DOI : 10.2478/aucft-2025-0014

EVALUATION OF THE FEASIBILITY OF COFFEE VARIETY CLASSIFICATION FROM BEAN IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

– Research paper –

Michał ZIMKA, Katarzyna PENTOŚ*

Institute of Agricultural Engineering, Wrocław University of Environmental and Life Sciences, Chełmońskiego 37b, 51-630 Wrocław, Poland

Abstract: The study evaluates the feasibility of classifying Arabica and Robusta coffee beans using convolutional neural networks (CNNs). A custom CNN model (CNN_Coffee_Classifier) was developed and its performance was compared to that of three state-of-the-art architectures: MobileNet, ResNet50, and ResNet101. The models were trained both from scratch and using transfer learning on a dataset comprising 495 standardized images. The aforementioned dataset was derived from both public repositories and original photographs. Given the limited number of images in the training set, the models were trained using 10-fold cross-validation to ensure robust results. The custom CNN demonstrated an average accuracy of 75% (with 95% confidence interval of [0.68, 0.83]), while models employing transfer learning - particularly ResNet101 - exhibited superior performance, achieving up to 95% accuracy (with 95% confidence interval of [0.92, 0.96]). The findings substantiate the hypothesis that CNNs, particularly when integrated with transfer learning, provide a robust methodology for automated coffee variety classification, even when operating with a modest dataset.

Keywords: Arabica, Robusta, machine learning, transfer learning, MobileNet, ResNet50, and ResNet101

INTRODUCTION

Coffee is one of the most widely consumed beverages globally, transcending cultural, economic, and geographical boundaries. The United States Department of Agriculture (USDA) has estimated that global coffee production in the 2024/2025 season will total 174.9 million 60-kilogram bags, underscoring its substantial magnitude and economic importance (Coffee: World Markets and Trade, 2024). The coffee industry is supported by major producers like Brazil, Vietnam, and Colombia, and caters to an ever-growing consumer base, with the European Union and the United States leading in consumption. This burgeoning demand has driven advancements in production, processing, and quality assurance, underscoring coffee's central role in the global economy.

Coffee's popularity is partly due to the diversity of its species, each offering unique flavour profiles and characteristics. Arabica, the most prized species, is known for its refined aroma and delicate acidity, while Robusta's bold, bitter notes and higher caffeine content make it indispensable in espresso blends (Zhang et al., 2024). Lesser-known species like Liberica and Excelsa contribute to specialized markets with their distinctive flavor complexities (Lachenmeier & Wee Ting Lee, 2023). Recent attention to Stenophylla and Eugenioides highlights their potential as resilient and genetically valuable species in the face of climate change (Lahai et al., 2025). This variety enriches consumer experiences but also presents challenges in ensuring accurate classification and consistent quality.

Beyond its sensory appeal, coffee is celebrated for its potential health benefits, supported by a rich composition of bioactive compounds like antioxidants and phenolic acids. Studies link moderate coffee consumption to reduced risks of chronic diseases, including cardiovascular conditions, neurodegenerative disorders, and certain cancers (Konieczka et al., 2020). Additionally, caffeine's metabolic and ergogenic effects contribute to weight management and physical performance. These benefits, however, depend on maintaining the quality and integrity of coffee throughout its production and distribution chains.

Received: 25.06.2025
Accepted in revised form: 20.09.2025

* Corresponding author. E-Mail address: katarzyna.pentos@upwr.edu.pl

To make coffee beans suitable for consumption, they must undergo a roasting process. Under high temperatures, raw green beans undergo significant chemical and physical transformations that give them their distinctive flavour and aroma. This process is essential for enhancing the sensory qualities of coffee, such as its fragrance, acidity, and flavour intensity.

Ensuring the quality and consistency of coffee involves rigorous classification processes, particularly in high-volume operations. Manual inspection methods, while traditional, are increasingly inadequate to meet the scale and precision required by modern coffee production. Convolutional Neural Networks (CNNs), a class of deep learning algorithms adept at image-based tasks, offer a transformative solution. CNNs excel in extracting hierarchical features and detecting intricate patterns, making them well-suited for applications like classifying coffee bean species, verifying roast levels, and analyzing blend proportions (Kamilaris & Prenafeta-Boldú, 2018). Their ability to automate quality control processes enhances efficiency, scalability, and accuracy in large-scale roasteries. Despite their promise, the use of CNNs in coffee classification must address the risks associated with misclassification. Errors in identifying coffee species or roast levels can result in economic losses, regulatory penalties, and damage to brand reputation (Martins et al., 2018; Santoso et al., 2025). For instance, premium Arabica beans misclassified as Robusta may lead to underpricing, while inaccuracies in blend ratios can necessitate costly recalls. Furthermore, consumer dissatisfaction from inconsistent product quality can erode trust and loyalty, emphasizing the importance of precision in automated systems. The application of CNNs in coffee classification extends beyond operational efficiency. It represents a broader trend of leveraging computer vision technologies to enhance the quality and safety of food products. By standardizing classification processes, CNNs not only support the coffee industry's quest for excellence but also align with consumer demands for transparency and consistency. As the industry continues to grow, integrating advanced machine learning systems becomes essential in maintaining the balance between tradition and innovation.

The integration of vision systems with machine learning algorithms has emerged as a significant technological advancement in the field of agricultural quality assessment. This development has found application in the evaluation of quality parameters in beans of coffee and other plants. The classification system for the recognition of peaberry and normal coffee beans was developed by Gope and Fukai (2020). A comparison of CNN and Support Vector Machine (SVM) was conducted, leading to the conclusion that an optimal CNN structure, trained with images of 256×256 size, yielded a classification system of high accuracy. Chang and Liu (2024) presented a multiscale defect-detection model for coffee bean defect detection based on deep-learning techniques. The employment of CNN facilitated the precise identification of defect-free beans and those exhibiting defects such as immaturity, partial sourness, and mild insect damage. The development of custom CNNs is not an essential prerequisite. In many cases, a superior solution is to utilise specialised, pre-trained architectures and to train them on the target dataset. It is also possible to train such architectures on the target dataset from the outset, without transfer learning. Unal et al. (2022) subjected the four CNNs (SqueezeNet, Inception V3, VGG16, and VGG19) with transfer learning to a series of tests in order to develop an optimal tool for the classification of three types of coffee beans according to their coffee varieties. The highest level of accuracy was obtained in the case of the SqueezeNet model. In addition to the utilisation of machine learning for the evaluation of coffee beans, there are studies that employ this technology for the assessment of fruit quality. CNN with transfer learning was employed by Tamayo-Monsalve et al. (2022) for the classification of cherry coffee fruits into five groups according to stages of ripening. Their system was based on multispectral image data and was found to be an effective tool for coffee fruit classification based on maturation stage. Bazame et al. (2021) developed a system that has been shown to be effective in detecting coffee fruits and classifying them according to their maturation stage. This system is based on YOLOv3-tiny network technology.

A review of the literature indicates a lack of studies that compare different approaches to developing models that classify coffee varieties based on images of roasted beans. The present study hypothesises that the employment of pretrained CNN models will facilitate more precise classification of coffee beans than models trained from scratch. The main contributions of this work are summarized as follows:

(i) We propose a custom CNN architecture (CNN_Coffee_Classifier) for the classification of Arabica and Robusta coffee beans.

(ii) We compare the accuracy of the proposed architecture with three specialised CNN architectures (MobileNet, ResNet50, and ResNet101) trained from scratch and with the use of transfer learning.

MATERIALS AND METHODS

Image dataset

The primary portion of the analyzed dataset consists of images sourced from the dataset titled "Robusta or Arabica Dataset" published at RoboFlow platform (*Robusta or Arabica Object Detection Dataset by Aefnattanon, n.d.*). This dataset includes 400 images of the two main coffee bean species: Arabica (200 images) and Robusta (200 images). The images illustrating this dataset are presented in Figure 1.

To enrich the dataset and increase its diversity, additional coffee bean images were captured in domestic conditions using a Nikon D5000 camera. These images differ from those sourced from RoboFlow, offering higher resolution and darker tonalities, which contribute to the diversity of the dataset and improve the models' ability to generalize. The final dataset consisted of 495 images in the JPEG format, with 247 depicting Arabica beans and 248 depicting Robusta beans. All images, including those sourced from RoboFlow and captured with the Nikon D5000, were standardized to dimensions of 416x416 pixels. Images from RoboFlow were already of the required size, whereas the images captured with the Nikon D5000 were rescaled while preserving their proportions to avoid distortion of their original features. The diversity of images, stemming from different sources, significantly enhances the dataset's value. Images captured with the Nikon D5000 camera feature higher resolution and darker tonalities compared to those from RoboFlow, enriching the dataset and enabling better adaptation of machine learning (ML) models to varying conditions. Such data diversity is critical for improving models' generalization capabilities, increasing their practical applicability.

The development of CNN models

Convolutional Neural Networks are advanced deep learning algorithms designed to analyze visual data such as images. With a hierarchical structure of layers, CNNs can automatically recognize patterns, ranging from simple edges to complex shapes. Key components like convolutional, pooling, and fully connected layers enable efficient feature extraction and classification. Due to their local connectivity and weight sharing, CNNs are ideal for applications such as object recognition, medical image analysis, and quality control in industrial settings (Crespo et al., 2025; Fu et al., 2025; Pai et al., 2025).

In this study, a custom CNN model for the classification of Arabica and Robusta coffee beans was developed (CNN_Coffee_Classifier). The accuracy of this model was then compared with that of three other popular CNN-based classification models, namely MobileNet, ResNet50 and ResNet101. In the case of these models, two strategies were adopted: the first entailed the training of a CNN structure with random values of initial weights; the second entailed the training of a CNN structure that had been pre-trained with other data sets (transfer learning).

The training of all models was conducted using the same dataset. Due to the limited number of images in the training set, the 10-fold cross-validation method was employed to train the models. The dataset was divided into ten subsets, with training performed ten times. Each time, nine out of ten subsets were used for training and the remaining subset was used for validation. The models were then evaluated based on the average error metric value across the ten cases. Confusion matrices present the mean outcomes derived from ten models. This approach allows for a more accurate assessment of the models' quality.

The original data and software used in the study are openly available in repository: https://github.com/gzim324/CNN_Coffee_Classifier-model

Architecture and parameters of CNN_Coffee_Classifier

The custom CNN architecture was developed with consideration of several key parameters that influenced the classification quality. The details are shown in Table 1.

The architecture of the most advanced model is illustrated in Table 2. This architecture provides insights into the design and components of the custom CNN model.

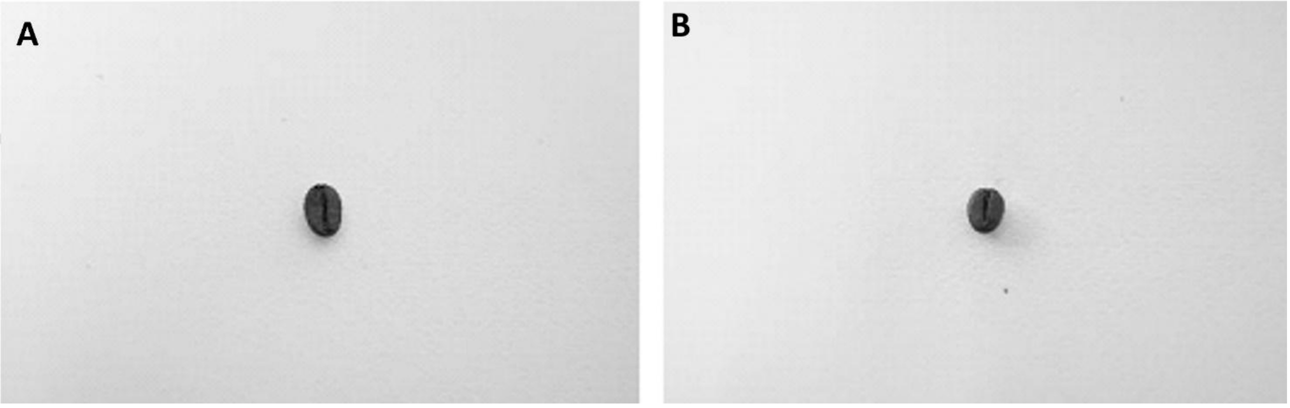

Figure 1. Example images of Arabica (A) and robusta (B) from Robusta or Arabica dataset.

Table 1. Parameters and hyperparameters relevant to the development of custom CNN model (CNN_Coffee_Classifier)

Parameter	Values	Description
Number of epochs		Every model was allowed to train for maximum 50 epochs, with early stopping (patience = 3) to prevent over-fitting and unnecessary computation.
Batch size	32 64	Two values were examined to balance gradient-estimate stability against GPU memory footprint.
Drop-out rate	0 0.3 0.5	
Kernel size	2 × 2 3 × 3 5 × 5	Convolutional kernels were compared to gauge the receptive-field effect on feature extraction.
Optimizer	Adam AdaDelta SGD	The optimizers represent adaptive, second-order–like, and momentum-based methods.
Activation function	ReLU	Preliminary investigation indicated no tangible gain from sigmoid, tanh, ELU or Leaky variants for tested dataset.
Number of filters	64 128 256	The architecture consisted of three consecutive Conv2D layers with 64, 128, and 256 filters, followed by a dense layer of 256 neurons and a soft-max output.

Table 2. The summary of the best custom model (No. 1 in table 5)

Layer (type)	Output shape	Number of parameters
InputLayer	(416, 416, 3)	
Conv2D	(412, 412, 64)	4,864
MaxPooling2D	(206, 206, 64)	-
Dropout	(206, 206, 64)	-
Conv2D	(202, 202, 128)	204,928
MaxPooling2D	(101, 101, 128)	-
Dropout	(101, 101, 128)	-
Conv2D	(97, 97, 256)	819,456
MaxPooling2D	(48, 48, 256)	-
Dropout	(48, 48, 256)	-
Flatten	589,824	
Dense	256	150,995,200
Dense	2	514
Total params: 152,024,962 Trainable params: 152,024,962 Non-trainable params: 0		

Models based on the state-of-the-art ML-based methods

MobileNetV2 is a lightweight CNN architecture designed for efficient deployment on mobile and embedded devices. The model has been constructed on the basis of the original MobileNet framework, with the incorporation of inverted residual blocks and linear bottlenecks (Sandler et al., 2018). These alterations reduce computational expense whilst maintaining representational capacity, thus facilitating high accuracy with a considerably reduced number of parameters and diminished latency in comparison to conventional CNNs. MobileNetV2 achieves an optimal balance between performance and efficiency, rendering it particularly suitable for real-time image classification and object detection tasks in environments where resources are limited.

The ResNet50 architecture is a deep convolutional neural network comprising 50 layers and is based on the concept of residual learning (He et al., 2015). The ResNet50 model was developed by substituting the 2-layer blocks in the 34-layer ResNet with 3-layer bottleneck blocks. ResNet50 addresses the vanishing gradient problem, which is commonly encountered in very deep networks, by incorporating shortcut connections that enable the direct propagation of feature maps across layers. These residual connections facilitate the learning of identity mappings, thus improving convergence during training and allowing the construction of substantially deeper networks without performance degradation. The primary benefit of the ResNet50 architecture is its reduced memory requirements in comparison to alternative CNN models. The ResNet101 is a deeper variant of the residual network family, consisting of 101 layers. It extends the design principles introduced in ResNet50 by increasing the number of bottleneck blocks, thereby enhancing the network's capacity to learn more complex feature representations. The increased depth allows ResNet101 to achieve higher accuracy on complex image recognition tasks. However, this enhancement is accompanied by an escalation in the computational and memory demands of the system.

The following hyperparameters were defined for all three models: dropout - 0.3, maximum number of epochs - 20, bach size - 16, loss function - crossentropy, training aborted if no improvement after 5 epochs (early stopping), optimiser - Adam.

Research platform

The research was conducted on a dedicated testing platform, comprising both hardware (Table 3) and software components (Table 4).

Table 3. Details of the hardware environment

	Hardware		
	CPU	RAM	GPU
Name	Intel	SK Hynix	Intel
Model	Core i7-1068NG7	16GB LPDDR4X	Iris Plus

Table 4. Details of the software environment

	Software						
Name	MacOS	Python	TensorFlow	Keras	NumPy	Scikit-learn	Jupyter
Version	15.2	3.9.18	2.12	2.12	1.23.5	1.5.1	1.0.0

Evaluation metrics

The error metrics described in equations 1-4 were used to evaluate the quality of the CNN models:

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

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

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

$$F1 - score = \frac{2 \cdot Precission \cdot Recall}{Precission+Recall} \quad (4)$$

where TP (True Positive) and FP (False Positive) are the number of correctly and incorrectly classified images of Robusta beans. TN (True Negative) and FN (False Negative) are the number of correctly and incorrectly classified images of Arabica beans. In addition, the ROC AUC score, the area under the ROC curve, was calculated for each model. This metric varies between 0 and 1 and represents the likelihood that a randomly selected positive example is given a higher rank than a randomly selected negative example (Vujovic, 2021).

Statistical analysis

In addition to the mean value and standard deviation of performance metrics from 10 validation runs, 95% confidence intervals were calculated using the Bootstrap method with 1000 resamplings. The performance of the models was compared by means of statistical significance tests. The analysis of variance (ANOVA) test is only conducted subsequent to the verification of the assumptions concerning the normality of distribution (Shapiro–Wilk test) and the homogeneity of variance (Levene test). In instances where these assumptions were not met, the non-parametric Kruskal–Wallis test was employed as an alternative. The significance level was set at $p < 0.05$.

RESULTS AND DISCUSSION

CNN_Coffee_Classifier model

The experimental process yielded models exhibiting a range of average accuracies from 0.24 to 0.75. The training hyperparameters and average error metrics for the test set for the top three models are detailed in Table 5.

As demonstrated in Table 5, the optimal model accuracy was attained through the utilisation of the Adam optimiser. As the classification process involved two distinct coffee beans, the most significant metric in terms of model quality is accuracy, which is defined as the percentage of examples that have been correctly classified. The mean accuracy of model no. 1 was found to be 0.75. The remaining two models demonstrated an average accuracy of 0.53. The ROC AUC score is also a crucial metric as it indicates the probability that the model will correctly distinguish between positive and negative examples. In the case of CNN_Coffee_Classifiers, the most accomplished model attained a value of 0.80 for this indicator. This finding suggests that the model's practical utility is limited (Çorbacıoğlu & Aksel, 2023). The results of the statistical analysis are presented in Table 6. The analysis indicates that model no. 1 demonstrates statistical significance in its deviation from the other models with regard to quality. However, no significant differences were observed between the models no. 2 and no. 3.

The confusion matrices shown in Figure 2 demonstrate that model no. 1 achieves the same level of accuracy in recognising both coffee species. Model no. 2 had greater difficulty classifying Robusta beans, with 56% being classified incorrectly. Model no. 3, on the other hand, was less accurate at recognising Arabica beans, classifying 60% of them as Robusta.

Table 5. Characteristics of the three best classifiers

No.	Training hyperparameters				Error metrics (mean ± standard deviation and 95% confidence interval)				
	Optimizer	Drop-out	Kernel	Batch size	Accuracy	Precision	Recall	F1	ROC AUC
1	Adam	0	(5, 5)	32	0.75±0.11 [0.68, 0.83]	0.82±0.06 [0.78, 0.87]	0.75±0.11 [0.68, 0.82]	0.73±0.14 [0.64, 0.82]	0.80±0.16 [0.70, 0.89]
2	SGD	0.3	(3, 3)	32	0.53±0.09 [0.48, 0.60]	0.51±0.22 [0.38, 0.65]	0.54±0.09 [0.49, 0.60]	0.46±0.13 [0.38, 0.54]	0.63±0.18 [0.52, 0.74]
3	AdaDelta	0	(2, 2)	64	0.53±0.11 [0.46, 0.60]	0.59±0.17 [0.47, 0.69]	0.53±0.11 [0.46, 0.60]	0.49±0.11 [0.42, 0.56]	0.58±0.23 [0.43, 0.72]

Table 6. Statistical test results for comparisons of models presented in Table 5

Model1	Model2	*p*-Value				
		Accuracy	Precision	Recall	F1	ROC AUC
Adam	AdaDelta	0.0005	0.0034	0.0005	0.0016	0.0536
SGD	AdaDelta	0.9920	0.9999	0.9850	0.8927	0.8508
Adam	SGD	0.0006	0.0011	0.0008	0.0005	0.1569

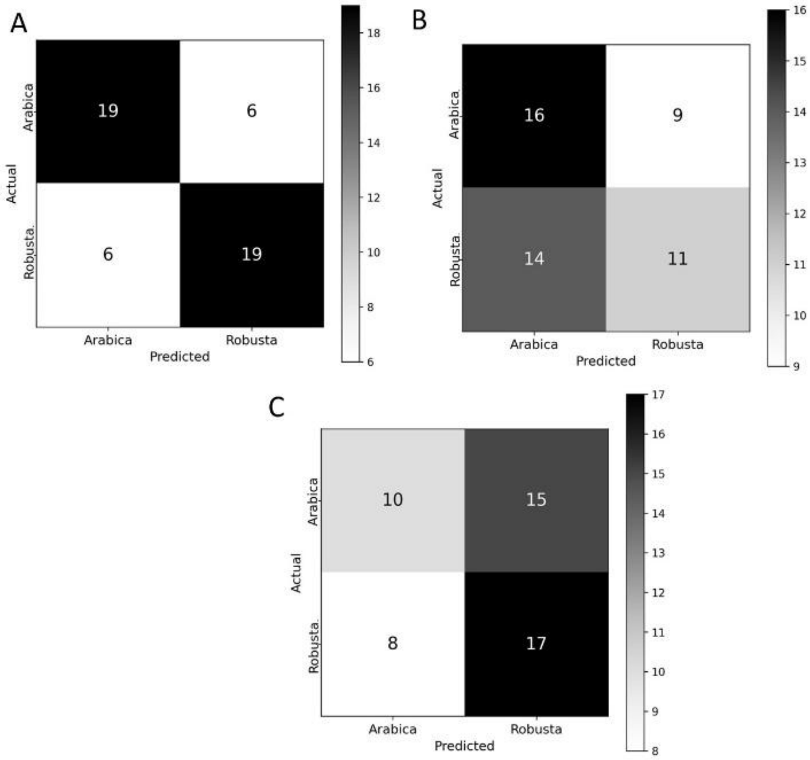

Figure 2. Confusion matrices of the proposed custom models: model no. 1 (A), model no. 2 (B), and model no. 3 (C) described in Table 5

Models based on the state-of-the-art ML-based methods

The results obtained for classifiers based on MobileNet, ResNet50 and ResNet101 when training these models with random initial weights are presented in Table 7. Confusion matrices with detailed information about classification results are depicted in Figure 3. The findings obtained for all structures examined indicate that classifiers exhibiting a remarkably low level of accuracy were developed when these models were subjected to training "from scratch". All models recognised Robusta beans slightly better. However, even in this case, 28–32% of the beans were classified as Arabica. Therefore, these models are regarded as wholly ineffectual in practical applications. Statistical analysis did not reveal any statistically significant differences in the quality of models trained "from scratch" (see Table 8).

The MobileNet, ResNet50 and ResNet101 structures were also trained using the transfer learning method. The results are presented in Table 9 and Figure 4. The employment of the transfer learning method resulted in the acquisition of models of a notably high calibre. The developed models demonstrated a higher level of classification accuracy than the custom models. The optimal model was identified for the ResNet101 architecture. In this particular instance, the classification of only one sample from each group was incorrect. A marginally inferior outcome was achieved for the ResNet50 network, which misclassified two samples of Robusta and one sample of Arabica beans. The MobileNet architecture yielded a set of classifiers that, on average, misclassified four samples. The statistical analysis (see Table 10) indicates that there are no statistically significant differences in the quality of models obtained using transfer learning.

Table 7. Error metrics of models trained using random initial weights (mean ± standard deviation and 95% confidence interval)

Model	Accuracy	Precision	Recall	F1	ROC AUC
MobileNet	0.66±0.07 [0.61, 0.70]	0.64±0.07 [0.60, 0.68]	0.73±0.07 [0.69, 0.78]	0.68±0.06 [0.65, 0.71]	0.72±0.06 [0.68, 0.76]
ResNet50	0.66±0.06 [0.62, 0.71]	0.65±0.07 [0.61, 0.69]	0.71±0.08 [0.66, 0.77]	0.67±0.05 [0.64, 0.71]	0.71±0.06 [0.67, 0.74]
ResNet101	0.65±0.07 [0.60, 0.69]	0.65±0.08 [0.60, 0.70]	0.69±0.10 [0.63, 0.75]	0.66±0.06 [0.63, 0.70]	0.68±0.05 [0.65, 0.71]

Table 8. Statistical test results for comparisons of models presented in Table 7

Model1	Model2	*p*-Value				
		Accuracy	Precision	Recall	F1	ROC AUC
MobileNet	ResNet101	0.9436	0.9752	0.5360	0.7105	0.3048
ResNet50	ResNet101	0.9675	0.9851	0.8344	0.8840	0.6509
MobileNet	ResNet50	0.9965	0.9851	0.8709	0.9456	0.8138

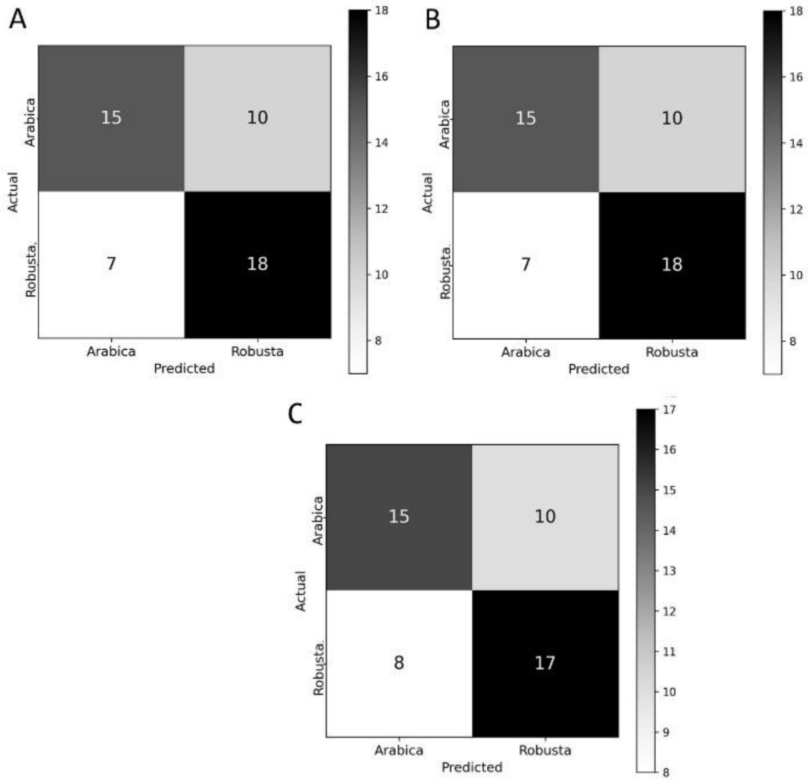

Figure 3. Confusion matrices of DL-based models trained using random initial weights: MobileNet (A), ResNet50 (B), and ResNet101 (C)

Table 9. Error metrics of models trained using transfer learning (mean ± standard deviation and 95% confidence interval)

Model	Accuracy	Precision	Recall	F1	ROC AUC
MobileNet	0.91±0.06 [0.87, 0.95]	0.90±0.09 [0.84, 0.95]	0.94±0.07 [0.88, 0.99]	0.91±0.05 [0.88, 0.95]	0.99±0.01 [0.98, 0.99]
ResNet50	0.92±0.04 [0.90, 0.95]	0.94±0.05 [0.91, 0.97]	0.91±0.09 [0.85, 0.96]	0.92±0.05 [0.89, 0.95]	0.99±0.01 [0.992, 0.999]
ResNet101	0.95±0.03 [0.92, 0.96]	0.95±0.04 [0.93, 0.98]	0.94±0.07 [0.89, 0.98]	0.94±0.04 [0.92, 0.96]	0.99±0.01 [0.98, 0.99]

Table 10. Statistical test results for comparisons of models presented in Table 9

Model1	Model2	*p*-Value				
		Accuracy	Precision	Recall	F1	ROC AUC
MobileNet	ResNet101	0.2887	0.1792	0.9999	0.7581	0.7268
ResNet50	ResNet101	0.8018	0.9391	0.9999	0.9999	0.1291
MobileNet	ResNet50	0.6423	0.3197	0.6229	0.9999	0.9999

The study utilised a relatively modest training set. The classification objects were Arabica and Robusta beans, which are comparatively simple in structure. Arabica beans characteristically exhibit a larger, more elongated and oval shape, whilst Robusta beans demonstrate a smaller and rounder morphology. Arabica beans exhibit a more curved furrow when compared to the straighter furrow of Robusta beans (Shara et al., 2021). Nevertheless, the disparity in morphology between Arabica and Robusta beans is not substantial, which may impede the efficacy of their automatic classification by vision systems utilising machine learning algorithms. It is noteworthy that all models with potential practical applications (models employing transfer learning) achieved comparable accuracy for both recognised classes.

Our findings demonstrate that the implementation of deep learning methodologies facilitates the development of precise models for the classification of coffee beans according to their species. The study compares different approaches to constructing such models, including the development of CNN models with custom structures and the utilisation of state-of-the-art DL-based methods. The findings suggest that in the context of an undersized training set, the implementation of a transfer learning approach is particularly advantageous.

The models based on transfer learning developed in this research were characterised by a high degree of accuracy. This finding is consistent with other examples of the use of ML methods to classify coffee beans according to their species. Buonocore et al. (2022) proposed CNN model based on the YOLO algorithm for the purpose of distinguishing Arabica and Robusta beans. The precision and recall of this model were found to be 0.99. Korkmaz et al. (2025) employed a range of CNN structures and transfer learning for the recognition of three coffee bean species (Starbucks Pike Place, Espresso, and Kenya). The accuracy of the models ranged from 0.90 for DenseNet201 to 0.93 for InceptionV3. Hendrawan et al. (2021) trained the AlexNet network to classify three types of Indonesian Arabica coffee beans (Gayo Aceh, Kintamani Bali, and Toraja Tongkonan). The model demonstrated an accuracy of 0.99.

In the present study, the accuracy of custom models was found to be lower than that of models employing transfer learning. This finding aligns with the observations reported by other researchers in the field. Only a small number of studies have indicated that custom models are comparable or more accurate than transfer learning models (Abasi et al., 2023; Manikandakumar & Karthikeyan, 2022). Some researchers have optimised specialised structures by modifying selected elements. This approach has typically resulted in an improvement in the quality of the modified model compared to the original structure. Wu et al. (2025) modified the YOLOv7-w6-pose model to develop an accurate tool for detecting mature soybean stem nodes. The modifications included changes in the backbone network, replacing Conv layers with max pooling layer_2 and RepConv in the neck network, and replacing the original detection head with the YOLOv7 detection head. The modified model demonstrated an increase of 2.21% in the recall rate and 2.6% in average accuracy compared to the original YOLOv7-w6-pose model. He et al. (2025) developed a new approach based on YOLOv11 for detecting maize leaf diseases. They improved the model's accuracy by replacing the C3k2 backbone network with the RepLKNet network in the backbone layer and incorporating a Selective Kernel Attention (CBAM) module in the neck network. They also used the DynamicHead module and the WIoU loss function, as well as the DynamicATSS label assignment strategy. These modifications increased precision from 87.7% to 92.6%, recall from 78% to 85.4% and F1-score from 82.6% to 88.9%.

The findings of this study demonstrate that CNNs, particularly when combined with transfer learning, constitute a robust methodology for coffee classification. This outcome emphasizes the efficacy of transfer learning for data sets that are limited in size. However, it is imperative to exercise caution when interpreting the results, given the limitations of the present study. The dataset, which consists of 495 images, is the primary factor influencing the generalisation of the results. Despite the utilisation of 10-fold cross-validation and statistical rigour, which serves to enhance reliability, this limitation remains unresolved. In the future, it will be important to significantly expand and diversify the dataset in order to confirm our findings on a larger scale. Another aspect is the potential lack of diversity in the current dataset, which could have led to implicit "information leakage". In future research, it will be crucial to implement more advanced data augmentation techniques to increase the diversity and resilience of models to changes in visual conditions, such as lighting or background. Furthermore, in order to ensure greater transparency in the decision-making process, further work should focus on the interpretability of models (explainable AI), which will allow for a better understanding of which image features are key to classification. Such a step will enable a more accurate understanding of the mechanisms of the network and may contribute to further optimisation.

CONCLUSIONS

The automation of the food quality assessment process has become a reality, largely due to the rapid advancements in artificial intelligence-assisted vision systems. The employment of these techniques facilitates the expeditious identification of specific characteristics pertaining to vegetables and fruits, whilst ensuring a level of accuracy that is commensurate with the capabilities of human observers. In the present work, a custom convolutional neural network model was proposed for the classification of Arabica and Robusta coffee beans. The model achieved an accuracy of 75%. Furthermore, a comparison was made between the custom model and the well-known transfer learning approach. This approach yielded models that exhibited superior accuracy. However, it should be noted that the custom model featured a much simpler structure. It can be concluded that in cases where the training dataset is limited, transfer learning should be selected as the preferred approach. Further work will focus on augmenting the training dataset to enhance the quality of the custom model.

REFERENCES

1. Abasi, A. K., Makhadmeh, S. N., Alomari, O. A., Tubishat, M., & Mohammed, H. J. (2023). Enhancing Rice Leaf Disease Classification: A Customized Convolutional Neural Network Approach. *Sustainability, 15*(20), 15039. https://doi.org/10.3390/SU152015039
2. Bazame, H. C., Molin, J. P., Althoff, D., Martello, M., Bazame, H. C., Molin, J. P., Althoff, D., & Martello, M. (2021). Detection, classification, and mapping of coffee fruits during harvest with computer vision. *CEAgr*, *183*, 106066. https://doi.org/10.1016/J.COMPAG.2021.106066
3. Buonocore, D., Carratù, M., & Lamberti, M. (2022). Classification of coffee bean varieties based on a deep learning approach. *18th IMEKO TC10 Conference on Measurement for Diagnostic, Optimisation and Control to Saupport Sustainability and Resilience 2022*, 14–19. https://doi.org/10.21014/TC10-2022.002
4. Chang, S. J., & Liu, K. H. (2024). Multiscale Defect Extraction Neural Network for Green Coffee Bean Defects Detection. *IEEE Access*, *12*, 15856–15866. https://doi.org/10.1109/ACCESS.2024.3356596
5. *Coffee: World Markets and Trade*. (2024). https://apps.fas.usda.gov/psdonline/circulars/coffee.pdf
6. Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, *23*(4), 195–198. https://doi.org/10.4103/TJEM.TJEM_182_23
7. Crespo, A., Moncada, C., Crespo, F., & Morocho-Cayamcela, M. E. (2025). An efficient strawberry segmentation model based on Mask R-CNN and TensorRT. *Artificial Intelligence in Agriculture*, *15*(2), 327–337. https://doi.org/10.1016/J.AIIA.2025.01.008
8. Fu, H., Lu, J., Li, J., Zou, W., Tang, X., Ning, X., & Sun, Y. (2025). Winter Wheat Yield Prediction Using Satellite Remote Sensing Data and Deep Learning Models. *Agronomy 15*(1), 205. https://doi.org/10.3390/AGRONOMY15010205
9. Gope, H. L., & Fukai, H. (2020). Normal and Peaberry Coffee Beans Classification from Green Coffee Bean Images Using Convolutional Neural Networks and Support Vector Machine. *International Journal of Computer and Information Engineering*, *14*(6), 189–196. https://doi.org/10.5281/ZENODO.7085081
10. He, J., Ren, Y., Li, W., & Fu, W. (2025). YOLOv11-RCDWD: A New Efficient Model for Detecting Maize Leaf Diseases Based on the Improved YOLOv11. *Applied Sciences, 15*(8), 4535. https://doi.org/10.3390/APP15084535
11. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proc. IEEE Comp. Society Conference on Computer Vision and Pattern Recognition*, *2016-December*, 770–778. https://doi.org/10.1109/CVPR.2016.90
12. Hendrawan, Y., Rohmatulloh, B., Ilmi, F. I., Fauzy, M. R., Damayanti, R., Al Riza, D. F., Hermanto, M. B., & Sandra. (2021). AlexNet convolutional neural network to classify the types of Indonesian

coffee beans. *IOP Conference Series: Earth and Environmental Science*, *905*(1), 012059. https://doi.org/10.1088/1755-1315/905/1/012059
13. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/J.COMPAG.2018.02.016
14. Konieczka, P. P., Aliaño-González, M. J., Ferreiro-González, M., Barbero, G. F., & Palma, M. (2020). Characterization of Arabica and Robusta Coffees by Ion Mobility Sum Spectrum. *Sensors,* 20(11), 3123. https://doi.org/10.3390/S20113123
15. Korkmaz, A., Talan, T., Koşunalp, S., & Iliev, T. (2025). Comparison of deep learning models in automatic classification of coffee bean species. *PeerJ Computer Science*, *11*, e2759. https://doi.org/10.7717/PEERJ-CS.2759/SUPP-2
16. Lachenmeier, W., & Wee Ting Lee, K. (2023). Liberica Coffee Development and Refinement Project in Sarawak Malaysia. *Proceedings, 89*(1), 15. https://doi.org/10.3390/ICC2023-14849
17. Lahai, P. M., Aikpokpodion, P. O., Bah, A. M., Lahai, M. T., Meinhardt, L. W., Lim, S., Ahn, E., Zhang, D., & Park, S. (2025). Unveiling the Genetic Diversity and Demographic History of Coffea stenophylla in Sierra Leone Using Genotyping-By-Sequencing. *Plants*, *14*(1), 50. https://doi.org/10.3390/PLANTS14010050
18. Manikandakumar, M., & Karthikeyan, P. (2022). Weed Classification Using Particle Swarm Optimization and Deep Learning Models. *Computer Systems Science and Engineering*, *44*(1), 913–927. https://doi.org/10.32604/CSSE.2023.025434
19. Martins, V. D. C., Godoy, R. L. D. O., Gouveâ, A. C. M. S., Santiago, M. C. P. D. A., Borguini, R. G., Braga, E. C. D. O., Pacheco, S., & Nascimento, L. D. S. D. M. Do. (2018). Fraud investigation in commercial coffee by chromatography. *Food Quality and Safety*, 2(3), 121–133. https://doi.org/10.1093/FQSAFE/FYY017
20. Pai, P., Amutha, S., Basthikodi, M., Ahamed Shafeeq, B. M., Chaitra, K. M., & Gurpur, A. P. (2025). A twin CNN-based framework for optimized rice leaf disease classification with feature fusion. *Journal of Big Data*, *12*(1), 1–27. https://doi.org/10.1186/s40537-025-01148-z
21. *Robusta or Arabica Object Detection Dataset by aefnattanon*. (n.d.). Retrieved May 17, 2025, from https://universe.roboflow.com/aefnattanon/robusta-or-arabica
22. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474
23. Santoso, I., Sholilah, A., Rucitra, A. L., Lu'ayya, N. M., & Choirun, A. (2025). Sustainable Coffee Supply Chain Risk Mitigation Analysis Using the Failure Mode and Effect Analysis. *BIO Web of Conferences*, 165, 02002. https://doi.org/10.1051/BIOCONF/202516502002
24. Shara, P. N., Thomy, Z., Anhar, A., Harnelly, E., Ramlan, R. R., Hasan Kreung Kalee, T., & Banda Aceh, D. (2021). Morphological characterization of some Coffea arabica L. varieties in Gayo Experimental Garden Bener Meriah. *Journal of Physics: Conference Series*, *1882*, 12092. https://doi.org/10.1088/1742-6596/1882/1/012092
25. Tamayo-Monsalve, M. A., Ruiz, E. M., Pulgarin, J. P. V., Ortiz, M. A. B., Arteaga, H. B. A., Rubio, A. M., Alzate-Grisales, J. A., Garzon, D. A., Cano, V. R., Arias, S. O., Osorio, G., & Soto, R. T. (2022). Coffee Maturity Classification Using Convolutional Neural Networks and Transfer Learning. *IEEE Access*, *10*, 42971–42982. https://doi.org/10.1109/ACCESS.2022.3166515
26. Unal, Y., Taspinar, Y. S., Cinar, I., Kursun, R., & Koklu, M. (2022). Application of Pre-Trained Deep Convolutional Neural Networks for Coffee Beans Species Detection. *Food Analytical Methods*, *15*(12), 3232–3243. https://doi.org/10.1007/s12161-022-02362-8
27. Vujovic, Z. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications, 13*(6). https://doi.org/10.14569/IJACSA.2021.0120670
28. Wu, Q., Liu, H., Zhu, H., Wang, C., Wang, H., Han, Z., Zhao, L., & Liu, F. (2025). YOLO_SSP: An Auto-Algorithm to Detect Mature Soybean Stem Nodes Based on Keypoint Detection. *Agronomy*, 15(5), 1128. https://doi.org/10.3390/AGRONOMY15051128

29. Zhang, Q., Ma, D., Yang, Y., Hu, F., Fu, X., Li, G., Zhang, X., Wang, N., Liu, D., Wu, R., Bi, X., Lou, Y., Li, Y., Yu, H., Yan, W., & Li, Y. (2024). Population Genetic Characteristics of the Cultivated Coffea arabica with Whole-Genome Resequencing. *Horticulturae*, 10(11), 1153. https://doi.org/10.3390/horticulturae10111153