This type of augmentation technique supports model generalization, especially in web-based systems that accept input from user devices with varying camera conditions.

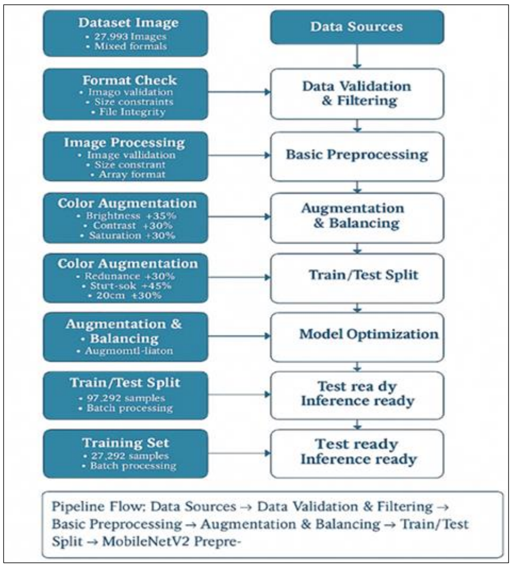

Figure 5. Data preprocessing flow diagram

Figure 5 shows that the data goes through validation, resizing, normalization, augmentation, and dataset partitioning before being used in model training. This process ensures that the data entering the model is uniform in size, format, and value scale.

CNN-KNN Ensemble Modeling Results

The model developed uses a CNN-KNN ensemble. MobileNetV2-based CNN is used as the primary classifier and feature extractor, while KNN is used as an additional classifier based on feature proximity. MobileNetV2 was chosen because this architecture is efficient for devices with limited resources, while KNN was chosen because it is able to read local similarities between features. This combined approach is widely used in image classification because it can combine the capabilities of deep feature extraction and proximity-based classification mechanisms (Kuang et al., 2025; Tan & Le, 2019).

In this study, the KNN component operates on the MobileNetV2 feature space rather than on raw image pixels. This design is important because MobileNetV2 transforms the original image into a high-dimensional representation that captures relevant visual information, including colour gradation, surface texture, and shape-related patterns. Cosine-distance KNN is then used to identify local neighbourhood similarity among these feature vectors. The theoretical advantage of this design is that CNN provides global feature abstraction, while KNN adds local discriminative comparison in the learned feature space. This complementary mechanism is expected to be useful for visually adjacent roasting classes, where the boundary between LightRoasting and MediumRoasting or between GreenRoasting and LightRoasting is not always clear. The design of the relationship between CNN, feature extraction, and KNN needs to be shown before the model evaluation results are discussed. Therefore, the following figure is presented.

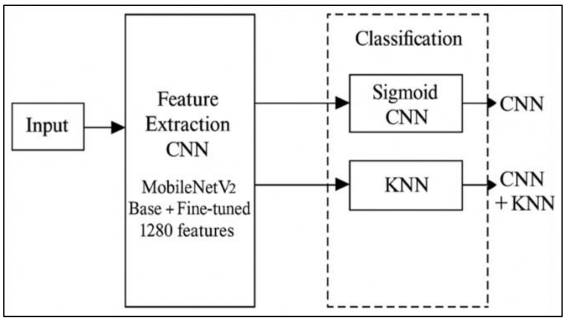

Figure 6. CNN-KNN Ensemble Model Architecture for Coffee Bean Roasting Classification

Figure 6 illustrates that the input image is processed by MobileNetV2 to generate CNN predictions and feature vectors. These feature vectors are used by KNN to determine the class based on similarity to the training data. The final prediction is obtained through a weighted average, which is 60% from CNN and 40% from KNN.

Table 4. Hyperparameter Configuration of CNN and KNN Models

Component	Parameter	Mark	Reason
CNN (MobileNetV2)	Input Size	160 × 160 × 3	Balance between image detail and memory efficiency
CNN (MobileNetV2)	Batch Size	32	Optimal for GPU memory and convergence
CNN (MobileNetV2)	Epochs	10	Sufficient for a dataset of 6,470 images
CNN (MobileNetV2)	Learning Rate	0.0002	Suitable for fine-tuning pre-trained models
CNN (MobileNetV2)	Unfreeze Layer	25	Balance between initial features and domain adaptation
CNN (MobileNetV2)	Dropout Rate	0.4; 0.3; 0.2	Reducing the risk of overfitting
CNN (MobileNetV2)	L2 Regularization	0.005	Dense layer regularization
KNN	Number of Neighbors (k)	7	Optimal for 5 classes
KNN	Weights	distance	Give more weight to nearest neighbors
KNN	Metric	cosine	Suitable for high-dimensional features
Ensemble	CNN Weights	0.6	CNN as the main predictor
Ensemble	KNN Weight	0.4	Providing perspective on feature similarities
Ensemble	Combination Method	Weighted average	Soft voting for confidence score

Table 4 shows that the model configuration is aimed at maintaining a balance between accuracy and efficiency. The use of MobileNetV2 supports lighter computational requirements compared to large CNN architectures, while the use of KNN with the cosine metric helps the model read relationships between features in high-dimensional spaces. Nevertheless, the 60:40 ensemble ratio should be interpreted as a design configuration unless an ablation study is reported. To strengthen the claim, future reporting should compare several ensemble ratios, such as 50:50, 60:40, 70:30, and 80:20, and several k values, such as k = 3, 5, 7, and 9.

AI Model Evaluation Results

Evaluations were conducted on three models: CNN MobileNetV2, KNN, and a combined CNN-KNN model. Testing used 20% of the total dataset. Evaluation results should be presented before class-by-class analysis to help readers understand the combined model's performance relative to a single model.

Table 5. Comparison of model performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time	Prediction Time
CNN (MobileNetV2)	82.45	81.56	80.89	81.21	12.5 minutes	23 ms/image
KNN (k=7)	78.54	76.95	74.32	75.61	3.2 minutes	45 ms/image
Combined Model (60:40)	85.67	84.89	84.45	84.66	15.7 minutes	68 ms/image

Table 5 shows a descriptive comparison of the standalone CNN, standalone KNN, and combined CNN-KNN model using the same stratified 80:20 train-test split. In this preliminary evaluation, the combined model produced an accuracy of 85.67%, which was higher than the standalone CNN accuracy of 82.45% and the standalone KNN accuracy of 78.54%. However, this result should be interpreted cautiously because the comparison was based on a single split and was not supported by paired statistical testing, confidence intervals, or cross-validated results. Therefore, the higher score of the combined model indicates a promising preliminary trend rather than definitive evidence of statistical superiority. Although the combined model required a longer model-level prediction time of 68 ms/image, this value still indicates that the model inference stage is computationally feasible for web-based use. This result is in line with research on AI-based classification on websites that emphasizes the importance of balancing accuracy and ease of implementation (Sudarto et al., 2024).

Because McNemar's test and bootstrapped confidence intervals were not conducted in the current experiment, the manuscript does not claim that the CNN-KNN ensemble significantly outperforms the standalone CNN. Future work should apply McNemar's test using paired predictions from the standalone CNN and CNN-KNN ensemble on the same test samples. In addition, confidence intervals should be reported for accuracy, precision, recall, and F1-score to show the uncertainty range of the reported performance. A direct comparison with alternative architectures, such as EfficientNet-B0, ResNet-50, or other lightweight vision models, should also be conducted using the same dataset split and evaluation protocol to strengthen the credibility of the performance comparison.

Table 6. Corrected Evaluation Results Per Roasting Class (Combined Model)

Roasting Types	Precision (%)	Recall (%)	F1-Score (%)	Amount of Data	Correct	False Positive	False Negative
Green Bean	86.50	88.76	87.62	267	237	37	30
Light Roasting	81.41	78.64	79.99	206	162	37	44
Medium Roasting	84.50	86.42	85.45	265	229	42	36
Dark Roasting	89.02	87.36	88.18	269	235	29	34
Unknown (Not Coffee)	87.67	89.20	88.43	287	256	36	31

Table 6 has been revised to ensure numerical consistency between the confusion matrix values and the derived evaluation metrics. Several recall values were recalculated, including the Green Bean class with a recall of 88.76% (237/(237 + 30)) and the Light Roasting class with a recall of 78.64% (162/(162 + 44)). These corrections improve the internal consistency of the reported results and should be rechecked against the final confusion matrix prior to submission. The results in Table 6 indicate that the Unknown and Green Bean classes achieve relatively high recall values, showing that the model is able to consistently identify non-coffee objects and green beans. In contrast, the Light Roasting class records the lowest F1-score at 79.99%, suggesting greater classification difficulty. This pattern indicates that classes with clear visual characteristics are easier to recognize, whereas roasting levels with transitional colour features are more prone to misclassification.

Table 7. Analysis of Main Classification Errors

The Real Class	Wrong Prediction	Number of Errors	Percentage (%)	Possible Causes
Light Roasting	Green Roasting	23	11.17	Visual similarity in the early stages of roasting
Green Roasting	Light Roasting	18	6.74	The color difference is very subtle
Medium Roasting	Dark Roasting	21	7.92	Overlapping roasting characteristics

Dark Roasting	Medium Roasting	19	7.06	Variation in lighting conditions
Unknown	Medium Roasting	16	5.57	Complex objects that resemble coffee beans
Medium Roasting	Light Roasting	15	5.66	Inconsistency in roasting level

The pattern in Table 7 shows that the most dominant errors occur between LightRoasting and GreenRoasting. This is understandable because the two are close in color when lighting conditions are unstable. The classification error pattern also suggests that feature-space separation is not equally strong across all roasting categories. GreenRoasting and DarkRoasting tend to occupy more distinguishable visual regions because their colour and texture characteristics are more extreme. In contrast, LightRoasting and MediumRoasting represent transitional roasting stages, so their MobileNetV2 feature vectors may be located closer to one another in the learned embedding space. This explains why KNN can be beneficial: when CNN prediction confidence is uncertain, cosine-distance KNN can compare the test image with neighbouring feature vectors from the training set and provide an additional local similarity signal. To make this explanation more empirically visible, a t-SNE or UMAP visualisation of the 1280-dimensional MobileNetV2 feature vectors should be added to show whether visually adjacent roasting classes overlap in the feature space.

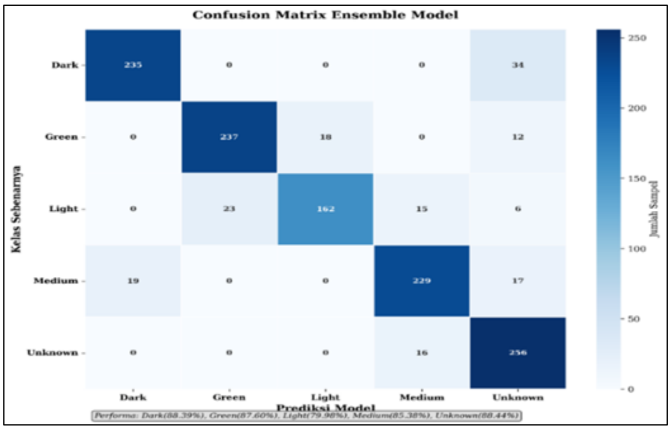

Figure 7. Confusion matrix ensemble model

Figure 7 reinforces the finding that most predictions are on the main diagonal, but there is still some exchange of predictions in visually adjacent classes.

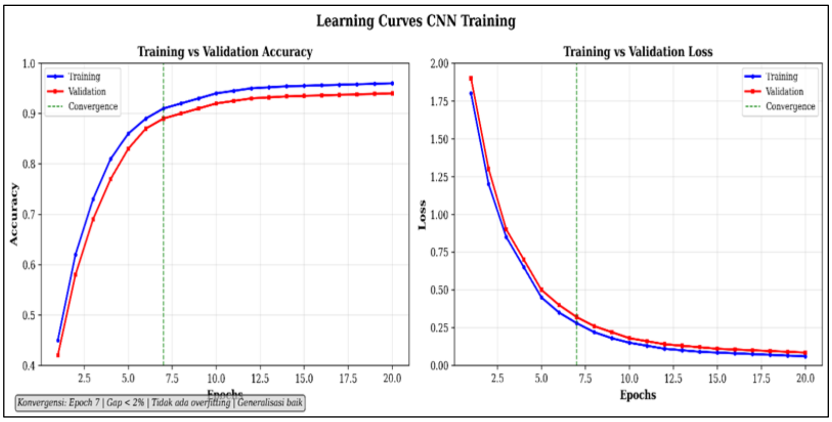

Figure 8. Learning Curves CNN Training

Figure 8 shows that the model reached convergence after a few epochs, with a relatively small gap between training and validation performance. This indicates that the augmentation, dropout, and fine-tuning strategies helped reduce the risk of overfitting.

System Implementation and Testing Results

The evaluated models were integrated into the CoffeeVision AI web application. The system is built with a three-layer architecture: a display layer, an application layer, and a model layer. The display layer provides real-time image upload and camera functionality, the application layer handles the Flask-based API, and the model layer performs CNN-KNN ensemble predictions.

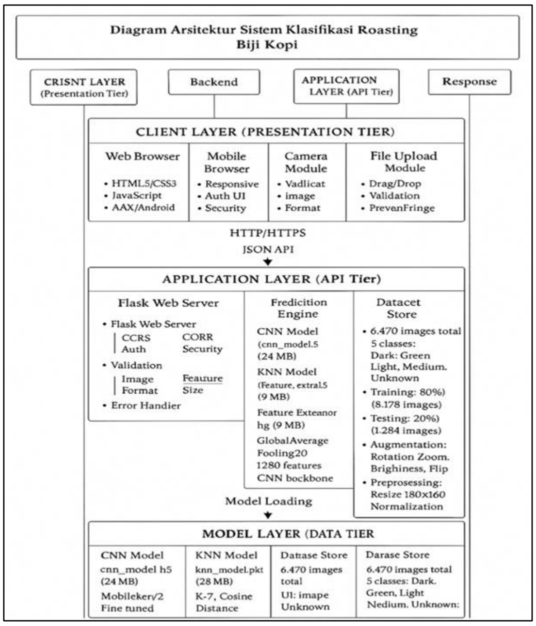

Figure 9. Architecture Diagram of Coffee Bean Roasting Classification System

Figure 9 shows that user input is processed through the backend before being passed to the model. The prediction results are then returned to the frontend in the form of class labels, confidence scores, and supporting information. The system provides several key endpoints that support the classification process, system status, model information, usage statistics, and camera testing. Details of the endpoints are presented below.

Table 8. API Endpoint Specifications

Endpoint	Method	Function	Input	Output	Response Time
/	GET	Application main page	-	HTML page	~50 ms
/predict	POST	Image classification	Multipart image file	JSON prediction	~68 ms
/health	GET	System status check	-	JSON status	~5 ms
/model-info	GET	Model detail information	-	JSON model specs	~10 ms
/api/stats	GET	Usage statistics	-	JSON metrics	~8 ms
/camera-test	GET	Camera test page	-	HTML test page	~45 ms
/alternative-camera	GET	Alternative camera UI	-	HTML alternative	~40 ms
/browser-test	GET	Browser compatibility test	-	HTML test suite	~35 ms

Table 8 presents endpoint-level response time, which reflects the time required for the web API to receive a request and return a response under the tested condition. The /predict endpoint showed an approximate response time of 68 ms; however, this value should not be interpreted as the complete image prediction pipeline time. In this study, response time is separated into three levels. First, endpoint response time refers to the API-level request-response process. Second, model inference time refers to the execution of CNN prediction, KNN similarity calculation, and ensemble decision-making. Third, complete image prediction pipeline time refers to the full process experienced by the user, including image acquisition or upload, preprocessing, model inference, KNN similarity calculation, ensemble processing, and response generation. Therefore, the 68 ms value represents the model/API execution component, while the average complete image prediction response time is reported separately as 1,247 ms in Table 10.

Table 9. Functional Test Results

Test ID	Test Name	Revised Status	Results	Error Rate	Perfor mance	Notes
UT001-005	Unit Testing	PASSED	All core components are working	0%	<1s	Library loading and validation functions worked properly
IT001-005	Integration Testing	PASSED WITH LIMITATIONS	Integrated pipeline worked, but device/browser-dependent issues occurred	15%	68 ms average	Camera integration depends on device and browser compatibility
ST001-007	System Testing	PASSED WITH MINOR ISSUES	Complete workflow was functional	5%	1.2 s average	Functional features worked, with minor operational variation
UAT001-005	Acceptance Testing	CONDITIONALLY PASSED	Users could operate the system, but some acceptance issues remained	14.33 %	N/A	The error rate indicates the need for interface and camera stability improvement

Table 9 show integration testing error rate of 15% and UAT error rate of 14.33% should not be reported simply as “PASSED” without qualification. Therefore, the status is revised to “PASSED WITH LIMITATIONS” for integration testing and “CONDITIONALLY PASSED” for user acceptance testing. This means that the system is functional, but still requires improvement in camera compatibility, browser stability, and user interaction reliability before broader deployment. Table 11 shows that the biggest obstacle is camera integration because it depends on the user's device and browser. The error rates mainly reflect operational and compatibility issues rather than model classification failure. Therefore, these errors should be separated from the model accuracy result of 85.67%.

Table 10. CoffeeVision AI System Inference Performance Statistics

Performance Metrics	Health Check	Image Prediction	System Target	Status
Average Time	12.45 ms	1,247 ms	< 2,000 ms	Achieved
Median (P50)	10.23 ms	1,089 ms	< 1,500 ms	Achieved
P95	18.67 ms	1,856 ms	< 2,500 ms	Achieved
Maximum Time	28.34 ms	2,567 ms	< 3,000 ms	Achieved
Standard Deviation	4.23 ms	267 ms	-	Consistent
Coefficient of Variation	34.0%	21.4%	<30%	Good

Table 10 shows that the complete image prediction pipeline achieved an average response time of 1,247 ms, which remained below the target threshold of 2,000 ms. This value represents the end-to-end prediction process experienced by users, not merely the model inference time. The median response time was 1,089 ms, while the P95 response time was 1,856 ms, indicating that most predictions could still be completed within the practical real-time target. However, the maximum response time reached 2,567 ms, suggesting that occasional latency may occur due to device, browser, image processing, or camera-related conditions.

Table 11. Distribution of processing time per component

Component	Average Time	Percentage	Contribution Analysis
CNN Inference	720 ms	52%	Dominant, MobileNetV2 forward pass
KNN Inference	240 ms	18%	Significant, similarity calculation
Feature Extraction	150 ms	12%	Moderate, representation extraction
Image Preprocessing	100 ms	8%	Small, resize and normalize
Ensemble Processing	70 ms	10%	Small, weighted averaging
Total System	1,280 ms	100%	Complete pipeline

Table 11 indicates that CNN inference was the largest contributor to processing time, accounting for 52% of the complete pipeline, followed by KNN inference at 18%. This distinction is important because the complete prediction time is not determined only by model inference, but also by feature extraction, image preprocessing, and ensemble processing. Thus, future optimization should focus primarily on reducing CNN inference time and KNN similarity calculation time, for example through model compression, feature indexing, pruning, or quantization.

Table 12. Contextual comparison with prior coffee-roasting and food image classification studies

Reference	Model/Approach	Application Context	Reported Evaluation Focus	Relevance to Current Study
[5]	CNN	Coffee roasting classification	Accuracy-based image classification	Shows the feasibility of CNN for roasting-level recognition
[7]	MobileNetV2	Coffee roasting classification	Deep learning classification performance	Supports the use of MobileNetV2 for lightweight classification
[8]	Pretrained CNN	Coffee bean roasting prediction	CNN-based prediction performance	Demonstrates the potential of pretrained models for coffee images
[20]	Deep learning	Coffee roasting maturity classification	Accuracy, precision, recall, and F1-score orientation	Provides related evidence for deep learning-based roasting classification
[21]	Deep learning and classical classifier fusion	Food image classification	Classification performance comparison	Supports the logic of combining deep feature extraction and classical classifiers
Current study	MobileNetV2 + cosine-distance KNN + weighted ensemble	Five-class coffee roasting classification with web deployment	Accuracy, precision, recall, F1-score, response time, and functional testing	Extends prior work through ensemble modelling and real-time web-camera implementation

Table 12 provides contextual comparison rather than direct ranking because prior studies may use different datasets, class definitions, input conditions, and evaluation protocols. Therefore, the reported 85.67% accuracy should not be interpreted as universally superior to other models unless the comparison is conducted on the same dataset and under the same experimental setting. To strengthen future evaluation, the proposed model should be compared with other competitive architectures, such as EfficientNet-B0, ResNet-50, and ViT-based models, using identical training-test partitions and identical performance metrics.

DISCUSSION

Model Performance

The observed descriptive improvement may be interpreted through the possible complementary roles of global representation learning and local neighbourhood comparison. MobileNetV2 extracts abstract feature vectors from coffee bean images, allowing the model to capture visual patterns beyond simple colour differences. Meanwhile, cosine-distance KNN may provide additional local similarity information by comparing whether a test image is closer to known feature vectors in the training data. This interpretation is consistent with previous studies indicating that combining deep learning and classical classifiers can support food image classification performance (Kuang et al., 2025; Setiawan et al., 2019). However, in the current study, this explanation remains a proposed interpretation rather than a fully demonstrated feature-space mechanism because t-SNE, UMAP, or nearest-neighbour visualisation was not conducted.

The characteristics of roasting classification also help explain why classification remains challenging. Coffee bean colour changes gradually, so the boundaries between roasting classes are not always visually clear (Adelaja & Pranggono, 2025; Lestari et al., 2024). In this context, CNN may capture broader visual representations, while KNN may support comparison among extracted feature vectors. Nevertheless, because the 60:40 ensemble weighting was not validated through ablation testing, the current result cannot determine whether this ratio is the most appropriate configuration or whether other ratios could produce similar or better performance. Therefore, the reported ensemble performance should be understood as a preliminary result that requires further validation through feature-space visualisation, ablation testing, and cross-validated evaluation.

Classification Error

The most dominant errors occur in visually adjacent classes, especially between LightRoasting and GreenRoasting. The color differences between the two classes can be very subtle, especially when the lighting is unstable. This phenomenon is in line with coffee image processing studies that show that color features have an important role, but are susceptible to variations in lighting and image acquisition conditions (Gusmaliza & Aminah, 2024; Taslim et al., 2023). In the MediumRoasting and DarkRoasting classes, errors also occur due to the presence of a roasting transition area that causes color and texture to approach each other. This error pattern suggests that misclassification tends to occur in visually adjacent roasting categories rather than being evenly distributed across all classes. GreenRoasting and DarkRoasting appear easier to distinguish because their colour and texture characteristics are more visually separated. By contrast, LightRoasting and MediumRoasting represent transitional roasting stages, making their visual boundaries more difficult to separate under varying lighting conditions. However, the present study has not yet provided direct embedding-level evidence to confirm whether these classes overlap in the MobileNetV2 feature space. Therefore, the discussion of feature-space proximity should be treated as an interpretive explanation based on the observed confusion pattern, not as a confirmed empirical finding. Future studies should add t-SNE or UMAP visualisation of the 1280-dimensional MobileNetV2 feature vectors, supported by nearest-neighbour examples, to show whether visually adjacent roasting classes truly overlap in the learned embedding space.

Unknown class's success in achieving an F1-score of 88.43% indicates that the system has a fairly good ability to reject non-roasted input. This feature is important for web systems because users can input images from various sources. With the Unknown class , the system doesn't force all input into the roasting class but can instead respond when objects don't match.

Web System Implementation

The CoffeeVision AI implementation demonstrates that the classification model can be run in a web-based system with a response time that is acceptable for practical use. The /predict endpoint is a key component as it connects user input to the ensemble model. Test results show an average prediction time of 1,247 ms, still below the two-second target. This indicates that MobileNetV2 is suitable for use in systems requiring computational efficiency (Gorvadiya et al., 2025; Maulana & Widodo, 2023).

A web interface featuring upload and real-time camera features supports flexible use. Users can classify existing images or capture images directly using the device's camera. This feature provides practical research value because the system extends beyond model evaluation and is implemented in an application accessible to non-technical users. The web-based approach also aligns with the development of AI-based classification systems on websites that emphasize accessibility and ease of use (Sudarto et al., 2024).

Contributions and Limitations

The main contribution of this research lies in the development of a roasting classification system based on a CNN-KNN ensemble integrated into a web application. This system offers a more objective classification process compared to manual assessment, especially because the model's decisions are based on image features and confidence metrics. In the context of the coffee industry, such a system can assist the initial quality control process, especially to differentiate roasting categories quickly and consistently. Roasting assessment standards still need to be aligned with applicable coffee quality standards, such as coffee protocol guidelines and chemical roasting processes (Arumsari et al., 2021; Specialty Coffee Association, 2020). This study also contributes to applied coffee quality assessment by combining three elements in one system: MobileNetV2-based visual feature extraction, cosine-distance KNN-based feature-neighbourhood classification, and web-based real-time camera implementation. The contribution is practical because the system can support preliminary roasting-level screening for small and medium coffee businesses. However, the system should not yet be claimed as a replacement for professional roasting assessment or laboratory-based quality evaluation.

Another important limitation concerns the interpretability of the ensemble mechanism. Although cosine-distance KNN is expected to complement MobileNetV2 by adding local feature-neighbourhood information, the current study does not yet include feature-space visualisation, nearest-neighbour analysis, or ablation testing of the 60:40 ensemble ratio. Therefore, the complementary role of CNN and KNN should be interpreted as a preliminary and theoretically plausible explanation rather than a fully confirmed empirical finding. In addition, the model remains sensitive to image quality, as poor lighting, complex backgrounds, and inappropriate shooting angles can reduce classification accuracy. The performance of the real-time camera feature is also still affected by device and browser compatibility. For further development, future studies should include t-SNE or UMAP visualisation, nearest-neighbour case examples, systematic ablation of ensemble weights, more diverse field-condition datasets, and lightweight optimization techniques such as pruning or quantization to make the system more efficient for deployment (Gorvadiya et al., 2025).

CONCLUSION

This study successfully achieved its research objective by developing a web-based, real-time coffee bean roasting classification system using a MobileNetV2–KNN ensemble model. The proposed system was able to classify five roasting-related image classes, including DarkRoasting, GreenRoasting, LightRoasting, MediumRoasting, and Unknown, addressing the need for a more detailed and objective preliminary roasting assessment. Based on the evaluation of 6,470 coffee bean images, the system demonstrated satisfactory performance, achieving an accuracy of 85.67%, precision of 84.89%, recall of 84.45%, and an F1-score of 84.66%.

In addition to classification performance, the system met the practical objective of real-time deployment, as indicated by an average image prediction response time of 1,247 ms when implemented in a web-based application with camera input. These results confirm that the integration of MobileNetV2 as a feature extractor and KNN as a similarity-based classifier can support consistent and accessible preliminary roasting-level classification without requiring specialized software installation. Overall, the findings demonstrate that the proposed system fulfills its intended purpose as a functional preliminary prototype for objective visual coffee bean roasting classification.

REFERENCES

Adelaja, O., & Pranggono, B. (2025). Leveraging deep learning for real-time coffee leaf disease identification. *AgriEngineering*, *7*(1), Article 13. https://doi.org/10.3390/agriengineering7010013

Alfiantama, A., Ilham, M., & Putra, A. F. S. (2024). Klasifikasi tingkat roasting biji kopi menggunakan convolutional neural network. *Jurnal Sistem dan Teknologi Informasi*, *5*(1), 10–18. https://doi.org/10.29407/182k1t17

Anastácio, L. M., et al. (2023). Relationship between physical changes in the coffee bean due to roasting profiles and the sensory attributes of the coffee beverage. *European Food Research and Technology*, *249*(2), 327–339. https://doi.org/10.1007/s00217-022-04118-4

Arisandi, D., Sutanto, K., & Perdana, N. J. (2024). Implementasi aplikasi pemesanan pada kedai kopi berbasis web. *Jurnal Panrita Abdi*, *8*(3). https://doi.org/10.20956/pa.v8i3.26742

Arnita, Y. M., Marpaung, F., Hidayat, M., & Widianto, A. A. (2022). A comparative study of convolutional neural network and K-nearest neighbours algorithms for food image recognition. *Computational Technologies*, *27*(6), 88–99. https://doi.org/10.25743/ICT.2022.27.6.008

Arumsari, A., Surya, R., Irmasuryani, S., & Sapitri, W. (2021). Analisis proses roasting pada kopi. *Jbk*, *1*(2), 98–101. https://doi.org/10.35508/jbk.v1i2.5586

Bachrun, R. K. A. (2022). *Klasifikasi tingkat roasting biji kopi dengan ekstraksi fitur HSV menggunakan support vector machine* (Thesis). Universitas Islam Negeri Sunan Ampel.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Gorvadiya, J., Chagela, A., & Roy, M. (2025). Energy efficient pruning and quantization methods for deep learning models. In *Proceedings of the 2025 International Conference on Sustainable Energy Technologies and Computational Intelligence (SETCOM)* (pp. 1–6). https://doi.org/10.1109/SETCOM64758.2025.10932458

Gusmaliza, D., & Aminah, S. (2024). Sistem identifikasi kualitas biji kopi robusta berbasis image processing dengan support vector machine. *Edumatic*, *8*(2), 744–753. https://doi.org/10.29408/edumatic.v8i2.28008

Kuang, Z., Gao, H., Zhao, J., Wang, L., & Sun, L. (2025). FFFNet: A food feature fusion model with self-supervised clustering for food image recognition. *Applied Sciences*, *15*(17), Article 9542. https://doi.org/10.3390/app15179542

Laili, N. (2024). Pengaruh normalisasi fitur terhadap akurasi klasifikasi KNN pada data citra kopi. *Jurnal Teknologi Informasi dan Sistem Komputer*, *14*(1), 34–40.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lestari, P., Zulfauzi, Z., & Santoso, B. (2024). Klasifikasi tingkat kematangan roasting kopi menggunakan deep learning. *Jurnal Digital Teknologi Informasi*, *7*(1), 1–5. https://doi.org/10.32502/digital.v7i1.5655

Liu, Y., Chen, J., & Wang, L. (2021). Visual color analysis for food image classification. *Journal of Food Science and Technology*, *56*(4), 2147–2155.

Maulana, M., & Widodo, R. (2023). Optimasi model machine learning untuk aplikasi mobile pada klasifikasi tingkat roasting kopi. *Jurnal Sistem Komputer*, *11*(1), 45–54.

Permana, A., Hidayat, R., & Silalahi, B. (2024). Effect of feature normalization on KNN classification accuracy. *Jurnal Teknologi dan Sistem Komputer*, *12*(1).

Pratama, N., Sunge, A. S., & Budiarto, E. (2025). Penerapan model MobileNetV2 untuk prediksi tingkat roasting biji kopi berbasis gambar pada bot Telegram. *RIGGS*, *4*(2), 4571–4576. https://doi.org/10.31004/riggs.v4i2.1272

Putra, F. S., Jayanta, J., & Santoni, M. M. (2020). Penentuan level kematangan kopi berdasarkan hasil roasting menggunakan metode deteksi RGB dan klasifikasi minimum distance. In *Proceedings of Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*. https://conference.upnvj.ac.id/index.php/senamika/article/view/564

Rizal, M. A. (2019). Klasifikasi mutu biji kopi menggunakan metode K-nearest neighbor berdasarkan warna. *Jurnal Teknik Informatika Universitas PGRI Palembang*.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4510–4520). https://doi.org/10.1109/CVPR.2018.00474

Saputra, R. A., Suryana, A., & Raharja, S. (2023). Penerapan model pretrained CNN untuk prediksi tingkat roasting biji kopi. *Jurnal Sistem dan Teknologi Informasi*, *5*(1), 45–56.

Setiawan, M. R., Sari, Y. A., & Adikara, P. P. (2019). Klasifikasi citra makanan menggunakan K-nearest neighbor dengan fitur bentuk simple morphological shape descriptors dan fitur warna grayscale histogram. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *3*(3), 2726–2731. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4810

Siregar, Y. H., Kuala, S. I., & Yulianti, L. E. (2025). Aplikasi pengolahan citra dalam diferensiasi biji kopi non-fermentasi dan fermentasi. In *Prosiding Seminar Nasional Sains dan Teknologi "SainTek"*, *2*(1), 597–606. https://conference.ut.ac.id/index.php/saintek/article/view/5056/1906

Specialty Coffee Association. (2020). *Coffee standards and protocols*.

Sudarto, M. F., et al. (2024). Klasifikasi tingkat kematangan biji kopi berbasis AI pada website. *e-Proceeding of Applied Science*, *10*(1).

Taslim, A., Sudin, S., & Suratin, M. D. (2023). Identifikasi mutu roasting biji kopi menggunakan fitur warna dan backpropagation. *Jurnal Sains Komputer dan Teknologi Informasi*, *6*(1), 90–93. https://doi.org/10.33084/jsakti.v6i1.5600

Zhang, W., & Wang, J. (2022). Confidence-aware classification system for food images based on deep learning. *Computers in Industry*, *134*, Article 103565. https://doi.org/10.3390/rs14174161