



Application of Pre-Trained Deep Convolutional Neural Networks for Coffee Beans Species Detection

Yavuz Unal¹ · Yavuz Selim Taspinar² · Ilkay Cinar³ · Ramazan Kursun⁴ · Murat Koklu³

Received: 12 April 2022 / Accepted: 29 July 2022 / Published online: 3 August 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Coffee is an important export product of the tropical countries where it is grown. Therefore, the separation of coffee beans in the world in terms of the quality element and variety forgery is an important situation. Currently, the use of manual control methods leads to the fact that the parsing processes are inconsistent, time-consuming, and subjective. Automated systems are needed to eliminate such negative situations. The aim of this study is to classify 3 different coffee beans by using their images, through the transfer learning method by utilizing 4 different Convolutional Neural Networks-based models, which are SqueezeNet, Inception V3, VGG16, and VGG19. The dataset used in the models' training was created specially for this study. A total of 1554 coffee bean images of Espresso, Kenya, and Starbucks Pike Place coffee types were collected with the created mechanism. Model training and model testing processes were carried out with the obtained images. In order to test the models, the cross-validation method was used. Classification success, Precision, Recall, and F-1 Score metrics were used for the detailed analysis of the models of performances. ROC curves were used for analyzing their distinctiveness. As a result of the tests, the average classification success of the models was determined as 87.3% for SqueezeNet, 81.4% for Inception V3, 78.2% for VGG16, and 72.5% for VGG19. These results demonstrate that the SqueezeNet is the most successful model. It is thought that this study may contribute to the subject of coffee beans of separation in the industry.

Keywords Coffee beans · Deep learning · CNN · Transfer learning

Introduction

Coffee is one of the major commercial products in world trade after petroleum products (Nair 2010). Given its importance in world markets, maintaining quality control of coffee beans gains importance. Because coffee prices vary depending on the types and quality of the coffee (Zhang et al. 2018; Janandi and Cenggoro 2020).

In order to identify the types of coffee beans and classify them according to their physical state, manual techniques are often applied in some food industries. This technique is a manual visual examination carried out by a trained expert. This somewhat subjective process can be tedious and time-consuming. In addition, expert's mood swings, mental states, and emotions such as stress may negatively affect the classification process (Arboleda 2019). Therefore, the use of computer vision applications instead of manual techniques stands out as a better option to classify products and separate them from unwanted substances. Computer vision applications can automatically extract useful information from a particular object in an image and analyze it. A computer

✉ Yavuz Unal
yavuz.unal@amasya.edu.tr
Yavuz Selim Taspinar
ytaspinar@selcuk.edu.tr
Ilkay Cinar
ilkay.cinar@selcuk.edu.tr
Ramazan Kursun
rkursun@selcuk.edu.tr
Murat Koklu
mkoklu@selcuk.edu.tr

¹ Department of Computer Engineering, Amasya University, Amasya, Turkey

² Doganhisar Vocational School, Selcuk University, Konya, Turkey

³ Department of Computer Engineering, Selcuk University, Konya, Turkey

⁴ Guneysinir Vocational School, Selcuk University, Konya, Turkey

vision algorithm, in general, consists of two main stages for defect or class detection. First, the segmentation of objects from the background is performed accurately. Second, diverse physical features of objects are extracted, and, by using classical machine learning methods, these features are classified or the defective ones are detected.

Coffee is a product grown in more than 60 tropical countries around the world. It constitutes a large part of the exports of the countries where they are grown. Coffee beans should be classified so that their quality can be determined before export. Determination of the variety and quality of coffee is carried out according to size, shape, color, acidity of aroma, and taste. The shape of coffee can vary depending on where it is grown. However, due to the similarity in taste, color, and shape, it is quite difficult for the consumer to identify the type of coffee (Waliyansyah and Hasbullah 2021). Food counterfeiting activities and counterfeiting are increasing in coffee commodities. In imitation, it is an attempt to change the appearance made for the purpose of making a large profit by beautifying the appearance of coffee. Coffee beans are usually classified according to expert opinion and tradition. Since these measurements may be inconsistent, the classification of coffee via digital images is of great importance. Unlike human beings, the results of classifications made on digital images have the ability to be more precise, objective, and sensitive (Vogt 2020; Adiwijaya et al. 2022a). In addition, it has superior features compared to manual methods that are boring, time-consuming, expensive, and more inefficient. Due to the mentioned problems, image processing and classification techniques have started to be used for coffee bean classification and determination. However, studies in this area are limited. Therefore, it is necessary to develop this field by carrying out work and integrating it into production facilities (Ansar et al. 2021).

When the literature is examined, it is seen that the researches are carried out with different types of coffee beans, and the classification and defect detection processes of coffee beans are made by different machine learning methods. In some of these studies, processes were performed by using Artificial Neural Networks (ANN) and Naive Bayes (NB) algorithms to classify according to CIE $L^*a^*b^*$ color values (Oliveira et al. 2016; Ropelewska et al. 2022), ANN with utilizing various morphological features, K-Nearest Neighbor (KNN) algorithms (Arboleda et al. 2018) and 22 different machine learning algorithms (Arboleda 2019), Multi-Layer Perceptron (MLP) algorithm with shape, size, and color features (Pizzaiia et al. 2018), and Support Vector Machine (SVM), Deep Neural Network (DNN), Rule inference (Koklu et al. 2014), and Random Forest (RF) algorithms (Santos et al. 2020).

In addition to all these studies, deep learning methods ensure extracting features from images, performing classification, and defect detection based on these features.

Examining the studies carried out by using deep learning methods in the literature, it is seen that there are different studies utilizing Convolutional Neural Networks (CNN) and their architectures (Janandi and Cenggoro 2020; Pinto et al. 2017; Huang et al. 2019, 2020; Rivalto et al. 2020; Fuentes et al. 2020; Kuo et al. 2019). Moreover, there are studies that extract features from images of coffee beans with the help of CNN and compare the results by classifying these features with machine learning methods. In the next chapter, the details of the studies carried out on the coffee bean in the literature will be given.

The objectives of the study can be listed as follows:

- To determine which classification technique is more successful.
- To examine the distinguishability of coffee varieties with images of coffee beans.
- To study to what extent coffee beans are classified.
- To examine the availability of models developed for the classification of coffee beans in automated systems.

The contributions of this study to the literature are given below.

1. A total of 1554 images of 3 different coffee types, which are Espresso, Kenya, and Starbucks Pike Place coffee beans, were obtained through a specially developed mechanism, and a dataset was created.
2. In order to distinguish coffee beans, 4 different CNN-based models were used.
3. The coffee beans were classified with the models created, and different performance metrics were obtained for each model and each class. The average classification success of each model was compared, and the model having the highest success was determined.
4. It is foreseen that the separation of coffee beans will be ensured by the methods performed in this study.

This study consists of five sections. In Section "Introduction", the developed work is contextualized. In Section "Related Works", the coffee classification process is described. In Section "Material and Methods", the developed context is explained. In Section "Experimental Results", the results are presented. Lastly, conclusions and perspectives for future studies take place in Section "Conclusion".

Related Works

In the literature, there are studies conducted with different coffee beans and in which classification processes are carried out with different methods. Researchers, in these studies, have used different machine learning algorithms and

different deep learning models for classification processes. Some of the studies conducted on coffee beans in the literature are given below.

de Oliveira et al. (2016) used Artificial Neural Networks (ANN) and NB classifier to classify green coffee beans according to CIE $L^*a^*b^*$ color values. Coffee beans were classified into 4 groups as whitish, red green, green, and bluish green. A classification accuracy of 100% was obtained with NB (de Oliveira et al. 2016).

Pinto et al. (2017) utilized Convolutional Neural Network to classify beans according to 6 different defect types (black, sour, fade, peaberry, damage, and normal) by using coffee bean images. As a result of the classification, accuracy values of 72% and above were obtained according to the defect types (Pinto et al. 2017).

In the study of Arboleda et al. (2018), various morphological features were obtained from 195 training images and 60 test images in order for the automatic classification of coffee beans. Classification processes were performed with Artificial Neural Networks (ANN) and K-Nearest Neighbor (KNN). The highest classification accuracy was obtained from ANN with 96.66% (Arboleda et al. 2018).

Pizzaia et al. (2018) used Multi-Layer Perceptron (MLP) algorithm by extracting shape, size, and color features in order to classify arabica coffee beans as good and bad. As a result of the study, 94.10% classification accuracy was obtained (Pizzaia et al. 2018).

Fukai et al. (2018) aimed to develop an automatic coffee bean classification system for coffee bean producers by using state-of-the-art machine learning techniques and Raspberry Pi in their studies. As the first step of the system development, they performed classification processes with approximately 13 thousand images of 5 different types of coffee beans through deep Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). Classification results were compared, and it was stated that the results obtained with CNN are higher than those obtained with SVM. Additionally, Raspberry Pi camera module and CNN results were compared, and it was concluded that higher results were obtained with CNN. The classification accuracies obtained through CNN for coffee beans varied between the range of 75% and 95% (Fukai et al. 2018).

Arboleda (2019) extracted 4 morphological features of Liberica, Robusta, and Excelsa green coffee bean types. These features were classified by 22 different machine learning algorithms, and their performances were compared. The results of the study showed that the highest accuracy rate in the classification of green coffee beans was achieved in the Coarse Tree Algorithm with 94.10% (Arboleda 2019).

Huang et al. (2019), in their study aiming to separate imperfect and perfect coffee beans with Convolutional Neural Network, obtained 1000 coffee bean images in the imperfect class and 1000 coffee bean images in the perfect

class. They augmented the data by using rotation and flip methods in the obtained data set. At the end of this process, the network was trained on 72 thousand images, 36 thousand of which are imperfect and 36 thousand are perfect. The authors stated that they obtained an accuracy value of 93.34% as a result of the classification (Huang et al. 2019).

Kuo et al. (2019) proposed a control scheme to detect defects in coffee beans, named Hough Circle-Assisted Deep-Network Scheme. This scheme using a deep learning-based YOLOv3 network targets small circular objects and can accurately detect imperfect beans. With the proposed scheme, high control accuracy and precision were achieved in the detection of defects in coffee beans (Kuo et al. 2019).

Rivalto et al. (2020) used a deep learning-based Convolutional Neural Network to identify the types of coffee beans. Their study were carried out on 617 images of 4 different coffee beans grown in Indonesia. As a result of the training, it was stated that coffee types were classified with an accuracy of 74.26% (Rivalto and Pranowo, and A.J. Santoso. 2020).

Gope and Fukai (2020), in their study, aimed to distinguish the Peaberry, a bean that is relatively rounder and has a different flavor than flat coffee beans, from the others. In accordance with this purpose, classification operations were performed by using deep Convolutional Neural Networks (CNN) and Support Vector Machine (SVM). In the study, images of 1900 normal and 1438 Peaberry bean were used. In addition, classification processes were carried out with 4 different image sizes (32×32 , 64×64 , 128×128 , and 256×256) in the training. As a result, it was stated that higher accuracy values (97% and above) were achieved in the classification results performed with CNN compared to SVM, and the image sizes did not affect the classification results (Gope and Fukai 2020).

Santos et al. (2020) aimed to analyze the importance of coffee beans' shape and color characteristics via different machine learning techniques such as Support Vector Machine (SVM), Deep Neural Network (DNN), and Random Forest (RF) to evaluate the defects of coffee beans. For this purpose, an algorithm written in Python was used in order to extract shape and color features from images of coffee beans. Among the variables used, gmean from RGB (Red, Green, and Blue) color space and Vmean from HSV (Hue, Saturation, and Value) color space stand out as the most relevant features for classification models. Moreover, it was stated that accuracy values of 88% and above were obtained as a result of the classification processes of coffee beans performed with SVM, DNN, and RF machine learning techniques (Santos et al. 2020).

Janandi and Cenggoro (2020) developed a deep learning-based mobile application to automatically classify the quality of coffee beans via a mobile phone camera. One hundred sixty coffee bean images were collected to be utilized in

the models created for the study in which they performed classification processes by using ResNet-152 and VGG16 deep learning architectures. At the end of the study, with the ResNet-152 architecture, the highest accuracy value of 73% was achieved (Janandi and Cenggoro 2020).

Huang et al. (2020) used CNN to classify coffee beans according to their quality. They obtained a total of 2000 images, 1000 of each coffee bean labeled as good and bad. Considering that these images may be insufficient in the training of the CNN model, rotation and flip procedures were used, and data were augmented 36 times. In conclusion, an accuracy value of 94.63% was achieved in the classification of coffee beans as good and bad (Huang et al. 2020).

In their study, Fuentes et al. (2020) aimed to classify the coffee fruit as ripe or unripe according to their color and shape by using deep learning algorithms. Within the scope of the study, the algorithm was trained with a total of 196 coffee fruit images, 108 ripe and 88 unripe. As a result, it was stated that the classification process was performed with 97.6% accuracy (Fuentes et al. 2020).

Waliyansyah and Hasbullah (2021) used NB, Tree, SVM, and Logistic Regression machine learning methods to classify two different coffee beans. With these methods, digital image processing was performed on a total of 58 coffee bean images from two types to obtain various features and perform classification operations. As a result of the study, they stated that coffee beans were successfully classified and that the most successful method was SVM (Waliyansyah and Hasbullah 2021).

Suyoto et al. (2021) performed texture analysis and feature extraction on grayscale images of 120 coffee bean images to determine the diversity of coffee beans. The obtained attributes indicated that they achieved an accuracy rate of 87.27% in detecting coffee bean varieties using the SVM algorithm (Suyoto et al. 2021).

Jumarlis et al. (2022) used the GLCM (Gray Level Co-Occurrence Matrix) and KNN methods to detect the defects of coffee beans. In addition, they have designed a website

where farmers can perform defect detection using images of coffee beans. As a result of the study, they noted that the proposed system for detecting defects in coffee beans has achieved an accuracy of 90% (Jumarlis et al. 2022).

In their study, Lee and Jeong (2022) developed a CNN-based model to predict normal and defective beans from two-class coffee bean images. In the experimental results, they stated that they obtained an average classification accuracy of 90.44% (Lee and Jeong 2022).

Adiwijaya et al. (2022a) used the KNN machine learning method to classify coffee bean quality. They focused on the color features of coffee bean images for use in classification. In the study conducted using 90 coffee beans in three different classes, they stated that they achieved a classification accuracy of 83% (Adiwijaya et al. 2022b).

Material and Methods

In this section, information about the coffee dataset, machine learning models, and performance metrics used in the study will be given.

Coffee Dataset

The coffee plant is a type of plant that grows in certain parts of the world. Even if coffee cultivation is made from the same type of coffee, it can change the aroma of the coffee according to the soil, climate, precipitation, and the way the coffee is harvested. The three types of coffee beans used in the study are coffee beans grown in different regions. The origin of Espresso coffee bean in the study is Ethiopia, the origin of Kenya coffee bean is Kenya, and the origin of Starbucks Pike Place coffee bean is Mexico, Costa Rica, and Colombia (Seninde and Chambers 2020). In Fig. 1, the countries of origin of the coffee beans are shown on the map.

In this study, images of three different coffee beans were used with the aim of recognizing the coffee type via deep

Fig. 1 Origin countries of coffee beans used in the study



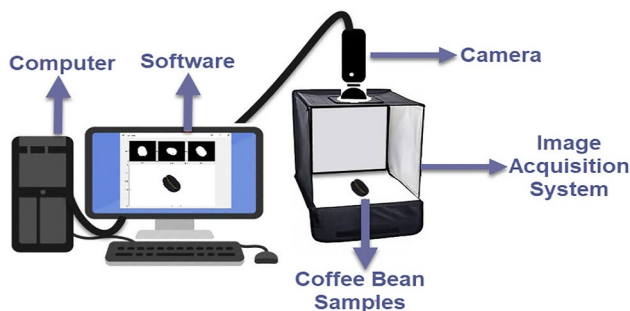


Fig. 2 Coffee image acquisition system

learning. In order to acquire the dataset, images of espresso, Kenya, and Starbucks Pike Place coffee beans were collected through a specially created mechanism. The setup created for the study has a camera, a box where the images will be captured, a computer, and a Computer Vision System (CVS) that allows the images to be saved in the desired color and resolution. The system used to take images of coffee beans is shown in Fig. 2.

A total of 1554 coffee bean images were obtained with the created image acquisition mechanism. Five hundred thirty of these images are espresso coffee bean, 502 of them

are Kenya coffee bean, and 522 of them are Starbucks Pike Place bean images. Images are RGB and have a resolution of 400×400 pixels. Figure 3 gives the sample coffee bean images obtained.

The distribution of images in the coffee bean image dataset by classes is shown in Fig. 4.

Convolutional Neural Network

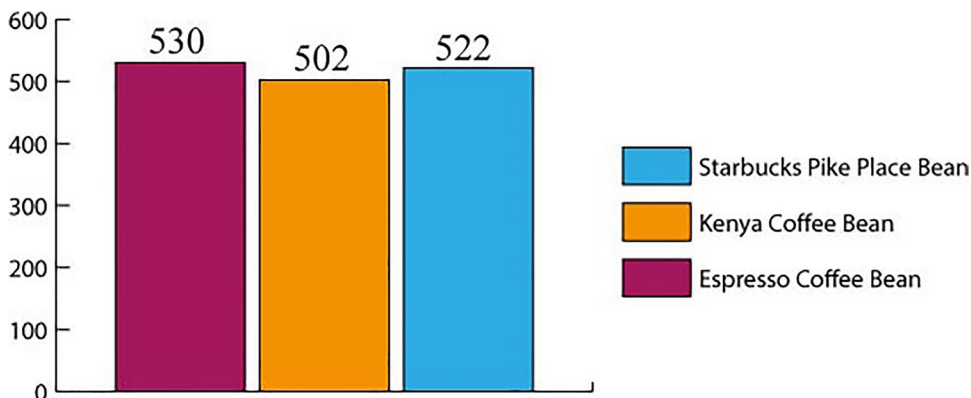
CNN, which is a deep learning method that includes many layers with complex structure, is frequently used in solving image processing problems (Albawi et al. 2017). The CNN method can work as an end-to-end classifier. Enabling feature extraction from the data given as input to the CNN network thanks to its layers, this method ensures learning and classification with the extracted features (Koklu et al. 2021).

In order to be able to perform these processes, convolution layer where various features are obtained by applying step-by-step filters to the image, the pooling layer where large-scale data coming from this layer is simplified in order to facilitate learning, the activation layer to prevent values from being out of the applicable data range, the fully connected layer which is the artificial neural network layer for

Fig. 3 Sample images of coffee beans used in the study



Fig. 4 Distribution of images in the dataset by classes



performing the classification, and lastly softmax layer where the classes are separated exist (Guo et al. 2016).

By applying various filters step by step on the regions on the image, image features are extracted from each region of the convolution layer. It is possible to increase or decrease the number of features to be obtained by determining the number of steps and filters desired to be obtained in this layer. However, this number should be set at an optimum level since a large number of features will make learning difficult for the network (Kandel and Castelli 2020).

In the pooling layer, the processes of reducing the large number of data coming from the convolution layer and the complexity are performed. Additionally, it is ensured that the image is transferred to the next layer by reducing its size without deteriorating its properties. Pooling layer types are mentioned in the literature. In this layer also, it is necessary to make optimum adjustments in a way that the classification is not affected (Singh et al. 2022).

After the required layers, it is ensured that the data is within the certain intervals by adding the activation layer. Then, in the fully connected layer as the classification layer, the features are reduced to the neural network level and learning processes are performed to make extraction. At the end of this process, the Softmax activation function is utilized to separate the classes. In this layer, an output is obtained by performing the labeling process (Guo et al. 2016).

Transfer learning approaches are influenced by the human learning model. In the learning process, in order to solve a problem they have never encountered, people benefit from the solution of problems they experienced before (Ying et al. 2018). Within the scope of this study, the CNN network is trained with the transfer learning method by using previously trained models (Deepak and Ameer 2019). SqueezeNet, Inception V3, VGG16, and VGG19 are the transfer learning models used in the study.

SqueezeNet SqueezeNet is a smaller CNN architecture that uses fewer parameters than other CNN models. SqueezeNet has fifteen layers consisting of five different layers: two convolution layers, three maximum pooling layers, eight fire layers, one global average pooling layer, and the softmax layer with one output layer. SqueezeNet consists of fire layers, which are convolution layers compressed with only 1×1 filters. Fire layers form compression and expansion processes between convolution layers (Ucar and Korkmaz 2020; Lee et al. 2019; Taspinar et al. 2021a).

Inception V3 Inception V3 was trained to recognize 1000 different objects for the ImageNet 2014 competition. This network structure which consists of 48 layers has an image input size of 299×299 . Inception V3 consists of symmetric and asymmetric building blocks, including convolution

layers, maximum pooling layers, average pooling layers, dropout layers, and fully connected layers (Demir et al. 2019; Sinha and Clarke 2017).

VGG16 This network structure was proposed by Zisserman and Simonyan in 2014. The network, based on the AlexNet deep network, is more successful in image recognition and classification problems when the data set is defined correctly. It contains 13 convolutional layers (Bicakci et al. 2020; Koklu et al. 2022a).

VGG19 The main layers of the VGG19 architecture consist of 16 convolutional, five pooling, and three fully connected layers. This architecture has a total of 24 main layers. Since VGG19 has a deep network, filters used in the convolutional layer are used to reduce the number of parameters. The size of the filter selected in this architecture is 3×3 pixels. VGG19 architecture contains approximately 138 million parameters (Mateen et al. 2019; Taspinar et al. 2021a).

Confusion Matrix and Evaluation Metrics

A three-class confusion matrix was used to evaluate the performance of the classification models used in the study. Table 1 shows an example of a three-class confusion matrix.

Accuracy, Precision, Recall, and F-1 Score values which are frequently used performance measures were obtained from the confusion matrix. In order to obtain these values, there are 4 values in the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP and TN indicate the number of correctly predicted positive and negative samples, while FP and FN indicate the number of incorrectly predicted positive and negative samples. Performance metrics are calculated with the formulas given in Table 2 by using these 4 values. Table 3 shows the calculation of TP, TN, FP, FN values for each class (Taspinar et al. 2021b; Koklu et al. 2022b; Koklu and Taspinar 2021).

Cross-Validation

Cross-validation is a method utilized for objective measurement of the classification models’ accuracy. In this method,

Table 1 Coffee beans confusion matrix

		Predicted class		
		Espresso	Kenya	Starbucks Pike Place
True class	Espresso	T_1	F_{12}	F_{13}
	Kenya	F_{21}	T_2	F_{23}
	Starbucks Pike Place	F_{31}	F_{32}	T_3

Table 2 Formulas for performance metrics

Performance metrics	Formula
Accuracy	$((TP + TN)/(TP + TN + FP + FN))*100$
Precision	$TP/(TP + FP)*100$
Recall	$TP/(TP + FN)*100$
F1-Score	$2TP/(2TP + FP + FN)*100$

the dataset is divided into equal parts according to the determined number value. $1/k$ part of the dataset is allocated for testing, and the $k - 1$ part is allocated for training. This process continues until each piece of the data set is used as the test segment, that is, this process is repeated k times. The overall classification success of the model is calculated by taking the arithmetic average of the classification successes obtained as a result of these processes (Arlot and Celisse 2010; Koklu and Tutuncu 2017). In our study, the k value was determined as 10. Figure 5 shows how the cross-validation method works.

Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

AUC-ROC curve is used as a performance measure for classification problems. ROC is a probability curve, while AUC represents the degree or metric of scalability and shows how

much the model can distinguish between classes (Sam et al. 2019).

Experimental Results

The obtained images of coffee beans were classified by using four different pre-trained CNN models. The training of the models was carried out on a virtual server having a 2-core Intel Xeon CPU, 12 GB ram, and 16 GB Nvidia Tesla T4 GPU provided by Colab through the Keras library in the Colab environment. In Fig. 6, the flowchart showing the processes of obtaining experimental results is given.

The networks to be used for Transfer Learning were chosen among the networks trained with ImageNet: SqueezeNet, Inception V3, VGG16, VGG19. In this study, the weights of these previously trained networks (SqueezeNet, Inception V3, VGG16, VGG19) were transferred and used in the classification of coffee beans.

In Table 4, the parameters for the pre-trained deep convolutional neural networks used in the study are given.

The success of the models was measured with the accuracy, Precision, Recall, and F-1 Score metrics that are calculated by using the values in the confusion matrix of each CNN model. The cross-validation method was utilized to ensure the reliability of the test results of all models. In Table 5, the confusion matrix obtained from the SqueezeNet model is given.

Table 3 TP, TN, FP, and FN values by classes

Espresso	Kenya	Starbucks Pike Place
$TP_1 = T_1$	$TP_2 = T_2$	$TP_3 = T_3$
$TN_1 = T_2 + T_3 + F_{23} + F_{32}$	$TN_2 = T_1 + T_3 + F_{21} + F_{23}$	$TN_3 = T_1 + T_2 + F_{12} + F_{21}$
$FP_1 = F_{21} + F_{31}$	$FP_2 = F_{12} + F_{32}$	$FP_3 = F_{13} + F_{23}$
$FN_1 = F_{12} + F_{13}$	$FN_2 = F_{21} + F_{23}$	$FN_3 = F_{31} + F_{32}$

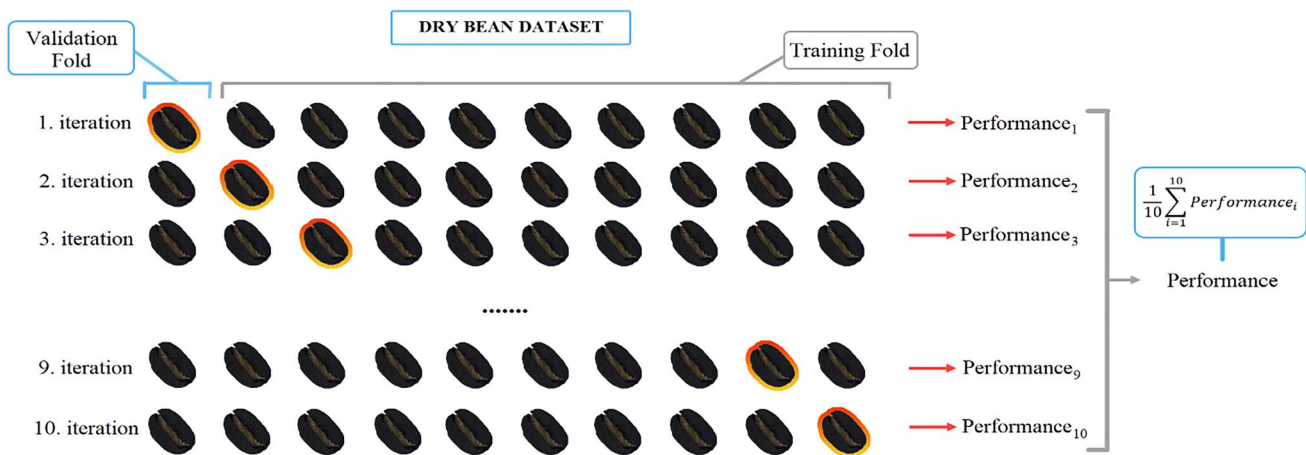


Fig. 5 Cross-validation process

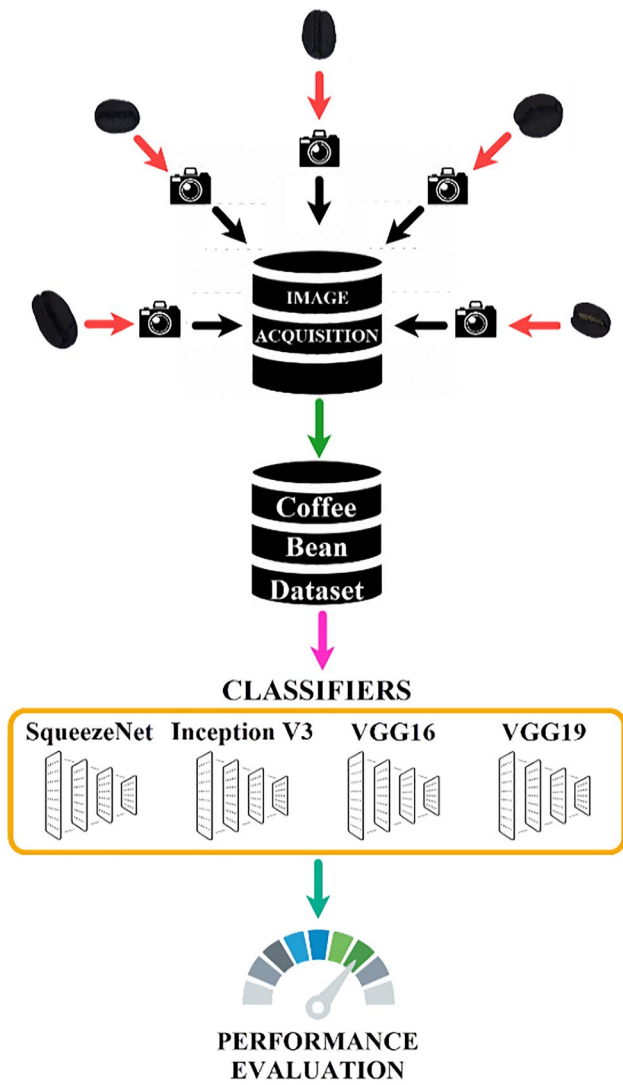


Fig. 6 Flow diagram of classification process

Table 4 Model parameters

Optimizer	Adam
Leaning rate	0.0003
Batch size	10
Epoch	200

According to Table 5, the SqueezeNet model classified 489 of the images of Espresso coffee beans correctly and misclassified 41 of them. While it correctly classified 449 images and incorrectly classified 53 images of Kenya coffee beans, 418 of the Starbucks Pike Place coffee bean images were classified correctly, and 104 of them were classified incorrectly. Starbucks Pike Place is the class in which the largest number of coffee bean images is misclassified. Table 6 shows the performance metrics calculated by using the confusion matrix data of the SqueezeNet model.

Table 5 Confusion matrix of SqueezeNet model

SqueezeNet		Predicted class		
		Espresso	Kenya	Starbucks Pike Place
Actual class	Espresso	489	1	40
	Kenya	1	449	52
	Starbucks Pike Place	55	49	418

Table 6 Performance metrics of the SqueezeNet model (%)

	Accuracy	Precision	Recall	F-1 score
Espresso	93.8	89.7	92.3	91
Kenya	93.4	90	89.4	89.7
Starbucks Pike Place	87.4	82	80.1	81

Table 7 Confusion matrix of Inception V3 model

Inception V3		Predicted class		
		Espresso	Kenya	Starbucks Pike Place
Actual class	Espresso	457	3	70
	Kenya	5	436	61
	Starbucks Pike Place	90	60	372

According to Table 6, SqueezeNet model achieved the highest classification success in the Espresso class. The class with the highest precision value is Kenya, while the Espresso class has the highest recall and F-1 score values.

The Inception V3 model was trained by using the images of coffee beans under the same conditions as the other models. The tests of the model that emerged as a result of the training were carried out. Table 7 shows the confusion matrix obtained as a result of the Inception V3 model of tests.

According to Table 7, the Inception V3 model classified 457 images correctly and classified 73 images incorrectly in the Espresso class. With the same model, in Kenya class, 436 images were classified correctly, and 66 images were classified incorrectly, while 372 images of the Starbucks Pike Place class were correctly classified and 150 of them were incorrectly classified. As a result, the highest number of misclassification belongs to the Starbucks Pike Place class. Table 8 shows the performance metrics calculated by using the confusion matrix data of the Inception V3 model.

According to Table 8, the Inception V3 model achieved the highest classification success in Kenya class. Again, with Inception V3, the highest Precision, Recall, and F-1 Score values were obtained in Kenya class.

Table 8 Performance metrics of the Inception V3 model (%)

	Accuracy	Precision	Recall	F-1 score
Espresso	89.2	82.8	86.2	84.5
Kenya	91.7	87.4	86.9	87.1
Starbucks Pike Place	81.9	74	71.3	72.6

Table 9 Confusion matrix of VGG16 model

VGG16		Predicted class		
		Espresso	Kenya	Starbucks Pike Place
Actual class	Espresso	445	6	79
	Kenya	9	409	61
	Starbucks Pike Place	83	77	362

Table 10 Performance metrics of the VGG16 model (%)

	Accuracy	Precision	Recall	F-1 Score
Espresso	88.6	82.9	84	83.4
Kenya	88.7	83.1	81.5	82.3
Starbucks Pike Place	79.2	69	69.3	69.1

As a result of the model VGG16's test process, the confusion matrix data in Table 9 were obtained. According to Table 9, 445 of the images were correctly classified, and 85 images were incorrectly classified in Espresso class, while 409 images were correctly and 70 images were incorrectly classified in Kenya class. Again, in the Starbucks Pike Place class, 362 images were correctly classified, and 160 images were incorrectly classified. So, the highest number of misclassifications belonged to the Starbucks Pike Place class. Table 10 shows the performance metrics calculated by using the confusion matrix data of the VGG16 model.

According to Table 10, Kenya class, the model VGG16 has the highest classification success and Precision value. Same model has the highest Recall and F-1 Score values in the Espresso class.

The confusion matrix data given in Table 11 was obtained as a result of testing the VGG19 model. According to Table 11, the VGG19 model correctly classified 397 of the images in the Espresso class and incorrectly classified 133 of them, while the number of images it correctly classified is 418 and the number of the images it incorrectly classified is 84 in Kenya class. With the same model, 312 of the images in the Starbucks Pike Place class were classified correctly, and 210 of them were classified incorrectly. As with other models, the highest number of misclassifications belongs to the Starbucks Pike Place class. Table 12 shows

Table 11 Confusion matrix of VGG19 model

VGG19		Predicted class		
		Espresso	Kenya	Starbucks Pike Place
Actual class	Espresso	397	15	118
	Kenya	20	418	64
	Starbucks Pike Place	152	58	312

Table 12 Performance metrics of the VGG19 model (%)

	Accuracy	Precision	Recall	F-1 Score
Espresso	80.4	69.8	74.9	72.2
Kenya	89.9	85.1	83.3	84.2
Starbucks Pike Place	74.8	63.2	59.8	61.4

Table 13 Average performance metrics of all models (%)

	Accuracy	Precision	Recall	F-1 Score
SqueezeNet	87.3	87.2	87.3	87.2
Inception V3	81.4	81.3	81.4	81.3
VGG16	78.2	78.3	78.2	78.3
VGG19	72.5	72.5	72.5	72.5

the performance metrics calculated by using the confusion matrix data of the VGG19 model.

According to Table 12, the model VGG19 has the highest classification success, Precision value, Recall value, and F-1 Score value in Kenya class. In Table 13, the average classification success, Precision, Recall, and F-1 Score values of SqueezeNet, Inception V3, VGG16, and VGG19 CNN models used in the study are given.

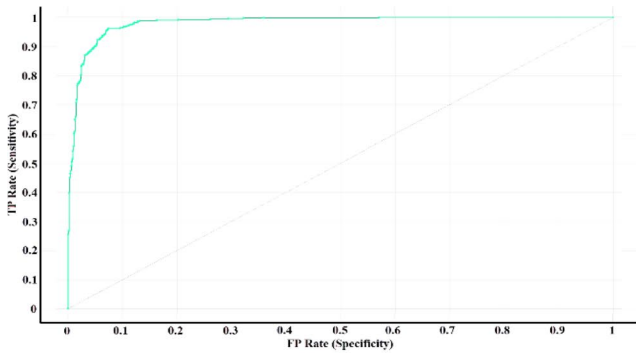
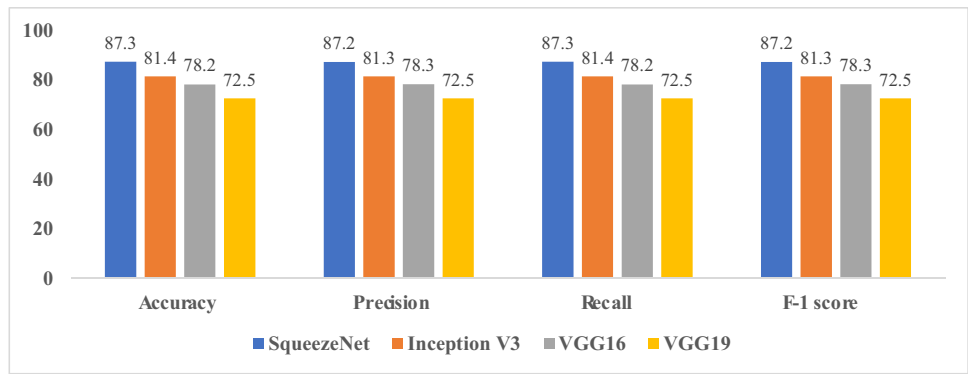
Comparison of performance metrics of all models used for classification in the study is shown in Fig. 7.

According to Table 13, the SqueezeNet model has the highest value in all performance metrics. ROC curves provide information about the distinctiveness of the models. Figure 8 gives the ROC curves of all models.

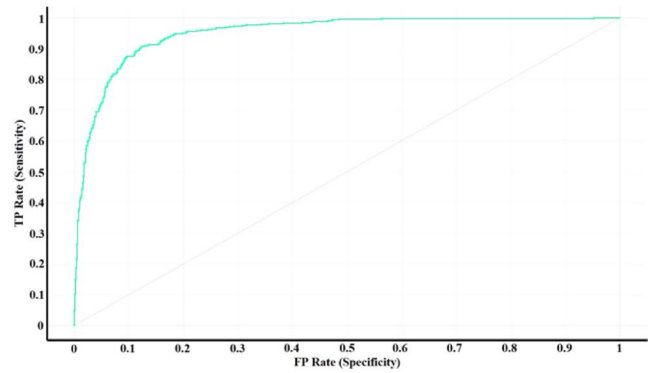
Conclusion

Within the scope of this study, a total of 1554 coffee bean images obtained with the image acquisition mechanism that is created specifically for the study were used. These images were of three different types of coffee beans: Espresso, Kenya, and Starbucks Pike Place. Using the images of coffee beans, four different CNN models were trained with the transfer learning method. The tests were carried out after the training processes of the models

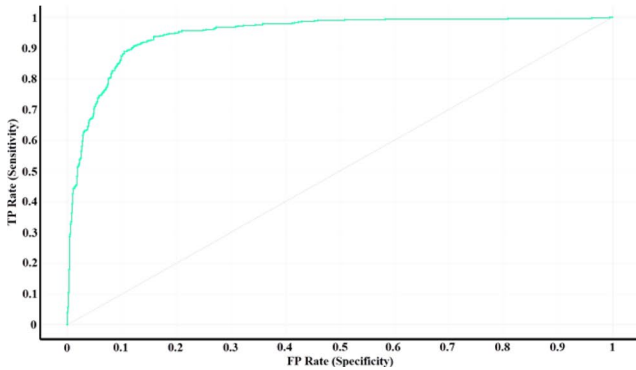
Fig. 7 Comparison of the performances of the models



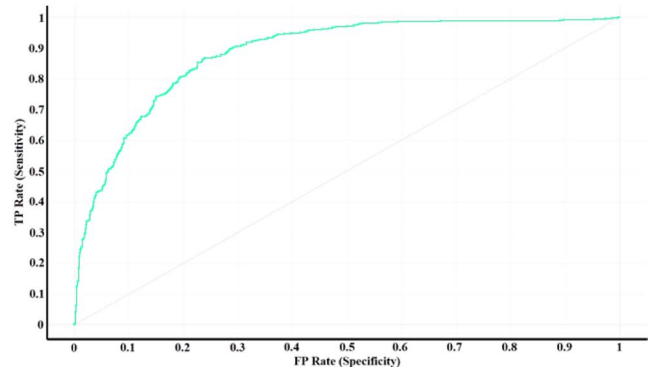
(a)



(b)



(c)



(d)

Fig. 8 ROC curves of all models ((a): SqueezeNet, (b): Inception V3, (c): VGG16, (d): VGG19)

used in the study, which are SqueezeNet, Inception V3, VGG16, and VGG19. In order to evaluate the performances of the models more objectively, tests were performed by the cross-validation method. The *k* value was determined as 10 in the cross-validation method. While the highest classification success rates were obtained with the SqueezeNet model, these rates were calculated as 93.8% for Espresso, 93.4% for Kenya, and 87.4% for Starbucks Pike Place classes. The average classification successes

of the models are 87.3%, 81.4%, 78.2%, and 72.5%, respectively for SqueezeNet, Inception V3, VGG16, and VGG19. It is concluded that SqueezeNet is the model with the highest classification success, while the model with the lowest classification success is VGG19. Furthermore, SqueezeNet is the model also with the highest average Precision, Recall, and F-1 Score values. When the ROC curves are examined, it is seen that SqueezeNet model has the highest distinctiveness.

When the obtained coffee bean images are examined, it is understood that the view of Espresso and Starbucks Pike Place coffee beans are very similar. This situation can also be understood from the values in the confusion matrices. Due to the similarity of their appearances, these two beans sometimes cannot be distinguished.

Within the scope of this study, different types of coffee beans were classified and distinguished. The use of such practices in the production, packaging, and trade stages of coffee will make it possible to distinguish coffee beans. It is envisaged to reduce the processing time and labor cost to a minimum. This will also improve the quality-based export of coffee beans. Thanks to these models, it will be able to ensure that all products have the same standard. Quality control will be facilitated. Decision-making errors caused by the mental and physical condition of the specialist, such as fatigue and vision, will be prevented. It will be able to allow different companies to conduct fair quality control without any bias.

Author Contribution All authors contributed to the design and study of the article. YU, literature research and preparation of the draft; YST, MK, analysis of data and result; IC, RK, arranged the materials, methods and data.

Funding This project was supported by the Scientific Research Coordinator of Selcuk University with the project number 22111002.

Data Availability The dataset used in the study can be accessed from the link; https://www.muratkoklu.com/datasets/Coffee_Image_Datas.et.zip

Declarations

Competing Interests The authors declare no competing interests.

Consent for Publication Not applicable.

Compliance with ethics requirements This study does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest Yavuz Unal declares that he has no conflict of interest. Yavuz Selim Taspinar declares that he has no conflict of interest. Ilkay Cinar declares that he has no conflict of interest. Ramazan Kur-sun declares that he has no conflict of interest. Murat Koklu declares that he has no conflict of interest.

References

Adiwijaya NO, Romadhon HI, Putra JA, Kuswanto DP (2022a) The quality of coffee bean classification system based on color by using k-nearest neighbor method. *J Phys: Conf Ser* 2157(1):012034. <https://doi.org/10.1088/1742-6596/2157/1/012034>

- Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET). Ieee. <https://doi.org/10.1109/ICEngTechnol.2017.8308186>
- Ansar, Sukmawaty, Murad, Muttalib SA, Putra RH, Abdurrahim (2021) Design and performance test of the coffee bean classifier. *Processes* 9(8). <https://doi.org/10.3390/pr9081462>
- Arboleda ER, Fajardo AC, Medina RP (2018) Classification of coffee bean species using image processing, artificial neural network and K nearest neighbors. in 2018 IEEE International Conference on Innovative Research and Development (ICIRD). IEEE <https://doi.org/10.1109/ICIRD.2018.8376326>
- Arboleda ER (2019) Comparing performances of data mining algorithms for classification of green coffee beans. *Int J Eng Adv Technol* 8(5):1563–1567
- Adiwijaya NO, Romadhon HI, Putra JA, Kuswanto DP (2022b) The quality of coffee bean classification system based on color by using k-nearest neighbor method. in *Journal of Physics: Conference Series*. IOP Publishing. <https://doi.org/10.1088/1742-6596/2157/1/012034>
- Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. *Statistics Surveys* 4:40–79. <https://doi.org/10.1214/09-SS054>
- Bicakci M, Ayyildiz O, Aydin Z, Basturk A, Karacavus S, Yilmaz B (2020) Metabolic imaging based sub-classification of lung cancer. *IEEE Access* 8:218470–218476. <https://doi.org/10.1109/ACCESS.2020.3040155>
- de Oliveira EM, Leme DS, Barbosa BHG, Rodarte MP, Pereira RGFA (2016) A computer vision system for coffee beans classification based on computational intelligence techniques. *J Food Eng* 171:22–27. <https://doi.org/10.1016/j.jfoodeng.2015.10.009>
- Demir A, Yilmaz F, Kose O (2019) Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3. in 2019 Medical Technologies Congress (TIPTEKNO). IEEE. <https://doi.org/10.1109/TIPTEKNO47231.2019.8972045>
- Deepak S, Ameer P (2019) Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology Medicine* 111:103345. <https://doi.org/10.1016/j.combiomed.2019.103345>
- Fuentes MS, Zelaya NAL, Avila JLO (2020) Coffee fruit recognition using artificial vision and neural networks. in 2020 5th International Conference on Control and Robotics Engineering (ICCRE). IEEE. <https://doi.org/10.1109/ICCRE49379.2020.9096441>
- Fukai H, Furukawa J, Katsuragawa H, Pinto C, Afonso C (2018) Classification of green coffee beans by convolutional neural network and its implementation on raspberry pi and camera module
- Gope HL, Fukai H (2020) Normal and peaberry coffee beans classification from green coffee bean images using convolutional neural networks and support vector machine. *Int J Comput Inform Eng* 14(6):189–196
- Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS (2016) Deep learning for visual understanding: a review. *Neurocomputing* 187:27–48. <https://doi.org/10.1016/j.neucom.2015.09.116>
- Huang N-F, Chou D-L, Lee C-A, Wu F-P, Chuang A-C, Chen Y-H, Tsai Y-C (2020) Smart agriculture: real-time classification of green coffee beans by using a convolutional neural network. *IET Smart Cities* 2(4):167–172. <https://doi.org/10.1049/iet-smc.2020.0068>
- Huang NF, Chou DL, Lee CA (2019) Real-time classification of green coffee beans by using a convolutional neural network. In 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC). IEEE. <https://doi.org/10.1109/ICISPC.2019.8935644>
- Janandi, R. and T.W. Cenggoro (2020) An implementation of convolutional neural network for coffee beans quality classification in a mobile information system. In 2020 International Conference

- on Information Management and Technology (ICIMTech). IEEE. <https://doi.org/10.1109/ICIMTech50083.2020.9211257>
- Jumarlis M, Mirfan M, Manga AR (2022) Classification of coffee bean defects using gray-level co-occurrence matrix and k-nearest neighbor. *ILKOM Jurnal Ilmiah* 14(1)
- Kandel I, Castelli M (2020) How deeply to fine-tune a convolutional neural network: a case study using a histopathology dataset. *Appl Sci* 10(10):3359
- Kuo CJ, Wang DC, Chen TT, Chou YC, Pai MY, Horng GJ, Hung MH, Lin YC, Hsu TH, Chen CC (2019) Improving defect inspection quality of deep-learning network in dense beans by using hough circle transform for coffee industry. in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE. <https://doi.org/10.1109/SMC.2019.8914175>
- Koklu M, Kahramanli H, Allahverdi N (2014) A new accurate and efficient approach to extract classification rules. *J Faculty Eng Arch Gazi Univ* 29(3):477–486
- Koklu M, Cinar I, Taspinar YS (2021) Classification of rice varieties with deep learning methods. *Comput Electron Agric* 187:106285
- Koklu M, Unlarsen MF, Ozkan IA, Aslan MF, Sabanci K (2022b) A CNN-SVM study based on selected deep features for grapevine leaves classification. *Measurement* 188:110425. <https://doi.org/10.1016/j.measurement.2021.110425>
- Koklu M, Taspinar YS (2021) Determining the extinguishing status of fuel flames with sound wave by machine learning methods. *IEEE Access* 9:86207–86216. <https://doi.org/10.1109/ACCESS.2021.3088612>
- Koklu M, Cinar I, Taspinar YS (2022a) CNN-based bi-directional and directional long-short term memory network for determination of face mask. *Biomed Signal Process Control* 71:103216
- Koklu M, Tutuncu K (2017) Classification of chronic kidney disease with most known data mining methods. *Int J Adv Sci Eng Technol* 5(2):14–18
- Lee HJ, Ullah I, Wan W, Gao Y, Fang Z (2019) Real-time vehicle make and model recognition with the residual SqueezeNet architecture. *Sensors* 19(5):982. <https://doi.org/10.3390/s19050982>
- Lee JY, Jeong YS (2022) Prediction of defect coffee beans using CNN. in 2022 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE. <https://doi.org/10.1109/BigComp54360.2022.00046>
- Mateen M, Wen J, Song S, Huang Z (2019) Fundus image classification using VGG-19 architecture with PCA and SVD. *Symmetry* 11(1):1. <https://doi.org/10.3390/sym11010001>
- Nair KP (2010) The agronomy and economy of important tree crops of the developing world. Elsevier. <https://doi.org/10.1016/B978-0-12-384677-8.00001-1>
- Pizzaia JPL, Salcides IR, de Almeida GM, Contarato R, de Almeida R (2018) Arabica coffee samples classification using a Multilayer Perceptron neural network. in 2018 13th IEEE International Conference on Industry Applications (INDUSCON). IEEE. <https://doi.org/10.1109/INDUSCON.2018.8627271>
- Pinto C, Furukawa J, Fukai H, Tamura S (2017) Classification of Green coffee bean images basec on defect types using convolutional neural network (CNN) in 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA). IEEE. <https://doi.org/10.1109/ICAICTA.2017.8090980>
- Rivalto A, Pranowo, Santoso AJ (2020) Classification of Indonesian coffee types with deep learning. in AIP Conference Proceedings. AIP Publishing LLC. <https://doi.org/10.1063/5.0000678>
- Ropelewska E, Sabanci K, Aslan MF, Azizi A (2022) A novel approach to the authentication of apricot seed cultivars using innovative models based on image texture parameters. *Horticulturae* 8(5). <https://doi.org/10.3390/horticulturae8050431>
- Suyoto RZH, Komarudin M, Nama GF, Yulianti T (2021) Classification of Civet and Canephora coffee using support-vector machines (SVM) algorithm based on order-1 feature extraction. in IOP Conference Series: Materials Science and Engineering. IOP Publishing. <https://doi.org/10.1088/1757-899X/1173/1/012006>
- Santos F, Rosas J, Martins R, Araújo G, Viana L, Gonçalves J (2020) Quality assessment of coffee beans through computer vision and machine learning algorithms. *Coffee Sci* 4:525. <https://doi.org/10.25186/v15i.1752>
- Seninde DR, Chambers E (2020) Coffee Flavor: a Review. *Beverages* 6(3):44
- Singh D, Taspinar YS, Kursun R, Cinar I, Koklu M, Ozkan IA, Lee HN (2022) Classification and analysis of pistachio species with pre-trained deep learning models. *Electronics*, 11(7). <https://doi.org/10.3390/electronics11070981>
- Sinha R, Clarke J (2017) When technology meets technology: retrained ‘Inception V3’ classifier for NGS based pathogen detection. In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE. <https://doi.org/10.1109/BIBM.2017.8217942>
- Sam SM, Kamardin K, Sjarif NNA, Mohamed N (2019) Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet Inception-v1 and Inception-v3. *Procedia Comput Sci* 161:475–483. <https://doi.org/10.1016/j.procs.2019.11.147>
- Taspinar YS, Cinar I, Koklu M (2021) Classification by a stacking model using CNN features for COVID-19 infection diagnosis. *Journal of X-ray science and technology*, 2021(Preprint), p. 1–16
- Taspinar YS, Koklu M, Altin M (2021a) Fire detection in images using framework based on image processing, motion detection and convolutional neural network. *Int J Intelligent Syst Appl Eng* 9(4):171–177
- Taspinar YS, Cinar I, Koklu M (2021b) Prediction of computer type using benchmark scores of hardware units. *Selcuk Univ J Eng Sci* 20(1):11–17
- Ucar F, Korkmaz D (2020) COVIDiagnosis-Net: deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med Hypotheses* 140:109761. <https://doi.org/10.1016/j.mehy.2020.109761>
- Vogt MAB (2020) Developing stronger association between market value of coffee and functional biodiversity. *J Environ Manage* 269:110777. <https://doi.org/10.1016/j.jenvman.2020.110777>
- Waliyansyah RR, Hasbullah UHAA (2021) Comparison of tree method, support vector machine, Naïve Bayes, and logistic regression on coffee bean image. *EMITTER Int J Eng Technol* 9(1):126–136
- Ying W, Zhang Y, Huang J, Yang Q (2018) Transfer learning via learning to transfer. in International Conference on Machine Learning. PMLR
- Zhang C, Liu F, He Y (2018) Identification of coffee bean varieties using hyperspectral imaging: influence of preprocessing methods and pixel-wise spectra analysis. *Sci Rep* 8(1):1–11. <https://doi.org/10.1038/s41598-018-20270-y>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.