jotaf

Journal of Tekirdag Agricultural Faculty
Tekirdağ Ziraat Fakültesi Dergisi

Eylül/September 2024, 21(4)
Başvuru/Received: 14/02/23
Kabul/Accepted: 01/08/24
DOI: 10.33462/jotaf.1250018

http://dergipark.gov.tr/jotaf
http://jotaf.nku.edu.tr/

ARAŞTIRMA MAKALESİ **RESEARCH ARTICLE**

Classification of Open and Closed Pistachio Shells Using Machine Vision Approach

Makina Görme Yaklaşımı Kullanılarak Açık ve Kapalı Antep Fıstığı Kabuklarının Sınıflandırılması

Khaled Adil Dawood IDRESS[1*], Yeşim Benal ÖZTEKİN[2], Omsalma Alsadig Adam GADALLA[3]

Abstract

Pistachio nuts are a type of nut that is widely consumed around the world due to their high nutritional value and pleasant taste. Pistachios are usually sold in their shells, either open or closed. However, closed-shell pistachios are not well received by consumers, resulting in a lower commercial value. It is essential to be able to distinguish between open and closed pistachio shells in order to ensure quality control during production processes and processing. This can be done manually or by using mechanical devices. Manual inspection and categorization of pistachio nuts have traditionally been done by workers, but this process is inefficient in terms of time and money. Mechanical separation of open and closed-shell pistachio can damage the kernels of open-shell nuts due to the needle mechanism used in the sorting process. This study aims to classify pistachio nuts using a machine vision-based system and evaluate its applicability in terms of classification accuracy. The system is evaluated on the Antep pistachio species, which can be distinguished from other pistachio varieties, such as Siirt and Urfa pistachios, based on their shape, size, and taste properties. The machine vision system in this study classifies pistachio nuts into closed and open shell classes in a completely automated manner. In this study, 1,000 Antep pistachio nuts images were obtained and examined, including 500 open and 500 closed nuts. The images were pre-processed and prepared for feature extraction. From the images, a total of 14 color features were extracted. Although the single feature was used, promising classification accuracy rates of 95.6%, 94.8%, and 93.6% from the Random Forest, Support Vector Machine (SVM), and Logistic Regression were achieved, respectively. The performances of classifiers were compared to each other. Almost similar performances were detected. These results demonstrate that the Random Forest classifier is the most effective algorithm for classifying open and closed Antep pistachio nuts.

Keywords: Pistachio, Image processing, Color feature, Logistic regression, Random forest, Support vector machine

[1]*Sorumlu Yazar/**Corresponding Author:** Khaled Adil Dawood Idress, Ondokuz Mayis University, Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, Samsun, Turkey. E-mail: adilkhaled850@gmail.com OrcID: 0000-0002-1631-6232
[2]Yeşim Benal Öztekin, Ondokuz Mayis University, Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, Samsun, Turkey. E-mail: yurtlu@omu.edu.tr OrcID: 0000-0003-2387-2322
[3]Omsalma Alsadig Adam Gadalla, Ondokuz Mayis University, Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, Samsun, Turkey. E-mail: omsalmagadalla@gmail.com OrcID: 0000-0001-6132-4672
Atıf: Idress. Kh .A. D., Öztekin, Y. B., Gadalla,. O. A. (2024). Makina görme yaklaşımı kullanılarak açık ve kapalı antep fıstığı kabuklarının sınıflandırılması. *Tekirdağ Ziraat Fakültesi Dergisi*, 21(4): 854-864.
Citation: Idress. Kh .A. D., Öztekin, Y. B., Gadalla,. O. A. (2024). Classification of Open and Closed Pistachio Shells Using Machine Vision Approach. Journal of Tekirdag Agricultural Faculty, 21(4):854-864.
*This study is summarized from Khaled Adil Dawood Idress MSc. thesis.

Öz

Antep fıstığı, besin değeri yüksek ve hoş tadı nedeniyle dünya çapında yaygın olarak tüketilen bir fıstık türüdür. Genellikle Antep fıstığı açık veya kapalı kabuklu olarak satılmaktadır. Ancak, kapalı kabuklu Antep fıstığı tüketiciler tarafından tercih edilmemekte ve bu da fıstığın ticari değerinin düşmesine neden olmaktadır. Üretim süreçleri ve işleme sırasında kalite kontrolünü sağlamak için açık ve kapalı uçlu Antep fıstığı kabuklarını ayırt edebilmek esastır. Bu manuel olarak veya mekanik cihazlar kullanılarak yapılabilir. Manuel sınıflandırma işlemi işçiler tarafından yapılmakta olup bu şekilde yapılan ayırma işlemi zaman ve maliyet açısından verimsiz sayılmaktadır. Mekanik sınıflandırma işleminde ise fıstığının mekanik olarak ayrılması, ayıklama işleminde kullanılan iğne mekanizması nedeniyle açık kabuklu somunların çekirdeklerine zarar verebilmektedir. Bu çalışma, Antep fıstığı tanelerinin makina görme tabanlı bir sistem kullanılarak sınıflandırılmasını ve sınıflandırma doğruluğu açısından uygulanabilirliğinin değerlendirilmesini amaçlamaktadır. Bu çalışmada kullanılan sistem, Siirt ve Urfa fıstıklarından şekil, boyut ve tat özellikleri bakımından farklı olan Antep fıstığı çeşidi için yapılan değerlendirmeleri içermektedir. Bu çalışmadaki makina görme sistemi, Antep fıstığı tanelerini tamamen otomatik bir şekilde kapalı ve açık uçlu kabuk sınıflarına ayırabilmektedir. Bu çalışmada 500 açık ve 500 kapalı uçlu tane olmak üzere 1.000 adet Antep fıstığı tane görüntüsü elde edilmiş ve incelenmiştir. Görüntüler işlenerek özellik çıkarma için hazırlanmıştır. Görüntülerden toplam 14 renk özelliği çıkarılmıştır. Tek özellik kullanılmasına rağmen, Rastgele Orman, Destek Vektör Makinesi ve Lojistik Regresyon modellerinden sırasıyla % 95.6,% 94.8 ve% 93.6'lık umut verici sınıflandırma doğruluk oranları elde edilmiştir. Sınıflandırıcıların performansları birbirleriyle karşılaştırılmıştır. Sınıflandırıcılar arasında yaklaşık benzer performanslar elde edilmiştir. Bu sonuçlar, Rastgele Orman sınıflandırıcısının, açık ve kapalı kabuklu uçlu olarak Antep fıstığı tanelerini sınıflandırmak için en etkili algoritma olduğunu göstermektedir.

Anahtar Kelimeler: Fıstık, Görüntü işleme, Renk özelliği, Lojistik regresyon, Rastgele orman, Destek vektör makinesi

1. Introduction

Pistachio (*Pistacia vera* L.) is one of the most popular tree nuts in the world, second only to the hazelnut in terms of fat. It is also rich in minerals and vitamins as it contains phosphorus, potassium, calcium, magnesium, and iron and contains vitamins such as E, C, B1, and B2 (Aktaş, Kızıldeniz, & Ünal, 2022). Pistachio nuts (*Pistacia vera* L.) are economically important for Turkey, Iran, China, the United States, and Southern European countries. Pistachio nuts are widely known as green gold due to their health-promoting nutrients and economic importance (Hosseinpour-Zarnaq, Omid, Taheri-Garavand, Nasiri, & Mahmoudi, 2022). Turkey is one of the main pistachio nut-producing countries in the world. In Turkey, there are eight main domestic pistachio vera varieties such as Uzun, Kırrmızı, Halebi, Siirt, Beyazben, Sultani, Değirmi, and Keten Gömleği. Also, there are five foreign varieties named Ohadi, Bilgen, Vahidi, Sefidi, and Mümtaz (Ak & Acar, 2001). Generally, pistachio nuts are grown in 56 provinces of Turkey. With its fluctuating production structure, Turkey ranks fourth behind Iran, the USA, and China, and with its annually increasing exports, Turkey ranks fourth behind the USA, Iran, and Germany (FAOSTAT, 2022). Iran's non-oil exports heavily rely on pistachio production, which has played a crucial role in the country's economy. Until 2010, Iran had produced over half of the world's pistachios and had supplied over 70% of global pistachio exports (Yaghoubi & Niknami, 2022). Harvested pistachio nuts contain many empty, undeveloped, closed shells due to factors such as unsuitable climate, incomplete pollination, lack of nutrition, and disease. However, pistachio nuts with closed shells have low consumer acceptance, leading to lower added commercial value. In the production and processing of pistachio nuts, it is very important to be able to classify opened and closed pistachio nuts accurately. This can be done manually and with mechanical devices. Distinguishing open and closed pistachio nuts can be achieved through both manual and mechanical methods. However, manual sorting is time-consuming and can be influenced by various factors, such as the operator's age, fatigue, visual acuity, and room conditions) (Omid, 2011). Mechanical devices to separate closed-shell pistachios from open-shell nuts can cause damage to the nuts of open-shell due to the needle mechanism used in the sorting process. In order to address these challenges, researchers have been working to develop new, non-destructive sorting methods in recent decades, including acoustic signal analysis and machine vision, to sort nuts quickly and accurately without causing any damage. Mechanical sorting is one of the popular techniques for sorting closed and open pistachio shells. This sorting method is done using a device called a "Pinpicker." Nevertheless, this technique damages open pistachio nuts, leading to a reduction in quality (C. Pearson, A. Doster, & J. Michailides, 2001). Pearson created a system for sorting open and closed-shell pistachio nuts in real time using acoustic signals. The system recorded audio signals, analyzed three static features in the time domain, and used linear discriminant analysis to differentiate the nuts. The acoustic sorter was found to have higher accuracy in classifying the nuts compared to a commercially available mechanical sorting system (Hosseinpour-Zarnaq et al., 2022). Decision tree and fuzzy logic classifiers were used to sort two Iranian pistachio nuts, Akbari (Ak) and Kaleghouchi (Ka). 240 Ak and 120 Ka pistachios fell down a chute onto a stainless steel plate, and the acoustic signal of these hits was recorded with a microphone and stored on a computer via a sound card. In this study, statistical parameters were extracted from these time-domain signals using the J48 algorithm. Feature selection and classification were performed using 200 Ak and 100 Ka as the training set and 40 Ak and 20 Ka as the test set. The results of the study showed that the proposed system achieved a classification accuracy of 93.3% for the training set and 96.67% for the test set (Jalali & Mahmoudi, 2013).

In recent years, there have been several studies on the classification of open and closed pistachio nuts using machine learning algorithms. Improved k-NN Classifier used in a study carried out by (Ozkan, Koklu, & Saraçoğlu, 2021) to classify open and closed pistachio nuts. In this study, shape and morphological features were extracted from images. These features were then used as inputs to train and test the machine learning algorithm. The results of the study showed that the proposed technique achieved a classification accuracy of 94.18%. Ghezelbash et al. (2013) developed a system that adopts a combination of two flat mirrors and one low-cost camera to obtain appropriate 3-dimensional images from pistachios which are processed to detect closed-shell nuts. The experimental results for the three varieties of pistachio nuts showed an average removal accuracy of 92.7 and 86.7% for open – and closed pistachio nuts, respectively. Omid (2011) proposed an expert system based on an acoustic emission signal and a fuzzy logic classifier for sorting open and closed pistachios nuts, and the overall accuracy of the sorting system for testing data sets was 95.56% (Hosseinpour-Zarnaq et al., 2022). Rahimzadeh and Attar (2022) proposed a computer vision system aimed to distinguish between open and closed pistachios of various varieties. Their approach involved employing Convolutional Neural Network (CNN)-based models, such as ResNet50, ResNet152, and VGG16, to extract relevant features from pistachio images and execute classification tasks. The resulting average classification accuracies attained by these models were 85.28%, 85.19%, and 83.32%, respectively (Rahimzadeh & Attar, 2022). The audio signal processing technique is one of the most popular and non-destructive methods used to separate closed pistachios from open pistachios. However, this method also has disadvantages. Represented the negative effect of ambient noise, and the moisture content of nuts affects the impact of sound (Farhadi, Abbaspour-Gilandeh, Mahmoudi, & Mari Maja, 2020). J48 Decision Tree, Naïve Bayes, and Multi-Layer Perceptron (MLP) were used in a study conducted by (Ataş & Doğan, 2015) to classify open and closed pistachio nuts. In this study, the J48 decision tree was utilized as a main classifier. The classification performance of J48 was also compared to other classifiers, including Naïve Bayes and Multi-Layer Perceptron (MLP). The results of the study showed that the proposed system using the J48 decision tree achieved a simple and interpretable classifier along with a satisfactory classification accuracy performance of 94.5%. Traditional machine learning algorithms were utilized to assess the effectiveness of color features in categorizing cashew kernels as either white or scorched. The outcomes indicated that the color features were effective in distinguishing between the two categories. Furthermore, all traditional machine learning methods tested yielded promising results in this classification task (BAİTU, GADALLA, & ÖZTEKİN). Hosseinpour-Zarnaq et al. (2022), developed a one-dimensional convolutional neural network (CNN) model for sorting (open-closed) pistachio nuts using acoustic emission signals. A total of 1600 pistachios from two pistachio varieties (Akbari and Ahmad Aghaei) were used. The overall accuracy of the CNN classifier was 98.75% (Hosseinpour-Zarnaq et al., 2022). In another study conducted by Singh et al. (2022), a pre-trained dataset consisting of a total of 2148 images was employed. Among these images, 1232 were of the Kirmizi type, and 916 were of the Siirt type. The study employed three distinct convolutional neural network models to classify these images. The models were trained using the transfer learning method, utilizing AlexNet as well as pre-trained VGG16 and VGG19 models. The classification results from this study indicated the following success rates for the respective models: AlexNet achieved a success rate of 94.42%, VGG16 achieved a notably high 98.84% success rate, and VGG19 achieved a slightly lower success rate of 98.14% (Singh et al., 2022). Lisda et al (2023) conducted a study on classifying two pistachio nuts varieties using the Inception-V3 and ResNet50 models. The study utilized a dataset comprising 2148 photos, with 916 images of the Siirt type and 1232 images of the Kirmizi type. The classification results revealed that the Inception-V3 model achieved an impressive accuracy of 96%, while the ResNet50 model achieved an accuracy of 86% (Lisda, Kusrini, & Ariatmanto, 2023).

This study aims to evaluate the performance of various machine learning algorithms, including Logistic Regression, Support Vector Machines, and Random Forest for classifying open and closed pistachio nuts. The algorithms were written using the Python programming language and run in Google Colab environment.

2. Materials and Methods

2.1. Image acquisition

This study was carried out at the Faculty of Agriculture's laboratory in the Department of Agricultural Machinery and Technologies Engineering at Ondokuz Mayis University in Samsun. One thousand samples of open and closed pistachio nuts were obtained from the Ministry of Agriculture and Forest Pistachio Research Institute in Gaziantep. The nuts were examined and stored at room temperature. The imaging system was equipped with a high-resolution Guppy camera model Pro F-032, which had a CCD progressive Sony ICX424 sensor, 4.9 mm x 3.7 mm, consisting of 656 horizontal by 492 vertical cells. The images were captured at a resolution of 80x80 pixels. In this experiment, a black surface was used as a background. Because good lighting is crucial for getting high-quality images, two 8-watt fluorescent lamps were used as the light source in the image acquisition chamber. The images of both open and closed pistachio nuts were taken under the same conditions (camera, position, and background). The image capture system used in this study (*Figure 1*) consisted of a camera connected to a computer equipped with an image capture card.

Figure 1. Image capturing system

2.2. Image pre-processing and segmentation

After the images were captured, the segmentation process was done to extract the object of interest from the background. In this study, the Canny edge detection approach was applied to segment the images. The canny edge detection algorithm was developed by John F. Canny (1986) and is widely used in image processing due to its good balance of speed and accuracy. It is a multi-stage algorithm that involves detecting edges through the use of image gradients, suppressing non-maximums, hysteresis thresholding, and edge tracking. Various image segmentation methods were tested to find out the one that effectively separated the foreground from the background. After testing different image segmentation methods, Canny edge detection was found to give the best performance in segmenting the foreground from the background. An example of raw and segmented images is shown in (*Figure 2*).

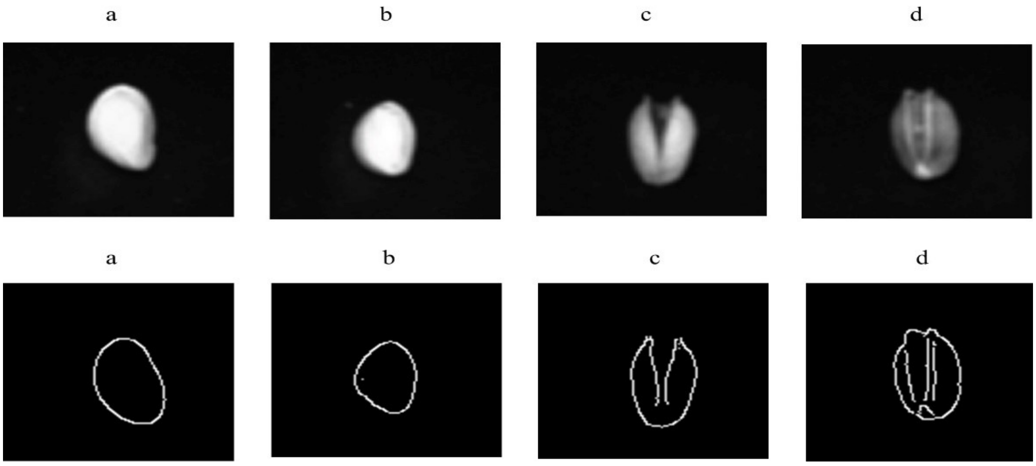

Figure 2. Raw image segmentation

2.3. Feature extraction and selection

Extracting appropriate features from an image is crucial for an accurate classification process. In order to enhance the accuracy of pistachio classification, this study focused on extracting color features from both RGB and CIELAB color spaces. Therefore, Pistachio images in RGB and CIELAB (L*a*b*) color spaces were split into R (red), G (green), B (blue), L (l), A (a), and B (b). Moments features representing mean, variance, and range were captured from the split color. The motivation behind using color features from both RGB and CIELAB color spaces lies in their ability to capture different aspects of color information, providing a more comprehensive representation of the data. In this study, 18 color features were extracted, and after evaluating the effectiveness of all features in predicting the category of the pistachio nuts, only 14 features were found to have a significant effect, and the other four irrelevant features were left out. Utilizing irrelevant features could detrimentally impact algorithm performance. Therefore, to enhance the efficiency of a classification model and decrease the running time, it is highly recommended to use only the important features and get rid of irrelevant features. This procedure is commonly referred to as feature selection. In this particular study, feature selection was executed in Python through the Boruta library, which employs a wrapper approach to identify relevant features by constructing an XGBoost classifier. The list of accepted features and those rejected are shown in (*Table 1*).

Table 1. List of accepted and rejected features.

No.	Feature	Status	No.	Feature	Status
1.	Red Mean	Accepted	10.	L _ Mean	Accepted
2.	Green Mean	Accepted	11.	A_ Mean	Accepted
3.	Blue Mean	Accepted	12.	B _ Mean	Accepted
4.	Red Variance	Accepted	13.	L_ Variance	Accepted
5.	Green Variance	Accepted	14.	A_ Variance	Rejected
6.	Blue Variance	Accepted	15.	B_ Variance	Rejected
7.	Red Range	Accepted	16.	L _ Range	Accepted
8.	Green Range	Accepted	17.	A _ Range	Rejected
9.	Blue Range	Accepted	18.	B _ Range	Rejected

2.4. Performance evaluation

The performance of the classification can be assessed by counting the number of cases that were accurately identified as being part of the class (true positives), the number of cases correctly recognized as not being part of the class (true negatives), the number of cases mistakenly categorized as belonging to the class (false positives), and the number of cases that were not properly identified as part of the class (false negatives) (Cinar & Koklu, 2022).

Calculation formulas for success criteria, such as accuracy, error rate, recall, specificity, and precision, were calculated using the confusion matrix for binary classification performance measurements, and their equations are given in *(Table 2)* (Hossin & Sulaiman, 2015).

Table 2. Performance measurements and calculation equations for binary classification.

No.	Performance Metrics	Explanation	Equation
1.	Accuracy	It measures the ratio of true prediction on all samples included in the assessment.	$\frac{tp+tn}{tp+fp+tn+fn} \times 100$ (Eq.1).
2.	Error Rate	It measures the ratio of false prediction on all samples included in the assessment.	$\frac{fp+fn}{tp+fp+tn+fn} \times 100$ (Eq.2).
3.	Recall	It is used to measure the ratio of correctly classified positive values	$\frac{tp}{tp+fn} \times 100$ (Eq.3).
4.	Specificity	It measures the ratio of correctly classified negative values.	$\frac{tn}{tn+fp} \times 100$ (Eq.4).
5.	Precision	It measures the ratio of accurately classified positive samples to estimated total positive samples.	$\frac{tp}{tp+fp} \times 100$ (Eq.5).

3. Results and Discussion

In this study, a total of 1,000 Antep Pistachio nuts images, including 500 open and 500 closed nuts, were used as a dataset. 75% of this dataset was used for training the model, and the rest 25% was used for testing the model. In this experiment, three traditional ML algorithms, Logistic Regression, Support Vector Machine (SVM), and Random Forest models, were tested. The correct compilation and training of the algorithms were done. The algorithm's performance was tested using a set of test data to evaluate the classification accuracy. Classifiers' performances were compared to each other using the confusion matrix. It is a table of the predicted classes and the actual classes that were observed. The confusion matrix allows seeing where the model is making correct predictions, where it is making incorrect predictions, and how accurate the model is overall. It also allows for identifying which classes are being confused with each other. This information can be used to improve the model by adjusting parameters or changing features. Confusion matrices are very important because they provide an easy way to visualize and understand how well a classification model is performing. The used classifiers and the confusion matrix for each model are presented below.

3.1. Logistic Regression classifier

Logistic regression is a powerful classification technique that can be used to classify opened and closed pistachio nuts. It is a supervised learning algorithm that uses a linear model to predict the probability of an outcome. In this study, the outcome was whether a pistachio nut is open or closed**.** The logistic regression classifier was able to achieve an accuracy of 93.6% in the test dataset. This indicates that the model accurately classified open and closed pistachio nuts with high accuracy. The achieved accuracy of 93.6% in the test dataset is quite impressive. It indicates that the model was able to identify the different classes of nuts accurately. The confusion matrix for the logistic regression classifier is shown in *(Figure3)*.

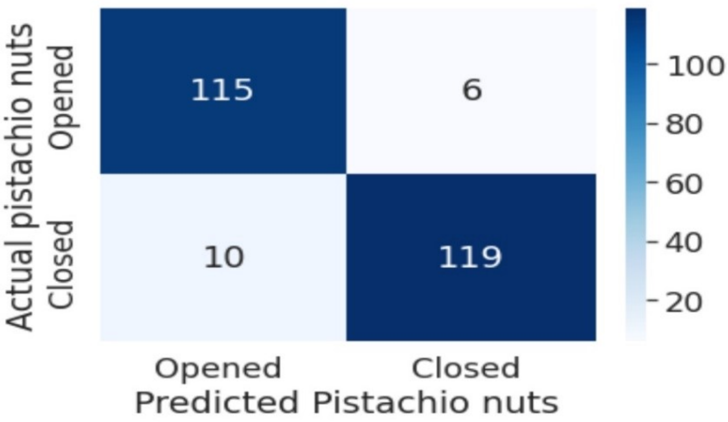

Figure 3. Logistic Regression confusion matrix

3.2. Support Vector Machine classifier

In this experiment, an SVM classifier was used to classify opened and closed pistachio nuts. The dataset was split into training and test datasets, with the training dataset used to train the model and the test dataset used to evaluate its performance. The model achieved an accuracy of 94.8% on the test dataset, indicating that it was able to accurately classify opened and closed pistachio nuts with high accuracy. The confusion matrix of the Support Vector Machine classifier is shown in *(Figure 4)*.

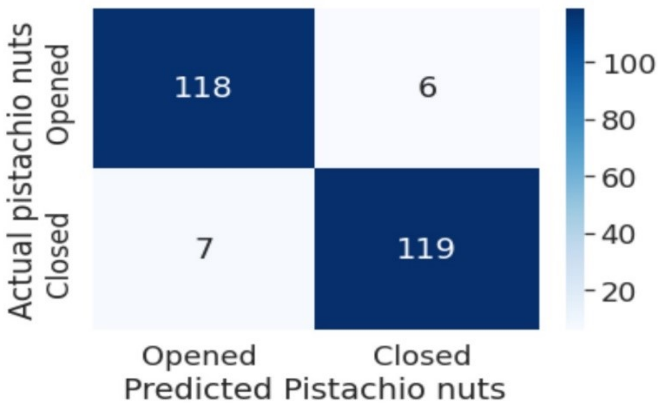

Figure 4. Support Vector Machine confusion matrix

3.3. Random Forest

Using a random forest classifier to classify open and closed pistachio nuts is an effective approach for this task. In this study, a random forest made of 100 trees was used. The dataset was split into training and test datasets. The training dataset was used to train the model, and the test dataset was used to test its performance. The achieved accuracy of 95.6% in the test dataset is an excellent result and indicates that the model was performing well. The confusion matrix of the random forest is shown in *(Figure5).*

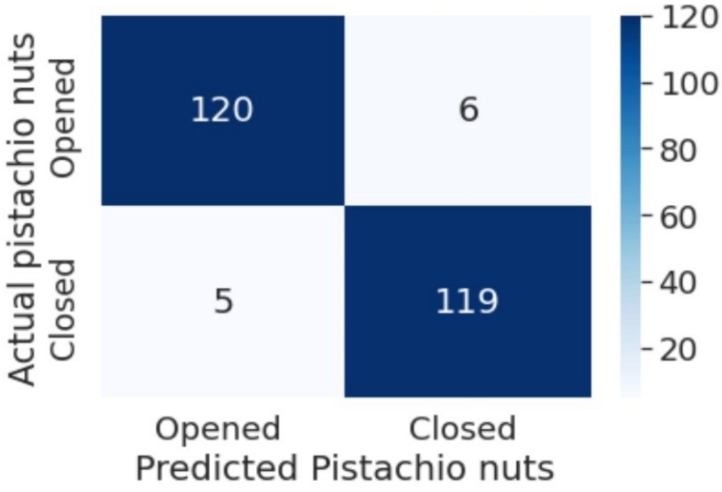

Figure 5. Random Forest confusion matrix

3.4. Performance measurements

Performance measurement in a confusion matrix is important because it provides a comprehensive overview of the performance of a model. It allows us to evaluate the accuracy of the model, as well as identify areas where it may be underperforming. It also helps us to identify potential sources of error, such as misclassification or false positives and negatives. By understanding these metrics, we can adjust our models to improve their performance and ensure that they are providing accurate results. Performance measurement values of all algorithms used in classification are given in *(Table3).*

Table 3. Performance measurements values of all algorithms used in classification

Performance Metrics	Logistic Regression	Support Vector Machine	Random Forest
Accuracy (%)	93.6	94.8	95.6
Error (%)	6.4	5.2	4.4
Specificity (%)	92.2	95.2	95.2
Precision (%)	92	94.4	96
Recall (%)	95	94.4	95.2

From the performance measurement values given in *(Table 3),* it can be seen that the classification accuracy for all algorithms is above 90%. It seems that the best classification accuracy belongs to the random forest algorithm, with 95.6%. The lowest classification accuracy belongs to the logistic regression algorithm, with 93.6%. Classifiers' performances were compared to each other. Almost similar performances were detected.

many research efforts have been carried out to enhance the accuracy of classifying open and closed pistachio nuts and different types of pistachio varieties through the application of machine learning and deep learning algorithms. Presented explicitly in *(Table 4)* are studies focused on the classification of two aspects: open-closed pistachio nuts and the differentiation of distinct pistachio varieties. *(Table 4)* also gives the number of images in the data set, the varieties that were classified, the models used, and their accuracy rates.

Table 4. Comparison of accuracy rates of classifying open-closed pistachio nuts and different types of pistachio varieties using traditional machine learning and deep learning models

Author/ authors	Model architecture	Images numbers	Varieties	Accuracy
Jalali and Mahmoudi, (2013)	J48 Decision tree, Fuzzy logic	360	Akbari and Kaleghouchi	96.67%
Hosseinpour-Zarnaq et al., (2022)	CNN model	1600	Akbari and Ahmad Aghaei (open or closed)	98.75%
Singh, et al. (2022)	AlexNet, VGG16 and VGG19	2.148	Kırmızıı and Siirt	94.42%, 98.84%, and 98.14%.
Lisda et al (2023)	Inception-V3 and ResNet50	2.148	Kırmızı and Siirt	96% and 86%
Rahimzadeh and Attar (2022)	ResNet50, ResNet152, and VGG16	3.927	different pistachio types (open-mouth or closed-mouth)	85.28%, 85.19%, and 83.32%
Ozkan et al. (2021)	Improved k-NN	2.148	Kırmızı and Siirt (open or closed)	94.18 %
Ataş and Doğan (2015)	J48 Decision Tree, Naïve Bayes, and Multi-Layer Perceptron (MLP)	200	different pistachio types (open – closed)	95.5%, 94.5% and 94.5%
Our proposed machine learning models	Logistic Regression, Support Vector Machine, and Random Forest	1.000	Antep Pistachio	93.6%, 94.8%, and 95.6%

From the previous studies, it has been observed that the highest classification accuracy for distinguishing different varieties of pistachios using deep learning algorithms was attained by Singh et al. (2022), who accomplished an impressive accuracy rate of 98.84%. On the other hand, the highest accuracy of classification of open-closed pistachio nuts using deep learning model was achieved by Hosseinpour-Zarnaq et al. (2022), which achieved an impressive accuracy of 98.75 using CNN model. Also, we have noticed that the highest accuracy of classification of open-closed pistachio nuts using a traditional machine learning model was achieved in a study conducted by Jalali and Mahmoudi (2013), which they were able to achieve a classification accuracy of 96.67%% using J48 Decision Tree model, which is close to the result that we have achieved with the Random Forest model (95.6%). These results indicate that the image segmentation method employed in our research and the extraction of color features from the images effectively captured the distinctive patterns, resulting in improving the performance of the classification.

4. Conclusion

In this study, the classification of closed and open-shell pistachio nuts was conducted using machine-learning algorithms based on color features. A total of 1,000 Antep pistachio nut images, including 500 open and 500 closed nuts, were obtained and examined. The images were pre-processed and prepared for feature extraction. From the images, a total of 14 color features were extracted. Three models were applied and studied; logistic regression, SVM, and random forest. The single-color feature was used. For classification, models were created, and the classification process was performed. Accuracy rates of 95.6%, 94.8%, and 93.6% from the random forest, SVM, and logistic regression, were achieved, respectively. Also, the classifiers` performances were compared to each other almost similar performances were detected. The result of this study showed that the random forest model has the highest accuracy among the studied models. Considering false negative issues, the random forest model was found more applicable than other models and suggested to be used due to its higher accuracy.

This study proved that only color features can be used for the classification of closed and open-shell pistachio nuts with high accuracy. This research can be useful for the pistachio industry to improve the quality of products, reduce labor costs, and speed up the classification process. Future studies should focus on the use of additional features such as texture, shape, size, and morphological p to improve the accuracy of classification.

Acknowledgment

This work was supported by Ondokuz Mayıs University (Project No: PYO.ZRT. 1904.23.003), Turkey.

Ethical Statement

There is no need to obtain permission from the ethics committee for this study.

Conflicts of Interest

We declare that there is no conflict of interest between us as the article authors.

Authorship Contribution Statement

Khaled Adil Dawood Idress; Writing, Designing, Data Collection, and Processing; Omsalma Alsadig Adam Gadalla; Literature Search:, Review and Editing: Yeşim Benal Öztekin.

References

Ak, B. and Acar, I. (2001). Pistachio Production and Cultivated Varieties Grown in Turkey. Project on Underutilized Mediterranean Species. Pistacia: Towards a Comprehensive Documentation of Distribution and Use of Its Genetic Diversity in Central & West Asia, North Africa and Mediterranean Europe. *Report Of The Ipgri Workshop*, 14-17 December 1998, Irbid, Jordan.

Aktaş, H., Kızıldeniz, T. and Ünal, Z. (2022). Classification of pistachios with deep learning and assessing the effect of various datasets on accuracy. *Journal of Food Measurement and Characterization,* 16(3): 1983-1996. https://doi.org/10.1007/s11694-022-01313-5

Ataş, M. and Doğan, Y. (2015). Classification of Closed and Open Shell Pistachio Nuts by Machine Vision. *International Conference on advanced Technology Sciences*, Antalya, Türkiye.

Baitu, G. P., Gadalla, O. A. A. and Öztekin, Y. B. (2023). Traditional machine learning-based classification of cashew kernels using colour features. *Journal of Tekirdağ Agricultraul Faculty,* 20(1): 115-124.

Çınar, I. and Koklu, M. (2022). Identification of rice varieties using machine learning algorithms. *Journal of Agricultural Sciences*, 28(2): 307-325. https://doi.org/10.15832/ankutbd.862482

Faostat (2022). Food and Agriculture Data. https://www.fao.org/faostat/en/

Farhadi, M., Abbaspour-Gilandeh, Y., Mahmoudi, A. And Mari Maja, J. (2020). An integrated system of artificial intelligence and signal processing techniques for the sorting and grading of nuts. *Applied Sciences,* 10(9): 3315.

Ghezelbash, J., Borghaee, A. M., Minaei, S., Fazli, S. and Moradi, M. (2013). Design and implementation of a low cost computer vision system for sorting of closed-shell pistachio nuts. *African Journal of Agricultural Research*, 8(49): 6479–6484. https://doi.org/10.5897/ajar10.1162

Hosseinpour-Zarnaq, M., Omid, M., Taheri-Garavand, A., Nasiri, A. and Mahmoudi, A. (2022). Acoustic signal-based deep learning approach for smart sorting of pistachio nuts. *Postharvest Biology and Technology,* 185: 111778. https://doi.org/10.1016/j.postharvbio.2021.111778

Hossin, M. and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process,* 5(2): 01-11. https://doi.org/10.5121/ijdkp.2015.5201

Jalali, A. and Mahmoudi, A. (2013). Pistachio nut varieties sorting by data mining and fuzzy logic classifier. *International Journal of Agriculture and Crop Sciences*, 5(2): 101–108.

Lisda, L., Kusrini, K. and Ariatmanto, D. (2023). Classification of pistachio nut using convolutional neural network. *Inform: Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi,* 8(1): 71-77.

Omid, M. (2011). Design of an expert system for sorting pistachio nuts through decision tree and fuzzy logic classifier. *Expert Systems with Applications, 38*(4), 4339-4347. https://doi.org/10.1016/j.eswa.2010.09.103

Ozkan, I. A., Koklu, M. and Saraçoğlu, R. (2021). Classification of pistachio species using improved k-nn classifier. *Progress in Nutrition,* 23(2): e2021044. https://doi.org/10.23751/pn.v23i2.9686

Pearson, T. C., Doster, M. A. and Michailides, T. J. (2001). Automated detection of pistachio defects by machine vision. *Applied Engineering in Agriculture,* 17(5), 729. https://doi.org/10.13031/2013.6905

Rahimzadeh, M. and Attar, A. (2022). Detecting and counting pistachios based on deep learning. *Iran Journal of Computer Science,* 5(1): 69-81.

Singh, D., Taspinar, Y. S., Kursun, R., Cinar, I., Koklu, M., Ozkan, I. A. and Lee, H. N. (2022). Classification and analysis of pistachio species with pre-trained deep learning models. *Electronics*, 11(7): 1–14. https://doi.org/10.3390/electronics11070981.

Yaghoubi, M. and Niknami, M. (2022). Challenges of precision agriculture application in pistachio orchards: factor analysis from Iranian Agricultural Experts' Perspective. *Journal of Tekirdağ Agricultrual Faculty,* 19(3): 473-482.