



Automatic classification of walnut (*Juglans Regia* L.) species using deep learning methods

Halil Kilif¹ · Ilkay Cinar² · Nurettin Dogan²

Received: 20 December 2024 / Accepted: 29 May 2025 / Published online: 3 June 2025
© The Author(s) 2025

Abstract

Walnut is an agricultural product with high economic value on a global scale. Walnut species classification is essential for research, conservation, and quality control, yet traditional methods rely heavily on manual identification, which is a time-consuming procedure and subject to human mistakes. With technological developments in the agricultural sector, making use of deep learning algorithms in the classification of products such as fruits, vegetables and grains were becoming increasingly widespread. In this study, it is aimed to automatically classify walnut species using pre-trained deep learning models. Thus, it is aimed at reducing the loss of time, workload and error rates in sorting processes. For this purpose, a dataset consisting of images belonging to Chandler, Kaman1, Fernor, Yalova3 and Maras18 walnut species was created. The dataset consists of 2540 images in total. The images were trained using VGG16, VGG19, ResNet-50, DenseNet-121, and Xception models. The classification was performed using pre-trained deep learning architectures, including VGG16, VGG19, ResNet-50, DenseNet-121, and Xception. Among these models, ResNet-50 delivered the best performance with an accuracy of 97.95% on the original dataset, while the Xception model excelled with 98.54% accuracy when trained with a weighted loss function and 98.27% accuracy with data augmentation. These findings highlight the effectiveness and reliability of ResNet-50 and Xception models for automated walnut species classification. The results underscore the potential of deep learning technologies in improving agricultural practices by offering faster, more accurate, and less labor-intensive alternatives to traditional methods. In comparison, machine learning algorithms such as SVM, RF, and k-NN achieved lower accuracies, with SVM performing best among them at 90.10%. The study provides an important contribution to the use of deep learning technologies in agricultural production processes and suggests solutions that can increase the efficiency of traditional manual methods.

Keywords Agriculture · Deep Learning · Transfer Learning · Walnut Classification · Walnut Species

Introduction

With the development of technology in the agricultural sector, incorporating deep learning algorithms in the classification of various agricultural products is becoming

increasingly widespread. The efficiency provided by deep learning models, especially in the sorting and classification processes of products such as fruits, vegetables and grains, offers great advantages compared to manual processes. Many studies conducted in this context reveal the potential of deep learning-based approaches in the classification of agricultural products.

For instance Kamilaris and Prenafeta-Boldú, comprehensively discussed the use of deep learning in various fields in agriculture and emphasized that the classification of products such as fruits, vegetables and grains is one of the most effective application areas of this technology [28]. The classification performance of deep learning models in agriculture has also achieved significant success, especially in the detection of plant diseases. In this direction Ferentinos, showed that deep learning models reached high accuracy rates in the detection of plant diseases and that this

✉ Ilkay Cinar
ilkay.cinar@selcuk.edu.tr

Halil Kilif
halil_kilif@icloud.com

Nurettin Dogan
nurettin.dogan@selcuk.edu.tr

¹ Department of Computer Engineering, Selcuk University Graduate School of Natural and Applied Sciences, Konya, Türkiye

² Department of Computer Engineering, Selcuk University Faculty of Technology, Konya, Türkiye

technology can be used effectively in fruit trees [17]. In addition Too et al., compared various deep learning models and achieved successful results in the detection and classification of diseases in agricultural products [49].

In the field of fruit classification Muresan and Oltean, stated that deep learning methods provide high accuracy in distinguishing fruit types and that these techniques can increase efficiency in agricultural processes [38]. Similarly Koirala et al., worked on fruit detection and yield predictions using deep learning models such as Faster R-CNN, YOLO, and SSD and revealed that these models are especially successful in real-time detection [34].

The effective use of deep learning in vegetable and grain products is also noteworthy. Zhang et al., has shown that the classification of corn leaf diseases using deep learning has high accuracy rates [56]. In addition Picon et al., addressed the detection of plant diseases with deep learning algorithms using vegetable and plant images taken in natural environments and proved that deep learning is effective even in difficult field conditions [40].

According to the results of studies, it is seen that deep learning algorithms have significant potential in the classification of agricultural products, especially in the detection and sorting processes of products such as fruits, vegetables and grains. These studies, in which various deep learning models are used, provide significant contributions to the agricultural sector with both their high accuracy rates and the potential to reduce manual labor.

Walnut is considered a hard-shelled fruit. It is in the *Juglans* genus of the *Juglandaceae* family. There are 18 different walnut species in this genus with general characteristics. *Juglans Regia L.*, also known as *Anatolian*, *Iranian* and *English* walnuts, is considered the walnut species with the highest fruit quality [44]. Türkiye ranks fourth in walnut production in the world [45]. Walnut planting and production are carried out all over the country. Therefore, there are many walnut trees and varieties in the country. Among these, there are high-quality and productive species as well as non-standard, low-quality and inefficient species [18].

Walnut is a versatile resource used in various sectors. Its green fruits are evaluated in the food and pharmaceutical industries before their shells harden. At the same time, its leaves, shells and roots are used in the paint industry, while walnut oil is a sought-after product in both technology and painting. Walnut trees are an important raw material source for the furniture sector. Walnut is the only type of fruit that contains silver. The only organ that needs silver ions is our brain. Silver, which is effective in protecting brain health and learning processes, is taken into our body through walnuts [6].

Walnuts contain significant amounts of protein, fat and minerals [47]. Especially with Omega-3, Omega-6, vitamin E, and fatty acids, they play a protective role against

cardiovascular diseases, some types of cancer and the negative effects of aging [2].

The process of sorting the harvested walnuts according to their quality before marketing is an important process. However, traditional walnut classification is done manually by people according to the physiological characteristics of the walnuts [11]. This method may lead to misclassifications, and the classification process may become longer and more inefficient as the amount of walnuts to be separated increases [39].

Nowadays, with the increase in Artificial Intelligence studies, the sorting-classification process of many agricultural products, which is done manually by humans, has started to be done autonomously by robots. The use of robots in these processes will reduce the need for human power and increase efficiency. In classifying walnut species, the prerequisite for the use of autonomous robots or systems is to correctly identify the walnut species [43].

The study aimed based on classification of walnut species accurately according to their visual features. In this direction, a walnut dataset consisting of 600 images for Chandler, 500 for Kaman1, 540 for Fernor, 500 for Yalova3 and 400 for Maras18, which are walnut species commonly found in Türkiye, was created. In this way, a walnut dataset involving 5 classes and 2540 images will be added to the literature.

In the second part of this study, similar studies, datasets used and results that emerged as a result of the literature review are mentioned. In the third part, the original dataset created is explained. The method of image acquisition and pre-processing performed on the image are processed in detail. Information about the deep learning models to be used in the study is given. Transfer learning is explained in detail. Metrics such as Confusion Matrix, Cross-validation, accuracy, precision, sensitivity, F1 score, which are methods of measuring the test performance of the models after training, are explained, and the performance metrics of the ROC Curve and AUC, where the area under it is calculated, are mentioned. In the fourth part, the division of the dataset, the system information where the experiment was performed, the parameters used in the models, experimental test results, metrics, and graphics are given in detail for each model and the resulting results are compared. In the fifth part, the results of the deep learning models used in the study are compared; the contributions of this study to the literature are mentioned and suggestions are made for future studies.

Literature review

Since classifying agricultural products such as walnuts is a lengthy and problematic process with traditional methods, machine learning, and deep learning techniques have made significant progress in this field in recent years. Such algorithms aim to increase efficiency in agricultural production

processes by allowing fast and accurate classification of walnut species.

In recent years, parallel to the advances in the field of artificial intelligence and deep learning, there have been many studies on correctly identifying and classifying different types of products in the field of agriculture, as in many areas of daily life. However, the number of studies on classifying different types of walnuts in the literature is less and limited. When these studies are examined, it is seen that; diseases in the walnut tree [3] or in the leaf [33], classification of walnut kernels (kernels) are made, and studies are carried out to determine whether walnuts are empty or full, etc.

An attempt was made to classify the Pikan, Yalova and Datça walnut species using MATLAB and fuzzy logic. The general classification success rate for 252 images belonging to three walnut species was 88.49% [39].

A dataset of 1751 leaf images of 18 various walnut species was created to determine the variety of walnut as of walnut tree leaves in Türkiye. On a dataset expanded with data augmentation methods, the VGG16 algorithm showed 90.55% correct classification success with different deep learning algorithms [29]. In another study conducted by the same authors with the same dataset, the features extracted by deep learning algorithms were classified with 92.59% accuracy by the KNN machine learning algorithm [30].

Another study used image processing techniques to determine the shape and size characteristics of walnut species. Using SPSS software, length, width, thickness, and shape indexes were calculated for ten different walnut species using images obtained from various sources [16]. In a study conducted using 3552 walnut slider images belonging to 10 species in Türkiye, the EfficientB0 deep learning model was able to classify different types of sliders with 91.87% accuracy [31].

500 images belonging to 3 different walnut species in China were labeled for the YoloV3, YoloV4 and YoloV5 object detection algorithms and object classification and detection were performed with the specified models. The YoloV5 algorithm performed the highest classification and object detection of Xin2, Xinguang and Wen185 local walnut species through an average accuracy of 99.5%, 98.4% and 97.1%, respectively [43].

In another study Zhang et al., proposed a YOLOX-based model to separate kernels from walnut shells [57]. Chen et al., published a study on separating walnut kernels according to physiological indicators using deep learning methods [11]. In another study, an approach based on fuzzy logic and audio processing was proposed to separate empty, half-full and full walnuts [32].

In a study conducted in 2024, deep learning models were used to classify three different alfalfa varieties based on images obtained under controlled conditions; it was reported that models trained with transfer learning achieved higher accuracy compared to those trained from scratch [21]. In another study, the disease status of cotton plants was classified using an improved deep learning architecture, and an accuracy of 97.98% was achieved [42]. Zhu et al. [58] aimed to develop a smart pruning system for dormant walnut trees to reduce labor costs and improve efficiency. Researchers used 3D models created from photos with Neural Radiance Fields (NeRFs) and an improved deep learning model (Walnut-PointNet) to accurately identify tree parts and locate where branches should be pruned. The system achieved accuracy in tree segmentation (over 93%) and detected pruning points with nearly 90% success, completing the process for a whole tree in just 16 s [58].

From the studies it is seen that Machine Learning and mostly Deep Learning algorithms are used. The results of deep learning algorithms in areas such as object recognition and image classification have caused them to be frequently preferred in studies on agricultural products. Although studies on walnut classification are few, the methods used in these studies, the number of species examined, the number of images in the datasets and the data on the results obtained are summarized in Table 1.

Table 1 presented the Walnut dataset sizes, algorithms used, and success rates of different studies in literature. When these studies are examined, it is established that deep learning algorithms achieve high accuracy rates in the classification of agricultural products. However, it is also noteworthy that the size of the dataset and species diversity within it significantly affect the classification results.

This study achieved higher accuracy rates compared to the studies conducted by Okan and Koklu, Karadeniz et al.,

Table 1 An overview of walnut classification studies

No	Number of Species	Number of Samples in the Dataset	Used Algorithms	Results	References
1	3	252	Fuzzy Logic	%88.49	[39]
2	18	1751	Deep Learning Algorithms	%90.55	[29]
3	3	500	YoloV3, YoloV4, YOLOv5	%99.5	[43]
4	10	3552	Deep Learning Algorithms	%91.87	[31]

and Kayaalp and Altunalan. Despite Sanlin attaining an accuracy rate of 99.5%, factors including the restricted sample size of the dataset, reduced class diversity relative to this study, the exclusive focus on local walnut varieties, and the YOLO method's primary optimization for object recognition indicate that it may lack efficacy in classification tasks necessitating intricate feature extraction. This work used an extensive dataset of five walnut types, both indigenous and foreign, which allowed deep learning and machine learning algorithms to more effectively identify minor distinctions among the kinds and enhanced the models' generalization capabilities.

Using pre-trained deep learning models, and machine learning algorithms we aim to develop a larger and more balanced dataset to contribute to existing literature. Thus, the classification accuracy will be improved, and results will be more extrapolatable. It is envisaged that this approach will lead to the development of more comprehensive and effective agricultural product classification solutions.

Material and methods

This section includes knowledge regarding the walnut dataset, deep learning models, machine learning algorithms, performance metrics and enhancements on results used in the study. Transferring the learning (parameters and weights) from pre-trained deep learning algorithms, freezing the transferred parameter weights in the newly trained model, adding new layers, training the model on the walnut dataset with new layers using the cross-validation method, evaluating the resulting results with performance measurement metrics frequently used in literature, with statistical evaluation as paired T-tests and the selection processes of the best model results. The illustration for the workflow given in Fig. 1.

Walnut dataset

For this study, 5 different walnut species that are widely produced in Türkiye were used. These are the local species

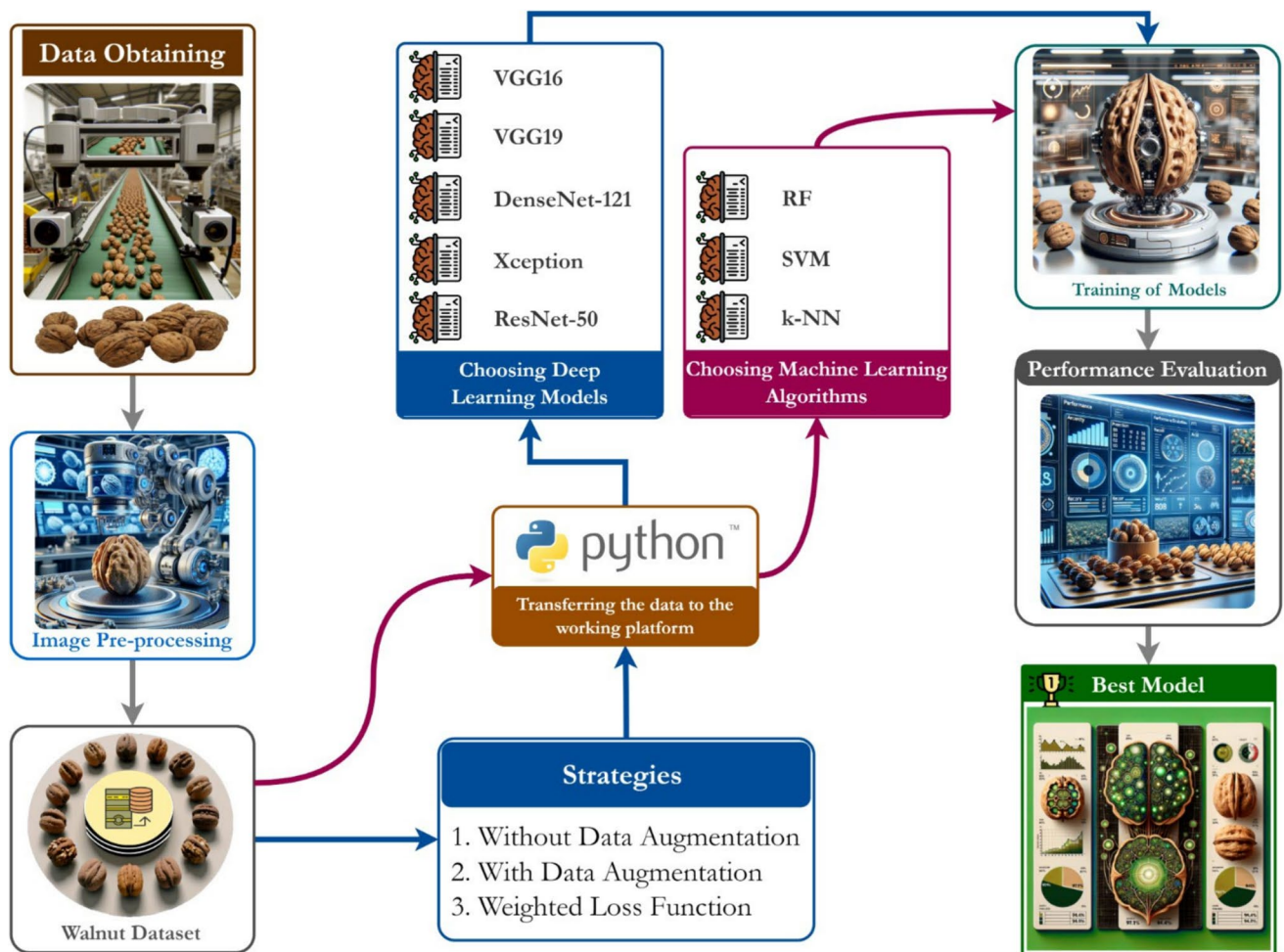


Fig. 1 Flow diagram for the walnut species classification

“Kaman1, Yalova3 and Maras18, the American Chandler and French Fernor” species. The Kaman1 species was obtained from the Kaman district of Kırşehir, and the Yalova3 species was obtained from Muğla province. The Chandler and Fernor species were taken from the gardens in the Central district of Aksaray province, and the Maras18 species was taken from the Ağaçören district of Aksaray province. The products of all species were harvested in 2023. The system used to capture walnut images is given in Fig. 2.

During the walnut photography process by authors, the camera was kept fixed to eliminate differences in resolution, shooting angle, and potential shadows in the background, ensuring a standardized image quality for all walnuts. Samples were selected from five different walnut varieties, with careful attention to maintaining equal total weight for each variety, even though the number of walnuts varied among them. Each walnut was photographed only once. All images were captured under identical conditions, with the same lighting, from a fixed distance, and at the same angle. The setup used to take the images is illustrated in Fig. 2. The images were obtained at a resolution of 5184×3456 using a Canon EOS 600D camera with a 135 mm lens. The images obtained were three-channel RGB and saved in ".jpg" format.

The walnut dataset contains 2540 images in total. The dataset was created by taking 600 images from the *Chandler* species, 540 images from the *Fernor* species, 500 images each from the *Kaman1* and *Yalova3* species, and 400 images from the *Maras18* species. The images obtained were reduced to 600

$\times 600$ pixels and making their backgrounds black as a preprocessing step. Sample images of the dataset are given in Fig. 3.

According to data from previous studies by Akca and Ozongun, Çavuşoğlu and Uzun, Sutyemez, and Yaşar et al., the walnut species examined in our study show differences in their average shell weights, kernel weights, and kernel ratios [1, 10, 48, 52]. Additionally, their oil, protein, and mineral contents are found in varying proportions [5, 50, 53].

Based on the analysis results made under laboratory conditions, walnuts can be easily separated and distinguished from each other with their protein, oil and mineral differences, however, as seen in Fig. 4, separating them from each other and correctly determining their types in the environment where they are mixed together in their shelled form requires expertise. However, when the number of walnuts to be separated increases, it will not be practical to perform these operations manually. Therefore, this study aims to classify the productive type or types among many walnut types accurately and quickly using today's technologies.

Convolutional neural network (CNN)

Convolutional neural networks, first developed in 1989 to read numbers on postal checks, have reached great accomplishment in topics such as image classification and object recognition [36]. These architectures, which perform learning and classification by extracting image features as a result of operations performed on the layers from the input image

Fig. 2 System used for image acquisition

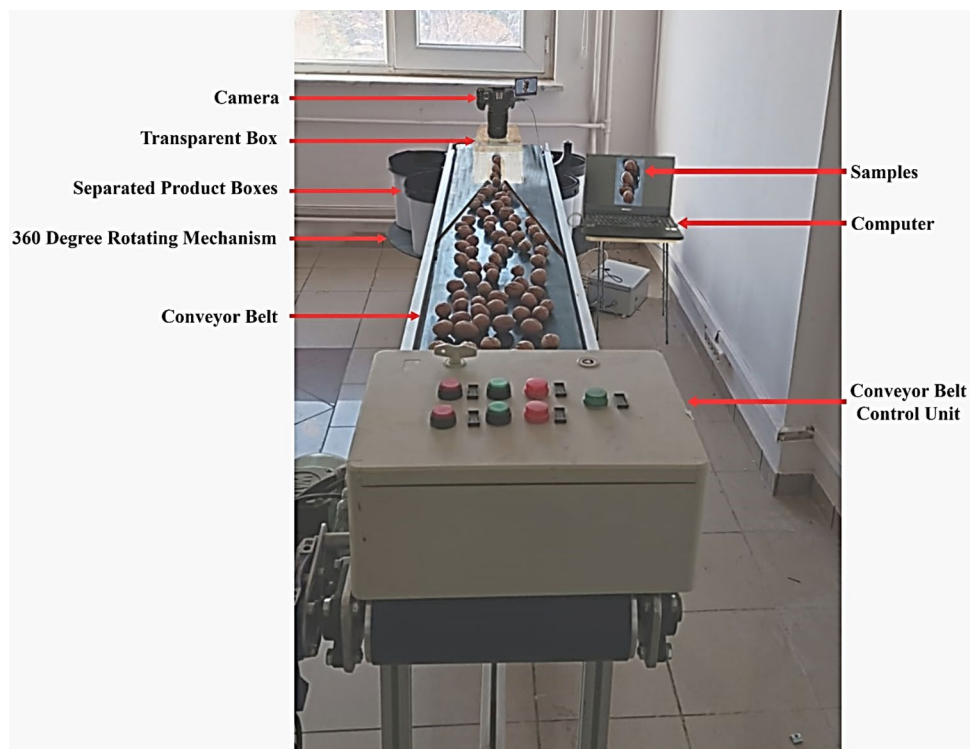


Fig. 3 Walnut species image samples used in this study

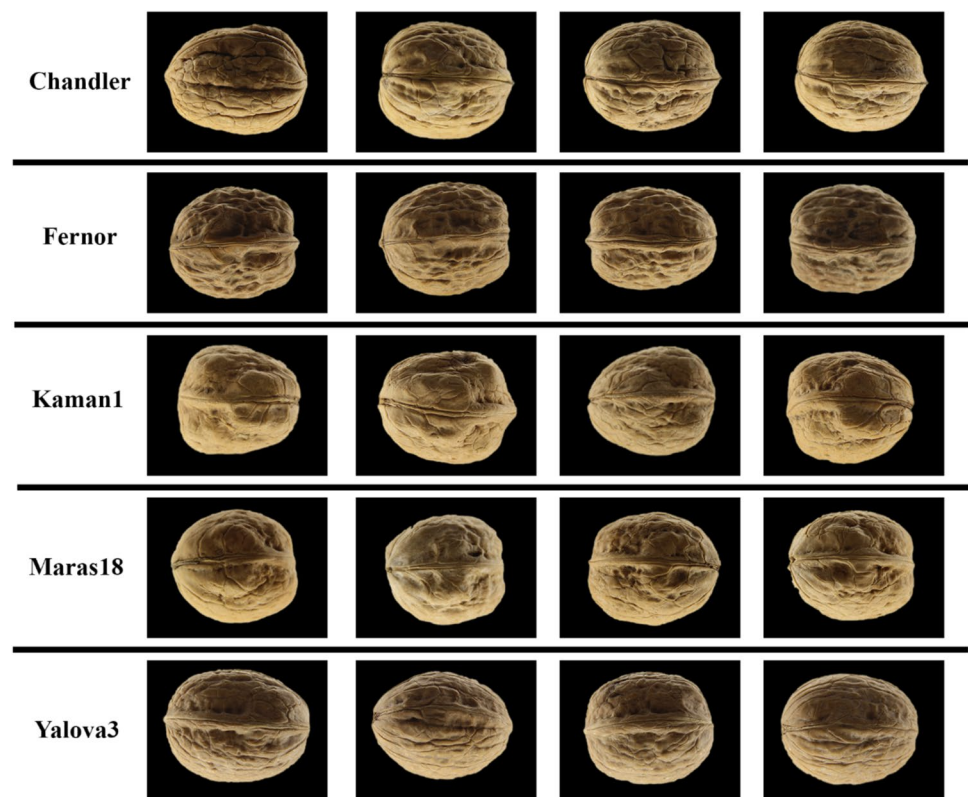


Fig. 4 A sample image containing 4 kernels from each of the 5 walnut species used in the study (walnuts numbered 1, 4, 5, 14 belong to Chandler, walnuts numbered 2, 10, 16, 18 belong to Yalova3, walnuts numbered 3, 11, 15, 19 belong to Fernor, walnuts numbered 6, 8, 12 and 20 belong to Kaman1 and walnuts numbered 7, 9, 13, 17 belong to Maras18)

given to the network, consist of convolution, pooling, activation, fully connected and classifier layers [20].

VGGNet

This model, developed by Karen Simonyan and Andrew Zisserman, reduced the top-5 error rate to 7.3% in the ImageNET competition held in 2014. The input image in all models of this convolutional neural network, designed in 6 different models with 11, 13, 16 and 19 layers, is 224×224 pixels [46].

ResNet

ResNet, developed by He et al. [23], came first in the ImageNet competition held in 2015 with an error rate of 3.57%. It has 5 different models consisting of 18, 34, 50, 101 and 152 layers [23].

DenseNet

The deep learning model introduced in 2016 with the article titled “Densely Connected Convolutional Networks” is referred to as “*Densely Connected Convolutional Networks*”. It is a convolutional neural network model that connects each layer to all the layers following it, thus providing the feature maps in the previous layers as input to the next layers. It has 4 different architectures: 121, 169, 201 and 261 [25].

Xception

Announced by François Chollet in 2016, this model is an improved and more performant version of the InceptionV3 model. The Xception model, consisting of 36 layers, aims to provide higher performance with fewer parameters [12].

Transfer learning

The most important reasons why convolutional neural networks are used effectively and achieve successful results in many areas such as classification and object recognition are their hierarchical layered structures and the deep features they learn as a result of training with datasets that usually contain millions of images [51]. However, when designing convolutional neural networks for a classification task such as determining the type of walnut, it is difficult to achieve successful results because there are not millions of data for the end-to-end training of the network. Therefore, it is a common practice to adapt convolutional neural networks, which are trained on the ImageNet dataset containing millions of images and have been successful in classification tasks, to our own classification problems by training them on datasets containing relatively fewer images [54]. Although there are different techniques in the literature for transferring learning, the most commonly used method is to freeze all or some of the layers of the previously trained convolutional neural network model and fix the parameter weights and train a few layers that will be newly added to the model with our own dataset [26]. In this study, transfer learning was applied using pre-trained models. This approach allowed the model to retain general features while adapting to the specific characteristics of our dataset.

Machine learning algorithms

To evaluate the classification performance of deep learning models, several widely used machine learning algorithms were also employed within the scope of comparative analysis. These algorithms are frequently referenced in the literature and are known to perform effectively across diverse data structures. They represent classical supervised learning approaches and serve as baseline methods for performance benchmarking against deep learning techniques. For the machine learning algorithms feature extraction taken place with VGG19. A total of 4096 features were extracted from each image and the classified using the following algorithms. The choice of VGG19 over VGG16 is based on its increased depth and enhanced feature extraction capability. The extracted features were inputted into the SVM, k-NN and RF algorithms.

Support vector machines (SVM)

Support Vector Machines are supervised learning algorithms that aim to find the optimal separating hyperplanes between different classes. SVM performs particularly well on high-dimensional datasets and can also handle non-linear classification tasks effectively through the use of kernel functions. Due to its robust performance, it has been widely applied in areas such as image classification, bioinformatics, and text mining [13].

k-Nearest neighbors (k-NN)

The k-Nearest Neighbors algorithm is a simple and intuitive classification method that assigns a class to a data point based on the majority class of its closest neighbors. As a non-parametric, instance-based learning approach, it does not involve an explicit training phase. k-NN performs well on small and low-dimensional datasets; however, it may become computationally expensive and less efficient when applied to large-scale data due to its reliance on distance calculations at inference time [14].

Random forest (RF)

Random Forest is an ensemble learning method that builds multiple decision trees and aggregates their outputs to improve predictive accuracy and generalization. Each tree in the forest is trained on a different subset of the data, and final predictions are typically made through majority voting. This approach reduces the risk of overfitting that is often associated with individual decision trees and enhances the model's robustness and performance on unseen data [8].

Performance metrics

Confusion matrix

It is used to measure the accuracy of the model by comparing the predicted values of the trained model with the actual values [55]. In the case of binary classification, the confusion matrix is a 2×2 matrix containing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). In the confusion matrix, the rows represent the true class of the sample, and the columns represent the classes predicted by the model [9]. The 5×5 confusion matrix used in this study is given in Table 2.

The following performance metrics can be calculated with the values obtained from the confusion matrix [24, 41].

Accuracy It is obtained by dividing the sum of all correct predictions (true positives and true negatives) of the trained

Table 2 5 × 5 Confusion Matrix

		Predicted Class				
Actual Class		Chandler	Fernor	Kaman1	Maras18	Yalova3
	Chandler	T ₁	F ₁₂	F ₁₃	F ₁₄	F ₁₅
	Fernor	F ₂₁	T ₂	F ₂₂	F ₂₃	F ₂₄
	Kaman1	F ₃₁	F ₃₂	T ₃	F ₃₃	F ₃₄
	Maras18	F ₄₁	F ₄₂	F ₄₃	T ₄	F ₄₄
	Yalova3	F ₅₁	F ₅₂	F ₅₃	F ₅₄	T ₅

model in the dataset by the total number of examples in the dataset. This metric shows the overall performance of the model and covers success in both positive and negative classes.

Precision It measures the accuracy of the model in positive predictions. It is calculated by dividing the number of true positive examples by the sum of all positive examples (true positives and false positives) predicted by the model. This metric shows how many true positive predictions the model makes.

Sensitivity It measures how well the model detects examples that are true positives. It is obtained by dividing the true positive examples by the sum of all positives (true positives and false negatives) predicted by the model. This metric expresses how well the model captures examples that belong to the positive class.

Specificity It is the metric that reveals how successful the model is in detecting examples that belong to the negative class.

F1-Score Is a metric that balances the performance of the model in terms of both Precision and Recall. The F1 Score is calculated by taking the harmonic meaning of these two metrics. It is useful in situations where both false positives and false negatives are important.

The formulas for performance measurement metrics are given in Table 3 [35], and the calculation of TP, TN, FP and FN values according to the classes from the confusion class is given in Table 4 [27].

Receiver operating characteristic (ROC) curve

A method for measuring the classification performance of the trained model for each class in the test dataset and provides a visual representation of this [7], also called ROC analysis, uses the relationship between the sensitivity and specificity metrics of the model. Sensitivity measures the

Table 3 Performance metrics formula

Metrics	Equations
<i>Accuracy</i>	$= (TP + TN) / (TP + TN + FP + FN)$
<i>Precision</i>	$= TP / (TP + FP)$
<i>Recall</i>	$= TP / (TP + FN)$
<i>Specificity</i>	$= TN / (TN + FP)$
<i>F1 Score</i>	$= 2 \times ((Precision \times Recall) / (Precision + Recall))$

rate of true positives (TP) classified by the model, while specificity measures the rate of true negatives (TN) classified [19].

The ROC graph is a two-dimensional graph representing the Sensitivity (True Positive Rate-TPR) on the y-axis depending on the Specificity (False Positive Rate-FPR) on the x-axis. The ROC curve is expressed as a curve with the starting point origin (0,0) and the ending point (1,1). The point (0,1) indicates a perfect classification. ROC curves measure the classification ability of the trained model on the examples [37].

Cross validation

Cross Validation is a method used to measure the predictive ability of the trained model. In cross validation, the dataset is divided into k subsets. While k-1 of the sets are used for training the model, 1/k sets are used for testing the model. This process is repeated k times until each set is used as a test set. The overall success of the trained model is calculated by taking the arithmetic average of the k-times repeated processes [4]. In this study, the 10-Fold Cross Validation method was used.

Paired T-test

An often-used statistical tool for evaluating the performance of two systems is the paired t-test. Demšar, addresses in detail how this test ought to be used to assess machine/deep

Table 4 Calculating TP, TN, FP and FN values according to classes from Confusion Matrix

Chandler	Fernor	Kaman1	Maras18	Yalova3
TP = T1	TP = T2	TP = T3	TP = T4	TP = T5
TN = T2 + T3 + T4 + T5 + F22 + F23 + F24 + F32 + F33 + F34 + F42 + F43 + F44 + F52 + F53 + F54	TN = T1 + T3 + T4 + T5 + F13 + F14 + F15 + F31 + F33 + F34 + F41 + F43 + F44 + F51 + F53 + F54	TN = T1 + T2 + T4 + T5 + F12 + F14 + F15 + F21 + F23 + F24 + F41 + F42 + F44 + F51 + F52 + F54	TN = T1 + T2 + T3 + T5 + F12 + F13 + F15 + F21 + F22 + F24 + F31 + F32 + F34 + F51 + F52 + F53	TN = T1 + T2 + T3 + T4 + F12 + F13 + F14 + F21 + F22 + F23 + F31 + F32 + F33 + F41 + F42 + F43
FN = F12 + F13 + F14 + F15	FN = F21 + F22 + F23 + F24	FN = F31 + F32 + F33 + F44	FN = F41 + F42 + F43 + F44	FN = F51 + F52 + F53 + F54
FP = F21 + F31 + F41 + F51	FP = F12 + F32 + F42 + F52	FP = F13 + F22 + F43 + F53	FP = F14 + F23 + F33 + F54	FP = F15 + F24 + F34 + F44

Table 5 Detailed information about the system that is used to make applications

Processor (CPU)	Intel® Core i7™ 10750 2.60 GHz
Graphics card	NVIDIA GeForce RTX 2060
RAM	32 GB DDR4 2933 MHz
Operating system	Ubuntu 22.04
CUDA Version	12.2
cuDNN Version	8.9.7

learning models' performance over several datasets. Two algorithms evaluated on the same dataset have their mean performances statistically significantly differently determined by this test. Considered in literature as a trustworthy comparison tool, the paired t-test has been used in this work to evaluate if the acquired performance measures show a statistically significant variance [15].

Weighted loss functions

Weighted loss functions are techniques developed to mitigate the negative effects of class imbalance on model performance. He and Garcia define this approach as follows: "A weighted loss function penalizes classification errors at different rates in order to make the learning algorithm more

sensitive to infrequent classes." [22]. Weighted loss was applied to better evaluate the impact of class imbalance, and the contribution of this function to model performance was analyzed. The general formula for computing class weight is given at Eq. 1.

$$Weight_i = \frac{Total\ Samples}{Number\ of\ Classes \times Samples\ in\ Class_i} \quad (1)$$

Experimental setup

The system information on which the applications are performed is given in Table 5, and the parameter information commonly used in all deep learning models is given in Table 6.

The decision to train each model for 10 epochs was based on early experimental results. During these initial trials, validation accuracy typically reached its highest point between the 6th and 8th epochs. After that, performance either plateaued or began to decline, particularly in deeper architectures like ResNet-50 and Xception, suggesting the onset of overfitting. To avoid this and to maintain efficiency, the number of epochs was capped at 10 for all models. This also ensured a fair comparison across different architectures. We observed peak performance before 10 epochs and designed

Table 6 Parameter information used in deep learning models

Input image dimensions	224 × 224 px
Optimizer	Adam
Learning rate	0.0001
Activation	Softmax
Batch size	16
Number of training rounds (Epochs)	10
Data Augmentation, Multiplication	Method Value
	Image rotation20
	Image horizontal shift0.2
	Image vertical shift0.2
	Image random skew0.15
	Image random zoom in and out0.2
	Image horizontal flipYes

our setup to ensure consistency and avoid overfitting. Training the models beyond this point offered little to no improvement and often resulted in unnecessary computational load without real performance gains. To enhance the ability of the models, data augmentation techniques were applied. These techniques were implemented solely on the training data, with no modifications made to the validation set. The parameters used for model training were determined through extensive experimental evaluations and are summarized in Table 6.

Table 7 shows the machine learning algorithm's parameter details used for developing and testing our application in this study.

Table 7 Parameter information used in machine learning models

Model	Parameters
Random Forest	number of trees = 500, max_depth = 5,
SVM	kernel = 'rbf', C = 1.0, gamma = 'auto', numerical tolerance = 0.001, Iteration limit = 100
k-NN	n_neighbors = 10, metric = 'euclidean', weight = 'uniform'

Experimental results

This section evaluates the performance of Walnut dataset depending on the different training strategy used. Firstly applying deep learning models (VGG16, VGG19, ResNet50, DenseNet121, and Xception) without data augmentation on the dataset, followed by data augmented training strategy, after that using a weighted loss function on the pre-trained deep learning and lastly machine learning algorithms (SVM, k-NN, and Random Forest) for the classification of walnut species.

Performance evaluation of deep learning models on dataset without augmentation

To assess the impact of techniques on model performance, pre-trained deep learning models were retrained without augmentation applied to the training data. The confusion matrix for all applied models presented in Fig. 5. Figure 6 visualizes the ROC curve results. The resulting performance metrics are presented in Table 8.

Monitor observations throughout the training process revealed clear signs of overfitting, particularly in models trained without data augmentation, most notably the Xception and VGG16 models. By exposing the models to varied transformations of the input data, data augmentation enhanced robustness and led to more stable and meaningful validation performance.

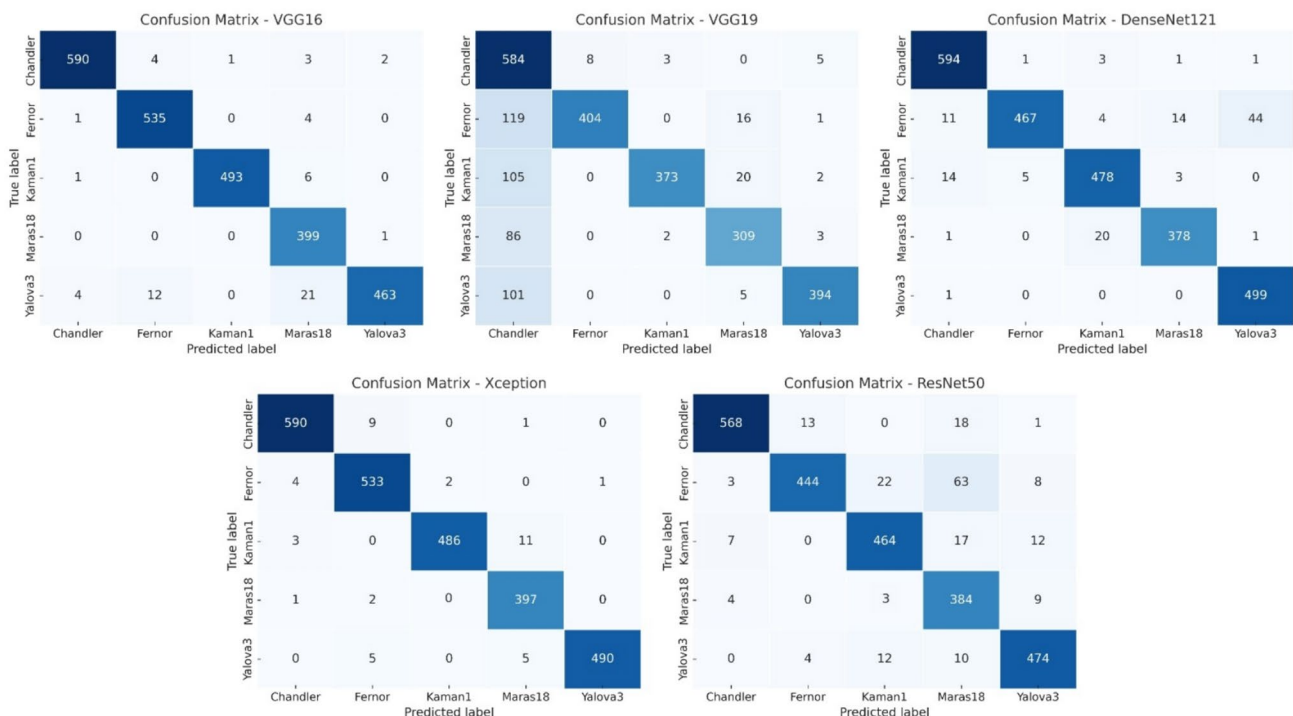


Fig. 5 Confusion Matrix Results for Walnut Dataset Without Augmentation Strategy

Performance evaluation of deep learning models on the dataset with data augmentation

The deep learning models (VGG16, VGG19, DenseNet-121, Xception and ResNet50) was trained and the resulting Confusion Matrix for each model given in Fig. 7. ROC curves and AUC values are given in Fig. 8.

According to the data in Table 9, the DenseNet-121 model fell behind the other models with low accuracy and F1 score values. Although the VGG16 and VGG19 models generally achieved compatible results, they could not reach the success achieved by the Xception and ResNet-50 models. The Xception model achieved successful results by reaching high accuracy. The ResNet-50 model was the model that showed the best classification performance on the entire dataset by reaching the highest accuracy, precision, sensitivity and F1 score values among the examined models.

Figure 9 shows the correct classification rates obtained by all models in the test dataset. It was determined that the Chandler class achieved high accuracy in ResNet-50, VGG19 and Xception models, but remained at lower accuracy results in DenseNet-121 and VGG16 models. It is seen that the Yalova3 class generally achieved high accuracy in all models, however, the models had difficulty distinguishing the Maras18 class. While the class that the VGG16 model made the most successful prediction was Fernor, the DenseNet-121 model fell behind the other models in all classes.

Table 8 Performance of Deep Learning Models Trained Without Data Augmentation

Model	Accuracy	Precision	Recall	F1-Score
VGG16	0.9763	0.9797	0.9764	0.9748
VGG19	0.8126	0.7801	0.8126	0.7831
DenseNet-121	0.9512	0.9638	0.9512	0.9469
Xception	0.9827	0.9850	0.9827	0.9830
ResNet-50	0.9189	0.9436	0.9189	0.9182

Performance evaluation of deep learning models with effect of using weighted loss function

The effect of applying a weighted loss function to mitigate class imbalance is examined. To assess whether the class imbalance in the dataset influences the model's ability to distinguish between classes, the models were retrained using a weighted loss function with the specified hyper-parameters.

Although the analysis revealed that the weighted loss function could slightly improve class-wise accuracy in models such as ResNet-50 and Xception, this approach generally resulted in negative outcomes in terms of overall accuracy and class balance across many models. Therefore, the final models in this study were trained using an unweighted loss function, which yielded more stable, accurate, and applicable results. The resulting confusion matrix, ROC curve and performance metrics are presented in Fig. 10, Fig. 11 and Table 10 in order.

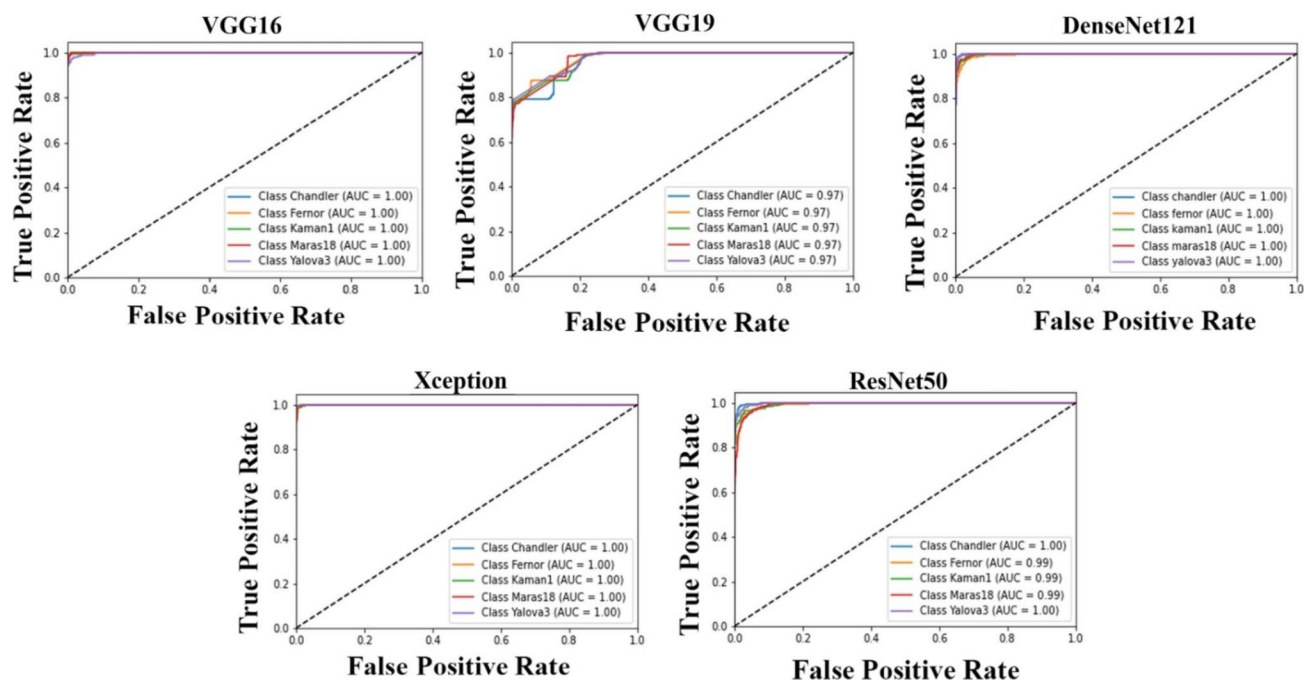


Fig. 6 Results of ROC curves for Walnut Dataset Without Augmentation Strategy

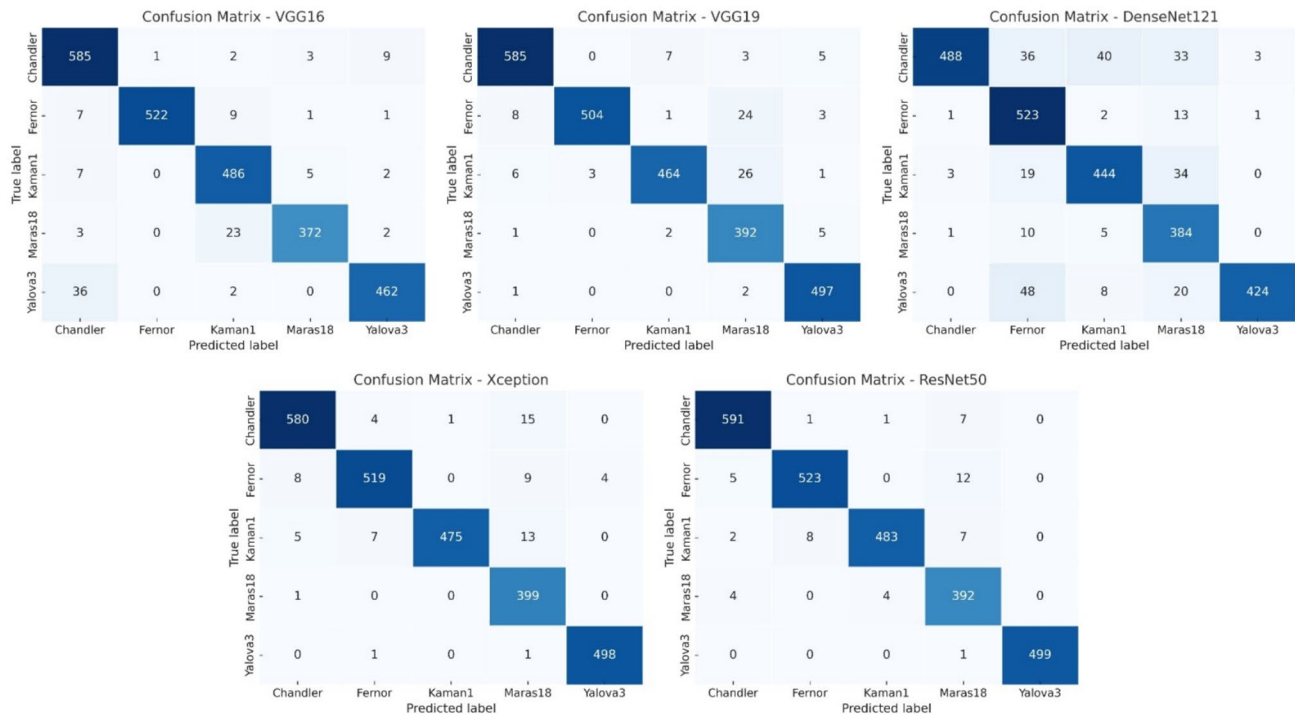


Fig. 7 Confusion Matrix Results for Walnut Dataset with Augmentation Strategy

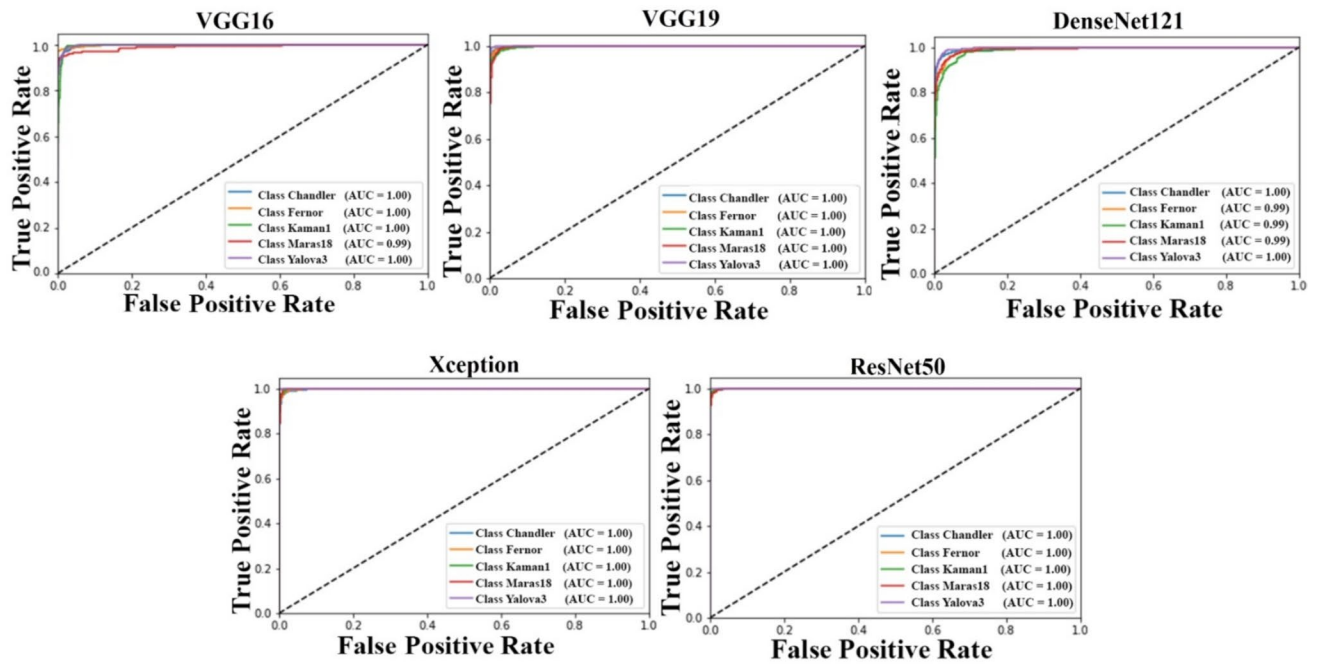


Fig. 8 Results of ROC curves for Walnut Dataset with Augmentation Strategy

Table 11 presents a comparative analysis of each model's class-wise discrimination performance based on accuracy values obtained before and after applying the weighted loss

function. This comparison was conducted to analyze the effect of the weighted loss function, particularly on classes with relatively fewer samples.

Table 9 Average results of the models on the used walnut dataset with data augmentation

Model	Accuracy	Precision	Recall	F1-Score
DenseNet-121	0.8909	0.9420	0.8909	0.8928
VGG16	0.9555	0.9696	0.9555	0.9533
VGG19	0.9614	0.9652	0.9614	0.9609
Xception	0.9728	0.9786	0.9728	0.9735
ResNet-50	0.9795	0.9817	0.9795	0.9796

Performance evaluation of machine learning algorithms on the dataset

To better contextualize the classification performance of deep learning models, several widely used classical machine learning algorithms were also applied to the same dataset. These included RF, SVM, and k-NN classifiers.

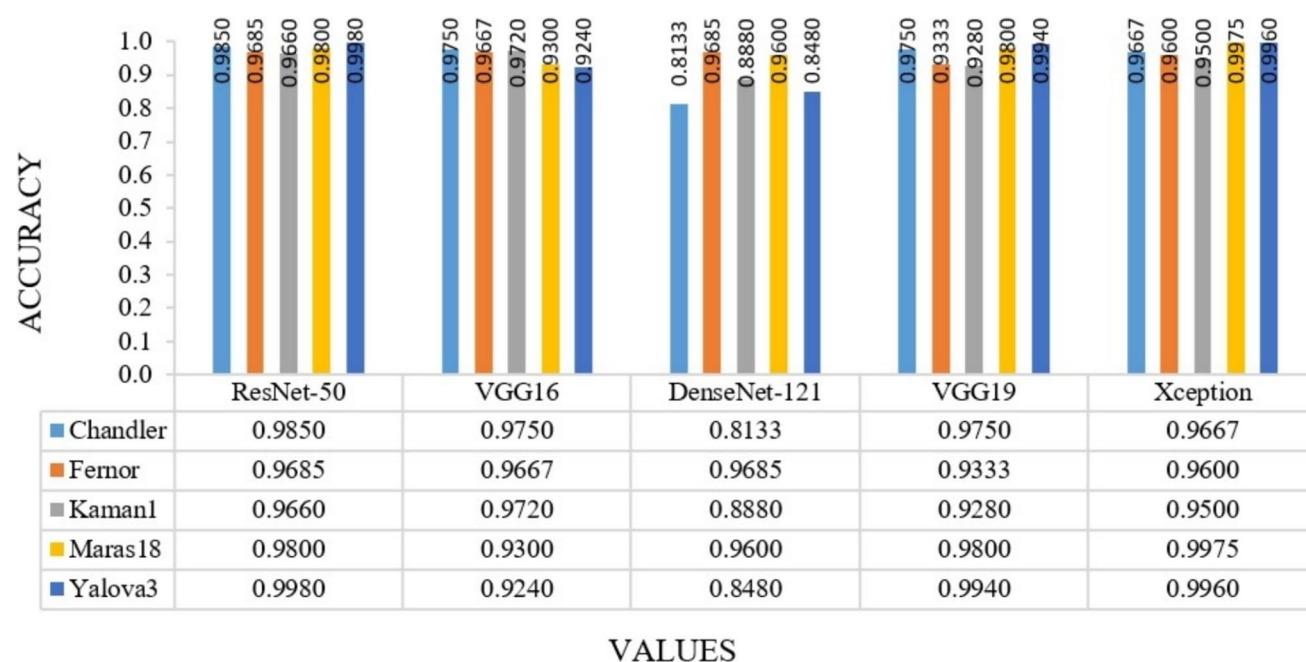
The comparison of confusion matrices for RF, SVM, and k-NN reveals noticeable differences in how each model handles class predictions as shown in Fig. 12. Overall, all three classifiers show strong diagonal dominance, indicating a generally good level of accuracy across all walnut species. However, some models handle specific classes better than others.

The RF model performs well across most classes, especially for Fernor and Yalova3, where misclassifications are minimal. It shows slightly more confusion in predicting Kaman1 and Chandler, with a moderate number of false positives spread across those classes.

The SVM model demonstrates a strong overall performance as well, particularly for Chandler and Fernor, where it records high true positive counts and fewer errors compared to k-NN. The SVM model seems to manage class boundaries more effectively, resulting in fewer misclassifications for Maras18 and Yalova3 than RF. In contrast, the k-NN model, while still performing reasonably well, exhibits more confusion between similar classes especially for Kaman1 and Maras18. It tends to misclassify a notable number of Kaman1 samples as Chandler, and likewise has difficulty separating Maras18 from Kaman1 and Chandler. This suggests that k-NN struggles more with fine-grained class distinctions, possibly due to its reliance on local feature similarity. While all three models are capable of performing multi-class classification on walnut varieties with high accuracy, SVM appears slightly more robust in maintaining class separation, whereas RF provides balanced performance. k-NN, though effective, is more sensitive to inter-class similarity and shows more overlap in its predictions.

The performance of these models was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The results, presented in Table 12, serve as a baseline for comparative analysis with deep learning architectures.

The results presented in Table 12 show that among the machine learning algorithms applied to the dataset, SVM achieved the highest performance with accuracy, precision, recall, and F1-score of 90.10%. This was followed by RF, which reached an accuracy of 83.40%, and k-NN, which recorded an accuracy of 80.10%. Although these results

**Fig. 9** Accuracy rates of algorithms on walnut species

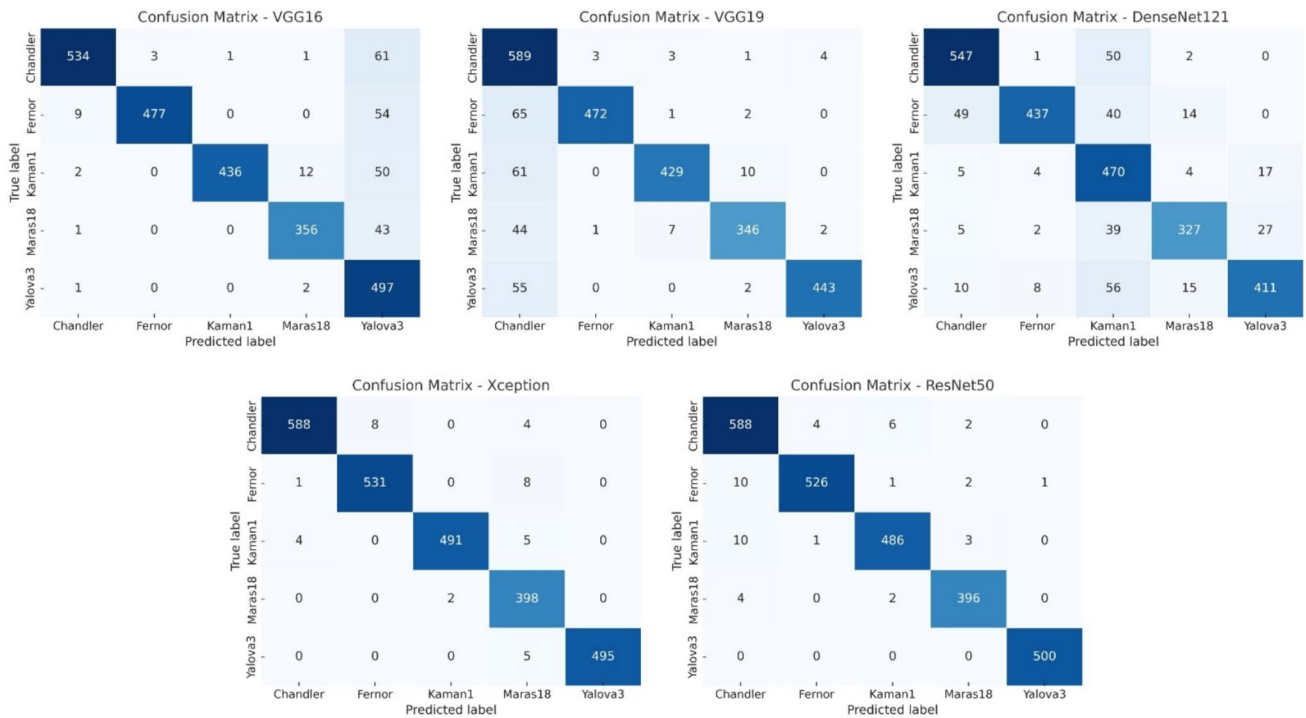


Fig. 10 Confusion Matrix Results for Walnut Dataset with Weighted Loss Function Strategy

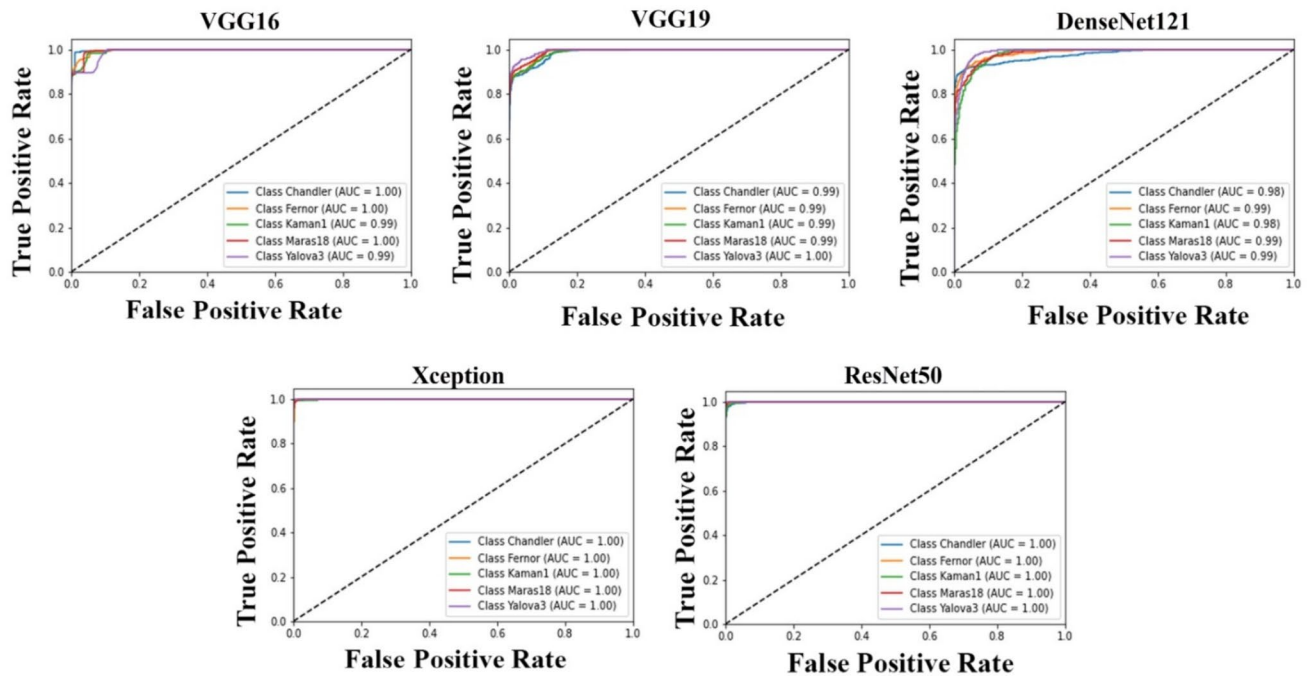


Fig. 11 Results of ROC curves for Walnut Dataset with Weighted Loss Function Strategy

reflect relatively strong performance by the machine learning models, especially SVM. These algorithms still fall short when compared to the deep learning models discussed in earlier part. Architectures such as Xception and VGG16

achieved accuracy and F1-scores exceeding 97%, highlighting their superior ability to capture complex patterns in high-dimensional feature spaces. Additionally, deep learning approaches demonstrated greater robustness to challenges

Table 10 Performance Metrics of Models Trained with a Weighted Loss Function

Model	Accuracy	Precision	Recall	F1-Score
VGG16	0.9055	0.8905	0.9055	0.8923
VGG19	0.8972	0.8830	0.8972	0.8830
DenseNet-121	0.8629	0.9311	0.8630	0.8533
Xception	0.9854	0.9880	0.9854	0.9858
ResNet50	0.9827	0.9844	0.9827	0.9827

such as class imbalance and variability within the data, ultimately delivering better improvement. Therefore, for the dataset used in this study, deep learning models clearly outperformed traditional machine learning methods.

Discussion and recommendations

The training times, total and trainable parameter counts, and FLOPS (Floating Point Operations Per Second) values of the models used in the system are presented in Table 13. According to the table, ResNet50, despite having the highest

Table 12 Performance Metrics of Machine Learning Algorithms on the Dataset

Model	Accuracy	Precision	Recall	F1-Score
RF	0.8340	0.8350	0.8340	0.8340
SVM	0.9010	0.9010	0.9010	0.9010
k-NN	0.8010	0.8120	0.8010	0.8000

FLOPS value, stands out with one of the shortest training times. This suggests that the model efficiently uses computational power and is well-optimized. DenseNet121 has the lowest total number of parameters, making it more memory-efficient. However, its relatively long training time indicates that it may not be as optimized in terms of computational efficiency. While VGG16 and VGG19 seem efficient in terms of FLOPS, their longer training times suggest architectural inefficiencies that may affect their overall performance. Xception is one of the most computationally efficient models, but its training time is still longer than some of the other models, indicating potential areas for further optimization.

ResNet50 is the most efficient model in terms of both training time and computational performance. DenseNet121

Table 11 Class-wise Accuracy of Models Trained with and Without Weighted Loss

Model	Chandler	Fernor	Kaman1	Maras18	Yalova3
VGG16	0.9750	0.9667	0.9720	0.9300	0.9240
VGG16 (with weight loss)	0.8900	0.8833	0.8720	0.8900	0.9940
VGG19	0.9750	0.9333	0.9280	0.9800	0.9940
VGG19 (with weight loss)	0.9817	0.8741	0.9533	0.8650	0.9911
DenseNet-121	0.8133	0.9685	0.8880	0.9600	0.8480
DenseNet-121 (with weight loss)	0.9117	0.8093	0.9400	0.8175	0.8474
Xception	0.9667	0.9600	0.9500	0.9975	0.9960
Xception (with weight loss)	0.9800	0.9833	0.9820	0.9950	0.9900
ResNet-50	0.9850	0.9685	0.9660	0.9800	0.9980
ResNet-50 (with weight loss)	0.9639	0.9741	0.9878	0.9851	1

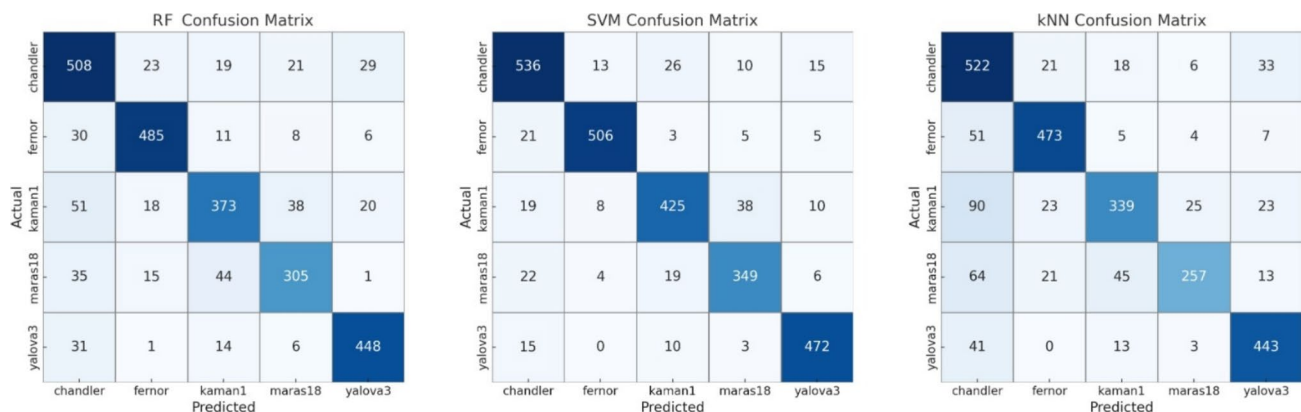
**Fig. 12** Confusion matrix visual representation for machine learning algorithms

Table 13 Comparison of Model Training Duration, Parameter Load, and Computational Complexity

Model	Training Period	Total Parameter Count	Trainable Parameter Count	FLOPS (Floating Point Operations Per Second)
VGG16	4170 s	14.717.253	14.717.253	71.762.772
VGG19	4680 s	20.026.949	20.026.949	86.297.184
DenseNet121	3420 s	7.042.629	6.958.981	80.056.895
Xception	3210 s	20.871.725	20.817.197	96.111.320
ResNet50	2540 s	23.597.957	23.544.837	105.756.149

offers memory efficiency due to its lower parameter count, while VGG16 and VGG19 appear to suffer from architectural inefficiencies due to their longer training times. Xception, although computationally efficient, could still benefit from improvements in training speed.

The average results obtained by the models on the entire dataset as a result of tenfold cross-validation are given in Table 14.

The deep learning models tested in this study showed strong performance overall, with some notable differences depending on the training strategy used.

Xception achieved the highest accuracy 98.27%, followed by closely by VGG16 at 97.63%, and DenseNet-121 at 95.12%. In contrast, VGG19 saw a drop in performance with accuracy falling to 81.26%. These results demonstrate that Xception and VGG16 consistently delivered strong and stable performance across all training settings, making them the most robust choices for this classification task. While VGG19's performance varied more noticeably depending on the training condition.

When using a weighted loss function, which helps address class imbalance, Xception and ResNet-50 stood out again with impressive accuracies of 98.54% and 98.27%, respectively. This suggests that these models are particularly effective at learning even when the dataset distribution is skewed.

Under the augmented dataset, where additional data was generated through transformations to improve prediction.

ResNet-50 achieved the highest accuracy on the original dataset, reaching 97.95%, closely followed by Xception at 97.28%, and VGG19 at 96.14%. VGG16 also performed well with 95.55%, while DenseNet-121 showed slightly lower accuracy at 89.09%.

Among the machine learning algorithms tested on the dataset, SVM delivered the best performance, correctly classifying the samples with an accuracy of 90.10%. RF also performed well, achieving an accuracy of 83.40%, while k-NN followed with an accuracy of 80.10%. These results indicate that while all three algorithms were effective, SVM was the most accurate in distinguishing between the different classes in this study.

Despite the high inter-class discriminative ability of the models, some walnut types are occasionally misclassified. This issue becomes more pronounced particularly among classes that share similar visual characteristics. Sample images of such misclassifications made by the models are presented in Figs. 13, 14, 15, 16 and 17. These visuals illustrate the challenges encountered by the models during the classification process and contribute to the overall error analysis.

Various deep learning and machine learning models were compared based on core classification metrics including accuracy, precision, recall, and F1-score, for the task of walnut classification. The results show substantial performance variability across models depending on the data processing strategy employed. These differences were statistically

Table 14 Average results of the models on the used walnut dataset

Models		Accuracy Without Augmentation	Accuracy for Weighted Loss Function	Accuracy with Augmentation
Deep Learning Models	VGG16	0.9763	0.9055	0.9555
	VGG19	0.8126	0.8972	0.9614
	DenseNet-121	0.9512	0.8629	0.8909
	Xception	0.9827	0.9854	0.9728
	ResNet-50	0.9189	0.9827	0.9795
Machine Learning Algorithms	Random Forest	0.8340	-	-
	SVM	0.9010	-	-
	k-NN	0.8010	-	-

assessed through paired T-tests to evaluate the significance of model performance gaps.

According to Table 8, which reflects the models trained with standard data, the ResNet-50 model demonstrated significantly superior performance compared to all other deep learning models. The F1-score difference with

DenseNet-121 was 0.0868 ($p = 0.0021$), indicating a highly significant result. Comparisons with VGG16 ($p = 0.0174$) and VGG19 ($p = 0.0312$) also revealed statistically significant differences, while the difference with Xception was less pronounced but still statistically meaningful ($p = 0.0448$).

Fig. 13 Some misclassified samples by the VGG16 model

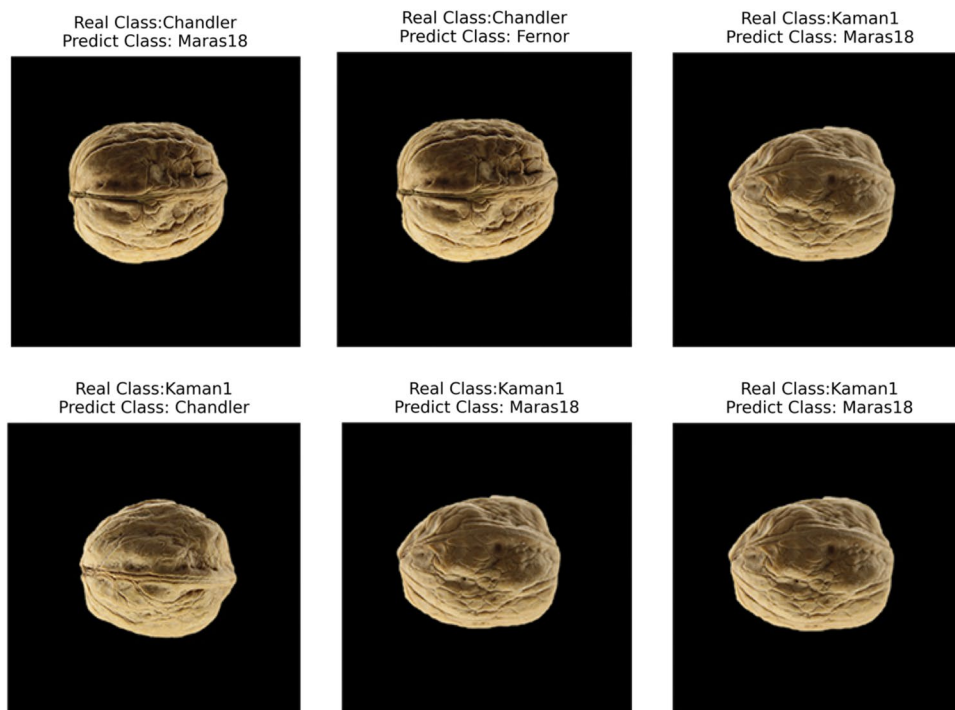


Fig. 14 Some misclassified samples by the VGG19 model

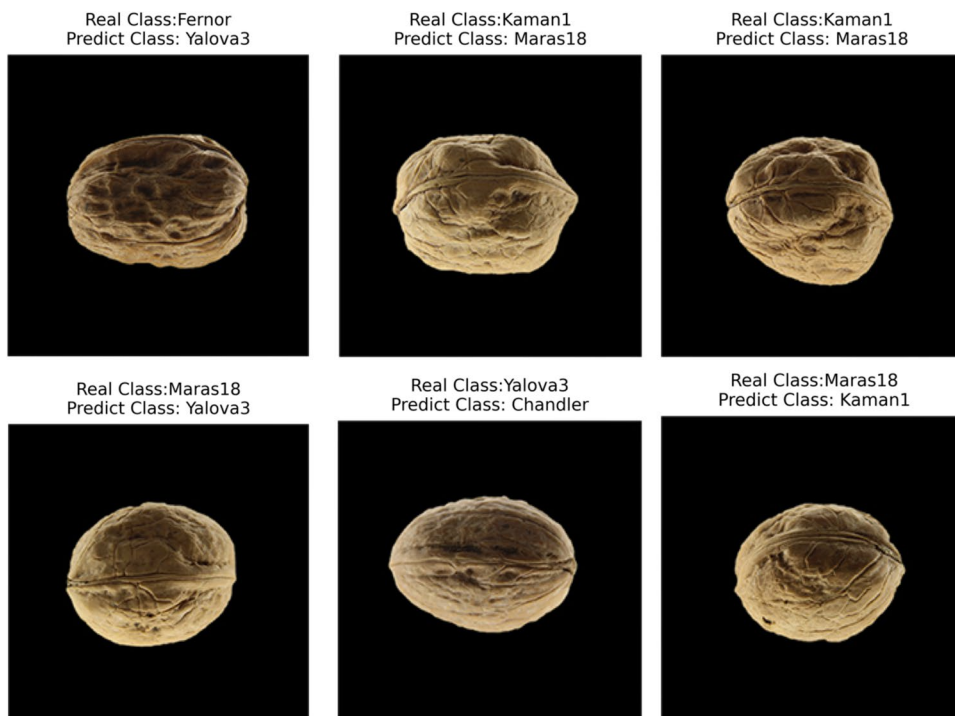


Fig. 15 Some misclassified samples by the DenseNet-121 model

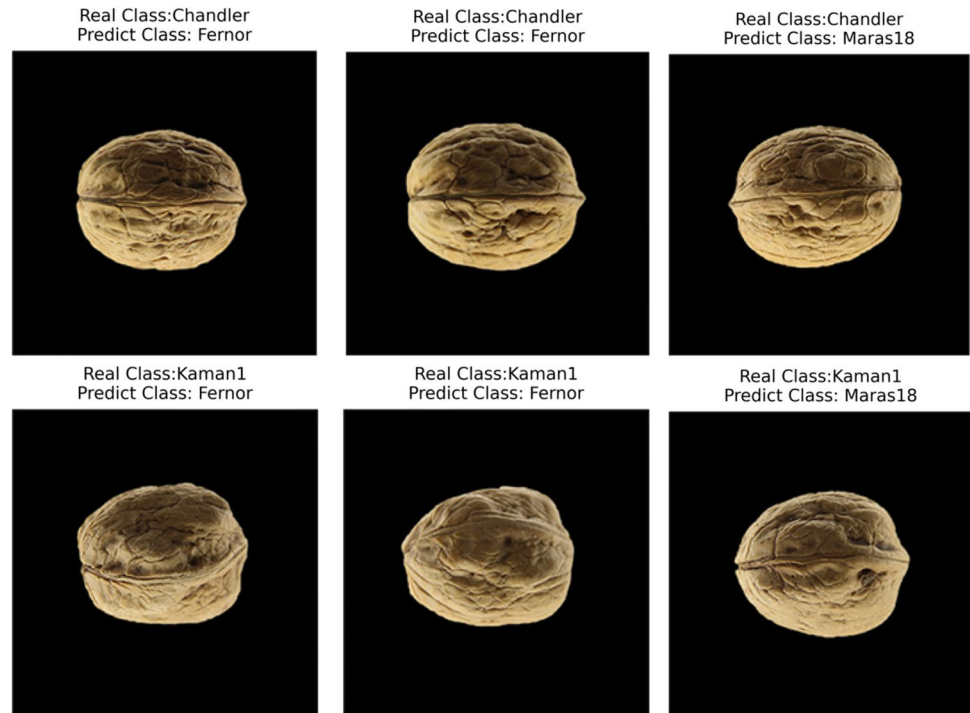
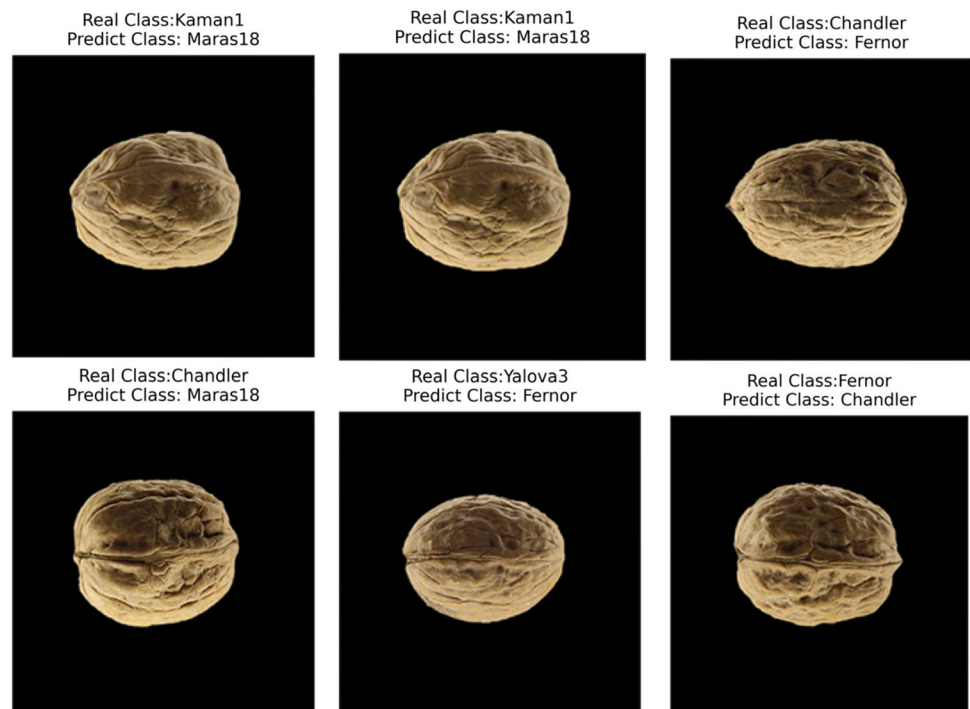


Fig. 16 Some misclassified samples by the Xception model



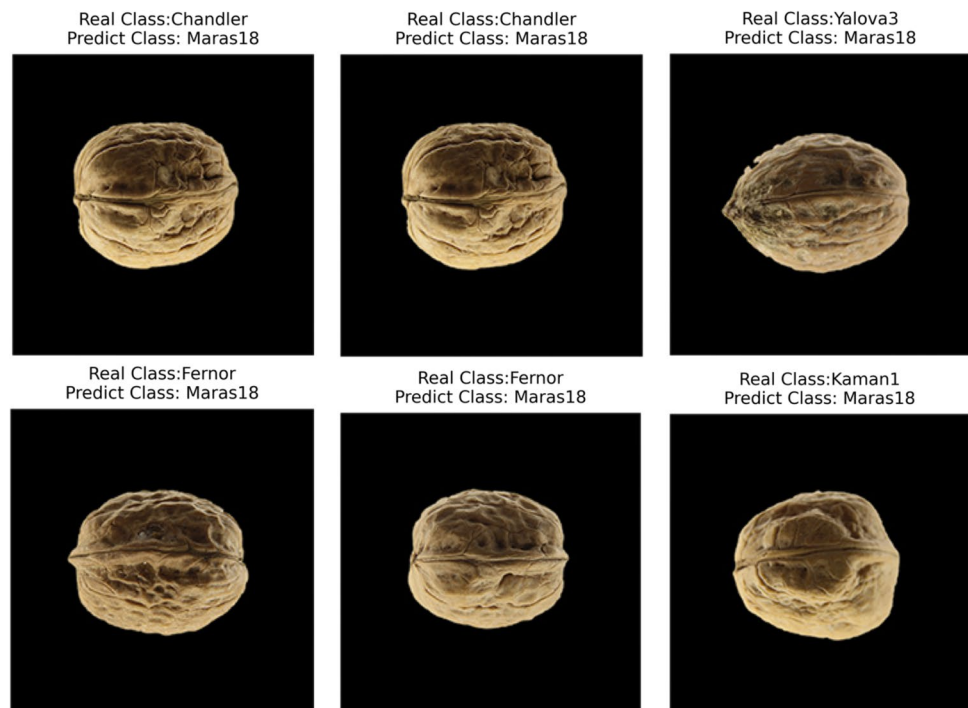
Thus, under standard conditions, ResNet-50 was identified as the top-performing deep learning model.

Based on Table 9, which involves models trained using data augmentation, the Xception model again achieved the best performance, with an F1-score of 98.30%. The performance gap with VGG19 was notably large (0.1999) and statistically highly significant ($p < 0.001$). Although VGG16

and DenseNet-121 exhibited near-optimal results, the differences were still statistically significant ($p < 0.05$). Notably, ResNet-50 underperformed in this setting (F1-score ≈ 91.82), further emphasizing Xception's statistical superiority when data augmentation is used.

In Table 10, which reflects the use of a weighted loss function, the Xception model achieved the highest

Fig. 17 Some misclassified samples by the ResNet-50 model



performance across all metrics, exceeding 98%. Although ResNet-50 performed comparably well (accuracy = 98.27%), the performance difference was not statistically significant ($p > 0.05$). However, the differences in F1-score between Xception and VGG16 (0.0935) and VGG19 (0.1028) were statistically significant ($p < 0.01$), confirming Xception's superiority under weighted classification settings.

As shown in Table 12, among the traditional machine learning algorithms, SVM (Support Vector Machine) achieved the best performance with an F1-score of 90.10%. Paired T-test comparisons confirmed that SVM significantly outperformed both Random Forest and k-NN ($p < 0.01$). Therefore, SVM is the most statistically successful classical method for this dataset. While model performance varied depending on the training strategy, the Xception model consistently emerged as the most statistically successful deep learning architecture, particularly when using weighted loss and data augmentation. ResNet-50 showed dominance only in the standard training setting, and SVM stood out among machine learning algorithms. These results demonstrate that for critical tasks such as walnut classification, both the selected model and data preprocessing strategy play crucial roles in achieving optimal performance.

Conclusion

This study is important in that it shows that deep learning models can be used for accurate and fast classification of walnut species. While the human factor is actively involved

in traditional classification methods, the transition to autonomous classification methods can be accelerated with this study. The fact that the models trained in the study can detect the differences between walnut species and classify with high accuracy rates reveals that time and labor can be saved in the field of agriculture.

As a result of this study, a fast and automatic classification method supported by artificial intelligence that can be used instead of manual separation processes that require expertise and take time has been proposed. According to the success rates obtained as a result of experimental studies; ResNet-50 and Xception models can be used in the separation of walnut species. The use of deep learning models in walnut classification can contribute to the increase and dissemination of the production of these species by determining higher quality and more efficient species in walnut production. The determination of higher quality walnut species, the production and dissemination of these species will increase the added value of walnuts by increasing the product quality.

Based on the findings of this study, ResNet-50 stands out as the most successful model in terms of accuracy, precision, recall, F1-score, and training time. However, its high FLOPS value (105.75 M) and large number of parameters (23.59 M) make it less suitable for embedded and mobile systems, where computational efficiency and memory usage are key concerns. In such scenarios, DenseNet-121 appears to be a more practical alternative, thanks to its lower parameter count (7.04 M) and relatively lower FLOPS (80.05 M), making it a more efficient option. If an even lighter model is needed, MobileNetV2 or EfficientNet-Lite could also be

considered as viable alternatives, offering lower computational costs while maintaining reasonable performance.

Within the scope of this study, Kaman1, Yalova3, Maras18, Chandler and Fernor species were used in training the models and successful classification results were obtained. In addition to these; the number of species recognized by the models can be increased by adding walnut species such as Pedro, Şebin, Yavuz, Hartley, Serr, Gültekin, Bilecik, Sütyemez, Şen, Franquette. Although the current dataset includes five widely cultivated walnut varieties, it is geographically limited. In future studies, we aim to incorporate data from additional regions and varieties to reduce geographic bias and enhance the model's applicability across diverse cultivation conditions. In addition, with the use of a mobile application that can be developed and a lighter convolutional neural network, the type of scanned walnut image can be determined, and it can contribute to the selection of the right walnut species by providing informative and guiding information about the species such as internal ratio, protein percentage, mineral amounts, suitable regional information for cultivation, etc.

Acknowledgements This study is based on Halil KILIF's unpublished master's thesis under the supervision of Ilkay CINAR.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability Data supporting the findings of this study are available upon request from the corresponding author.

Declarations

Conflicts of interest The authors undertake that there are no conflicts of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Y. Akca, S. Ozongun, Selection of late leafing, late flowering, laterally fruitful walnut (*Juglans regia*) types in Turkey. *N. Z. J. Crop. Hortic. Sci.* **32**(4), 337–342 (2004). <https://doi.org/10.1080/01140671.2004.9514313>
2. B. Akcay, H.O. Yılmaz, Bazı fonksiyonel besinlerin sağlık üzerindeki koruyucu etkileri. *Ank. Sağlık Bilimleri Derg.* **8**(2), 9–19 (2019)
3. A. Anagnostis, A.C. Tagarakis, G. Asimani, E. Papageorgiou, D. Kateris, D. Moshou, D. Bochtis, A deep learning approach for anthracnose infected trees classification in walnut orchards. *Comput. Electron. Agric.* **182**, 105998 (2021). <https://doi.org/10.1016/j.compag.2021.105998>
4. S. Arlot, A. Celisse, A survey of cross-validation procedures for model selection. (2010). <https://doi.org/10.1214/09-SS054>
5. E. Bakkalbaşı, Ö.M. Yılmaz, N. Artık, Türkiye'de yetiştirilen yerli bazı ceviz çeşitlerinin fiziksel özellikleri ve kimyasal bileşenleri. *Akademik Gıda* **8**(1), 6–54 (2010)
6. S. Bayazit, H. Tefek, O. Çalışkan, Türkiye'de ceviz (*Juglans regia* L.) araştırmaları. *Ziraat Fak. Derg.* **11**(1), 169–179 (2016)
7. A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* **30**(7), 1145–1159 (1997). [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
8. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
9. O. Caelen, A Bayesian interpretation of the confusion matrix. *Ann. Math. Artif. Intell.* **81**(3–4), 429–450 (2017). <https://doi.org/10.1007/s10472-017-9564-8>
10. Ş. Çavuşoğlu & Y. Uzun (eds.), *Pratik tarım ve sürdürülebilirliğin yansımaları* [in Turkish], pp. 3–26, (2022). Iksad Publications.
11. S. Chen, D. Dai, J. Zheng, H. Kang, D. Wang, X. Zheng, X. Gu, J. Mo, Z. Luo, Intelligent grading method for walnut kernels based on deep learning and physiological indicators. *Front. Nutr.* **9**, 1075781 (2023). <https://doi.org/10.3389/fnut.2022.1075781>
12. F. Chollet, Xception: deep learning with depthwise separable convolutions. 2017 IEEE Conf. Comput. Vis. Pattern Recog. (CVPR) 1800–1807 (2016). <https://doi.org/10.1109/CVPR.2017.195>
13. C. Cortes, V. Vapnik, Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995). <https://doi.org/10.1007/BF00994018>
14. T. Cover, P. Hart, Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**(1), 21–27 (1967)
15. J. Demšar, Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7**(Jan), 1–30 (2006)
16. S. Ercisli, B. Sayinci, M. Kara, C. Yildiz, I. Ozturk, Determination of size and shape features of walnut (*Juglans regia* L.) cultivars using image processing. *Sci. Hortic.* **133**, 47–55 (2012). <https://doi.org/10.1016/j.scienta.2011.10.014>
17. K.P. Ferentinos, Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **145**, 311–318 (2018). <https://doi.org/10.1016/j.compag.2018.01.009>
18. Y. Ferhatoğlu, Ceviz yetiştiriciliğinin genel prensipleri [in Turkish] (Publication No. 78). Atatürk Bahçe Kültürleri Merkez Araştırma Enstitüsü, Yalova, (1999)
19. P.A. Flach, ROC analysis. in *Encyclopedia of machine learning and data mining* (Springer, 2016), pp. 1–8. https://doi.org/10.1007/978-1-4899-7502-7_739-1
20. B. Gencturk, S. Arsoy, Y.S. Taspinar, I. Cinar, R. Kursun, E.T. Yasin, M. Koklu, Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models. *Eur. Food Res Technol.* 1–14. <https://doi.org/10.1007/s00217-023-04369-9>
21. Y. Gulzar, Z. Ünal, T. Kızıldeniz, U.M. Umar, Deep learning-based classification of alfalfa varieties: a comparative study using a custom leaf image dataset. *MethodsX* **13**, 103051 (2024). <https://doi.org/10.1016/j.mex.2024.103051>

22. H. He, E.A. Garcia, Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21**(9), 1263–1284 (2009). <https://doi.org/10.1109/TKDE.2008.239>
23. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. 2016 IEEE Conf. Comput. Vis. Pattern Recog. (CVPR) 770–778. <https://doi.org/10.1109/CVPR.2016.90>
24. E.E.-D. Hemdan, M.A. Shouman, M.E. Karar, Covidx-net: a framework of deep learning classifiers to diagnose covid-19 in x-ray images (2020), Preprint at <https://arxiv.org/abs/2003.11055>. <https://doi.org/10.48550/arXiv.2003.11055>
25. G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks. 2017 IEEE Conf. Comput. Vis. Pattern Recog. (CVPR) 2261 (2017). <https://doi.org/10.1109/CVPR.2017.243>
26. M. Iman, H.R. Arabnia, K. Rasheed, A review of deep transfer learning and recent advancements. *Technologies* **11**(2), 40 (2023). <https://doi.org/10.3390/technologies11020040>
27. H. Isik, S. Tasdemir, Y.S. Taspinar, R. Kursun, I. Cinar, A. Yasar, E.T. Yasin, M. Koklu, Maize seeds forecasting with hybrid directional and bi-directional long short-term memory models. *Food Sci. Nutr.* (2023). <https://doi.org/10.1002/fsn3.3783>
28. A. Kamilaris, F.X. Prenafeta-Boldú, Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **147**, 70–90 (2018). <https://doi.org/10.1016/j.compag.2018.02.016>
29. A.T. Karadeniz, E. Başaran, Y. Çelik, Identification of walnut variety from the leaves using deep learning algorithms. *Bitlis Eren Univ. Fen Bilimleri Derg.* **12**(2), 531–543 (2023). <https://doi.org/10.17798/bitlisfen.1263130>
30. A.T. Karadeniz, E. Başaran, Y. Çelik, Classification of walnut dataset by selecting CNN features with whale optimization algorithm. *Multimed. Tools Appl.* 1–16 (2024). <https://doi.org/10.1007/s11042-024-18586-1>
31. K. Kayaalp, A. Altınalan, Ceviz sürgün görüntüleri kullanılarak ceviz çeşitlerinin sınıflandırılması. *Uluborlu Mesleki Bilimler Derg.* **6**(2), 42–52 (2023)
32. R. Khakrangin, D. MohamadZamani, S.M. Javidan, Recognition of fill and empty walnuts using acoustic analysis and fuzzy logic. *J. Nuts* **12**(1), 17–30 (2021). <https://doi.org/10.22034/jon.2021.1918546.1103>
33. M.A. Khan, M. Ali, M. Shah, T. Mahmood, M. Ahmad, N. Jhanjhi, M.A.S. Bhuiyan, E.S. Jaha, Machine learning-based detection and classification of walnut fungi diseases. *Intell. Autom. Soft Comput.* **30**(3), 771–785 (2021). <https://doi.org/10.32604/iasc.2021.018039>
34. A. Koirala, K.B. Walsh, Z. Wang, C. McCarthy, Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of ‘MangoYOLO.’ *Precis. Agric.* **20**(6), 1107–1135 (2019). <https://doi.org/10.1007/s11119-019-09642-0>
35. D. Krstinić, M. Braović, L. Šerić, D. Božić-Štulić, Multi-label classifier performance evaluation with confusion matrix. *Comput. Sci. Inf. Technol.* **1** (2020). <https://doi.org/10.5121/csit.2020.100801>
36. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998). <https://doi.org/10.1109/5.726791>
37. M. Majnik, Z. Bosnić, ROC analysis of classifiers in machine learning: a survey. *Intell. Data Anal.* **17**(3), 531–558 (2013). <https://doi.org/10.3233/IDA-130592>
38. H. Muresan, M. Oltean, Fruit recognition from images using deep learning. *Acta Univ. Sapientiae Informatica* **10**(1), 26–42 (2018). <https://doi.org/10.48550/arXiv.1712.00580>
39. I. Ozkan, M. Koklu, Classification of walnut using image processing and fuzzy inference method. 8th international conference on advanced technologies (ICAT'19), 1, pp. 432–435, Sarajevo, Bosnia-Herzegovina (2019)
40. A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, A. Johannes, Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* **161**, 280–290 (2019). <https://doi.org/10.1016/j.compag.2018.04.002>
41. D.M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation (2020), Preprint at <https://arxiv.org/abs/2010.16061>. <https://doi.org/10.48550/arXiv.2010.16061>
42. C.K. Rai, R. Pahuja, Classification of diseased cotton leaves and plants using improved deep convolutional neural network. *Multimed. Tools Appl.* **82**(16), 25307–25325 (2023). <https://doi.org/10.1007/s11042-023-14933-w>
43. Z. Sanlin, Z. Liping, Z. Weiqiang, G. Zhuang, F. Ziqiang, Identification and localization of walnut varieties based on YOLOv5. *J. Chin. Agric. Mechanization* **43**(7), 167 (2022). <https://doi.org/10.13733/j.jcam.issn.2095-5553.2022.07.024>
44. S.M. Şen, *Ceviz yetiştiriciliği [in Turkish]*. (Eser Matbaası, Samsun, 1986), pp. 109–229
45. Y. Sesli, *Bazı ceviz (Juglans regia L.) çeşitlerinin tohum anacı olarak kullanılabilme potansiyellerinin belirlenmesi üzerine araştırmalar [Doctoral dissertation, Fen Bilimleri Enstitüsü, Adnan Menderes Üniversitesi] [in Turkish]* (2016)
46. K. Simonyan, A. Zisserman, A. Very deep convolutional networks for large-scale image recognition (2014), Preprint at <https://arxiv.org/abs/1409.1556>. <https://doi.org/10.48550/arXiv.1409.1556>
47. M. Şimşek, E. Gülsoy, Ceviz ve içediği yağ asitlerinin insan sağlığı açısından önemi üzerine yapılan bazı çalışmalar. *J. Inst. Sci. Technol./Fen Bilimleri Enstitüsü Derg.* **6**(4) (2016)
48. M. Sutyemez, New walnut cultivars: maras 18, sutyemez 1, and kaman 1. *HortScience* **51**(10), 1301–1303 (2016). <https://doi.org/10.21273/HORTSCI10972-16>
49. E.C. Too, L. Yujian, S. Njuki, L. Yingchun, A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* **161**, 272–279 (2019). <https://doi.org/10.1016/j.compag.2018.03.032>
50. B. Yalçın, A. Yıldırım, F. Yıldırım, C. Çelik, Bazı ceviz çeşitlerinin isparta ekolojisinde biyokimyasal ve mineral madde içerikleri. *Adnan Menderes Üniv. Ziraat Fak. Derg.* **18**(2), 285–291 (2021). <https://doi.org/10.25308/aduziraat.993686>
51. Z. Yang, W. Yu, P. Liang, H. Guo, L. Xia, F. Zhang, Y. Ma, J. Ma, Deep transfer learning for military object recognition under small training set condition. *Neural Comput. Appl.* **31**, 6469–6478 (2019). <https://doi.org/10.1007/s00521-018-3468-3>
52. A. Yaşar, İ.K. Özyurt, E. Kaplan, Comparison of some local and foreign walnut cultivars. *J. Agric. Fac. Gaziosmanpaşa Univ. (JAFAG)* **35**(3), 290–296 (2018). <https://doi.org/10.13002/jafag.4449>
53. E. Yıldız, A. Sümbül, Mineral content in kernel of some local and foreign walnut cultivars and genotypes. *Mustafa Kemal Univ. Tarım Bilimleri Derg.* **24**(3), 174–180 (2019)
54. Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, Z. Zhao, Deep transfer learning for modality classification of medical images. *Information* **8**(3), 91 (2017). <https://doi.org/10.3390/info8030091>
55. G. Zeng, On the confusion matrix in credit scoring and its analytical properties. *Commun. Stat.-Theory Methods* **49**(9), 2080–2093 (2020). <https://doi.org/10.1080/03610926.2019.1568485>
56. X. Zhang, Y. Qiao, F. Meng, C. Fan, M. Zhang, Identification of maize leaf diseases using improved deep convolutional neural networks. *Ieee Access* **6**, 30370–30377 (2018). <https://doi.org/10.1109/ACCESS.2018.2844405>
57. Y. Zhang, X. Wang, Y. Liu, Z. Li, H. Lan, Z. Zhang, J. Ma, Machine vision-based chinese walnut shell-kernel recognition

- and separation. *Appl. Sci.* **13**(19), 10685 (2023). <https://doi.org/10.3390/app131910685>
58. W. Zhu, B. Xiaopeng, X. Daochun, W. Li, Pruning branch recognition and pruning point localization for walnut (*Juglans regia* L.) trees based on point cloud semantic segmentation. *Agriculture* **15**(8), 817 (2025). <https://doi.org/10.3390/agriculture15080817>

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