X-CoffeeNet: A Novel Framework for Coffee Bean Image Classification Using Explainable Artificial Intelligence (XAI) and MobileNet

Mohamad Jamil [1,*], Suratman Sudjud [2], Sherly Asriany [3], and Muhammad Said [4]

[1] Department of Informatics Engineering, Universitas Khairun, Ternate, Indonesia
[2] Department of Agrotechnology, Universitas Khairun, Ternate, Indonesia
[3] Department of Architecture, Universitas Khairun, Ternate, Indonesia
[4] Department of Electrical Engineering, Universitas Khairun, Ternate, Indonesia
Email: jamil@unkhair.ac.id (M.J.); suratman.sudjud@unkhair.ac.id (S.S.); Sherly@unkhair.ac.id (S.A.); muhammad.said@unkhair.ac.id (M.S.)
*Corresponding author

Abstract—**This paper presents X-CoffeeNet, a novel deep learning framework for classifying coffee bean images by integrating MobileNetV2 and Explainable Artificial Intelligence (XAI) techniques. The model focuses on accurately classifying Arabica, Robusta, and Liberica beans while ensuring transparency in its decision-making process. The main novelty lies in combining a lightweight convolutional neural network with Local Interpretable Model-agnostic Explanations (LIME), allowing users to visualize and understand which parts of the image influence predictions. To improve generalization and reduce overfitting, the model uses image augmentation methods such as rotation, flipping, zooming, and brightness adjustment. MobileNetV2 is selected for its efficiency and low computational cost, making X-CoffeeNet suitable for deployment on mobile or embedded systems. Experimental evaluation on a dataset of 3,000 images shows that X-CoffeeNet achieves 100% accuracy, precision, recall, and F1-Score across all classes, highlighting its reliability and performance. Beyond coffee classification, the framework has broader applicability in agriculture and food industry automation where both accuracy and interpretability are essential.**

Keywords—**coffee bean classification, deep learning, Explainable Artificial Intelligence (XAI), MobileNet, image processing, computer vision**

I. Introduction

The coffee industry is among the most significant economic sectors worldwide. Coffee has grown into a popular beverage in numerous regions and serves as a key export commodity for various developing countries [1]. The quality of the coffee beans produced depends significantly on factors such as the type of bean, the processing method, and the environmental conditions in which the coffee is grown [2, 3]. Thus, classifying coffee beans by type and quality is crucial for ensuring that the resulting coffee products meet the desired quality standards [4]. However, the traditional process of classifying coffee beans is generally performed manually by trained experts, which requires a significant amount of time and is susceptible to human error. Furthermore, due to rising market demand, there is an urgent need to automate this process to enhance efficiency and accuracy. As a result, machine learning technologies, particularly deep learning, are increasingly being utilized in coffee bean image classification to expedite the process and lessen dependence on human expertise [5].

In recent years, deep learning has significantly contributed to the field of image processing, particularly in object classification and pattern recognition [6, 7]. A deep learning architecture frequently employed in image classification tasks is Convolutional Neural Networks (CNNs) [8, 9]. CNNs have proven to be highly effective at recognizing patterns and features in images, enabling them to identify various types of coffee beans based on visual image [10]. However, although CNNs can achieve high accuracy, these models are often black boxes, which makes their decisions challenging to understand and explain [11]. This is a significant issue in several applications, including the coffee industry, where transparency in decision-making is crucial. Coffee users and producers need to understand the reasons behind the decisions made by the model, allowing them to trust and utilize this technology on a broader scale. For this reason, a new approach is required that combines high accuracy with transparency, resulting in a model that is effective and capable of providing easily understandable explanations. This necessity has contributed to the emergence of Explainable Artificial Intelligence (XAI) [12].

XAI focuses on the development of machine learning models capable of explaining the decision-making process [13, 14]. Using XAI, users can figure out why the model classifies coffee beans in a certain method. This is especially crucial in real-world applications where user interpretation and trust are vital. By integrating XAI with

Manuscript received April 11, 2025; revised May 6, 2025; accepted June 23, 2025; published October 14, 2025.

deep learning, the model can clearly explain the key features that impact the classification results.

Moreover, MobileNet technology provides an efficient and lightweight solution for image classification tasks, which is particularly crucial for applications operating on mobile devices with resource limitation [15, 16]. MobileNet can create fast and accurate models while maintaining low resource usage, making it highly suitable for mobile device-based applications in coffee bean classification.

In response to these challenges and gaps, we propose X-CoffeeNet, a novel framework that integrates MobileNetV2 and Local Interpretable Model-agnostic Explanations (LIME)-based XAI to classify coffee bean images into three categories: Arabica, Robusta, and Liberica. The main contribution of this research is to improve the classification performance with 100% accuracy and increase transparency through visual and interpretable explanations for each prediction.

II. Related Works

Coffee bean image classification is a burgeoning field of research bolstered by advancements in machine learning and deep learning technologies. Numerous prior studies have utilized deep learning algorithms for coffee bean image classification, employing methods like Convolutional Neural Networks (CNNs) to enhance accuracy and efficiency in identifying coffee bean types.

In research by Zer *et al.* [17], CNN is employed to classify the shape of coffee beans into four categories: defect (damaged beans), longberry (long shape), peaberry (single bean), and premium (high-quality coffee beans). The results indicate an accuracy rate of approximately 90.63%, a precision of 88.23%, and a recall of 95.74%. Thai *et al.* [18] conducted research using CNNs to automatically identify defective coffee beans based on their visual characteristics. The results show an average accuracy of 38%, with a 1 to 3 Second processing time. Wallelign *et al.* [19] researched to design a CNN model for classifying raw coffee beans into 12 distinct quality grades. The results indicated that the model achieved an accuracy of 89.01% in classifying coffee.

Despite these advancements, many existing models suffer from limited interpretability and often lack robustness when applied to diverse datasets. Recent trends in precision agriculture highlight the importance of developing models that not only perform accurately but also operate efficiently in complex real-time environments. For instance, Roy *et al.* [20] proposed an improved object detection framework based on YOLOv4, enhanced with DenseNet, Spatial Pyramid Pooling, and a modified Path Aggregation Network, achieving high precision 90.33%, F1-Score 93.64%, and mean average precision 96.29% in detecting multiple plant diseases under challenging conditions at real-time speeds 70 FPS. This work exemplifies the integration of state-of-the-art deep learning architectures with practical agricultural applications, emphasizing both performance and computational efficiency.

Building upon these advancements, our study introduces X-CoffeeNet, a novel explainable AI framework leveraging MobileNetV2 and LIME, to address coffee bean classification with superior accuracy and interpretability compared to previous CNN-based models.

Meanwhile, using Explainable Artificial Intelligence (XAI) in image classification has become a major focus in recent studies. Bareedo *et al* [21] discuss the role and challenges of XAI to promote its wider application across various fields.

Local Interpretable Model-agnostic Explanations (LIME) is a popular XAI method that offers locally understandable explanations for individual predictions made by any machine learning model [22–25]. Although LIME has been widely applied in various fields, its use in coffee bean classification studies involving deep learning remains limited. Specifically, there are few explanations provided for model decisions related to coffee bean type and quality classification. This indicates that, despite its potential, LIME's application in the agricultural sector especially the coffee industry still has significant room for exploration. Therefore, further research integrating LIME with Convolutional Neural Networks (CNN) for coffee bean image classification could provide valuable contributions by improving model interpretability, boosting user trust, and enhancing the accuracy and efficiency of the classification process.

III. Research Method

The X-CoffeeNet model is built using the efficient and lightweight MobileNetV2 architecture, consisting of 53 layers with inverted residual and linear bottleneck structures, which optimizes performance with a parameter count of around 2.3 million and a computational complexity of around 300 million FLOPs. This architecture leverages depthwise separable convolution to accelerate the training and inference process, while maintaining high accuracy. The input image of 224×224×3 is processed through convolution and pooling blocks, then classified through a fully connected layer with a softmax activation function. The model is trained using Adam optimization and categorical crossentropy loss function, and is reinforced with augmentation techniques to reduce overfitting. This low complexity allows for deployment on resource-constrained edge devices. In addition to the efficient architecture, the model is also equipped with XAI-LIME to explain the contribution of pixels to predictions, increasing transparency and user trust.

The first stage in X-CoffeeNet involves collecting and labeling coffee bean image data of various types, such as Arabica, Robusta, and Liberica. The images are captured using a smartphone camera with a resolution of 50 megapixels. To ensure consistent image quality, all photos were taken under controlled lighting conditions using a diffuse LED light source to minimize shadows and reflections. The coffee beans were placed on a neutral-colored background to enhance contrast. These details are critical to maintain image consistency and support the reproducibility of the study. Next, they undergo a preprocessing stage that includes resizing to dimensions of

224×224 pixels, alongside data augmentation techniques like rotation and flipping to enhance the diversity and robustness of the model against input variations.

In the core stage, MobileNet is the backbone of a Convolutional Neural Network (CNN) model to perform visual feature extraction from coffee bean images. Utilizing a transfer learning approach, the MobileNet model can be fine-tuned on coffee-specific datasets to improve its classification performance. The output from MobileNet is subsequently passed to the fully connected layer and SoftMax function to classify the image into its corresponding coffee bean class. What sets X-CoffeeNet apart is its integration of interpretability methods from XAI. The framework facilitates using techniques like Local Interpretable Model-agnostic Explanations (LIME). LIME identifies areas of the image that most significantly contribute to the prediction by manipulating superpixels and analyzing the model's re-prediction outcomes. This enables the user to comprehend the reasoning behind the classification decision rather than passively accepting the result.

Fig. 1 illustrates a diagram of the X-CoffeeNet framework for coffee bean image classification. A detailed explanation is provided in the subsequent subsections.

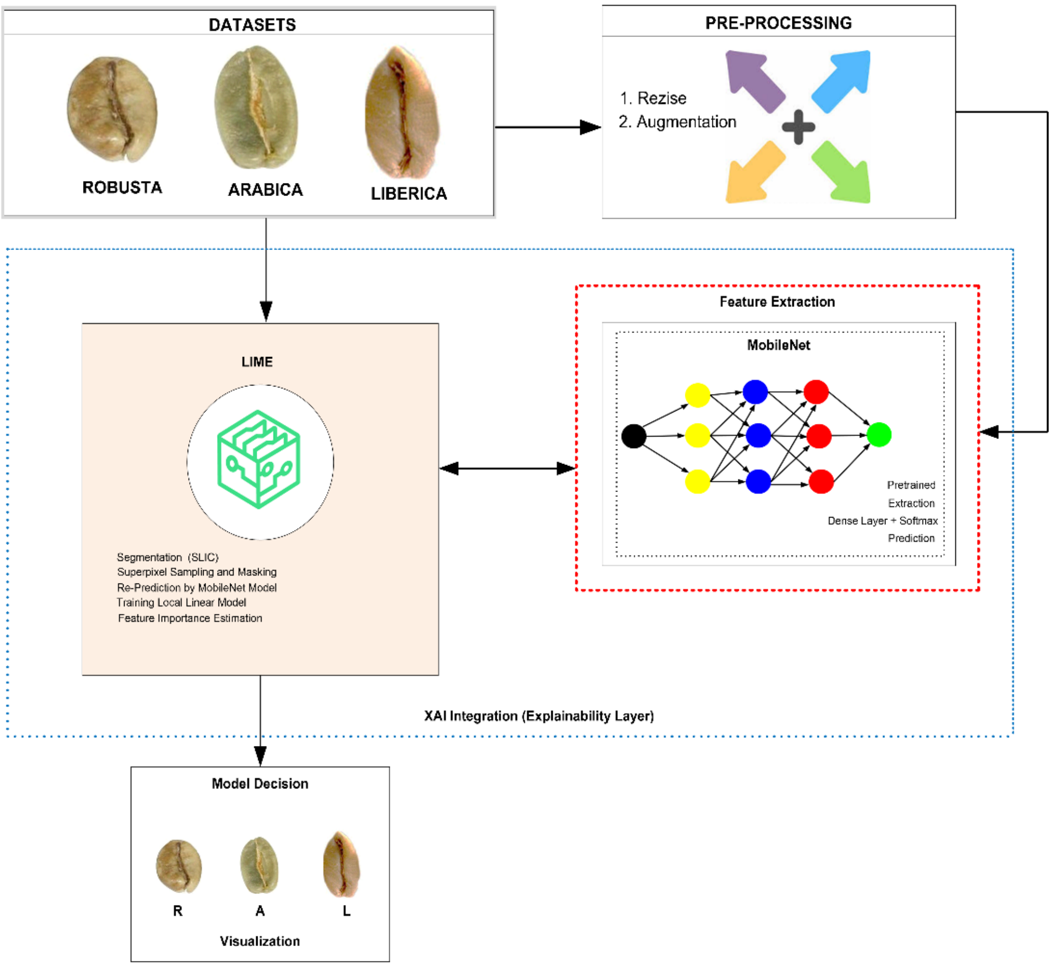

Fig. 1. X-CoffeeNet framework diagram for coffee bean image classification.

A. Coffee Bean Image Dataset

The dataset used in coffee bean image classification research consists of images categorized into three main types of coffee beans: Arabica, Robusta, and Liberica. The dataset contains 300 images (100 Arabica, 100 Robusta, and 100 Liberica). Each coffee bean type has distinct visual characteristics, including shape, size, color, and seed surface variations. Consequently, based on the provided images, this dataset will be utilized to classify coffee beans by their type.

B. Preprocessing

First, coffee bean images are resized to 224×224 pixels to meet MobileNet's input requirements, ensuring compatibility with the network's convolutional layer structure. The next step is data augmentation, which expands the variety of the dataset without the need to add new images manually. Techniques include random rotation, horizontal or vertical flipping, zooming, brightness adjustment, and shearing. Augmentation aims to enhance the model's generalization ability and prevent overfitting, especially when the original dataset is small. Moreover, the augmented image can be temporarily stored as batch input or directly fed into the training pipeline. In some instances, denoising or image enhancement may also be performed, particularly if the image is captured with a mobile device under varying lighting conditions.

C. Feature Extraction

In X-CoffeeNet, the feature extraction process employs MobileNet, a lightweight and efficient CNN architecture optimized for high performance on resource-constrained devices. MobileNet utilizes depthwise separable convolution, which divides the convolution process into two stages: depthwise (per-channel) and pointwise (1×1 convolution). This technique significantly reduces the number of parameters and computational complexity without compromising accuracy. A 224×224-pixel image is input into MobileNet, and the convolutional layer progressively extracts visual features from the image, ranging from low-level features (such as edges, color, and texture) to high-level features (such as seed surface patterns, crack shapes, or varietal characteristics). The output of the last convolutional block of MobileNet is a high-dimensional tensor representing the essential features of the image. This tensor is then forwarded to the global average pooling layer to reduce dimensionality while preserving significant spatial information. This stage also helps prevent overfitting and accelerates the classification process. The feature extraction results are then passed to the fully connected layer and finally to the SoftMax layer, which generates predictions in the form of probabilities for each coffee bean type class.

D. Classification

The classification stage in the X-CoffeeNet framework is the final step in the primary process before interpretation, where visual features extracted from the MobileNet model are utilized to identify the coffee bean type in the input image. After traversing all the MobileNet convolutional blocks and undergoing the global average pooling process, the coffee bean image is represented as a low-dimensional yet meaningful feature vector. This vector is then forwarded to the fully connected layer (dense layer), which functions as the decision layer based on the learned features. The final layer in this architecture is the softmax layer, which outputs probabilities for each coffee bean class.

E. Explainable Artificial Intelligence (XAI) Integration (Explainability Layer)

The XAI Integration, or Explainability Layer, is a key element differentiating X-CoffeeNet from traditional image classification systems. Its main objective is to provide a transparent and human-understandable explanation for the model's prediction results-specifically, why a coffee bean image is classified as a specific type by the MobileNet model. In this framework, LIME reveals which parts of the image most influence the classification decision. LIME works by breaking the image into small parts called superpixels. The system then creates hundreds of image variations by removing or blurring some of these superpixels. The model reclassifies these variations, and based on the differences in prediction results, LIME builds a simple local linear model to explain the primary prediction. The result is visualized as highlighted areas in the original image, where the most important parts the model considers are given a particular color or effect [26].

IV. EXPERIMENTAL RESULT AND DISCUSSION

This section is the core of the entire study, systematically describing the experimental process, the results obtained, and the analysis and discussion of the produced findings.

A. Experimental Setup

Experiments were conducted to evaluate the X-CoffeeNet model's performance in classifying three coffee bean types: Robusta, Arabica, and Liberica, using a dataset of 300 equally distributed images. The data was divided in a ratio of 80:20, with 240 images for training and 60 for testing. All data were processed using augmentation techniques such as rotation, flipping, zooming, and brightness shifting to enhance diversity and reduce the risk of overfitting. Table I shows the sample dataset and the training and test data composition from coffee bean images.

TABLE I. SAMPLE IMAGE DATASET OF THREE TYPES OF COFFEE BEANS

Class/Type	Ratio (80:20)		Image sample
	Training data (80%)	Test data (20%)	
Robusta	80	20	
Arabica	80	20	
Liberica	80	20	
Total	**240**	**60**	

Next, the image shown in Fig. 2 results from augmenting three types of coffee bean images. The original image serves as the input, and several visual variations are generated using a combination of transformations (rotation, zoom, and horizontal flip). This results in nine versions of the augmented image that illustrate changes in position, rotation, lighting, and scale.

This visualization demonstrates that the augmentation successfully creates significant visual diversity while maintaining the main features of the object, namely the shape and texture of the distinctive coffee beans. This is crucial to ensure that the model learns from realistic variations without losing the visual meaning of the class it represents.

After all the coffee bean images are processed through the augmentation stage, the next step in the experiment is the feature extraction process using the MobileNetV2 architecture. In this process, the MobileNetV2 model is utilized without the final classification layer (include_top = False), allowing only its convolutional part to capture important visual features of the image. Each image, measuring 224×224 pixels, is processed through the network, producing an output with a spatial dimension of 7×7 and a feature depth of 1280; thus, the total features per image amount to 7×7×1280 = 62,720.

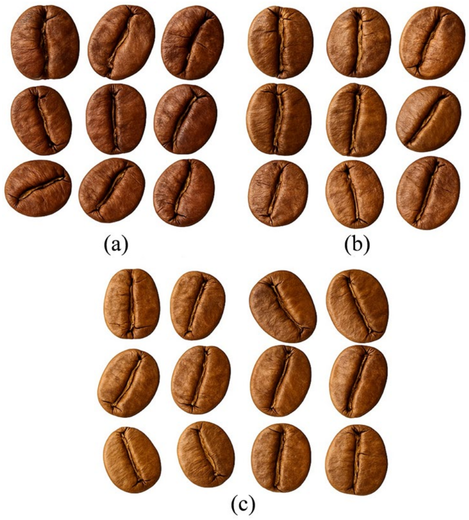

Fig. 2. Image augmentation of three types of coffee beans. (a) Examples of augementation of Arabica coffee bean images. (b) Examples of augementation of Robusta coffee bean images. (c) Examples of augementation of Liberica coffee bean images.

Based on Table II, it is evident that the feature extraction process successfully transformed the coffee bean image into a numerical representation that reflects the visual characteristics of each type of coffee. For instance, the arabica_001.jpg image has initial feature values such as F1 = 0.25123 and F2 = 0.43876, which differ significantly from the robusta_001.jpg image (F1 = 0.10781, F2 = 0.35942) and liberika_001.jpg (F1 = 0.0134, F2 = 0.0317). This variation indicates that these features accurately represent the differences in texture, color, and pattern of each coffee bean, making them suitable for classification.

TABEL II. SAMPLE FEATURE EXTRACTION RESULTS FROM THREE TYPES OF COFFEE BEAN IMAGES (4 INITIAL DIMENSIONS)

Image Name	**Label**	**Feature vector**			
		F1	**F2**	**F3**	**F4**
arabica_001.jpg	Arabica	0.25123	0.43876	0.13210	0.13210
robusta_001.jpg	Robusta	0.10781	0.35942	0.21863	0.0187
liberica_001.jpg	Liberica	0.0134	0.0317	0.0031	0.0256

In this process, the feature vector extracted from the coffee bean image is used as input for the classification model to distinguish between three classes: Arabica, Robusta, and Liberica, utilizing the reduced feature dimensions from MobileNetV2. Furthermore, performance evaluation is conducted using 20% of the total dataset as test data, aiming to measure the developed model's accuracy, precision, recall, and F1-Score. Table III displays the results of the image classification of the three types of coffee beans.

Based on Table III, the image classification results for three types of coffee beans Arabica, Liberica, and Robusta demonstrate exceptionally optimal performance. All evaluation metrics, including precision, recall, and F1-Score, yield a perfect value 1.00 for each class. This indicates that the model can classify all images with 100% accuracy without any prediction errors. High precision signifies that all positive predictions are correct, while high recall shows that the model successfully recognizes all actual data. A balanced F1-Score confirms that there is no trade-off between precision and recall. Furthermore, the macro average and weighted average values are also perfect, reflecting the model's consistent and unbiased performance across different classes. With support for each class set at 20, these results demonstrate that the features extracted from MobileNetV2 are highly representative for distinguishing between types of coffee beans, and the classification algorithm utilizes them effectively.

TABLE III. THE RESULTS OF THE IMAGE CLASSIFICATION OF THE THREE TYPES OF COFFEE BEANS

Label	**Precision**	**Recall**	**F1-Score**	**Support**
Arabica	1.00	1.00	1.00	20
Liberica	1.00	1.00	1.00	20
Robusta	1.00	1.00	1.00	20
Accuracy			**1.00**	60
Macro avg	1.00	1.00	1.00	60
Weighted avg	1.00	1.00	1.00	60

Fig. 3 presents the confusion matrix of the classification results.

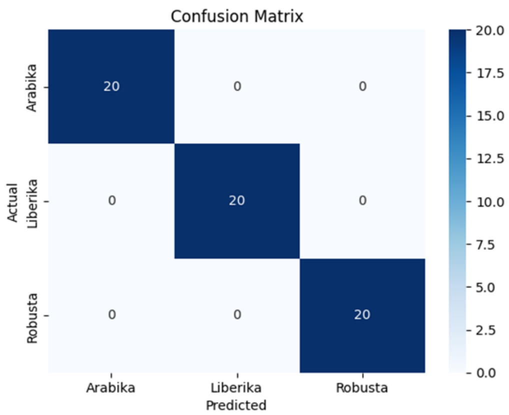

Fig. 3. Confusion matrix of three types of coffee beans image.

After completing the entire coffee bean image classification process, the next stage involves integrating the classification model with Explainable Artificial Intelligence (XAI) technology using the LIME method. LIME functions by altering (masking or randomizing) part of the original image and observing the changes in the model's predictions for the modified image. Based on these results, LIME can identify which aspects of the image contribute the most to the classification decision. Portions of the image that significantly influence the prediction results will be highlighted with contrasting colors, such as green or yellow.

Fig 4 presents the results of model integration using LIME.

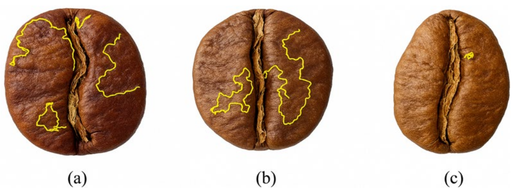

Fig. 4. The results of model integration using LIME. (a) Lime explanation for Arabica. (b) Lime explanation for Robusta. (c) Lime explanation for Liberica.

B. Comparison with State-of-the-Art Models

To provide a clearer understanding of the effectiveness of X-CoffeeNet, we present a comparative analysis with several state-of-the-art models that have been used for coffee bean image classification or related agricultural image tasks. The comparison focuses on key performance metrics such as accuracy, precision, recall, and F1-Score, as well as the presence of explainable features. Although some studies may have used different datasets or classification objectives, these models are relevant benchmarks to assess the classification performance and model transparency as shown in Table IV.

TABLE IV. COMPARATIVE PERFORMANCE OF X-COFFEENET AND OTHER COFFEE CLASSIFICATION MODELS

Model	Accuracy	Precision	Recall	F1-Score	Dataset Used	Explainability
X-CoffeeNet (Ours)	100%	100%	100%	100%	Custom (3-class, 300 images)	LIME
Zer *et al*. (2023) [17]	90.63%	88.23%	95.74%	-	4-class, coffee shape dataset	No
Thai *et al*. (2024) [18]	38%	-	-	-	Defective beans	No
Wallelign *et al*. (2019) [19]	89.01%	-	-	-	12-class grading	No
Hassan (2024) [10]	~95%	-	-	-	Pretrained models	No

C. Discussion

Based on the overall results of the experiments conducted, we can conclude that the X-CoffeeNet model provides excellent classification performance and demonstrates significant advantages over conventional classification models. One of the key findings of this experiment is the application of image augmentation techniques, which are crucial for introducing variations in the training data. Techniques such as rotation, flipping, zooming, and brightness shifting have proven effective in increasing the visual diversity of the data without compromising the quality or visual meaning of the object [27]. This directly affects the model's capacity to generalize and prevent overfitting, enhancing its robustness to real-world conditions.

Furthermore, feature extraction using MobileNetV2 enables the model to efficiently capture highly relevant visual characteristics from coffee bean images. With MobileNetV2, the model learns more complex and deeper patterns without requiring excessive parameters or computational resources, enhancing its efficiency compared to larger convolutional models like VGG16 or ResNet. This advantage is crucial for achieving exceptional classification performance, as experimental results indicate that the X-CoffeeNet model reaches 100% accuracy. The precision, recall, and F1-Score values all attaining 1.00 for each class (Arabica, Robusta, and Liberica) demonstrate that this model not only classifies images accurately but is also highly effective in recognizing each class without prediction errors.

Moreover, the integration with LIME for XAI enhances transparency for the model. LIME enables us to understand the reasons behind the model's decisions by analyzing and visualizing the parts of the image that most significantly influence the classification outcome. This offers deeper insight into the model's functionality and highlights the areas of the image deemed most crucial in the classification process.

Ultimately, when compared to traditional classification models such as Support Vector Machine (SVM), Random Forest, or K-Nearest Neighbors (KNN), X-CoffeeNet demonstrates superior performance in terms of efficiency, accuracy, and generalization capability. Traditional models often depend on feature engineering techniques, which necessitate a comprehensive understanding of the data [28], while X-CoffeeNet automatically extracts features from images using CNNs, which are more adept at capturing complex visual patterns. With the added benefit of LIME, this model also excels in the transparency and interpretability of its decisions.

V. CONCLUSION

This study proposed X-CoffeeNet, a novel and explainable deep learning framework for coffee bean image classification, integrating MobileNetV2 for efficient feature extraction and LIME for model interpretability. The model achieved 100% accuracy, precision, recall, and F1-Score across three coffee bean types—Arabica, Robusta, and Liberica—demonstrating a clear quantitative improvement over existing state-of-the-art models, which reported accuracies below 91%. These results confirm both the effectiveness and superiority of the proposed model in terms of classification performance and transparency. Compared to prior CNN-based approaches with limited interpretability and lower accuracy, X-CoffeeNet not only improves classification outcomes but also provides meaningful insights into model decision-making, which is essential for enhancing user trust and supporting real-world deployment in agricultural and food industries.

Despite its promising performance, the model has several limitations. The dataset used in this study is relatively limited in size and diversity, which may affect the model's generalizability to broader and more variable real-world scenarios. Additionally, the model currently supports only three coffee bean types and lacks evaluation on low-quality, mixed, or defective beans, which are commonly encountered in practical applications. Moreover, relying solely on LIME for explainability might restrict the depth of interpretability across different use cases.

Future research should address these limitations by incorporating larger and more diverse datasets, exploring hybrid and more advanced architectures such as Vision Transformers (ViT), and enabling real-time classification through IoT-based systems. Expanding the explainability dimension using complementary techniques like SHAP or integrated gradients can further enhance transparency, stakeholder trust, and acceptance among non-technical users.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Mohamad Jamil conceptualized the research, developed the methodology, designed the software, conducted formal analysis, and wrote the original draft. Suratman Sudjud contributed to data collection and analysis and reviewed and edited the manuscript. Sherly Asriany was responsible for research design and data interpretation and contributed to reviewing and editing the manuscript. Muhammad Said provided validation, conducted investigations, and assisted in reviewing and editing the manuscript. All authors have read and approved the final manuscript.

References

[1] M. Purnomo, Y. Yuliati, A. Shinta, and F. D. Riana, "Developing coffee culture among indonesia's middle-class: A case study in a coffee-producing country," *Cogent Soc. Sci.*, vol. 7, no. 1, 1949808, 2021. 10.1080/23311886.2021.1949808

[2] A. D. Bote and V. Jan, "Tree management and environmental conditions affect coffee (Coffea arabica L.) bean quality," *NJAS - Wageningen J. Life Sci.*, vol. 83, pp. 39–46, 2017. doi: 10.1016/j.njas.2017.09.002

[3] N. A. Febrianto and F. Zhu, "Coffee bean processing: Emerging methods and their effects on chemical, biological and sensory properties," *Food Chem.*, vol. 412, 135489, 2023. 10.1016/j.foodchem.2023.135489

[4] N. N. Amanina and G. W. Saraswati, "Classification of Arabica coffee green beans using digital image processing using the k-nearest neighbor method," *J. Appl. Intell. Syst.*, vol. 7, no. 2, pp. 111–119, 2022. doi: 10.33633/jais.v7i2.6449

[5] K. Przybył, M. Gawrysiak-Witulska, P. Bielska *et al.*, "Application of machine learning to assess the quality of food products—Case study: Coffee bean," *Appl. Sci.*, vol. 13, no. 19, 10786, 2023. doi: 10.3390/app131910786

[6] R. Archana and P. S. E. Jeevaraj, "Deep learning models for digital image processing: A review," *Artificial Intelligence*, vol. 57, no. 1. 11, 2024. doi: 10.1007/s10462-023-10631-z

[7] P. Tsirtsakis, G. Zacharis, G. S. Maraslidis, and G. F. Fragulis, "Deep learning for object recognition: A comprehensive review of models and algorithms," *Int. J. Cogn. Comput. Eng.*, vol. 6, pp. 298–312, 2025. doi: 10.1016/j.ijcce.2025.01.004

[8] L. Alzubaidi, J. Zhang, A. J. Humaidi *et al.*, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, no. 1, 53, 2021. doi: 10.1186/s40537-021-00444-8

[9] F. Badri, M. T. Alawiy, and E. M. Yuniarno, "Deep learning architecture based on Convolutional Neural Network (CNN) in image classification," *J. Ilm. Kursor*, vol. 12, no. 2, pp. 83–92, 2023. doi: 10.21107/kursor.v12i2.349

[10] E. Hassan, "Enhancing coffee bean classification: A comparative analysis of pre-trained deep learning models," *Neural Comput. Appl.*, vol. 36, no. 16, pp. 9023–9052, 2024. doi: 10.1007/s00521-024-09623-z

[11] T. Szandała, "Unlocking the black box of CNNs: Visualising the decision-making process with PRISM," *Inf. Sci. (Ny).*, vol. 642, no. May, 119162, 2023. doi: 10.1016/j.ins.2023.119162

[12] Z. Sadeghi, R. Alizadehsani, M. A. Cifci *et al.*, "A review of Explainable Artificial Intelligence in healthcare," *Comput. Electr. Eng.*, vol. 118, 109370, 2024. doi: 10.1016/j.compeleceng.2024.109370

[13] S. Ali, T. Abuhmed, S. El-Sappagh *et al.*, "Explainable Artificial Intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence," *Inf. Fusion*, vol. 99,. 101805, 2023. doi: 10.1016/j.inffus.2023.101805

[14] M. T. Hosain, J. R. Jim, M. F. Mridha, and M. M. Kabir, "Explainable AI approaches in deep learning: Advancements, applications and challenges," *Comput. Electr. Eng.*, vol. 117, 109246, 2024.

[15] A. T. Khan, S. M. Jensen, A. R. Khan, and S. Li, "Plant disease detection model for edge computing devices," *Front. Plant Sci.*, vol. 14, pp. 1–10, 2023. doi: 10.3389/fpls.2023.1308528

[16] H. Lokhande and S. R. Ganorkar, "Object detection in video surveillance using MobileNetV2 on resource-constrained low-power edge devices," *Bull. Electr. Eng. Informatics*, vol. 14, no. 1, pp. 357–365, 2025. doi: 10.11591/eei.v14i1.8131

[17] P. F. I. R. Zer, R. Rosnelly, and W. Wanayumini, "Classification of shape bean coffee using convolutional neural network," *Int. Conf. Inf. Sci. Technol. Innov.*, vol. 2, no. 1, pp. 6–10, 2023. doi: 10.35842/icostec.v2i1.28

[18] H. D. Thai, H. J. Ko, and J. H. Huh, "Coffee bean defects automatic classification realtime application adopting deep learning," *IEEE Access*, vol. 12, pp. 126503–126517, 2024. doi: 10.1109/ACCESS.2024.3452552

[19] S. Wallelign, M. Polceanu, T. Jemal, and C. Buche, "Coffee grading with convolutional neural networks using small datasets with high variance," *J. WSCG*, vol. 27, no. 2, pp. 113–120, 2019. doi: 10.24132/JWSCG.2019.27.2.4

[20] A. M. Roy, R. Bose, and J. Bhaduri, "A fast accurate fine-grain object detection model based on YOLOv4 deep neural network," *Neural Comput. Appl.*, vol. 34, no. 5, pp. 3895–3921, 2022. doi: 10.1007/s00521-021-06651-x

[21] A. Barredo Arrieta, N. Díaz-Rodríguez, J. D. Ser *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020. https://doi.org/10.1016/j.inffus.2019.12.012

[22] S. U. Hassan, S. J. Abdulkadir, M. S. M. Zahid, and S. M. Al-Selwi, "Local interpretable model-agnostic explanation approach for medical imaging analysis: A systematic literature review," *Comput. Biol. Med.*, vol. 185, 109569, 2025. doi: 10.1016/j.compbiomed.2024.109569

[23] P. S. R. Aditya and M. Pal, "Local interpretable model agnostic shap explanations for machine learning models," arXiv preprint, arXiv:2210.04533, 2022,

[24] N. B. Kumarakulasinghe, T. Blomberg, J. Liu, A. S. Leao, and P. Papapetrou, "Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models," in *Proc.-IEEE Symp. Comput. Med. Syst.*, 2020, pp. 7–12. doi: 10.1109/CBMS49503.2020.00009

[25] A. Stadlhofer and V. Mezhuyev, "Approach to provide interpretability in machine learning models for image classification," *Ind. Artif. Intell.*, vol. 1, no. 1, pp. 1–15, 2023. doi: 10.1007/s44244-023-00009-z

[26] M. R. Zafar and N. Khan, "Deterministic local interpretable model-agnostic explanations for stable explainability," *Mach. Learn. Knowl. Extr.*, vol. 3, no. 3, pp. 525–541, 2021. doi: 10.3390/make3030027

[27] K. Alomar, H. I. Aysel, and X. Cai, "Data augmentation in classification and segmentation: A survey and new strategies," *J. Imaging*, vol. 9, no. 2, 46, 2023. doi: 10.3390/jimaging9020046

[28] J. Ebiele, T. Ansah-Narh, S. Djiokap, E. Proven-Adzri, and M. Atemkeng, "Conventional machine learning based on feature engineering for detecting pneumonia from chest X-rays," *ACM Int. Conf. Proceeding Ser.*, 2020, pp. 149–155. doi: 10.1145/3410886.3410898