Received November 14th 2025; accepted December 24th 2025. Date of publication December 31st 2025
Digital Object Identifier: https://doi.org/10/25047/jtit.v12i2.457

Performance Comparison of CNN Transfer Learning Models for Coffee Bean Quality Classification

NUR MUHAMMAD FADLI[1], PRAWIDYA DESTARIANTO[2], HENDRA YUFIT RISKIAWAN[3], BEKTI MARYUNI SUSANTO[4], SATRIO ADI PRIYAMBADA[5], WAWAN HENDRIAWAN NUR[6], MUKHAMAD ANGGA GUMILANG[7]

[1,2,3,4,7]Jurusan Teknologi Informasi Politeknik Negeri Jember, Jember, Indonesia
[5,6]National Research and Innovation Agency, Bandung, Indonesia

CORESPONDING AUTHOR: MUKHAMAD ANGGA GUMILANG(email:angga.gumilang@polije.ac.id)

ABSTRACT Accurate and consistent sorting of roasted coffee beans is essential for meeting export quality standards SNI 01-2907-2008, yet manual grading remains time-consuming and subjective, especially when visual differences between roast levels are subtle. This study addresses a gap in the literature by providing a controlled, head-to-head comparison of three modern transfer-learning CNN architectures including Xception, MobileNetV2, and EfficientNet-B1 on a single, uniformly preprocessed dataset of 1.600 coffee-bean images across four classes (dark, medium, light, green). All models were trained under identical preprocessing and hyperparameter settings and evaluated using accuracy, precision, recall, F1-score, and confusion matrices to reveal per-class behavior. EfficientNet-B1 yielded the best test performance (100% accuracy; per-class precision/recall/F1 = 1.00), Xception followed closely (99.5% accuracy), and MobileNetV2 obtained 97% accuracy while offering a markedly smaller model footprint. This paper discuss the trade-off between predictive performance, and recommend EfficientNet-B1 for high-accuracy quality control and MobileNetV2 for resource constrained edge deployment experimental artifacts and per-class metrics are provided to support reproducibility and practical adoption.

KEYWORDS: Transfer Learning, Coffee Bean Classification, CNN, Xception, MobileNet V2, EfficientNet B1

I. INTRODUCTION

The development of technology makes computer vision have an important role to be studied and implemented for various sectors, including the agricultural and plantation sectors [6]. Computer vision covers various fields, including physics, art, photography, electronics, mathematics, and computer technology. Thus, computer vision has an important role in this research. Computer vision and image processing are interrelated. The main functions of computer vision are object detection, image segmentation, and image classification [7]. The combination of computer vision and machine learning (ML) can produce maximum output in this research.

Indonesia has long been a major player in the global coffee industry. In 2024, coffee commodity exports in Indonesia had a value of US$929 billion [1]. According to (Tempo 2024), Indonesia was the third largest coffee producer in the world after Brazil and Vietnam. And Indonesia was the fourth largest coffee exporter in the world after Brazil, Vietnam, and Columbia. East Java is one of the provinces that is the largest coffee producer in Indonesia, with annual coffee production of around 68,916 tons.

Bondowoso Regency as one of the regencies in East Java has excellent coffee, namely arabica and robusta coffee. Arabica coffee from Bondowoso's Ijen sub-district has successfully penetrated the export markets even to European countries [2]. In 2024 Bondowoso district produced 8,439 tons of coffee [3], an increase from 8,271 tons in 2023 [4].

Indonesia is a major coffee producer and exporter; proper grading and sorting of coffee beans directly affect export value and market acceptance. In many producing regions (including Bondowoso, East Java) the sorting process is still performed manually, which is labor-intensive, inconsistent across operators, and prone to human error [5]. This study focuses on the automation of coffee bean sorting by visual roast level (dark, medium, light, green) using computer vision and deep learning methods. Automating sorting can increase throughput, reduce subjectivity, and ensure consistent quality control for export standards such as SNI 01-2907-2008.

The application of ML which is part of Artificial Intelligence (AI) that functions to adapt human intelligence into computers, so that machines or computers are able to recognize, see, react, learn and solve problems like humans [7], [8]. In this research, ML is used to identify the quality of coffee beans to make it easier and more accurate in the sorting process. The most advanced AI specialized in ML is Deep Learning (DL) [9].

The DL architecture consists of several convolution layers that describe the learning features based on the trained data[10]. Convolutional Neural Networks (CNN) algorithm is a popular DL method to solve various problems in agriculture. The basic architecture of CNN is divided into several stages that consider a collection of small objects from the location of the shared object as well as the classification stage in a structured neural network [11], [12]. The application of the CNN algorithm is also known as a model that has high performance for image classification by applying the transfer learning (TL) method.

Transfer learning can be defined as a basic framework in ML and DL, which implements knowledge gained from previous models and applies it to the training of other data[13]. When applied to Neural Network architecture, Transfer Learning can provide excellent results. This includes the implementation process on a pre-trained model, the data is trained on complex and varied datasets, and applied to new datasets [13], [14], [15].

In research conducted by [16] with the application of transfer learning for CNN architectures including DenseNet, VGG19, Xception, MobileNet, Inception, InceptionResNet, ResNet101, ResNet152, ResNet50, and VGG16. The results showed that DenseNet achieved the highest accuracy of 0.989, followed by MobileNet and ResNet152 with accuracy of 0.982 and 0.980. DenseNet has the highest precision and F1 value among all models, with a precision value of 0.996 and F1 value of 0.992.

In addition to the research conducted by [16], this study refers to several previous studies relevant to coffee classification using several CNN architectures, such as that conducted by [17], which compares CNN vs Vision Transformer for coffee maturity classification Coffee beans (roasted, raw, non-coffee). Study [6] on transfer learning and fine-tuning of various pre-trained CNNs for coffee type classification (Arabica, Robusta, etc.). Additionally, [18] on CNN architectures using EfficientNet B0-B7 for coffee leaf disease classification, as detailed in Table 1.

TABLE 1. Comparative Research Study

Aspect	**Comparative Study**			
	[17]	[6]	[18]	This Study
Methods	CNN vs Vision Transformer for coffee maturity classification	Transfer learning and fine-tuning of various pre-trained CNNs	CNN architectures EfficientNet B0-B7 for coffee leaf disease classification	Comparison of CNN architectures: Xception, EfficientNet B1, MobileNet V2
Research Objects	Coffee beans (roasted, raw, non-coffee)	Type of coffee (Arabica, Robusta, etc.)	Disease-infected coffee leaves	Roasted coffee beans by quality color (Dark, Medium, Light, Green)
Type of Application	Determination of postharvest coffee maturity level	Specialty coffee prediction	Digital image-based coffee leaf disease detection	Visual coffee quality classification for export coffee production quality control
Technology	Xception, InceptionV3, ViT-B16, TensorFlow	MobileNet V2, ResNet, HRNet, EfficientNet, DenseNet	EfficientNet B0–B7, VGG16, ResNet50, Adam optimizer, LR 0.0001	Xception, EfficientNet B1, MobileNet V2, TensorFlow, Adamax Optimizer, LR 0.001
Evaluation	Xception: 96,67%; ViT-B16: 99,33%	MobileNet V2: 99% accuracy, F1: 99%	EfficientNet B1: 97% accuracy on test data; loss: 0.1328	EfficientNet B1: 100%, Xception: 99.5%, MobileNet V2: 97%

Despite numerous studies applying CNNs and transformers to coffee-related image tasks (maturity grading, type classification, leaf disease detection), there remains a lack of a focused, controlled comparison of modern CNN architectures specifically for roasted coffee bean quality classification under identical experimental conditions. Existing works often differ in dataset composition, preprocessing pipelines, or evaluation protocols, which limits direct comparability and practical guidance for industry adoption. Moreover, many studies report aggregate accuracy without detailed per-class performance or without analyzing trade-offs between predictive performance for deciding between edge and server deployments in real-world quality-control systems.

This study addresses these gaps by providing a head-to-head evaluation of three representative architectures is Xception, MobileNetV2, and EfficientNet-B1 on a single, consistently prepared dataset of roasted coffee beans. The main contributions are: (1) a controlled comparative benchmark using identical preprocessing and training protocols to enable fair model comparison; (2) comprehensive reporting of per-class metrics (precision, recall, F1-score) and confusion matrices to reveal class-specific strengths and weaknesses relevant to operational grading; and (3) an analysis of the trade-off between predictive performance and implementation considerations, delivering actionable. Collectively, these contributions advance empirical knowledge and practical guidance for applying deep learning to automated coffee bean grading.

Thus, this research aims to compare these three CNN architectures using transfer learning on a dataset of 1,600 coffee bean images (4 classes). The best-performing model is expected to inform more effective and accurate sorting processes in post-harvest operations.

.

II. METHOD

The flow of this research is carried out as the flowchart in Figure 1. There are 3 models that will be used in this research, namely Xception, MobileNet v2 and EffientNet B1. These three models were selected because, according to various studies, they offer high performance and are lightweight for use on devices such as mobile phones. At the same time, the selection of these three models represents different architectural characteristics in terms of depth, computational complexity, and efficiency. After sharing the dataset, all models will be subjected to the same preprocessing and trained with various parameters. Then all models will be tested and evaluated, so that the model with the best accuracy can be determined.

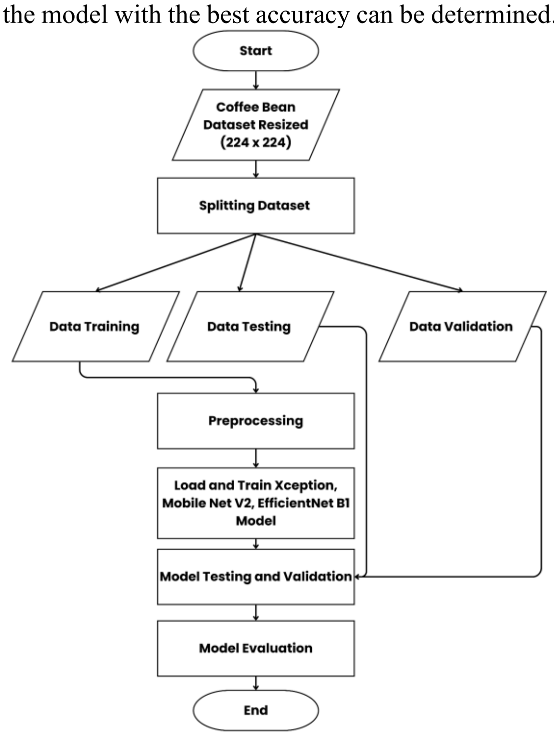


FIGURE 1. Research Stages

A. Dataset Collection

The dataset used in this study is the *Coffee Bean Dataset Resized (224 × 224)* from Kaggle [22]. It contains 1,600 images divided equally into four classes (dark, green, medium, light). We split the data into 1,200 images for training (including validation) and 400 images for testing; per-class distribution is shown in Table 2.

a b c d

FIGURE 2. Coffee bean images from the kaggle dataset (a) dark, (b) green, (c) medium, (d) light

TABLE 2. Dataset Distribution

Category	Training	Validation	Testing	Total
Dark	300	50	50	400
Green	300	50	50	400
Light	300	50	50	400
Medium	300	50	50	400
Total	1200	200	200	1600

B. Preprocessing

One of the most important stages in ML development is dataset preprocessing. Before training, the dataset must be preprocessed first to ensure compatibility. In this study, preprocessing is performed through image processing where all images are resized to 224x224 pixels to ensure all images have the same dimension in accordance with the standard input size required by the model.

Next, the image augmentation process is carried out using ImageDataGenerator by adjusting the brightness level randomly 80% to 120% of the original image value. The author uses the generator to set the batch data, image size, and brightness level. This augmentation process aims to make the developed model able to learn more detailed features, and increase generalization on previously unrecognized data [14].

C. Coffee beans Classification Process with Convolutional Neural Network (CNN)

This research was conducted using a transfer learning approach to classify coffee beans using CNN with Xception, MobileNet v2, and EfficientNet B1 architectures. The three architectures were chosen to represent complementary trade-offs relevant to practical deployment. Comparing these architectures yields practical guidance: which model maximizes accuracy, which supplies reasonable accuracy with small model size (edge use), and which strikes a balanced trade-off for production systems.

Xception, or “Extreme Inception”, is an advanced deep learning-based CNN architecture proposed by François Chollet in 2017. This architecture is an advance in the design of CNN, specifically developed for image classification [19], [20]. The Xception architecture was chosen because this model uses a convolutional approach known as depthwise separable convolution, which is a process of separating or dividing spatial convolution and the channel-wise convolution [20]. The approach used by this model does not employ non-linear activation of the two convolution stages. This factor contributes to the high performance of the model.

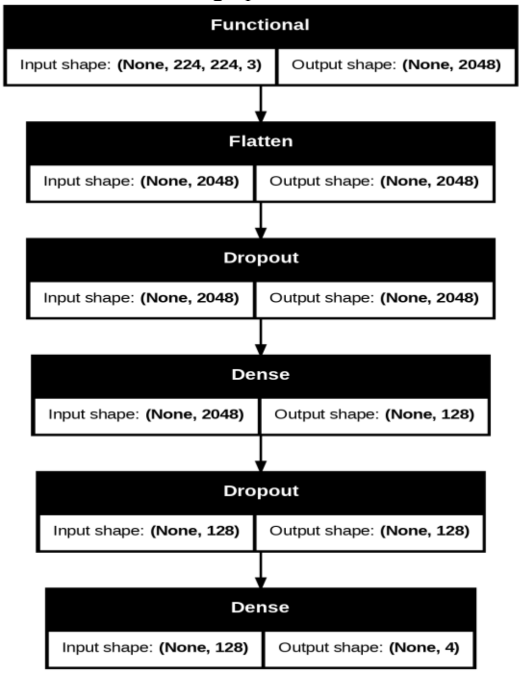


FIGURE 3. Xception Architecture

MobileNetV2 is a CNN architecture typically used for mobile devices or devices with low computational limitations. This architecture focuses on efficient processing speed without significantly reducing accuracy. It uses inverted residuals and linear bottlenecks to enable efficient information processing in narrow paths and reduce representation loss. These two combinations with depthwise separable convolutions result in a CNN architecture that is lightweight and effective when used on low-power devices [21].

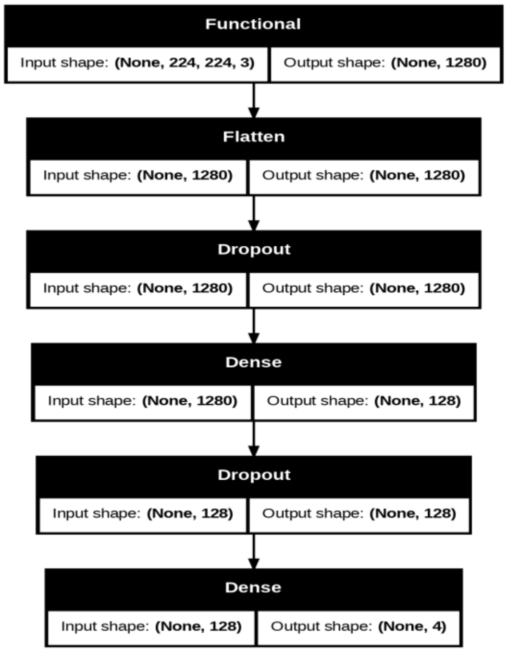


FIGURE 4. MobileNet V2 Architecture

EfficientNet is one of the CNN architectures developed in 2019, and is a superior architecture in image processing. The advantage of this architecture lies in its ability to achieve high accuracy without reducing computational efficiency. This is possible because this architecture applies compound scaling, which is a strategy that optimizes the depth, width, and resolution of the network simultaneously [6], [14].

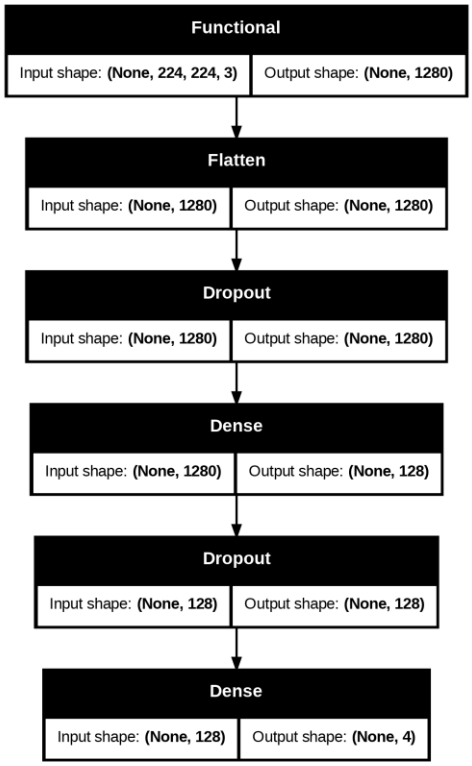


FIGURE 5. EfficientNet Architecture

The structure of EfficientNet is divided into two parts: the backbone network and the head network. The backbone network extracts important features from the input image and uses a combination of squeeze and excitation (SE) blocks and the best convolution layer (inverted residual blocks), which is used to recognize spatial relationships between channels. Meanwhile, the head network completes the final classification by applying a combination of fully connected layers and global average polling to produce an accurate final output [14].

D. Training Settings

The research was conducted using Google Colab with a T4 GPU to train, test, and validate the developed model for classifying coffee bean quality based on the roasting level. DL implementation requires very high computing power, so it requires hardware that is easy to use and has collaboration features [22].

TABLE 3. Performance Metrics

Hyperparameter	Values
Batch Size	16
Number of Epoch	5
Initial Learning Rate	0.001
Optimizer	Adamax
Loss Function	Categorical Crossentropy
Metrics	Accuracy

Table 3 describes the hyperparameters used in this study to set up the machine learning model to be developed. The selection of hyperparameters in this study aims to train stability and computational efficiency. A batch size of 16 was chosen to maintain a balance between GPU memory capacity and learning sensitivity to gradients, while the number of epochs was set to 5 because the model's accuracy had shown optimal convergence within that range without signs of overfitting. A learning rate of 0.001 was used because it is a stable and common value for optimization methods with Adamax, as it is more robust against large gradients and suitable for transfer learning models. The loss function used is categorical cross-entropy, suitable for multi-class classification problems such as those in the coffee bean dataset. The evaluation metric used is accuracy, as it directly represents how well the model classifies coffee bean categories in the test data.

E. Evaluation

The final stage of this study involved conducting measurements to evaluate the performance of the developed model, which were presented in tabular form for ease of reading. Accuracy refers to the correct results of the classification process and is expressed as a percentage [14]. Equation 1 describes the formula used for calculating accuracy, equation 2 describes the formula for calculating precision, and equation 3 describes the formula for calculating recall.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

When evaluating the model, there are terms True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). These four terms are used in a table called the confusion matrix, which describes the results of the model predictions in a structured and comprehensive manner. In assessing the performance of the model, three main matrices are generally used: accuracy, precision, recall, F1-Score and support. This evaluation data is also displayed in the form of a classification table for easier understanding.

TABLE 4. Confusion Matrix

		Predicted	
		Negative	***Positive***
Actual	***Negative***	TN	FP
	Positive	FN	TP

Table 4 illustrates the relationship between the actual labels and the predicted results of the model. True Positive (TP) and True Negative (TN) indicate that the model has correctly classified the data. Whereas False Positive (FP) and False Negative (FN) indicate that there is a mismatch or error between the model prediction and the actual label.

III. RESULT AND DISCUSSION

.

The models were implemented with the aim of solving classification problems on a previously prepared coffee image dataset. The models are trained and tested on the same dataset so that objective and fair evaluation results can be obtained when comparing them. The evaluation is based on the defined main matrix such as accuracy, precision, recall, F1-score and support calculated from the confusion matrix of the model test results. Each model is tested for its performance in correctly classifying images and the results are compared to determine which model has the best overall performance.

This research also applies transfer learning which can help reduce the need for large amounts of data to speed up training time.

a. Preprocessing Result

The results of preprocessing in image processing by resizing into 224x224 pixels to ensure that all images have the same size according to the standard input format of the Xception model. Next, the image augmentation process is carried out by using ImageDataGenerator with a random brightness level adjustment with a predetermined range between 80% to 120% of the original image value. The author uses the generator to set the data batch, image size, and brightness level. Figure 6 shows the result of image before and after image augmentation.

a b

FIGURE 6. Coffee bean images (a) before preprocessing, (b) after preprocessing

b. Model Performance Comparison

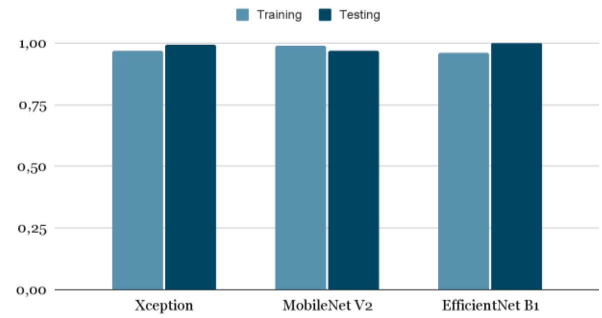
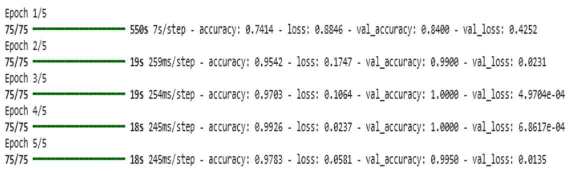
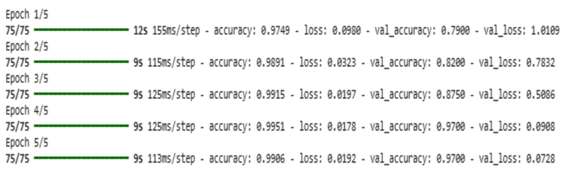


FIGURE 7. Model Evaluation On Training And Testing



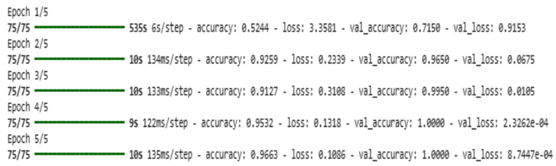
```
Epoch 1/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 550s 7s/step - accuracy: 0.7414 - loss: 0.8846 - val_accuracy: 0.8400 - val_loss: 0.4252
Epoch 2/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 19s 259ms/step - accuracy: 0.9542 - loss: 0.1747 - val_accuracy: 0.9900 - val_loss: 0.0231
Epoch 3/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 19s 254ms/step - accuracy: 0.9703 - loss: 0.1064 - val_accuracy: 1.0000 - val_loss: 4.9704e-04
Epoch 4/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 18s 245ms/step - accuracy: 0.9926 - loss: 0.0237 - val_accuracy: 1.0000 - val_loss: 6.8617e-04
Epoch 5/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 18s 245ms/step - accuracy: 0.9783 - loss: 0.0581 - val_accuracy: 0.9950 - val_loss: 0.0135
```

FIGURE 8. Xception Model Performance During the Training Process



```
Epoch 1/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 12s 155ms/step - accuracy: 0.9749 - loss: 0.0980 - val_accuracy: 0.7900 - val_loss: 1.0109
Epoch 2/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 9s 115ms/step - accuracy: 0.9891 - loss: 0.0323 - val_accuracy: 0.8200 - val_loss: 0.7832
Epoch 3/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 9s 125ms/step - accuracy: 0.9915 - loss: 0.0197 - val_accuracy: 0.8750 - val_loss: 0.5086
Epoch 4/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 9s 125ms/step - accuracy: 0.9951 - loss: 0.0178 - val_accuracy: 0.9700 - val_loss: 0.0908
Epoch 5/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 9s 113ms/step - accuracy: 0.9906 - loss: 0.0192 - val_accuracy: 0.9700 - val_loss: 0.0728
```

FIGURE 9. MobileNet v2 Model Performance During the Training Process



```
Epoch 1/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 535s 6s/step - accuracy: 0.5244 - loss: 3.3581 - val_accuracy: 0.7150 - val_loss: 0.9153
Epoch 2/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 10s 134ms/step - accuracy: 0.9259 - loss: 0.2339 - val_accuracy: 0.9650 - val_loss: 0.0675
Epoch 3/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 10s 133ms/step - accuracy: 0.9127 - loss: 0.3108 - val_accuracy: 0.9950 - val_loss: 0.0105
Epoch 4/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 9s 122ms/step - accuracy: 0.9532 - loss: 0.1318 - val_accuracy: 1.0000 - val_loss: 2.3262e-04
Epoch 5/5
75/75 ━━━━━━━━━━━━━━━━━━━━ 10s 135ms/step - accuracy: 0.9663 - loss: 0.1086 - val_accuracy: 1.0000 - val_loss: 8.7447e-04
```

FIGURE 10. EfficientNet B1 Model Performance During the Training Process

Figure 7 shows the results of comparing the accuracy of models previously trained and tested with batch sizes of 16 and 5 epochs. The model built with the Xception architecture achieved an accuracy of 97% on the training data and 99.5% on the test data, indicating that this dataset is capable of achieving very good accuracy. The model built with the MobileNet V2 architecture achieved an accuracy of 99% on the training data and 97% on the test data, which is slightly lower than the previous test. The EfficientNet B1 model also achieved very good results. It achieved 96% on the training data and the highest accuracy of 100% on the test data, even though its training results were slightly lower than those of Xception and MobileNet V2. The EfficientNet B1 model had the best generalization results among the other two models in classifying coffee beans into 4 classes. All three models showed excellent performance in learning visual patterns, but EfficientNet B1 has the best balance between training and test data for implementation in real-world applications.

Figures 8, 9, and 10 show that all three CNN architectures perform very well in classifying the quality of roasted coffee beans. Xception converges quickly, with validation accuracy reaching 100% by the third epoch, although this condition could potentially indicate overfitting. MobileNetV2 performed more stably with a highest validation accuracy of 97%, slightly lower than the other two models, but tended to be consistent and not extreme. Meanwhile, EfficientNetB1 experienced a slow initial learning phase with low accuracy and high loss, but then improved sharply to reach 100% validation accuracy in the fourth and fifth epochs with very small loss values. Overall, Xception and EfficientNetB1 showed the best performance on the validation data, while MobileNetV2 offered a balance between high accuracy and model stability

TABLE 5. Xception Model Performance Evaluation Result

Class	Precision	Recall	F1-Score	Support
Dark	1.00	1.00	1.00	39
Green	1.00	1.00	1.00	47
Light	0.98	1.00	0.99	57
Medium	1.00	0.98	0.99	57

TABLE 6. Mobilenet V2 Model Performance Evaluation Result

Class	Precision	Recall	F1-Score	Support
Dark	1.00	1.00	1.00	39
Green	1.00	0.98	0.99	47
Light	0.92	1.00	0.96	57
Medium	1.00	0.93	0.96	57

TABLE 7. EfficientnetB1 Model Performance Evaluation Result

Class	Precision	Recall	F1-Score	Support
Dark	1.00	1.00	1.00	39
Green	1.00	1.00	1.00	47
Light	1.00	1.00	1.00	57
Medium	1.00	1.00	1.00	57

Tables 5, 6, and 7 show the performance of three CNN models, namely Xception, MobileNet V2, and EfficientNet B1, in classifying coffee bean roasting levels. The EfficientNet B1 model showed the best performance with accuracy, precision, recall, and F1-score values of 100% across all classes, indicating that this model was able to classify each image correctly without error. Meanwhile, Xception also showed excellent performance with an accuracy of 99.5%, experiencing only a slight decline in the Light and Medium classes. MobileNet V2 has an accuracy of 97.5%, and although its overall performance is good, this model shows a decrease in F1-score in the Green and Medium classes. These results show that although all three models are capable of performing classification well, EfficientNet B1 is more stable and accurate for this dataset. Thus, high performance does not only depend on the model, but also on the suitability of the architecture to the complexity of the features in the data.

Figure 11 illustrates the confusion matrix results for the model built with the Xception architecture. Overall, the model can correctly classify all the data, but there is an error when a sample in the medium class is identified as the light class. The other three classes were identified correctly without error. The only error occurred between visually similar classes, but it was not significant in degrading the overall performance.

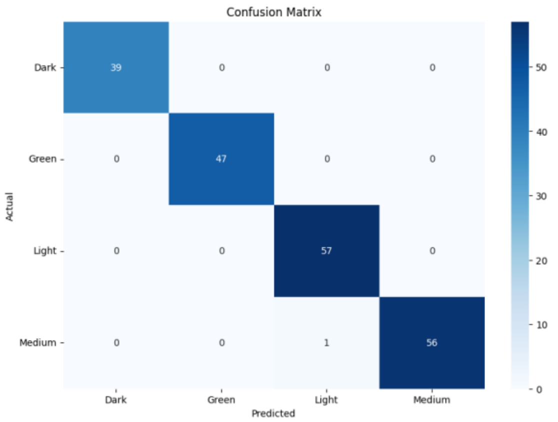


FIGURE 11. Xception Confusion Matrix Result

Figure 12 illustrates the confusion matrix results of the model built with the MobileNet V2 architecture. The model managed to classify most of the image data correctly, although there were some errors. When identifying and classifying the medium class sample, the model identified it as light and one green class sample was also identified as light. When performing classification on the light and dark classes, the model successfully performed the classification with perfect results. Although the model has an accuracy of 99% on the training data and 97% on the test data, the model has a weakness in identifying the light and medium classes.

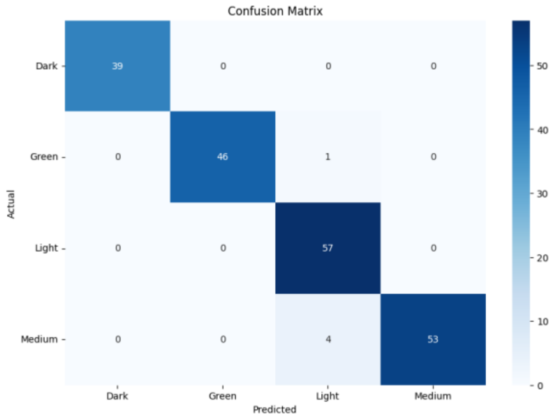


FIGURE 12. MobileNet V2 Confusion Matrix Result

Figure 13 illustrates the confusion matrix results of the model built with EfficientNet B1 architecture. As a result, this model successfully classifies and identifies all coffee bean samples into 4 classes perfectly. All test data containing 200 samples can be perfectly classified without any error. The results of the accuracy of this test are 100%, as can be seen in table 4. Thus these results can show that the EfficientNet B1 model in this study can distinguish image features between classes optimally. These results also show that the model developed with EfficientNet B1 architecture has the most stable and accurate performance among the three models tested.

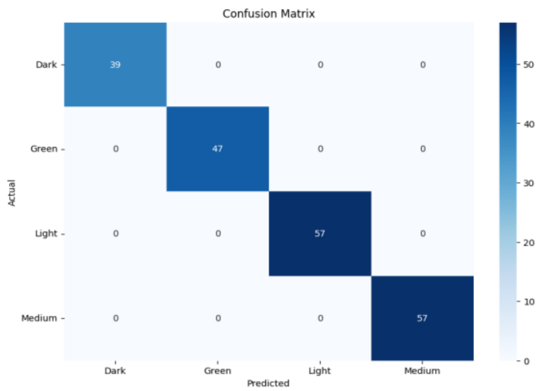


FIGURE 13. EfficientNet B1 Confusion Matrix Result

IV. Conclusion

This study aims to compare three pre-trained CNN architectures, namely Xception, MobileNet V2, and EfficientNet B1, for the automatic classification of roasted coffee beans into four visual quality classes using a consistent transfer learning protocol. The experimental results show that EfficientNet B1 achieved the best test performance (100% test accuracy, with no classification errors), Xception followed closely (99.5% test accuracy with one classification error), and MobileNet V2 offered competitive but lower accuracy (97% with several errors in similar visual classes). These results directly address the research objective by demonstrating that model architecture significantly influences the ability to capture subtle visual features relevant to agricultural assessment, and provide a practical contribution in the form of a reproducible and directly comparable benchmark that guides model selection according to implementation needs (EfficientNet B1 when maximum accuracy is required, MobileNet V2 when computational or edge implementation constraints dominate). For broader application, future research should evaluate robustness on larger and cross-location datasets, investigate advanced data augmentation and hyperparameter optimization, and measure inference latency on target hardware to support real-world adoption.

ACKNOWLEDGMENT

The author would like to thank all parties who have supported this research, including the Directorate of Vocational Higher Education, for the funding support provided for the Student Creativity Program in Science and Technology Application (PKM-PI). Thanks also go to RAISA Coffee School as a partner, the National Research and Innovation Agency (BRIN), and Jember State Polytechnic for their assistance and support in every stage of this research. Support from all parties has been instrumental in the completion of this research.

REFERENCE

[1] Badan Pusat Statistik, "Statistik Kopi Indonesia 2023." Accessed: Jul. 03, 2025. [Online].

[2] B. T. Firmansyah, F. Y. Ali, U. Setyoko, and A. L. Alwi, "Karakterisasi Morfologi Kopi Arabika (Coffea arabica L.) di Kawasan Desa Sempol Kecamatan Ijen Kabupaten Bondowoso," *Agropross Natl. Conf. Proc. Agric.*, pp. 254–260, Oct. 2024, doi: 10.25047/agropross.2024.698.

[3] Badan Pusat Statistik Kabupaten Bondowoso, "Kabupaten Bondowoso Dalam Angka 2025." Accessed: Jul. 03, 2025. [Online].

[4] Badan Pusat Statistik Kabupaten Bondowoso, "Kabupaten Bondowoso Dalam Angka 2024." Accessed: Jul. 03, 2025. [Online].

[5] Z. E. Fitri, B. A. Syahbana, A. Madjid, and A. M. N. Imron, "Penerapan Fitur Warna dan Tekstur untuk Identifikasi Kerusakan Mutu Biji Kopi Arabika (Coffea Arabica) di Kabupaten Bondowoso," *J. Ilm. Teknol. Inf. Asia*, vol. 15, no. 2, pp. 123–128, Sep. 2021, doi: 10.32815/jitika.v15i2.593.

[6] 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, Jun. 2024, doi: 10.1007/s00521-024-09623-z.

[7] W. Li *et al.*, "Intelligent metasurface system for automatic tracking of moving targets and wireless communications based on computer vision," *Nat. Commun.*, vol. 14, no. 1, p. 989, Feb. 2023, doi: 10.1038/s41467-023-36645-3.

[8] D. Novtahaning, H. A. Shah, and J.-M. Kang, "Deep Learning Ensemble-Based Automated and High-Performing Recognition of Coffee Leaf Disease," *Agriculture*, vol. 12, no. 11, p. 1909, Nov. 2022, doi: 10.3390/agriculture12111909.

[9] J. G. M. Esgario, R. A. Krohling, and J. A. Ventura, "Deep learning for classification and severity estimation of coffee leaf biotic stress," *Comput. Electron. Agric.*, vol. 169, p. 105162, Feb. 2020, doi: 10.1016/j.compag.2019.105162.

[10] S. H. Lee, H. Goëau, P. Bonnet, and A. Joly, "New perspectives on plant disease characterization based on deep learning," *Comput. Electron. Agric.*, vol. 170, p. 105220, Mar. 2020, doi: 10.1016/j.compag.2020.105220.

[11] H. A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, and J.-M. Kang, "A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Finetuned EfficientNet," *IEEE Access*, vol. 10, pp. 65426–65438, 2022, doi: 10.1109/ACCESS.2022.3184113.

[12] Y.-H. Wang and W.-H. Su, "Convolutional Neural Networks in Computer Vision for Grain Crop Phenotyping: A Review," *Agronomy*, vol. 12, no. 11, p. 2659, Oct. 2022, doi: 10.3390/agronomy12112659.

[13] M. Z. Khaliki and M. S. Başarslan, "Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN," *Sci. Rep.*, vol. 14, no. 1, p. 2664, Feb. 2024, doi: 10.1038/s41598-024-52823-9.

[14] C. B. Sanjaya, M. I. Rosadi, Moch. Lutfi, and L. Hakim, "Comparison of Transfer Learning Model Performance for Breast Cancer Type Classification in Mammogram Images," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 9, no. 1, pp. 130–136, Feb. 2025, doi: 10.29207/resti.v9i1.6177.

[15] C. Srinivas *et al.*, "Deep Transfer Learning Approaches in Performance Analysis of Brain Tumor Classification Using MRI Images," *J. Healthc. Eng.*, vol. 2022, no. 1, p. 3264367, 2022, doi: 10.1155/2022/3264367.

[16] C. C. Enriquez, J. Marcelo, D. R. Verula, and N. J. Casildo, "Leveraging deep learning for coffee bean grading: A comparative analysis of convolutional neural network models," *Trans. Sci. Technol.*, vol. 11, no. 1, pp. 1–6, Mar. 2024, Accessed: Jul. 04, 2025. [Online].

[17] M. A. Leonardi and A. Y. Chandra, "Analisis Perbandingan CNN dan Vision Transformer untuk Klasifikasi Biji Kopi Hasil Sangrai," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 3, p. 1398, Jul. 2024, doi: 10.30865/mib.v8i3.7732.

[18] M. I. Rosadi, L. Hakim, and M. Faishol A., "Classification of Coffee Leaf Diseases using the Convolutional Neural Network (CNN) EfficientNet Model," *Conf. Ser.*, vol. 4, no. 1, pp. 58–69, Dec. 2023, doi: 10.34306/conferenceseries.v4i1.627.

[19] D. Manu, P. M. Tshakwanda, Y. Lin, W. Jiang, and L. Yang, "Seismic Waveform Inversion Capability on Resource-Constrained Edge Devices," *J. Imaging*, vol. 8, no. 12, p. 312, Nov. 2022, doi: 10.3390/jimaging8120312.

[20] R. E. Mandiya, H. M. Kongo, S. K. Kasereka, K. Kyandoghere, P. M. Tshakwanda, and N. M. Kasoro, "Enhancing COVID-19 Detection: An Xception-Based Model with Advanced Transfer Learning from X-ray Thorax Images," *J. Imaging*, vol. 10, no. 3, p. 63, Feb. 2024, doi: 10.3390/jimaging10030063.

[21] Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability*, vol. 15, no. 3, p. 1906, Jan. 2023, doi: 10.3390/su15031906.

[22] R. T. Handayanto and H. Herlawati, "Prediksi Kelas Jamak dengan Deep Learning Berbasis Graphics Processing Units," *J. Kaji. Ilm.*, vol. 20, no. 1, pp. 67–76, Jan. 2020, doi: 10.31599/jki.v20i1.71.

[22] Sakdipat Ontoum, P. Sroison, and T. K, "Coffee Bean Dataset Resized (224 X 224)." 2022. Accessed: Nov. 07, 2025. [Online]. Available: https://www.kaggle.com/datasets/gpiosenka/coffee-bean-dataset-resized-224-x-224

Nur Muhammad Fadli, was born in Bondowoso, East Java in 2003. He is currently pursuing her undergraduate degree in Informatics Engineering at Politeknik Negeri Jember, Indonesia. His interests revolve around mobile development, internet of things and machine learning. As part of her academic research, He is also exploring traceability system for coffee products and disease detection in coffee leaves using machine learning. He can reached via email at nmuhammadfadli48@gmail.com.

PRAWIDYA DESTARIANTO, holds a Bachelor's degree in Computer Science from STMIK PPKIA Pradnya Paramita and a Master's degree in Engineering from Brawijaya University. He currently teaches at the Department of Information Technology at Jember State Polytechnic. His research interests include artificial intelligence, the Internet of Things, and information systems. He can be contacted via email at prawidya@polije.ac.id.

HENDRA YUFIT RISKIAWAN, was born in Jember, East Java, Indonesia, in 1983. He received his Bachelor degree from Bogor Agricultural University, Indonesia, in 2005 in Computer Science Education, and his Master's degree from Gadjah Mada University, Indonesia, in 2015, in Computer Science. His research interests include Information Systems, the Internet of Things, and machine learning. He can be contacted via email: yufit@polije.ac.id.

BEKTI MARYUNI SUSANTO, was born in Yogyakarta Province, Indonesia, in 1984. He received the Bachelor degree from the Yogyakarta State University, Indonesia in 2010 in Electrical Engineering Education and the Master degree from the STMIK Nusa Mandiri Jakarta, Indonesia, in 2012, in Computer Science. His research interests include cloud computing, internet of things, and machine learning. He can be contacted at email: bekti@polije.ac.id.

SATRIO ADI PRIYAMBADA, holds a Bachelor's degree in Information Systems from the Sepuluh Nopember Institute of Technology (ITS), a Master's degree in Management Information Systems from the Sepuluh Nopember Institute of Technology (ITS), and a Doctor of Philosophy - PhD in Human and Environmental Informatics from Kumamoto University. He is currently a Researcher at the Indonesian National Research and Innovation Agency (BRIN RI). His research interests include Process Mining, Data Mining, and Machine Learning. He can be contacted via email at satr014@brin.go.id.

WAWAN HENDRIAWAN NUR, holds a Bachelor's of Science degree from Padjadjaran University and a Master of Informatics degree from the Bandung Institute of Technology (ITB). He currently serves as the Geographic Information System (GIS) Administrator at the National Research and Innovation Agency (BRIN RI). His research interests include Geographic Information Systems (GIS) and Spatial Information Systems. He can be contacted via email at wawa020@brin.go.id.

MUKHAMAD ANGGA GUMILANG, holds a Bachelor's degree in Information Technology Education from Malang State University and a Master of Engineering (M.Eng.) degree from Gadjah Mada University. He currently teaches at the Department of Information Technology at Jember State Polytechnic while pursuing a doctoral program in the Electrical Engineering department at Gadjah Mada University. His research interests include artificial intelligence, social media analysis, and human-computer interaction. He can be contacted via email: angga.gumilang@polije.ac.id.

.