Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Large-Scale classification of roasted coffee bean varieties using deep learning and machine learning methods

Emrah Cenk Hastürk

chasturk1@gmail.com

Başkent University

Emre Sümer

Başkent University

Research Article

Keywords: Coffee bean classification, roasted coffee beans, deep learning, convolutional neural networks, image classification, coffee bean dataset

Posted Date: October 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10832816/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Large-Scale classification of roasted coffee bean varieties using deep learning and machine learning methods

E. Cenk Hastürk[1]· E. Sümer

Abstract
The aim of this study is to classify ready-to-consume roasted coffee beans using image processing and artificial intelligence algorithms. Coffee bean images were photographed using a specially designed photography setup. Coffee beans sourced from 15 suppliers were used to create the coffee bean image dataset. A total of 8,417 high-resolution images were obtained. Following the image processing stages, approximately 190,000 images were generated; these images were resized to 244x244 and 299x299 resolution pixels. The Stratified group k-Fold cross-validation method was used to measure the models' performance, with the number of groups set to 10 (k=10). In the deep learning approach, the ResNet-18, ResNet-50, MobileNetV3-Large, EfficientNet-B0, and ConvNeXt-Tiny models were used. The most successful model was EfficientNet-B0, which uses a 299x299-pixel input and achieves an average test accuracy of 99.86% ± 0.09. Classification studies were also conducted using machine learning, with Random Forest and XGBoost models tested. The eXtreme Gradient Boosting algorithm, using the XGB-Deep-1000 configuration, achieved a test accuracy of 97.98%, outperforming the Random Forest model. The success of a hybrid approach was also sought to be measured; within this scope, CNN models were used for feature extraction, whilst the Random Forest machine learning model was used for classification. Besides, a study was conducted on the classification of Arabica coffee species. Eight different types of Arabica beans were classified using various CNN models. In this classification study, which utilized 224x224-pixel images, EfficientNet-B0 achieved a test accuracy of 99.73%. This figure demonstrates that the models could distinguish even the subtle differences between Arabica bean varieties. In a nutshell, the contributions supplied by this study are below:

- First large-scale dataset was created for roasted coffee beans.
- The images are obtained under a controlled environment with a specifically designed photographic setup.
- Four main types of roasted coffee beans which are ready to consume were classified.
- Arabica type coffee beans classified within its species, which is a subtype classification.

Keywords: Coffee bean classification, roasted coffee beans, deep learning, convolutional neural networks, image classification, coffee bean dataset

1 Introduction

Coffee is a beverage produced by roasting and grinding tropical, evergreen coffee plants native to Africa. Coffee forms the basis of countless beverages such as espresso, mocha, and latte. [1] Coffee is at the heart of many social activities and communal rituals. Through its preparation, serving, and cultural roles, coffee enjoys popularity in almost all societies. Coffee consumption facilitates social interaction and strengthens supportive relationships between individuals. For example, 'Turkish coffee culture and tradition' was inscribed on the 'UNESCO List of Intangible Cultural Heritage' in 2013. [2] Turkish coffee is recognized as part of Turkey's cultural heritage. It is frequently featured in literature and songs. It is an essential part of celebrations and ceremonies.

Coffee is both one of the most widely consumed beverages in the world and one of the most important commodities traded internationally. [3] Global coffee production for the 2025/2026 period is expected to reach a record level of 178.7 million bags, an increase of 4.3 million bags compared to the previous year, due to the ongoing recoveries in Vietnam and Indonesia and record production in Ethiopia. Global coffee bean exports are expected to rise by a further 700,000 bags to reach 122.3 million. Global consumption is forecast to reach a record level of 169.4 million bags. In contrast, coffee prices, as measured by the International Coffee Organization

1 Emrah Cenk Hastürk ✉ chasturk1@gmail.com
Emre Sümer ✉ esumer@baskent.edu.tr
Department of Computer Engineering, Başkent University, Ankara, Turkey

(ICO) monthly composite price index, have risen by over 90% in the last two years. [4] [5]

1.1 Problem definition

Once processed to similar roast levels, coffee beans look very similar to one another. For this reason, it is difficult to distinguish beans by appearance. Experts find it difficult to identify high-priced coffees that have been mixed with a small number of low-priced beans. [6]

Deceptive practices in coffee production, such as blending beans of varying economic value, incorporating non-coffee substances, and mislabelling, are used to reduce costs and increase profits. Consequently, verifying coffee authenticity is critical for accurate value assessment and consumer protection. [7]

For fraud detection in roasted and ground coffee, ultraviolet-visible absorption spectroscopy [8], electrochemical methods [9] [10], chromatographic methods [11], liquid chromatography coupled with mass spectrometry [12] [13] gas chromatography coupled with mass spectrometry [13], nuclear magnetic resonance spectroscopy [12] [13], and various other techniques have been used to characterize and classify coffee beans. However, these methods require sample pretreatment, are time-consuming, and may produce subjective results. Therefore, research has been conducted on other analytical techniques that could be more reliable, reproducible, and widely applicable. [14]

The rise of specialty coffee has changed people's standards regarding coffee quality. [15] This shift in consumer expectations is increasing the need for more advanced quality control systems and creating pressure in this direction.

Considering the coffee cultivation, identifying the most suitable varieties is becoming increasingly important for sustainable production and marketing. However, due to weak organizational structures and the absence of a professional coffee seed industry, a large proportion of existing coffee farms (and even seed suppliers and plant nurseries) do not even know which varieties they are using. [16]

1.2 Objective of the study

The primary objective of this study is to automatically classify roasted coffee beans without human intervention using image processing techniques and artificial intelligence models. As part of the study, four main types of coffee beans—Arabica, Robusta, Liberica, and Excelsa—were photographed at high resolution in a controlled shooting environment. Classification algorithms were developed using these photographs.

1.3 Contributions of the study

The main benefits of automating the coffee classification process include reduced costs, faster evaluation, and standardized classification. [17]

Determining the types of coffee beans is crucial to ensure consumers have access to the correct product. As it protects consumers economically, this is particularly important for specialty coffee varieties with significant price differences. [16] In the long term, this accuracy fosters transparency and a sustainable environment of trust among coffee brands, thereby encouraging conscious consumption behaviour. The use of AI to classify coffee beans will have a positive impact on the gourmet coffee market. The automatic classification of bean types will enhance product standardization and reduce the risk of mislabelling.

The creation of a new, original, and well-labelled image dataset within the scope of this study will make significant contributions to both the academic literature and industry. Since many published datasets are small-scale, homogeneous, or narrowly scoped, their use across different models remains limited. With the dataset we have created; comparisons between different classification models can be made, and studies with high academic validity will be supported.

Unlike most previous studies that focus on green coffee beans or relatively small datasets, this study addresses the classification of roasted coffee beans using a large-scale image dataset collected under controlled conditions. In addition, the use of group-based cross-validation prevents image patches originating from the same source image from being distributed across training and test sets. Therefore, the main novelty of this study lies in combining a large-scale roasted coffee bean dataset with a leakage-aware validation strategy and multiple machine learning and deep learning approaches.

2 Literature review and previous studies

This section provides an overview of the literature. A semantic search was conducted on previous studies on the classification of coffee beans. The search was based on the themes addressed in this article. Information regarding coffee bean types was collected. Species names (Arabica, Robusta, Liberica, and Excelsa), regional types, total number of classes, and geographic origin information were extracted. The characteristics of the dataset were examined. The number of collected images, the number of images per coffee bean type/class, image acquisition methods, data augmentation processes used, and measures taken to address data imbalance were considered. Considering the importance of image processing methods, information on this topic was gathered. Since the main objective was classification, the classification methods applied were evaluated. The performance results obtained from the studies were examined. Performance metrics, the model with the best performance, and related metric details information were collected.

Forty studies meeting these criteria were selected, and conclusions were drawn from them. The analysis reveals growing academic and industrial interest in the automatic classification of coffee beans. The methods used have evolved over time. In early studies, classical machine learning algorithms were frequently preferred. Recent research has particularly used deep learning methods. It is worth noting that research has concentrated on specific regions. Many studies have focused on coffee varieties originating from Indonesia, the Philippines, and Ethiopia. The classifications range from simple tasks, such as binary classification (e.g., distinguishing between Arabica and Robusta types), to multi-class problems involving numerous regional varieties.

Three studies classifying the four main coffee bean types were identified. The study titled “Analysis of Shape Features by Applying Gain Ratio and Machine Learning for Coffee Bean Classification” [18] was examined. In this study, the work was conducted using 400 images, with 100 images from each coffee bean type. It is reported that the classification accuracy reached 99.5%. Upon examining the shared sample image, it appears that the coffee beans' roast levels differ.

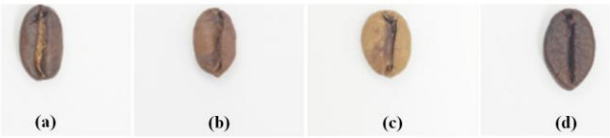

Fig. 1 Images of coffee beans from the study identified by reference number 28: (a) Arabica, (b) Excelsa, (c) Liberica, (d) Robusta.

The study titled “Convolutional Neural Network Architectures for Classification of Coffee Varieties in the Philippines” [19] was reviewed. The raw dataset consists of 500 samples for each coffee variety, a total of 2,000 images. To expand the dataset, data augmentation techniques such as horizontal flipping, vertical flipping, and rotation were applied. As a result of these operations, an additional 2,500 images were obtained for each coffee variety. Thus, 12000 images were included in the final dataset. The images were divided into training (70%), validation (20%), and test (10%) subsets. Unroasted (raw) coffee beans were used in the study. It is reported that the classification accuracy reached 90%.

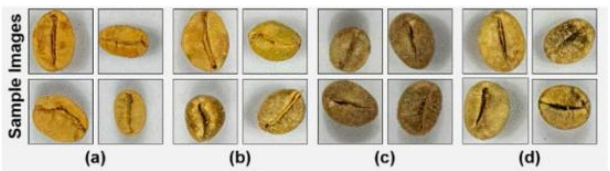

Fig. 2 Images of coffee beans from the study identified by reference number 29: (a) Arabica, (b) Excelsa, (c) Liberica, (d) Robusta.

The study titled “Enhanced Classification of Coffee Bean Varieties in the Philippine Market using MobileNetV1” [20] was reviewed. It is noted that the coffee beans used in the study were unroasted (green) beans. It is stated that 12,000 coffee bean image samples were used. The dataset used in the study “CNN Architectures for the Classification of Coffee Varieties in the Philippines” [19] was used. It is reported that the classification accuracy reached 94.33%.

Overall, previous studies have generally focused on green coffee beans, relatively small datasets, or coffee varieties from limited geographical regions. Furthermore, many studies have relied on conventional train-test splits, while the potential similarity between images originating from the same source has received limited attention. These limitations indicate a need for larger and more diverse roasted coffee bean datasets together with validation strategies that reduce the risk of data leakage.

3 Types of coffee beans

When evaluating coffee bean varieties, four distinct types emerge as the primary varieties: Robusta, Arabica, Liberica, and Excelsa. Global coffee farming, and consequently the global coffee supply chain, is primarily based on two varieties: Arabica (Coffea arabica) and Robusta (Coffea canephora). The other two species—Coffea liberica, also known as Liberica coffee (C. liberica var. liberica), and Excelsa coffee (C. liberica var. dewevrei)—account for a smaller share of global production. However, cultivation of these species is steadily increasing due to climatic challenges affecting Arabica and Robusta, as well as rising market demand. [21]

The Arabica coffee plant is a small tree that can grow to 2 to 8 meters tall. [22] Its scientific name is Coffea arabica L. Coffea arabica, commonly known as Arabica coffee, is a flowering plant species belonging to the Rubiaceae family. It is currently the dominant species, accounting for approximately 60% of global coffee production. [23]



Fig. 3 Flowers of arabica coffee, Aaron Davis © RBG Kew

The origins of the Robusta coffee bean, scientifically known as Coffea canephora, date back hundreds of thousands of years. It evolved in the humid, shaded evergreen forest ecosystems of tropical Africa [24], Coffea canephora is one of the two species commercially cultivated worldwide. It is easily distinguishable visually due to its large. [25]



Fig. 4 Flowers of robusta coffee, Timothy A. Gonsalves

Coffea liberica is described as a rare and even hard-to-find variety among commercially cultivated coffee species. This is because Coffea liberica accounts for only 2% of global coffee production. [26] This species is commercially cultivated in five countries: Indonesia, Malaysia, Vietnam, Ethiopia, and the Philippines. Coffea liberica accounts for 90% of the coffee grown in Malaysia. [27]



Fig. 5 Liberica coffee flowers and fruits [28]

Excelsa coffee is classified as a variety of Coffea liberica and is scientifically named C. liberica var. dewevrei. [21] Coffea liberica var. dewevrei (Excelsa coffee) is a coffee variety native to West Africa. This variety is considered one of the rarest coffee species in the world and was discovered in 1904. [29]



Fig. 6 Coffea liberica var. dewevrei (Excelsa coffee) is cultivated in Central Uganda. [30]

4 Creation of the data set

An original custom photography setup was used in this study. A wooden table was designed to stabilize the camera.

A center point was marked on the table, and a hole large enough to accommodate the camera's macro lens was drilled at that point. Supports(stands) were added to secure the camera. This ensured that the camera remained in the exact same position every time it was placed in the setup. A product tent was procured to maintain consistent conditions such as lighting, distance, and angle, to obtain strong lighting from all angles, and to create a high-quality shooting environment. The wooden photography table was constructed to fully enclose this product tent.

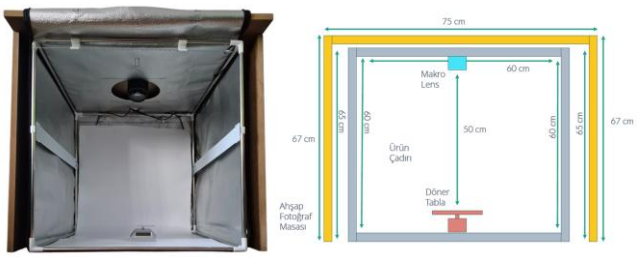

Fig. 7 Front view and dimensions of the photography setup

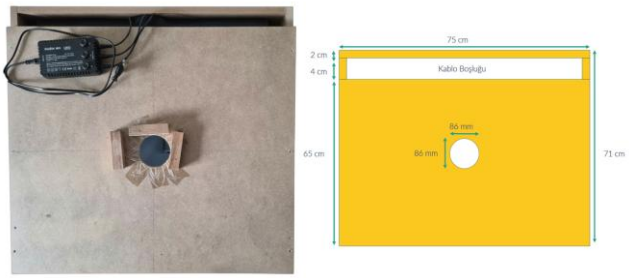

Fig. 8 Top view and dimensions of the photography setup

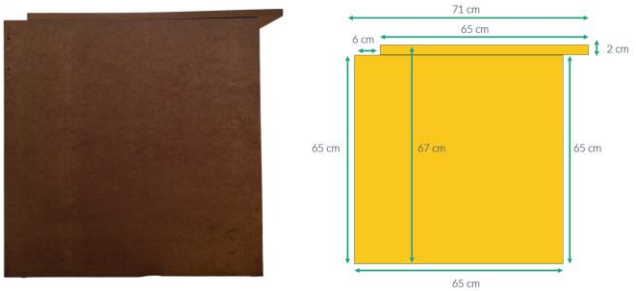

Fig. 9 Side view and dimensions of the photography setup

Images of coffee beans were acquired with a Nikon D600 SLR camera. The Nikon D600 SLR has a full-frame 35.9 × 24.0 mm CMOS sensor (Nikon FX format). The camera captures photos at a high resolution of 24.3 megapixels. This allows for detailed visualization of the coffee beans' texture, colour distribution, and morphological characteristics. A Nikon 105mm f/2.8 macro lens was used on the camera. The lens features a precise autofocus mechanism, which minimizes focus-related blur and image loss. A Godox LST60 model product tent was used for the coffee bean photography as part of this thesis. The product tent is a cube with 60 cm sides. Thanks to the three LED strips inside, a consistent, homogeneous colour temperature and light intensity were achieved. The light was evenly distributed over the coffee beans, preventing the formation of shadows. A colour temperature of approximately 5800K was achieved, which is very close to natural daylight.

To capture photographs of the coffee beans from different angles, a rotating turntable was placed inside the product tent. The coffee beans were placed on this turntable for imaging. In each shot, 6 photos were taken of the beans resting on the turntable, with a 4-second interval between frames. After the 6 frames, the beans were replaced, and a new series of 6 images was started. The imaging parameters were as follows:

- Focal length: 105 mm
- Aperture: f/9
- Shutter speed: 1/125 sec
- ISO: 100

The images were obtained from 15 different coffee beans. The coffee bean type and the number of images taken are listed below.

Table 1 Number of images per coffee bean

Coffee bean	Type	Number of images
Anisah Coffee Colombia	Arabica	510
Anisah Coffee Guatemala	Arabica	512
Anisah Coffee Kenya	Arabica	512
Brownberry Colombia Supremo	Arabica	517
Brownberry Guatemala Antigua Pastoral	Arabica	514
Brownberry Robusta Cherry AA	Robusta	510
Brownberry Sumatra Mandheling	Arabica	516
Brownberry Yirgacheffe Gr1 Natural	Arabica	511
Glowing Mumbai	Arabica	513
Schiller India Robusta	Robusta	507
Uganda Luwero Liberica	Excelsa	757
Uganda Luwero Liberica	Liberica	759
Vietnam 96B Liberica	Excelsa	506
Vietnam Excelsa	Excelsa	765
Zekahve Robusta	Robusta	508

Table 2 Number of images by coffee bean type

Coffee bean type	Number of images
Arabica	4105
Excelsa	2028
Liberica	759
Robusta	1525

5 Data preprocessing

The original images are 6016x4016 pixels in size and approximately 7.1 MB in file size. AI-based image classification models typically work with image dimensions of 224x224, 256x256, or 299x299. Therefore, these images were resized to dimensions compatible with AI models. If rectangular images are resized to a square format without any preprocessing, the image shape will change. This will result in images that are compressed on the left and right sides. This compression will distort the coffee shapes, affecting both training and results. To prevent this image distortion, all images were cropped before being resized to a square dimension, such as 224x224. This cropping process was performed in two different ways.

1. Three square images, each 4016x4016 pixels in size, were generated from each original photo.

Since the short side of the photo is 4016 pixels, the largest possible square images were created at 4016x4016 pixels. The images were cropped from the right, left, and center. This ensured that square images were obtained without losing any of the coffee bean images in each photo.

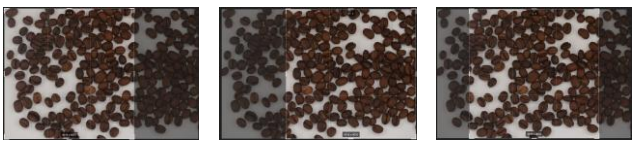

Fig. 10 4016x4016-pixel squares extracted from the original images

2. 24 square images, each 1004x1004 pixels in size, were generated from each original photo.

When evaluating the 6016 pixels along the X-axis, the value 1004 does not provide an exact division. Therefore, the images were cropped so that the minimum number of pixels overlap, to a degree that can be ignored. Starting from the top-left corner and proceeding to the bottom-right, 16 square images, each 1004x1004 pixel, were captured and saved. Starting from the top-right corner and proceeding to the bottom-left, 8 adjacent 1004x1004-pixel frames were captured and saved. Using this method, by overlapping only an 8-pixel-wide area along the horizontal axis, 24 1004x1004-pixel coffee bean images were obtained from the original image. An 8-pixel overlap area is quite small enough to be easily ignored. The size of the overlapped area is equivalent to the region highlighted in red below.



(a) **(b)** **(c,)**

Fig. 11 1004x1004-pixel square images extracted from the original images. (a) 16 frames taken from the left (b) 8 frames taken from the right (c) 8-pixel overlap

It is expected that each image used in the classification study will contain at least one meaningful coffee bean photograph. All images with a size of 4016x4016 were included in the study. It was determined that not all the cropped images contained meaningful coffee bean images. Therefore, a screening process was conducted to eliminate cropped images that did not contain meaningful content within 1004x1004 pixel images. Histogram analysis was utilized in the image screening process. Each image was first converted to grayscale. Histogram values were calculated, and it was decided to use images where pixel values in the 0–100 range accounted for at least 22%. The number of images included in the training after the image splitting and filtering processes is listed below.

Table 3 The number of coffee beans by type included in the training

Coffee bean type	3 square images	24 square images included in training	Total images
Arabica	12315	90126	102441
Excelsa	6084	34523	40607
Liberica	2277	10807	13084
Robusta	4575	29270	33845

6 Technical Infrastructure

The coffee bean classification studies were conducted using an HP Omen Max laptop. The computer has Intel® Core™ Ultra 9 275HX (up to 5.4 GHz with Intel® Turbo Boost Technology) processor, 32 GB DDR5 RAM and NVIDIA® GeForce RTX™ 5080 Laptop GPU (16 GB GDDR7 dedicated).

7 Training techniques and optimization

The Stratified Group K-Fold cross-validation method was chosen to measure model performance. Training was conducted in 10 groups by setting k=10. Consequently, the train and test data are split: 90% and 10%. A 15% portion of the training dataset was allocated for validation. To ensure the repetition of the study, a seed value was assigned to 42 (seed=42). Two factors were considered when selecting this method:
a. The number of images per coffee bean is not equal. Therefore, the stratified method was chosen.
b. Since we trained the model using patch images, it was necessary to ensure that data from the same source did not appear in the training, test, and validation sets. For this reason, a clustering process was performed to separate images within the same data group.

Table 4 Training, validation and test data sets

Coffee bean type	Total	Training (%76,5)	Validation (%13,5)	Test (%10)
Arabica	102441	78367	13830	10244
Excelsa	40607	31064	5482	4061
Liberica	13084	10009	1766	1308
Robusta	33845	25891	4569	3385
Total	189976	145331	25647	18998

In each fold of the k-fold cross-validation method, a maximum of 25 training epochs was run. Too many epochs would cause the model to overfit the images. To prevent this, an early-stopping method was applied. [31] [32] If the validation loss value remains below 0.001 for five consecutive epochs (δ=0.001), it is determined that the model is no longer improving, and training is terminated. Also, the training and validation datasets are completely independent.

Assigning different weighting factors to classes is considered. To prevent the model from neglecting classes with fewer data samples, Weighted Cross-Entropy Loss was used. Weights were assigned to each class in inverse proportion to the number of samples.

Minority classes are penalized more heavily in the error calculation. A Weighted Random Sampler ensures that image types with fewer data samples are more likely to be included in training. As a result, classes with fewer data samples are represented more frequently in training.

The frequency at which training data is presented to the model has been adjusted. In each batch, samples are drawn in greater proportion from the underrepresented classes. As a result, the class distribution within each batch is balanced.

If the model sees the same data in every epoch, it will result in overfitting. To prevent this, data augmentation has been applied. This encourages the model to focus on distinguishing features of the images rather than memorizing their patterns. The method does not increase the total amount of data but enhances data diversity. To create data diversity, images are:

- Horizontally flipped with a 50% probability.
- Vertically flipped with a 30% probability.
- Rotated randomly within a range of +15 degrees and -15 degrees. The rotation angle is randomly determined within these ranges.
- Random changes were applied to the brightness, contrast, saturation, and hue values. The randomness range was set to 0.2. The image's brightness, contrast, and saturation levels were each adjusted to a value between 0.8 and 1.2, independently of one another. The coefficient for hue changes was set with a randomness range of 0.1. Based on a randomly determined coefficient, the hue was updated to a value between 0.9 and 1.1.

With this process, data richness increased without an additional data-collection burden. Recognition rates for images taken in real-world, non-homogeneous environments will increase. Success rates in validation measurements and test result values have also increased with this data enrichment work. The model was constructed with a more robust structure to handle changes in color, angle, and image orientation.

Methods to minimize the loss function were employed. Multiple CNN models were used in this study. The Adam (Adaptive Moment Estimation) optimization algorithm was applied to each model. [33] In the Adam algorithm, a standard step size is not used for every parameter. The first moment determines the direction of the update, while the second moment determines its magnitude. There are situations where the Adam optimization method alone is insufficient. During model training, some weights may grow excessively. Growing weights lead the model to over-learn the training data, or in other words, to memorize it. To prevent overfitting, the weight decay (L2 regularization) method was preferred. [34] For weight regularization, a value of 0.0001—generally considered "reasonable" and widely used—was selected. (L2=0.0001) By using this coefficient, the strength of the penalty term added to the loss

function was adjusted, large weight values were suppressed, and the model's generalization ability was enhanced.

A learning rate scheduler (StepLR Scheduler) that reduces the learning rate at specific intervals was used. [35] The learning rate was halved every 5 epoch cycles (gamma=0.5). For this purpose, the values step=5 and gamma=0.5 were set. The learning rate starts at 0.001 at the beginning of the study. This value is reduced by the specified gamma (γ=0.5) at the end of every 5-epoch cycle. As the learning rate is reduced, the model takes smaller steps as it approaches the minimum point, resulting in more stable convergence. Accordingly, the learning rates that will be formed based on the number of epochs are as follows:

Table 5 Learning rates based on the number of epochs

Epoch	Learning rates
1 - 5	0.001
6 - 10	0.0005
11 - 15	0.00025
16 - 20	0.000125
21 - 25	0.0000625

To keep the magnitude of the gradients calculated during training within a certain range, the "gradient clipping" method was employed. [36] This method prevented large jumps in the model's steps. This ensures that the optimization proceeds in a more controlled manner toward the minimum point. The threshold value used for gradient clipping was set to 1.0 (max_norm=1.0). This is an upper limit. When gradient values are small, the training process is not interfered with. Clipping is applied when gradient values exceed the detected upper limit.

8 Classification results of deep learning models

The classification study utilized the following CNN models: Residual Network – 18 (ResNet–18) [37], Residual Network – 50 (ResNet – 50) [37], MobileNet V3 – Large [38], EfficientNet-B0 [39], and ConvNeXt-Tiny [40].

8.1 Classification results for a single batch (k=1) using 224x224-pixel images

A classification study was conducted using 224x224-pixel coffee bean images to classify four main coffee bean groups. The classification study was performed for a single batch only (k=1). The results and durations of the study are presented in the tables below:

Table 6 Results of all models on a single batch (k=1) using 224x224-pixel images and training durations

Model	Early Stopping Epoch	Accuracy	F1	Average time per epoch	Total training duration
ResNet-18	17	0,9955	0,996	~8.2m	~2h 19m
ResNet-50	25	0,9984	0,998	~8.0m	~3h 21m
MobileNetV3-Large	22	0,9981	0,998	~8.2m	~3h 10m
EfficientNet-B0	16	0,9979	0,998	~7.8m	~2h 05m
ConvNeXt-Tiny	25	0,9379	0,943	~10.6m	~4h 25m

When evaluating the results from classifying 224×224-pixel images, the ResNet-50 achieved the highest performance, with 99.84% accuracy and 0.9981 F1 score. The other models, except ConvNeXt-Tiny, also produced similar results with accuracy rates above 99%. The ConvNeXt-Tiny model's accuracy ranged between 93% and 96%. The ConvNeXt-Tiny performed worse despite reaching its maximum training epoch count of 25. This performance indicates that this architecture may require a longer training process or different hyperparameters.

8.2 Classification results for a single batch (k=1) using 299x299-pixel images

A classification study was conducted using 299x299-pixel coffee bean images to classify four main coffee bean groups. The classification study was performed for a single batch only (k=1). The results and durations of the study are presented in the tables below:

Table 7 Results of all models on a single batch (k=1) using 299x299-pixel images and training durations

Model	Early Stopping Epoch	Accuracy	F1	Average time per epoch	Total training duration
ResNet-18	12	0,9933	0,9934	~10.1m	~3h 31m
ResNet-50	16	0,9955	0,9957	~9.3m	~1h 51m
MobileNetV3-Large	21	0,9986	0,9987	~44.5m	~11h 52m
EfficientNet-B0	16	0,9993	0,9993	~9.8m	~3h 25m
ConvNeXt-Tiny	21	0,8733	0,8799	~9.6m	~2h 33m

With an input size of 299×299 pixels, EfficientNet-B0 achieved the highest performance across the entire dataset, with 99.93% accuracy and an F1 score of 0.9993. The MobileNetV3-Large and ResNet-50 models also demonstrated high performance. ConvNeXt-Tiny also showed relatively low performance at this size. This can be interpreted as an indication of the architecture's sensitivity to specific data structures and input sizes. It is observed that

the average runtime per epoch for the ResNet-50 model is significantly higher than that of the other models. The reason for this is that the total volume of activation tensors (data size) stored in memory with 299x299-sized inputs exceeds the GPU's cache capacity. Any data exceeding the GPU's cache is re-read from main memory.

8.3 EfficientNet-B0 model classification results for 224x224 pixel inputs

To measure the reliability of the results performed on a single fold and to evaluate the results of a model's performance across all folds, the EfficientNet-B0 model was run on 224x224 input images. This study was conducted with the expectation that the results of all groups would yield a low standard deviation. Achieving a low standard deviation indicates that the results of a study conducted with any given dataset can be generalized to the broader population. The EfficientNet-B0 model was trained using 189,977 images and 10-fold cross-validation. The results are presented in the table below:

Table 8 Training results of the EfficientNet-B0 model with 224x224 input images across 10 groups (k=10)

Fold	Early stopping epoch	Training duration	Accuracy	Precision	Recall	F1
1 / 10	16	~2h 35m	0.9972	0.9965	0.998	0.997
2 / 10	25	~4h 05m	0.9985	0.9982	0.999	0.999
3 / 10	22	~3h 45m	0.999	0.9987	0.999	0.999
4 / 10	21	~3h 39m	0.9988	0.9985	0.999	0.999
5 / 10	10	~2h 09m	0.9901	0.9861	0.992	0.989
6 / 10	21	~4h 30m	0.9978	0.9972	0.998	0.998
7 / 10	25	~5h 22m	0.9983	0.9981	0.998	0.998
8 / 10	23	~4h 56m	0.9986	0.998	0.999	0.999
9 / 10	18	~4h 04m	0.9965	0.995	0.998	0.997
10 / 10	19	~4h 19m	0.9975	0.9968	0.998	0.998

The standard deviation calculated based on the test accuracy range is approximately 0.002489. Based on this value, the intergroup standard deviation is 0.0025, which is very low. Therefore, it can be concluded that the result obtained from any selected group is representative of the entire set. Given the standard deviation, the average test accuracy was 99.73% ± 0.25. Early stopping was successful in 8 out of 10 groups. Two groups reached the full number of epochs. The average early stopping point was 20 epochs. Group 5 ended very early, so its performance value was lower than that of the other groups.

8.4 EfficientNet-B0 model 224x224 pixel inputs MCC and Cohen's Kappa scores

The data in this study is imbalanced. Arabica coffee beans are %53.9 of the total data set, whereas Liberica beans are %6.9 of the total data. Accuracy can be misleading under this imbalanced data situation. A model can predict the majority class and achieve a high accuracy. Therefore, MCC and Cohen's Kappa scores were calculated.

Cohen's Kappa (κ) measures the harmony between the model's predictions and the actual labels. It reveals the model's true contribution compared to a random classifier. κ=1 indicates perfect agreement, while κ=0 indicates performance equivalent to chance.

MCC (Matthews Correlation Coefficient) is a correlation coefficient reduced to a single scalar that takes into account all cells of the confusion matrix (TP, TN, FP, FN) in a balanced manner for multi-class problems. It takes values in the range [−1, +1]; +1 represents perfect classification, 0 represents random prediction, and −1 represents completely inverse classification. It is known to be a more reliable performance indicator than the F1 score, particularly for multi-class and imbalanced datasets.

Table 9 MCC and Cohen's Kappa scores of the EfficientNet-B0 model with 244x244 input images across 10 groups (k=10)

Fold	Early Stop Epoch	Accuracy	Precision	Recall	F1	Cohen's Kappa	MCC
1 / 10	12	0,9943	0,9919	0,997	0,994	0,991	0,991
2 / 10	21	0,9989	0,9989	0,999	0,999	0,9983	0,998
3 / 10	24	0,9982	0,9978	0,998	0,998	0,9971	0,997
4 / 10	25	0,9988	0,9985	0,999	0,999	0,9981	0,998
5 / 10	16	0,9983	0,9977	0,999	0,998	0,9972	0,997
6 / 10	19	0,9975	0,9968	0,998	0,998	0,9961	0,996
7 / 10	19	0,9983	0,998	0,998	0,998	0,9973	0,997
8 / 10	25	0,9984	0,9981	0,999	0,998	0,9975	0,998
9 / 10	12	0,9968	0,9954	0,998	0,997	0,9949	0,995
10 / 10	24	0,9989	0,9989	0,999	0,999	0,9983	0,998
Avg	—	0,9978	0,9972	0,998	0,998	0,9966	0,997
Std	—	—	—	—	—	0,0021	0,002

In the ten-fold cross-validation process, the EfficientNet-B0 model achieved a Cohen's Kappa average score of 0.9966 and an MCC average score of 0.9966. These values far exceed the "almost perfect agreement" threshold which is 0.80. The scores demonstrate the model's exceptional success in visually distinguishing between the four coffee bean varieties.

The model is capable of classifying the minority class, Liberica, with the same level of success as the majority class, Arabica. This outcome confirms the effectiveness of the WeightedRandomSampler and class-weighted loss function strategies applied during the training process.

In conclusion, it has been demonstrated that the EfficientNet-B0-based deep learning model offers a reliable and generalizable solution for coffee bean variety classification.

8.5 EfficientNet-B0 model classification results for 299x299 pixel inputs

The results of running the EfficientNet-B0 model on 299x299 input images were evaluated. The expectation was to achieve a low standard deviation in accuracy scores. Achieving a low standard deviation indicates that the results of experiments conducted with any given dataset can be generalized to other datasets. Statistical consistency is the goal. With this in mind, the EfficientNet-B0 model was trained on 189,645 images using 10-fold cross-validation. The results are shown in the table below:

Table 10 Training results of the EfficientNet-B0 model with 299x299 input images across 10 groups (k=10)

Fold	Early stopping epoch	Training duration	Accuracy	Precision	Recall	F1
1 / 10	25	~4h 11m	0.9993	0.9994	0.9993	0.9993
2 / 10	11	~1h 39m	0.9967	0.9951	0.9981	0.9966
3 / 10	21	~3h 05m	0.9994	0.9993	0.9996	0.9994
4 / 10	25	~3h 45m	0.999	0.9992	0.9985	0.9989
5 / 10	20	~3h 41m	0.9981	0.9972	0.9986	0.9979
6 / 10	21	~4h 02m	0.9995	0.9995	0.9994	0.9994
7 / 10	25	~3h 48m	0.9991	0.9987	0.999	0.9988
8 / 10	21	~4h 53m	0.9991	0.9986	0.9993	0.9989
9 / 10	21	~3h 12m	0.9988	0.9983	0.999	0.9986
10 / 10	16	~2h 29m	0.9973	0.9969	0.9977	0.9973

The standard deviation calculated based on the test accuracy area was approximately 0.000905. Based on this value, the standard deviation between groups is 0.0009, which is very low. Therefore, it can be concluded that the results obtained from any selected group will comfortably represent the entire dataset. The fact that the standard deviation decreases to 0.0009 for 299x299-pixel inputs indicates that increasing image resolution enhances the consistency of the test accuracy. Given the standard deviation, the average test accuracy is 99.86% ± 0.09.

8.6 The Arabica coffee beans classification results

The classification of four main types of roasted coffee beans forms the main focus of this study. Additionally, a study was conducted on the internal classification of eight types of Arabica beans to contribute to the research and evaluate classification performance. In the classification study of Arabica beans, images with a resolution of 224x224 were used. A total of 102,441 data images were used, divided into 78,368 training, 13,829 validation, and 10,244 test data. The study utilized the ResNet-18, ResNet-50, MobileNetV3-Large, and EfficientNet-B0 models. The models were run for a single class (k=1), and results were obtained. Information regarding model performance and training durations is provided in the tables below.

Table 11 Results of all models on a single batch (k=1) using 224x224-pixel Arabica bean images and training durations

Model	Early Stopping Epoch	Accuracy	F1	Average time per epoch	Total training duration
ResNet-18	8	0.9223	0.922	~3.4m	~27m
ResNet-50	25	0.995	0.995	~3.8m	~96m
MobileNetV3-Large	25	0.9944	0.995	~3.9m	~97m
EfficientNet-B0	25	0.9973	0.997	~3.7m	~93m

Among these four models, EfficientNet-B0 achieved the highest performance. In this challenging task of classifying coffee beans from the same bean family, it achieved a test accuracy of 99.73%. EfficientNet-B0's classification performance in the first epoch also stands out positively compared to the other models.

9 Classification results of machine learning models

In the classification study, the Random Forest (RF) [41] and eXtreme Gradient Boosting (XGBoost) [42] models were used. Different parameters for both models were tested.

9.1 Classification results of Random Forest (RF) model

Classification of the four main varieties of coffee beans was conducted with the Random Forest model. After extracting features from the images, the images were classified using the Random Forest model. The study was conducted on a total of 189,977 images, each with a resolution of 224x224 pixels. These images were divided into 145,333 training, 25,646 validation, and 18,998 test data samples. Training was performed using the Stratified Group KFold method, with the number of groups set to 10 (k=10). The AutoML

approach was applied during training. For each fold, all possible parameter configurations were tested, and the most accurate parameter combination was selected based on the F1 score. The 10-fold cross-validation was not fully executed for every parameter combination. To save time, the "successive halving" method was employed. Poor results were eliminated. Therefore, different parameter combinations were tested across varying numbers of folds. The 10-fold cross-validation was not performed for all combinations.

Table 12 Average performance of the Random Forest model parameter combinations used in the study

Model	Fold	Test Acc	Precision	Recall	F1	Std Acc
RF-Regularized-500	10	0.9274	0.9287	0.8758	0.8995	0.0018
RF-BalSub-500	10	0.9262	0.9309	0.8716	0.8978	0.0016
RF-Default-500	5	0.9214	0.9344	0.8604	0.8923	0.0016
RF-Large-1500	6	0.9205	0.9329	0.8592	0.891	0.0016
RF-Deep-1000	5	0.9198	0.9332	0.8596	0.8913	0.0017
RF-Log2-750	3	0.8059	0.8757	0.6645	0.7238	0.0016
AutoML-Best	10	0.9274	0.9287	0.8758	0.8995	0.0018

RF-Regularized-500 was identified as the best parameter combination and was run for 10-fold cross-validation. The performance metrics and runtime for each fold are detailed in the table below. RF-Regularized-500 achieved the highest test accuracy (92.74%) and the highest F1 score (0.8995).

Table 13 Results of the Random Forest model RF-Regularized-500 parameter set for 10 batch (k=10)

Fold	Accuracy	Precision	Recall	F1	Training duration (s)	Total time in minutes
1 / 10	0.9285	0.9313	0.8783	0.902	611.8	25.7
2 / 10	0.9288	0.9311	0.8824	0.9043	616.5	25.8
3 / 10	0.9276	0.9277	0.8782	0.9004	607.5	25.7
4 / 10	0.9276	0.9306	0.8746	0.8997	623.8	25.9
5 / 10	0.9261	0.9247	0.873	0.8963	773.4	28.4
6 / 10	0.9253	0.9306	0.8737	0.8989	615.5	25.8
7 / 10	0.9248	0.9243	0.8713	0.8949	602.1	25.6
8 / 10	0.9253	0.9246	0.8689	0.8938	614.4	25.8
9 / 10	0.9303	0.9323	0.8822	0.9048	607.8	25.7
10 / 10	0.9294	0.93	0.875	0.8996	614.8	25.8
Average	0.9274	0.9287	0.8758	0.8995	628.8	26

9.2 Classification results of eXtreme Gradient Boosting (XGBoost) model

A study was conducted to classify coffee beans into four main categories using the eXtreme Gradient Boosting model. First, features were extracted from the images, and the images were then classified using the Extreme Gradient Boosting model. The study was conducted on a total of 189,977 images, each with a resolution of 224x224 pixels. These images were divided into 145,333 training, 25,646 validation, and 18,998 test data sets. Training was performed using the Stratified Group KFold method, and the number of groups was set to 10 (k=10). The AutoML approach was applied during training. For each fold, all different parameter configurations were tested, and the most accurate parameter combination was selected based on the F1 score. To save time, the "successive halving" method was used. Parameters yielding poor results were eliminated. Therefore, different parameter combinations were tested by varying the number of folds. Not all combinations were tested across all 10 groups.

Table 14 Average performance of the eXtreme Gradient Boosting model parameter combinations used in the study

Model	Fold	Test Acc.	Precision	Recall	F1	Std Acc.
XGB-Deep-1000	10	0.9798	0.9656	0.9686	0.9671	0.001
XGB-Default-500	10	0.9778	0.9607	0.9687	0.9646	0.001
XGB-SubCol-750	10	0.9771	0.9597	0.9677	0.9636	0.0012
XGB-Large-2000	3	0.9766	0.9578	0.9686	0.963	0.0005
XGB-Regularized-500	3	0.9741	0.9529	0.9678	0.96	0.0002
XGB-LowLR-1500	3	0.9615	0.9297	0.9572	0.9421	0.0004
AutoML-Best	10	0.9798	0.9656	0.9686	0.9671	0.001

XGB-Deep-1000 has been identified as the optimal combination of parameters. The performance metrics and runtime for each fold of the XGB-Deep-1000 algorithm are detailed in the table below.

Table 15 Results of the eXtreme Gradient Boosting model XGB-Deep-1000 parameter set for 10 batch (k=10)

Fold	Accuracy	Precision	Recall	F1	Training duration (s)	Total time in minutes
1 / 10	0.9803	0.967	0.9698	0.9684	695.7	24.6
2 / 10	0.9801	0.9678	0.9709	0.9693	713.5	24.9
3 / 10	0.9803	0.9655	0.9697	0.9675	842.4	27.1
4 / 10	0.9792	0.9654	0.9665	0.966	854.6	27.3
5 / 10	0.9777	0.9636	0.9654	0.9645	850.1	27.2
6 / 10	0.9805	0.9653	0.9686	0.9669	859.4	27.4
7 / 10	0.9791	0.9642	0.9664	0.9652	861.6	27.4
8 / 10	0.9806	0.9654	0.968	0.9667	774.1	25.9
9 / 10	0.9816	0.9689	0.9726	0.9707	728.8	25.2
10 / 10	0.979	0.9632	0.9682	0.9656	857.6	27.3
Average	0.9798	0.9656	0.9686	0.9671	798.8	26.4

Among the various parameter combinations tested, the XGB-Deep-1000 model performed the best. With parameters set to n=1000, max_depth=8, and lr=0.05, the test accuracy reached 97.98%. The standard deviation of the XGB-Deep-1000 model's results was calculated to be 0.0010. This standard deviation value also indicates that the model produces consistent results.

10 Feature extraction using CNN and classification using a Random Forest model

In addition to studies using CNN and classical machine learning models, a classification study was conducted using a hybrid approach that involved CNN-based feature extraction, followed by classification with a machine learning model. The study was conducted on a total of 189,977 images with a resolution of 224x224 pixels. These images were divided into 145,333 training, 25,646 validation, and 18,998 test data sets. Training was performed using the StratifiedGroupKFold method, with the number of groups set to 10 (k=10). The Random Forest model was selected for classification as part of the classical machine learning approach. To determine the features to be used in the Random Forest model, experiments were conducted using the ConvNeXt-Tiny, MobileNetV3-Large, EfficientNet-B0, ResNet-50, and ResNet-18 CNN backbone models. Among all CNN backbone models, the RF-Regularized-500 parameter set demonstrated the best performance during feature selection. Therefore, it was decided to conduct the classification study using this configuration. The classification results obtained using the Random Forest RF-Regularized-500 model on a single group (k=1 fold) with different CNN backbone models are presented below.

Table 16 Classification results obtained using the Random Forest RF-Regularized-500 model on a single group (k=1 fold) with different CNN backbone models

CNN Backbone	Accuracy	Precision	Recall	F1
ResNet-18	0.8464	0.8744	0.8233	0.8165
ResNet-50	0.8813	0.8937	0.8216	0.8581
MobileNetV3-Large	0.9072	0.9255	0.8657	0.8918
EfficientNet-B0	0.9019	0.9151	0.855	0.8815
ConvNeXt-Tiny	0.9255	0.9341	0.8925	0.9115

Based on the results of CNN feature extraction performed for a single group (k=1) followed by classification using a Random Forest model, the two most successful CNN models for feature extraction were MobileNetV3_Large and ConvNeXt-Tiny.

10.1 Classification study conducted using MobileNetV3-Large and Random Forest models

A classification study of the four coffee bean types was conducted for all folds using the MobileNetV3_Large CNN model and the Random Forest RF-Regularized-500 parameters. The performance metrics and execution times for each fold are detailed in the table below.

Table 17 Classification results obtained using the Random Forest RF-Regularized-500 model for all folds (k=10 folds) with the MobileNetV3_Large backbone model

Fold	Accuracy	Precision	Recall	F1	Training duration (s)	Total time in minutes
1 / 10	0.9076	0.9244	0.8672	0.8923	657.2	15.5
2 / 10	0.9106	0.9277	0.8715	0.896	697.6	16.2
3 / 10	0.9115	0.9282	0.8739	0.8979	651.5	15.4
4 / 10	0.904	0.9229	0.8609	0.8878	477.7	12.5
5 / 10	0.9115	0.9263	0.8745	0.8974	478	12.5
6 / 10	0.9098	0.9285	0.8706	0.8961	481.4	12.5
7 / 10	0.91	0.9267	0.8706	0.8953	492.6	12.7
8 / 10	0.9076	0.9241	0.8667	0.8919	548.3	13.7
9 / 10	0.9104	0.9262	0.8746	0.8975	463.1	12.2
10 / 10	0.909	0.9243	0.8688	0.893	463.3	12.2
Average	0.9092	0.9259	0.8699	0.8945	541.2	13.5

The standard deviation calculated based on the test accuracy area is 0.0022. This value demonstrates that the model exhibited highly consistent and stable performance across the 10-fold split. Based on this information, the standard deviation across folds is quite low. Therefore, it can be concluded that the result obtained from any selected fold is representative of the entire dataset. Given the standard deviation, the average test accuracy is 90.92% ± 0.22.

10.2 Classification study conducted using ConvNeXt-Tiny and Random Forest models

A classification study of the four coffee bean types was conducted for all folds using the ConvNeXt-Tiny CNN model and the Random Forest RF-Regularized-500 parameters. The performance metrics and execution times for each fold are detailed in the table below.

Table 18 Classification results obtained using the Random Forest RF-Regularized-500 model for all folds (k=10 folds) with the ConvNeXt-Tiny backbone model

Fold	Accuracy	Precision	Recall	F1	Training duration (s)	Total time in minutes
1 / 10	0.9253	0.9339	0.8923	0.9114	168.6	8.4
2 / 10	0.9221	0.9308	0.8875	0.9071	170	8.4
3 / 10	0.9238	0.9306	0.8929	0.9103	170.6	8.5
4 / 10	0.9204	0.9285	0.8862	0.9054	170.2	8.4
5 / 10	0.924	0.9319	0.8898	0.909	171.3	8.5
6 / 10	0.9182	0.9246	0.8849	0.9031	172.1	8.5
7 / 10	0.9226	0.928	0.8882	0.9065	169.9	8.4
8 / 10	0.9217	0.9276	0.8892	0.9069	170.5	8.5
9 / 10	0.9214	0.9269	0.8858	0.9046	170.8	8.5
10 / 10	0.9189	0.929	0.8832	0.904	170.5	8.5
Avg	0.9219	0.9292	0.888	0.9068	170.5	8.5

The standard deviation of the test accuracy area is 0.0021. This value demonstrates that the model exhibited highly consistent and stable performance across all 10 folds. Based on this information, the interfold standard deviation is quite low. Therefore, it can be concluded that the result obtained from any selected fold is representative of the entire dataset. Given the standard deviation, the average test accuracy is 92.19% ± 0.21.

11 Conclusion and future scope

The objective of this study is to classify four main types of roasted coffee beans using image processing techniques and artificial intelligence algorithms. To achieve this goal, a unique dataset consisting of photographs of coffee beans was created. The dataset was compiled using data processing techniques, and the coffee beans were classified using classical machine learning models, artificial intelligence models, and hybrid approaches. A review of the literature reveals that most studies on coffee bean classification have been conducted on green coffee beans. Since morphological features such as shape, size, and color are preserved in unroasted coffee beans, classification studies on them are relatively easier. The roasting process makes the appearances of coffee beans more similar to one another. The fact that our study was conducted on roasted coffee beans distinguishes it from other studies.

Upon examining the results of classifying coffee beans, it has been demonstrated that roasted coffee beans can be classified with high accuracy using image processing and CNN models. Among the classification methods analysed, EfficientNet-B0 trained on images with an input size of 299x299 pixels, is the most successful model. The model achieved an average test accuracy of 99.86% ± 0.09, an F1 score of 0.9985, and a precision value of 0.9982. The model's low standard deviation indicates that classification results will be successful regardless of which portion of the data is used.

Among the two machine learning models applied, the XGBoost model stood out as the more successful model with a test accuracy of 97.98%. In our study on the classification of Arabica beans within their own category, eight subtypes of Arabica coffee beans were classified with 99.73% accuracy using the EfficientNet-B0 model. This result demonstrates that even subtle differences within the same species can be distinguished.

The high success results in classification are carried out by carefully conducted studies. A large number of high-resolution images were captured using a macro lens in a controlled lighting environment, at a fixed distance and angle, and with the same background, using a specially designed photographic setup. These images were resized to serve as input for classification algorithms and CNN models. The resizing process was carried out meticulously to avoid distortion in the images. Data augmentation has been applied to encourage the model to focus on distinguishing features of the images rather than memorizing their patterns.

Crops that are taken from the same image were never used in both the training and test/validation sets. They are grouped properly and the same group of images are used in either of these processes.

The number of images per coffee bean is not equal. Therefore, the Stratified Group K-Fold cross-validation method was chosen for training. Too many epochs would cause the model to overfit the images, so an early-stopping method was applied. Weighted Cross-Entropy Loss was used to prevent the model from neglecting classes with fewer data samples. Weighted Random Sampler is utilized to ensure image types with fewer data samples are more likely to be included in training. Adam (Adaptive Moment

Estimation) optimization algorithm was employed for dynamic parameter updates.

With all these practices, a high classification success rate was achieved.

The current dataset was constructed using various coffee bean types to the extent possible. A wide variety of coffees are grown in many different regions around the world under varying climatic conditions. In these geographical regions, not only do coffee bean types differ, but coffee processing techniques also vary. Coffee bean roasting levels and methods also vary according to processing techniques. If the dataset can be expanded, the models' generalization capacity can be increased and the necessary conditions for commercial use established.

Another issue is the scalability of the conducted studies across platforms with varying resource capabilities. The usability of the developed solutions on mobile platforms and embedded systems, as well as in resource-constrained environments, can be evaluated.

In this study, a classification of homogeneous bean types was performed. Methods to separate mixed bean types can be developed. Ensuring the active use of the developed models in the industry will support quality control systems. Additionally, mobile applications can be developed to ensure consumers have safe access to products. With the widespread adoption of the classification structure, fraud in the mass-market coffee sector will decrease, market oversight will increase, and evaluations in the specialty coffee sector can be conducted objectively. Organizations such as the Specialty Coffee Association can be supported through such studies.

References

[1] N. Myhrvold, "Encyclopaedia Britannica," 19 09 2025. [Online]. Available: https://www.britannica.com/topic/coffee.

[2] Ministry of Culture and Tourism, "UNESCO Intangible Cultural Heritage," 07 12 2013. [Online]. Available: https://ich.unesco.org/en/8com.

[3] International Coffee Organization (ICO), "International Coffee Day 2024 (ICD)," 2024. [Online]. Available: https://ico.org/international-coffee-day/.

[4] United States Department of Agriculture Foreign Agricultural Service, "Coffee: World Markets and Trade," Foreign Agricultural Service/USDA Global Market Analysis, Washington DC, 2025.

[5] United States Department of Agriculture, Foreign Agriculture Service, "Production - Coffee," United States Department of Agriculture, Foreign Agriculture Service, Washington DC, 2025.

[6] J.-J. Tsai, "Analysis and classification of coffee beans using single coffee bean mass spectrometry with machine learning strategy," *Food Chemistry,* vol. 423, 15 11 2023.

[7] M. Perez, "Current strategies to guarantee the authenticity of coffee," *Critical Reviews in Food Science and Nutrition,* pp. 539 - 554, 19 July 2021.

[8] A. Belay, "Measurement of caffeine in coffee beans with UV/vis spectrometer," *Food Chemistry,* vol. 108, pp. 310-315, 2008.

[9] R. Wada, "Electrochemical Analysis of Coffee Extractions at Different Roasting Levels Using a Carbon Nanotube Electrode," *Analytical Sciences,* vol. 37, pp. 377-380, 2021.

[10] I. Tomac, "Evaluation of antioxidant activity of chlorogenic acids and coffee extracts by an electrochemical DNA-based biosensor," *Food Chemistry,* vol. 325, 2020.

[11] A. Yashin, "Chromatographic methods for coffee analysis: a review," *Journal of Food Research,* no. 6, p. 60, 2017.

[12] M. Gigl, "Identification and quantitation of reaction products from quinic acid, quinic acid lactone, and chlorogenic acid with strecker aldehydes in roasted coffee," *Journal of Agricultural and Food Chemistry,* vol. 69, pp. 1027-1038, 2021.

[13] M. Farag, "Metabolite profiling of premium civet luwak bio-transformed coffee compared with conventional coffee types, as analyzed using

chemometric tools," *Metabolites,* vol. 13, p. 173, 2023.

[14] X. Wang, "Quantitative lipidomics reveals the effects of roasting degree on arabica coffee beans lipid profiles," *Food Control,* vol. 169, March 2025.

[15] N. Huang, "Smart agriculture: real-time classification of green coffee beans by using a convolutional neural network," *IET Smart Cities,* 31 July 2020.

[16] S. Pruvot-Woehl, "Authentication of Coffea arabica Varieties through DNA Fingerprinting and its Significance for the Coffee Sector," *Journal of AOAC International,* vol. 103, no. 2, p. 325–334, April 2020.

[17] J. P. L. Pizzaia, "Arabica coffee samples classification using a Multilayer Perceptron neural network," in *2018 13th IEEE International Conference on Industry Applications (INDUSCON)*, 2018.

[18] A. Septiarini, "Analysis of shape features by applying gain ratio and machine learning," *Coffee Science,* p. e192206, 28 06 2024.

[19] E. J. P. Bibangco, "Convolutional Neural Network Architectures for Classification of Coffee Varieties in the Philippines," in *2024 IEEE Symposium on Wireless Technology & Applications (ISWTA)*, Kuala Lumpur, Malaysia, 2024.

[20] R. Rama, "https://www.researchgate.net/publication/389205152_Enhanced_Classification_of_Coffee_Bean_Varieties_in_the_Philippine_Market_using_MobileNetV1," *PHILIPPINE JOURNAL OF SCIENCE, ENGINEERING, AND TECHNOLOGY,* vol. 1, no. 2, 25 April 2024.

[21] A. P. Davis, "Genomic data define species delimitation in Liberica coffee with implications for crop development and conservation," *Nature Plants,* no. 11, p. 1729–1738, 8 August 2025.

[22] Royal Botanic Gardens Kew, "Royal Botanic Gardens Kew," Royal Botanic Gardens, [Online]. Available: https://www.kew.org/plants/arabica-coffee. [Accessed 29 October 2025].

[23] Wikipedia, "Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Coffea_arabica. [Accessed 29 October 2025].

[24] A. R. F. Davis, "Six new species of coffee (Coffea) from northern Madagascar," *Kew Bull,* vol. 76, p. 497–511, 13 August 2021.

[25] World Coffee Research, "World Coffee Research," [Online]. Available: https://varieties.worldcoffeeresearch.org/robusta-2/history-of-robusta. [Accessed 31 October 2026].

[26] M. C. A. Lagman, "Variability in Philippine Coffea liberica Provides Insights into Development Amidst a Changing Climate," in *International Coffee Convention 2023*, Mannheim, Germany, 2023.

[27] M. A. M. A. Wahab, "FFTC Agricultural Policy," Food And Fertilizer Technology Center for the Asia and Pacific Region, 12 January 2016. [Online]. Available: https://ap.fftc.org.tw/article/1005. [Accessed 01 November 2025].

[28] NParks Flora & Fauna Web, "Coffea liberica," NParks Flora & Fauna Web, 28 04 2026. [Online]. Available: https://www.nparks.gov.sg/florafaunaweb/flora/2/8/2820. [Accessed 28 04 2026].

[29] Socfindo Conservation, "Socfindo Conservation," [Online]. Available: https://www.socfindoconservation.co.id/plant/918?lang=en. [Accessed 01 11 2025].

[30] D. AP, "A review of the indigenous coffee resources of Uganda and their potential for coffee sector sustainability and development," *Frontiers in Plant Science,* p. 1057317, 17 February 2023.

[31] L. Prechelt, Early Stopping — But When?, Berlin: Springer, 2012, pp. 53-67.

[32] G. Huang, L. Zhuang, L. Van Der Maaten and K. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 2017.

[33] J. B. Diederik P. Kingma, “Adam: A Method for Stochastic Optimization,” in *3rd International Conference for Learning Representations*, 2014.

[34] I. Goodfellow, Deep Learning, Massachusetts: The MIT Press, 2016.

[35] S. Ruder, “An overview of gradient descent optimization algorithms,” 15 06 2017. [Online]. Available: https://arxiv.org/abs/1609.04747. [Accessed 29 03 2026].

[36] R. Pascanu, “On the difficulty of training recurrent neural networks,” in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, 2013.

[37] K. He, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 2016.

[38] Zhu, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *ArXiv,* vol. abs/1704.04861, 2017.

[39] M. T. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *ArXiv,* vol. abs/1905.11946, 2019.

[40] Z. L. Mao, “A ConvNet for the 2020s,” *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 11966-11976, 2022.

[41] L. Breiman, “Random Forests,” *Machine Learning,* no. 45, pp. 5-32, October 2001.

[42] M. S. Balaji Krishnapuram, “Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,” in *Knowledge Discovery and Data Mining*, New York, 2016.