Food Analytical Methods (2022) 15:739–750
https://doi.org/10.1007/s12161-021-02154-6

Prediction of Pistachio (*Pistacia vera* L.) Mass Based on Shape and Size Attributes by Using Machine Learning Algorithms

Cevdet Saglam[1] · **Necati Cetin**[1]

Received: 3 August 2021 / Accepted: 26 October 2021 / Published online: 4 November 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Size, mass, and shape attributes play a significant role in the quality assessment and post-harvest technologies of agricultural products. Pistachio is widely consumed worldwide, and Turkey has 3rd place in world pistachio production. In this study, physical attributes of 6 different pistachio cultivars (Beyaz Ben, Keten gömleği, Kirmizi, Siirt, Tekin, Uzun) were determined and machine learning algorithms (Multilayer Perceptron (MLP), k-Nearest Neighbor (kNN), Random Forest (RF), Gaussian processes (GP)) were used for mass prediction of these pistachio cultivars. Siirt and Tekin cultivars had the greatest gravitational and dimensional attributes. Among the pistachio cultivars, Kirmizi and Uzun had the greatest shape index and elongation values. Keten gömleği and Beyaz cultivars had the lowest averages of mass and area attributes both for nuts and kernels. Kernel and nut mass of pistachio had significant correlations with volume, geometric mean diameter, and projected and surface area ($p < 0.01$). Present findings revealed that Gaussian processes had the greatest correlation coefficients (0.976 for nut mass and 0.948 for kernel mass prediction) and the lowest RMSE values (0.038 for nut and 0.029 for kernel mass prediction). This algorithm was respectively followed by Multilayer Perceptron and Random Forest algorithms. Present findings revealed that Gaussian processes, Multilayer Perceptron, and Random Forest algorithms could potentially be used for mass prediction of pistachio cultivars.

Keywords Pistachio · Mass · Multilayer Perceptron · Gaussian processes · Random Forest

Introduction

Pistachio (*Pistacia vera* L.) is a member of the Anacardiaceae family, and it is among the most widely consumed tree nuts worldwide. Pistachio (*Pistacia vera* L.) is grown in dry regions of central and west Asia and Mediterranean basin (Tomaino et al. 2010). Iran (574,987 tons) is the leading producer in the world, and it was respectively followed by the USA (272,291 tons), China (95,294 tons), and Turkey (78,000 tons) in 2017 (FAO 2019).

Appearance, shape, and size significantly influence consumer preference of agricultural products. Identification of such attributes will also aid in classification, drying, storage, packaging, and transportation processes. Shell thickness, and nut and kernel size and shape are significant attributes for classification and quality assessment of nut fruits (Warmund 2008; Mahawar et al. 2017; Çetin et al. 2020).

The most important procedures conducted when pistachios arrive at the processing plant are as follows: (a) dehulling, to separate the soft hull from nuts; (b) trash and blank separation, to remove blank pistachios and trashes such as small branches, remaining shells, and leaves; (c) unpeeled pistachio separation, to remove unpeeled and unripe nuts; (d) washing, which involves spraying water at high pressure on the pistachios to clean the nuts; (e) drying, to decrease moisture content of pistachios from 37 to 40% to appropriate level; (f) split nut separation, to separate split nuts from unsplit ones; (h) roasting; and (i) packaging (Kashaninejad et al. 2003).

Consumers generally prefer fruits with similar size and shape (Rashidi and Keshavarzpour 2012). Therefore, agricultural commodities are usually classified in size, and then processes and packed accordingly (Ashtiani et al. 2014). Mechanical or electronic mechanisms are used in such sizing or classification operations. Electronic sizing mechanisms are quite costly operations. Mechanical

✉ Necati Cetin
necaticetin@erciyes.edu.tr

[1] Department of Biosystems Engineering, Faculty of Agriculture, Erciyes University, 38039 Kayseri, Turkey

sizing mechanisms are highly time-consuming operations. Therefore, previous studies indicated that some easily determined geometrical parameters could be used to predict fruit mass (Tabatabaeefar and Rajabipour 2005; Khoshnam et al. 2007; Hassan-Beygi et al. 2010; Demir et al. 2020).

Physical dimensions, volume, mass, and projected and surface areas are generally used in fruit sizing operations (Ashtiani et al. 2014). Fruit physical attributes and relationships among these attributes should be investigated for design and optimization of post-harvest operations like transportation, sorting, packaging, and classification (Ercişli et al. 2012; Ashtiani et al. 2014). Such relations are used in design of mass-based sorting and in food processing operations. Fruit's main physical attributes and projected area could be used to predict fruit mass (Stroshine and Hamann 1994; Naderi-Boldaji et al. 2008; Ghabel et al. 2010).

Mass and volume predictions from easily determined physical attributes may facilitate design and operation of post-harvest sorting machines and processes (Gonzalez et al. 2017). Several researchers indicated that machine learning algorithms yielded better outcomes for prediction of fruit characteristics than the conventional methods (Moosavi and Sepaskha 2012; Demir et al. 2017; Kus et al. 2017; Shabani et al. 2017; Çetin et al. 2021).

There are some problems experienced in pistachio cultivation and processing and product catalogs about quality standards and product characteristics have not been created yet (Abidi 2016). Recently, some machine learning technologies have been developed and used to overcome these problems. These techniques offer solution-oriented approaches through rapid and simple simulations (Soares et al. 2013). Such powerful approaches are needed for accurate depiction and identification of descriptive characteristics used in quality assessment of foodstuffs and agricultural products (Omid et al. 2010). Machine learning offers non-linear models able to predict past and future values within an input-output linkage (Zhang et al. 2012).

Different studies were conducted to determine the shape, mass, and dimensional attributes of pistachios (Kashaninejad et al. 2006; Razavi et al. 2007; Abidi 2016; Mohammadi Moghaddam and Razavi 2019). There are a limited number of studies about the prediction of product mass from physical attributes (Soares et al. 2013; Rad et al. 2015, 2017; Gurbuz et al. 2018; Demir et al. 2020). However, there are not any studies on mass prediction of pistachio with machine learning algorithms. In this study, the size, mass, and shape attributes of 6 different pistachio cultivars were determined for the purpose of classification and discrimination. Besides, four different machine learning algorithms (MLP, kNN, RF, and GP) were used for mass prediction from physical attributes, and performances of these algorithms were compared.

Material and Methods

Sample Preparation

In this study, 6 different pistachio cultivars (Keten Gömleği, Beyaz Ben, Kirmizi, Tekin, Uzun, Siirt) were used as the material. Pistachio samples were collected from the orchards of Gaziantep province of Turkey in the 2019 harvest season. Foreign materials, and damaged and unripe nuts with irregular shapes were separated manually. Pistachios were kept in a fridge at 5 ± 0.5 °C for 2 weeks until a constant moisture level. A sufficient quantity of samples was taken for analyses and placed out for warm up for 2 h. To obtain the mass, each fruit was weighed on a precision electronic balance with ± 0.001 accuracy.

Shape and Dimensions

Three principal dimensions: length (L, mm) width (W, mm), and thickness (T, mm) of each cultivar were measured with a digital caliper with an accuracy of ± 0.01 mm on 100 fruits. The equations used for calculation of the volume (mm^3), geometric mean diameter (D_g, mm), projected area (PA, mm^2), surface area (S, mm^2), sphericity (φ, %) shape index (SI), roundness (R), aspect ratio (AR), and elongation (E) are given in Table 1.

Statistical Analysis and Cultivar Classification

Correlations between the physical attributes of kernels and nuts were determined using Pearson correlations. One-factor variance analysis was applied to the assessed parameters and the differences among the means were tested with Tukey's multiple comparison test at 95% significance level. Differences in pistachio cultivars were assessed using linear discriminant analysis. A scatter plot with canonical discrimination functions presented group centroids of pistachio cultivars provided from discriminant analysis (Fig. 1). Statistical analyses of the samples were performed by using SPSS v20.0 software (IBM SPSS® 2010) and PAST® v3.20 software (Hammer et al. 2001).

Machine Learning Algorithms and Validation Methodology

A prediction of machine learning was applied in the present study by using the machine learning software Weka® v3.8 (Hall et al. 2009). Machine learning algorithm is implemented in a laptop having a core i5 central processing unit (CPU) with 2.5 GHz, and 8 GB memory. In this study, 100 kernel and 100 nut samples were measured for each cultivar.

Table 1 Equations used to calculate the size and shape attributes of the pistachio cultivars

Variables	Equations	Literature
Shape index (SI)	$SI = (2 \cdot L)/(W + T)$	Ozkan and Koyuncu (2005)
Volume (V, mm^3)	$V = (\pi/6) \cdot D_g^3$	Volume of ellipse
Surface area (S, mm^2)	$S = \pi D_g^2$	Sayinci et al. (2015)
Sphericity (φ)	$\varphi = (D_g/L) \cdot 100$	Mohsenin (1986)
Geometric mean diameter (D_g, mm)	$D_g = (L \cdot W \cdot T)^{(1/3)}$	Mohsenin (1986)
Roundness (R)	$R = A_p/A_c$	Sayinci et al. (2015)
Elongation (E)	$E = L/W$	Firatligil-Durmuş et al. (2010)
Aspect ratio (R_a)	$R_a = \frac{W}{L}$	Omobuwajo et al. (1999)
Projected area (A)	$A_p = \frac{\pi}{4} D_g{}^2$	Afonso Junior et al. (2007)

L, length (mm); W, width (mm); T, thickness (mm); D_g, geometric mean diameter (mm); A_p, projected area (mm^2); A_c, the biggest circular area (mm^2)

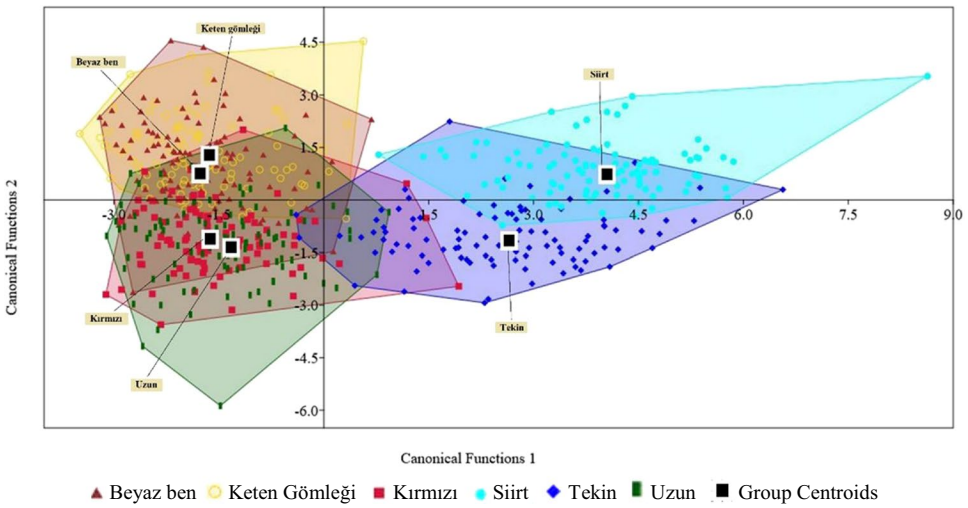


Fig. 1 Scatter plots of the pistachio cultivars in terms of the discriminant scores and group centroid functions 1 and 2

All measurements were carried out in 3 replications. In the mass prediction, 7800 values were used for both kernels and nuts.

Model performance was assessed through cross-validation method with the use of data not used in training and testing data sets. Cross-validation is largely used to assess classification and prediction performance of the models (Stegmayer et al. 2013). The *k*-fold cross-validation method was used in the present study to validate the prediction model. In this method, usually *k* values of 5 or 10 are used (Ataş et al. 2012) and 10 was selected in the present study since the present data set had 10 sub-sets. Training and testing procedures were carried out in 10 iterations. In each iteration, one sub-set was used for testing and the other sub-sets (9 sub-sets) were used for training. Each *k* sub-samples were used once for testing, respectively. Then, average of errors was taken for each iteration and expressed as error predict of the predictor. An evaluation matrix was generated for each algorithm. Performance of prediction algorithms was then compared by using specified evaluation matrices (Witten and Frank 2005; Stegmayer et al. 2013). For this purpose, 7020 data were used for calibration, and 780 data were used for validation.

Artificial Neural Networks (ANNs)

ANNs are machine learning models simulating the behaviors of the human brain. Models are composed of interconnected processing elements similar to biological neurons and weighted connections corresponding to brain snaps (Karray and Silva 2004). Multilayer Perceptron (MLP) is a commonly used ANN for classification purposes (Omid et al. 2010). MLPs are composed of input layers, hidden layers, and output layers. MLP is a feed-forward neural network (FFNN); in other words, data flow through input layers toward output layers in a single direction. In classification, MLP model is trained with the use of a backpropagation algorithm calculating the weights of activation function for each neuron (Karray and Silva 2004). Error minimization can be achieved by using a number of procedures such as gradient descent, gradient descent with a momentum, Levenberg–Marquardt, and conjugate gradient (Omid et al. 2010).

ANN is among the popular machine learning applications used in modeling, regression analysis, and classification of agricultural products (Marini et al. 2004; Mollazade et al. 2012). In the present study, learning rate: 0.3, momentum: 0.2, number of epochs: 500, 18–9-1 MLP structure model was taken into consideration. Applied MLP structure model is illustrated in Fig. 2.

The k-Nearest Neighbor (kNN)

The k-Nearest Neighbor (kNN) algorithm, so-called IBK, is a highly practical discriminant and classification algorithm for agricultural products. In this algorithm, at first, entire observation values were considered a cluster and these clusters are gradually compounded to form new clusters. Euclidean distance method is used to calculate distances (Nettleton et al. 2010; Romero et al. 2013). A low k may form a highly complex decision boundary; a higher k may end up with excessive generalization. Since a trained model was not generated, kNN classification requires more resources as the number of training samples is increasing. Therefore, selection of an accurate value is an important issue in classification (Maxwell et al. 2018).

Random Forest (RF)

In regression and classification, RF algorithm generates more than one decision tree (DT) with the use of bootstrap samples from the original training data to train each tree. Then, a set of trees is used for the classification of a sample based on the most frequent classification. Since each tree is a separate classifier in RF, this algorithm is used for group classification (Breiman 2001). Contrary to DT, a decision is made with the majority of an ensemble of trees constructed by RF in assigned class of data sets (Berhane et al. 2018). Since bootstrap and ensemble schemes could overcome overfitting problems inherited from DT, there is no pruning step in RF (Ali et al. 2012). Additionally, RF has high predictive accuracy and is robust against noise (Breiman 2001; Rodriguez-Galiano et al. 2012).

Gaussian Processes (GP)

GP is a probability distribution over the functions in Bayesian inference in which random variables are related to every single point within a range of values (Rasmussen and Williams 2006). In this technique, Bayesian Gaussian technique is used for non-linear regression. It is also required to specify a kernel function with a "noise" arrangement parameter or through check of closeness of fit. Moreover, the technique allows the selection of training data to be normalized or standardized before learning the regression (MacKay 1998). Gaussian processes are of great significance in statistical modeling since all parameters inherited from a normal distribution (normal distribution of error, variance, and covariance) could clearly be obtained (Rasmussen and Williams 2006).

Model Performance Evaluation

The performance of the MLR, kNN, RF, and GP models is assessed based on the following statistical indices: correlation coefficient (r), root mean square error (RMSE), mean absolute error (MAE), relative absolute error (RAE), and root relative squared error (RRSE):

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(M_i - \dot{M})\,(E_i - \dot{E})}{S_M S_E}$$

$$\mathrm{MAE} = \sum_{i=1}^{n}\frac{|\,E_i - M_i\,|}{n}$$

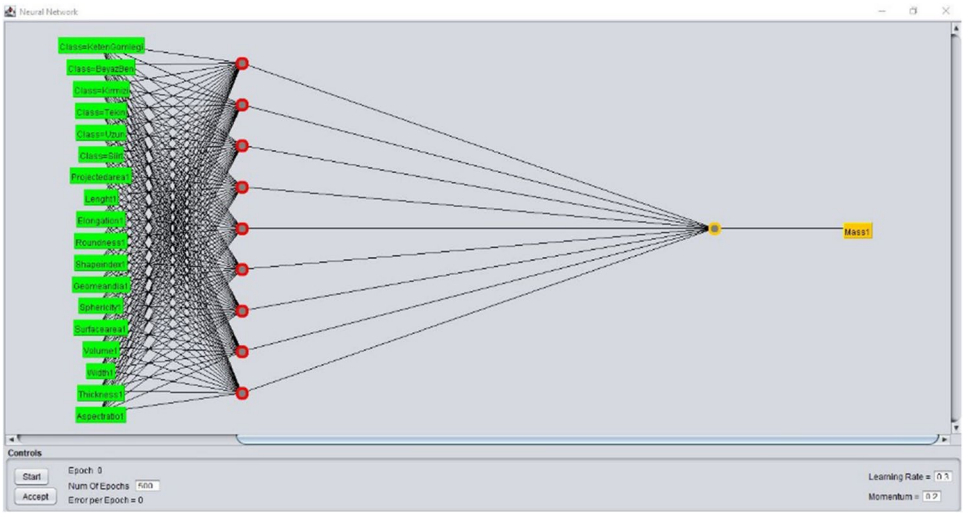


Fig. 2 Structure of the MLP model for prediction of the pistachio mass

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(E_i - M_i\right)^2}{n}}$$

$$RAE = \frac{\sum_{i=1}^{n} \mid E_i - M_i \mid}{\sum_{i=1}^{n} \mid \dot{M} - M_i \mid} \times 100$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}\left(E_i - M_i\right)^2}{\sum_{i=1}^{n}\left(\dot{M} - M_i\right)^2}} \times 100$$

where n is the number of data instances, M_i is the measured target value, E_i is the predicted target value, $\dot{M}$ is the mean of measured target values, $\dot{E}$ is the mean of predicted target values, S_M is the sum of measured target values, and S_E is the sum of estimated target values. The correlation coefficient (R) was analyzed for evaluating the goodness of the prediction according to rules given by Colton (1974). In the present study, taking into consideration correlation coefficients, from 0 to 0.25 means there was a little or no relationship; from 0.25 to 0.50, fair degrees of relationship; from 0.50 to 0.75, moderate to a good relationship; and from 0.75 to 1, very good to the excellent relationship.

Results and Discussion

Size and Shape Attributes of Pistachio Cultivars

Mass, volume, and dimensional attributes of 6 different pistachio cultivars are given in Table 2. The greatest nut and kernel mass values were obtained from Tekin (1.31 and 0.68 g) and Siirt (1.27 and 0.66 g) cultivars. Altuntas and Mutlu (2007) reported the average nut mass as 0.917 g and kernel mass as 0.452 g. The greatest nut and kernel volume values were obtained from Siirt and Tekin cultivars. Nut lengths varied between 15.76 and 23.60 and kernel lengths varied between 12.50 and 20.34 mm. The greatest nut and kernel widths were obtained from Tekin (12.21 and 9.41 mm) and Siirt (12.03 and 9.33 mm) cultivars. The greatest nut and kernel thickness were also obtained from Siirt and Tekin cultivars and the lowest values were obtained from the Kirmizi cultivar. Seferoglu et al. (2006) reported the nut length, width, and thickness of the Uzun pistachio cultivar in 7 different locations respectively as between 17.7 and 19.2 mm, between 10.3 and 11.4 mm, and between 9.1 and 10.8 mm. Similar values were also reported by Abidi

Table 2 Gravitational and dimensional attributes

Cultivars	Mass (M, g)		Volume (V, mm^3)		Length (L, mm)		Width (W, mm)		Thickness (T, mm)	
	Nut	Kernel	Nut	Kernel	Nut	Kernel	Nut	Kernel	Nut	**Kernel**
Keten Gömleği	0.82 ± 0.10^{c}	0.43 ± 0.07^{d}	1039.91 ± 111.82^{bc}	510.08 ± 79.08^{c}	18.81 ± 0.98^{c}	15.26 ± 0.65^{c}	10.91 ± 0.50^{b}	8.38 ± 0.55^{c}	9.66 ± 0.57^{b}	7.58 ± 0.70^{b}
Beyaz Ben	0.84 ± 0.10^{c}	0.43 ± 0.06^{d}	983.17 ± 111.22^{c}	489.84 ± 76.54^{c}	18.36 ± 0.97^{d}	15.08 ± 0.91^{c}	10.68 ± 0.43^{c}	8.19 ± 0.49^{c}	9.55 ± 0.52^{b}	7.53 ± 0.63^{b}
Kirmizi	0.91 ± 0.12^{b}	0.48 ± 0.07^{c}	1062.42 ± 156.55^{b}	569.81 ± 80.09^{b}	19.99 ± 1.13^{b}	16.58 ± 0.86^{b}	10.88 ± 0.62^{bc}	8.70 ± 0.56^{b}	9.27 ± 0.62^{c}	7.51 ± 0.54^{b}
Tekin	1.31 ± 0.15^{a}	0.68 ± 0.08^{a}	1529.78 ± 184.14^{a}	800.26 ± 105.69^{a}	20.65 ± 0.91^{a}	17.64 ± 0.80^{a}	12.21 ± 0.60^{a}	9.41 ± 0.59^{a}	11.55 ± 0.71^{a}	9.18 ± 0.66^{a}
Uzun	0.95 ± 0.11^{b}	0.51 ± 0.07^{b}	1092.58 ± 142.01^{b}	580.24 ± 77.60^{b}	20.09 ± 0.95^{b}	16.63 ± 0.79^{b}	10.99 ± 0.61^{b}	8.69 ± 0.49^{b}	9.41 ± 0.58^{bc}	7.64 ± 0.59^{b}
Siirt	1.27 ± 0.14^{a}	0.66 ± 0.10^{a}	1559.03 ± 194.94^{a}	797.26 ± 123.79^{a}	20.94 ± 0.86^{a}	17.96 ± 0.80^{a}	12.03 ± 0.56^{a}	9.33 ± 0.73^{a}	11.78 ± 0.87^{a}	9.04 ± 0.69^{a}
Mean	1.02 ± 0.02	0.53 ± 0.13	1211.15 ± 283.10	624.58 ± 156.89	19.81 ± 1.34	16.53 ± 1.35	11.28 ± 0.82	8.78 ± 0.73	10.21 ± 1.23	8.08 ± 0.97
Min–max	0.49–1.68	0.23–0.92	538.43–2077.73	307.65–1109.67	15.76–23.60	12.50–20.34	8.77–13.83	6.69–11.03	7.44–13.87	5.60–10.91
F values	323.56**	218.08**	288.43**	227.83**	110.24**	216.68**	138.89**	74.05**	304.19**	157.81**

Means indicated with different letters in the same column are significantly different ($p < 0.05$)

**Significant at $p < 0.01$

(2016) for the Mateur pistachio cultivar, by Kashaninejad et al. (2006) for the Ohadi pistachio cultivar, by Kola et al. (2018) for Kerman and Siirt pistachio cultivars, and by Saglam and Kup (2013) for Organik, Bozova, and Akziyaret pistachio cultivars.

The geometric mean diameter values could be used as valuable data especially in designing the grading process. Kashaninejad et al. (2006) stated that the average geometric mean diameter values of the Ohadi pistachio cultivar increased from 12.97 to13.90 for nuts and from 10.65 to 11.81 mm for kernels with increasing moisture contents. Kernel projected areas varied between 55.10 and 129.58 mm^2 and nut projected areas varied between 80.02 and 196.86 mm^2 (Table 3). Complying with the present findings, Polat et al. (2007) reported the average projected area of the Kirmizi pistachio cultivar as 132.6 mm^2 for nuts. The average surface area was calculated as 320.06 mm^2 for nuts and 220.39 mm^2 for kernels (Table 3). Similar values were also reported by Galedar et al. (2008) for wild pistachio nuts.

The shape attributes are given in Table 4. The average sphericity value of nuts and kernels was 66.43% and 63.80%, respectively. Razavi et al. (2007) indicated that sphericity of the Akbari pistachio cultivar increased from 0.682 to 0.683 for nuts with increasing moisture contents. The greatest nut and kernel shape index values were obtained from Kirmizi (1.99 and 2.05) and Uzun (1.97 and 2.04) cultivars. Similar values were also reported by Unver et al. (2016) and Eski et al. (2018) for walnut and almond cultivars, respectively. A value of roundness close to 1 indicates that cultivar was similar to circle. However, in this study, average roundness values were calculated as 0.44 for nuts and 0.41 for kernels. The average roundness value of 17 different hazelnut cultivars was reported as 0.813 (Sayinci et al. 2015), which was greater than the present values since present cultivars were mostly elliptical. Also, value of aspect ratios closes to 1 indicates increasing circularity of the product. As in roundness, values were not close to 1 in the aspect ratio. The average values of roundness and sphericity linearly decreased with increasing shape index values. The average of elongation had a relative trend related to the average of shape index. The highest elongation values of kernels were obtained from the Siirt cultivar (1.94). Similar to the present findings, Demir et al. (2019) reported elongation of almond cultivars as between 1.317 and 1.946 for kernels.

Relationships Between Groups of Physical Attributes

The correlation matrix for size, shape, and area attributes of pistachio cultivars is given in Table 5. The correlations greater than 0.70 were marked in bold. Kernel and nut mass of pistachio had significant correlations with volume, geometric mean diameter, and projected and surface area. Similarly, nut and kernel volume had a positive correlation with geometric mean diameter, projected area, surface area, and thickness. Çetin et al. (2020) reported significant positive correlations between mass and volume of nine different hazelnut cultivars. Additionally, some significant correlations were also observed in descriptive shape attributes of the pistachio cultivars: between sphericity and roundness, while a negative correlation was determined between shape index and sphericity values. Similar results were also reported by Demir et al. (2018) for walnut cultivars. At the same time, there were highly negative correlations between aspect ratio and elongation values as $R = -0.993$ for nuts and $R = -0.990$ for kernels.

Table 3 Area attributes of pistachio cultivars

Cultivars	Geometric mean diam (D_g, mm)		Projected area (PA, mm^2)		Surface area (SA, cm^2)	
	Nut	Kernel	Nut	Kernel	Nut	Kernel
Keten Gömleği	12.55 ± 0.45^{b}	9.89 ± 0.52^{c}	123.94 ± 8.90^{bc}	76.98 ± 7.99^{c}	495.75 ± 35.61^{bc}	307.91 ± 31.95^{c}
Beyaz Ben	12.32 ± 0.47^{c}	9.75 ± 0.52^{c}	119.37 ± 9.02^{c}	74.92 ± 7.89^{c}	477.48 ± 36.07^{c}	299.68 ± 31.55^{c}
Kirmizi	12.63 ± 0.63^{b}	10.26 ± 0.50^{b}	125.57 ± 12.45^{b}	82.91 ± 7.87^{b}	502.29 ± 49.80^{b}	331.64 ± 31.50^{b}
Tekin	14.27 ± 0.57^{a}	11.50 ± 0.51^{a}	160.26 ± 12.83^{a}	104.01 ± 9.17^{a}	641.04 ± 51.31^{a}	416.04 ± 36.67^{a}
Uzun	12.76 ± 0.55^{b}	10.33 ± 0.47^{b}	128.01 ± 11.07^{b}	83.94 ± 7.53^{b}	512.06 ± 44.29^{b}	335.75 ± 30.14^{b}
Siirt	14.36 ± 0.59^{a}	11.47 ± 0.59^{a}	162.28 ± 13.48^{a}	103.68 ± 10.71^{a}	649.11 ± 53.93^{a}	414.71 ± 42.83^{a}
Mean	13.15 ± 1.00	10.53 ± 0.87	136.57 ± 21.03	87.74 ± 14.59	546.29 ± 84.11	350.95 ± 58.38
Min–max	10.09–15.83	8.38–12.84	80.02–196.86	55.10–129.58	320.06–787.43	220.39–518.34
F values	279.19**	220.71**	286.29**	226.40**	286.28**	226.40**

Means indicated with different letters in the same column are significantly different ($p < 0.05$)

**Significant at $p < 0.01$

Table 4 Shape attributes of pistachio cultivars

Cultivars	Sphericity (*S*, %)		Shape index (SI)		Roundness (*R*)		Aspect ratio (AR)		Elongation (*E*)	
	Nut	Kernel	Nut	Kernel	Nut	Kernel	Nut	Kernel	Nut	Kernel
Keten Gömleği	66.84 ± 2.50^{b}	64.82 ± 2.84^{ab}	1.83 ± 0.10^{b}	1.92 ± 0.12^{bc}	0.45 ± 0.03^{b}	0.42 ± 0.04^{ab}	0.58 ± 0.04^{a}	0.55 ± 0.04^{a}	1.73 ± 0.11^{c}	1.83 ± 0.12^{d}
Beyaz Ben	67.18 ± 2.44^{b}	64.76 ± 2.82^{ab}	1.82 ± 0.10^{b}	1.92 ± 0.12^{bc}	0.45 ± 0.03^{b}	0.42 ± 0.04^{ab}	0.58 ± 0.03^{a}	0.54 ± 0.03^{ab}	1.72 ± 0.09^{c}	1.85 ± 0.11^{cd}
Kirmizi	63.22 ± 2.06^{c}	61.93 ± 2.07^{c}	1.99 ± 0.10^{a}	2.05 ± 0.10^{a}	0.40 ± 0.03^{c}	0.38 ± 0.03^{c}	0.55 ± 0.03^{c}	0.52 ± 0.03^{cd}	1.84 ± 0.10^{a}	1.91 ± 0.12^{ab}
Tekin	69.19 ± 2.72^{a}	65.22 ± 2.36^{a}	1.74 ± 0.10^{c}	1.90 ± 0.10^{c}	0.48 ± 0.04^{a}	0.43 ± 0.03^{a}	0.59 ± 0.03^{a}	0.53 ± 0.03^{bc}	1.69 ± 0.10^{c}	1.88 ± 0.11^{bc}
Uzun	63.54 ± 2.22^{c}	62.16 ± 2.25^{c}	1.97 ± 0.10^{a}	2.04 ± 0.11^{a}	0.40 ± 0.03^{c}	0.39 ± 0.03^{c}	0.55 ± 0.03^{c}	0.52 ± 0.03^{cd}	1.83 ± 0.10^{a}	1.92 ± 0.11^{ab}
Siirt	68.61 ± 2.15^{a}	63.89 ± 2.33^{b}	1.76 ± 0.08^{c}	1.96 ± 0.11^{b}	0.47 ± 0.04^{a}	0.41 ± 0.03^{b}	0.56 ± 0.04^{b}	0.52 ± 0.04^{d}	1.79 ± 0.12^{b}	1.94 ± 0.15^{a}
Mean	66.43 ± 3.29	63.80 ± 2.78	1.85 ± 0.14	1.97 ± 0.13	0.44 ± 0.04	0.41 ± 0.04	0.57 ± 0.04	0.53 ± 0.04	1.77 ± 0.12	1.89 ± 0.13
Min–max	57.84–75.83	55.70–72.73	1.51–2.27	1.61–2.40	0.33–0.58	0.31–0.53	0.47–0.71	0.40–0.64	1.41–2.13	1.56–2.52
F values	114.37**	33.58**	114.88**	32.76**	110.36**	33.72**	34.90**	12.90**	35.38**	12.51**

Means indicated with different letters in the same column are significantly different ($p < 0.05$)

** Significant at $p < 0.01$

Discrimination and Classification of Cultivars

The results of canonical discriminant functions are provided in Table 6. The higher the Eigenvalues, the greater the function explains dependent variable. Square of the canonical correlation values gives the effect size of functions. Function 1 and function 2 explained respectively 72.6 and 18.2% of variance between the cultivars and these functions explained 90.7% of total variation for nuts. Additionally, the first two functions explained 94.4% of total variation for kernels. Wilks' lambda gives the best prediction model. Wilks' lambda is significant for each predictor variable that is ideal; in this case, for 4 results for nuts, 3 results for kernels were significant. According to Wilks' lambda statistics, unexplained part of the differences between the groups was 6.6% for nuts and 16.4% for kernels. The standardized canonical discriminant function coefficients show the relative importance of 10 predictors.

For function 1, sphericity and length were the strongest predictors for both nuts and kernels. According to the loadings, function 2 had the highest loading of the shape index for nuts and volume for kernels. For function 3, sphericity for nuts and thickness for kernels had the greatest function coefficients.

Figure 1 shows the centroids of the 6 pistachio cultivars based on their canonical discriminant functions. Regarding the shape differences between principal components, sphericity was considered a significant distinguishing feature. The features of sphericity and roundness for Siirt and Tekin cultivars confirmed the location of canonical function 1. Canonical function 2 had the highest load for shape index for nuts, and for volume with the negative correlation for kernels. Keten Gömleği and Beyaz Ben cultivars together constituted a separate group in shape attributes. Kirmizi and Uzun were the cultivars comparatively located on the left of the canonical function 1 axis. Additionally, Tekin and Siirt cultivars on the right of the canonical function 2 axeis exhibited dissimilarity with its length and thickness features.

Results of the confusion matrix in LDA methods are presented in Table 7. LDA was able to classify pistachio cultivars with 73.17%. The greatest classification was obtained from Siirt and Tekin cultivars. However, among all cultivars, Kirmizi and Uzun had the worst correct classification. All cultivars had a prediction accuracy higher than 60% of correct classification.

Prediction Results of Machine Learning Algorithms

Models were built based on data from 6 pistachio cultivars. The results were discussed over the correlation coefficient of the predictions of mass. Results of mass prediction by MLP, kNN, Random Forest, and Gaussian Processes are shown in Table 8. The criteria for best prediction were higher

Table 5 Correlation matrix for some physical attributes of the pistachio cultivars

Variables	*M* (g)	*V* (mm^3)	*L* (mm)	*W* (mm)	*T* (mm)	D_g (mm)	PA (mm^2)	SA (mm^2)	S (%)	SI	*R*	AR	*E*
M (g)	1	**0.942**	**0.835**	**0.770**	**0.846**	**0.940**	**0.942**	**0.942**	**0.204**	**−0.201**	**0.202**	**−0.075**	**0.063**
V (mm^3)	**0.967**	1	**0.851**	**0.843**	**0.898**	**0.995**	**0.999**	**0.999**	0.275	−0.273	0.273	−0.005	−0.005
L (mm)	**0.752**	**0.749**	1	0.673	0.641	**0.863**	**0.858**	**0.858**	−0.257	0.258	−0.257	−0.399	0.389
W (mm)	**0.855**	**0.893**	0.598	1	0.608	**0.849**	**0.847**	**0.847**	0.335	−0.351	0.330	0.404	−0.416
T (mm)	**0.891**	**0.932**	0.528	**0.765**	1	**0.897**	**0.898**	**0.898**	0.492	−0.478	0.491	−0.029	0.026
D_g (mm)	**0.964**	**0.997**	**0.764**	**0.898**	**0.926**	1	**0.999**	**0.999**	0.266	−0.264	0.264	−0.010	0.001
PA (mm^2)	**0.967**	**0.999**	**0.757**	**0.897**	**0.930**	**0.999**	1	**1.000**	0.270	−0.269	0.268	−0.008	−0.002
SA (mm^2)	**0.967**	**0.999**	**0.757**	**0.897**	**0.930**	**0.999**	**1.000**	1	0.270	−0.269	0.268	−0.008	−0.002
S (%)	0.425	0.479	−0.215	0.541	0.676	0.465	0.473	0.473	1	**−0.996**	**0.997**	**0.741**	**−0.737**
SI	−0.408	−0.464	0.229	−0.536	−0.658	−0.451	−0.458	−0.458	**−0.996**	1	**−0.990**	**−0.759**	**0.759**
R	0.423	0.477	−0.213	0.539	0.674	0.464	0.471	0.471	**0.997**	**−0.991**	1	**0.735**	**−0.730**
AR	0.121	0.180	−0.411	0.380	0.315	0.163	0.172	0.172	**0.814**	**−0.820**	**0.815**	1	**−0.990**
E	−0.120	−0.179	0.411	−0.379	−0.315	−0.163	−0.171	−0.171	**−0.811**	**0.822**	**−0.810**	**−0.993**	1

M, mass; *V*, volume; *L*, length; *W*, width; *T*, thickness; D_g, geometric mean diameter; *PA*, projected area; *SA*, surface area; *S*, sphericity; *SI*, shape index; *R*, roundness; *AR*, aspect ratio; *E*, elongation

The correlation values in bold are significant at 0.01 level

Nut physical values in lower triangle; kernel physical values in upper triangle

Table 6 Discriminant analysis results

	Function 1	Function 2	Function 3	Function 4	Function 5	Function 1	Function 2	Function 3	Function 4	**Function 5**
Eigenvalue statistics of discriminant functions										
Eigenvalues	3.940	0.986	0.419	0.074	0.012	2.386	0.528	0.125	0.046	0.004
% of variance	72.6	18.2	7.7	1.4	0.2	77.3	17.1	4.0	1.5	0.1
% of cumulative variance	72.6	90.7	98.4	99.8	100.0	77.3	94.4	98.4	99.9	100.0
Canonical correlation	0.893	0.705	0.543	0.262	0.108	0.839	0.588	0.333	0.209	0.061
Significance test of canonical functions	**1–5**	**2–5**	**3–5**	**4–5**	**5**	**1–5**	**2–5**	**3–5**	**4–5**	**5**
Wilks' lambda	0.066	0.327	0.649	0.920	0.988	0.164	0.554	0.847	0.953	0.996
Chi-square	1605.174	661.179	255.824	49.083	6.990	1069.366	348.591	97.969	28.500	2.202
df	50	36	24	14	6	50	36	24	14	6
p (sigma)	0.000**	0.000**	0.000**	0.000**	0.322	0.000**	0.000**	0.000**	0.012	0.900
Standardized canonical discriminant function coefficients										
Mass	−0.411	1.590	1.553	−0.853	−0.227	−0.329	−0.204	1.335	−1.148	0.217
Volume	0.249	−2.419	−0.084	−2.495	4.280	−0.471	−3.319	−4.059	−1.304	−3.201
Length	3.169	−1.532	3.678	−0.054	−2.898	1.172	−0.179	−4.073	−4.863	0.153
Width	−1.962	1.758	−2.919	2.305	−0.925	−1.083	2.997	3.199	5.260	0.598
Thickness	−2.489	1.701	−3.454	2.601	−1.225	−1.004	2.626	7.056	5.333	3.747
Sphericity	5.314	1.007	6.588	−2.316	−3.017	2.575	1.317	−3.754	−5.332	−5.785
Shape index	1.764	4.646	0.860	1.117	0.455	1.053	2.932	5.584	2.261	−2.607
Roundness	−1.114	0.349	−1.141	1.431	1.725	−0.217	−0.744	1.244	1.758	0.383
Aspect ratio	−0.954	−0.840	−1.281	0.376	4.953	−0.542	−0.974	−0.999	−0.455	6.545
Elongation	−2.067	−1.635	−2.175	−0.577	4.559	−1.135	−0.971	−4.086	−0.004	4.404

**Highly significant ($p < 0.01$)

Table 7 Classification results (% correctly classified: 73.17)

Cultivars	Keten Gömleği	Beyaz Ben	Kirmizi	Tekin	Uzun	Siirt	Total
Keten Gömleği	67	25	5	1	2	0	100
Beyaz Ben	16	69	7	0	8	0	100
Kirmizi	6	6	60	3	25	0	100
Tekin	0	0	3	87	2	8	100
Uzun	2	7	28	1	62	0	100
Siirt	1	0	0	5	0	94	100
Total	92	107	103	97	99	102	600

correlation coefficient (R) and lower root mean square error (RMSE), mean absolute error (MAE), relative absolute error (RAE), and root relative squared error (RRSE). As seen in Table 8, the correlation coefficient for nut mass prediction was 0.973, 0.950, 0.963, 0.967, 0.972, and 0.976 for MLP, 1-NN, 3-NN, 5-NN, RF, and GP, respectively. These values for kernel mass prediction were 0.931, 0.890, 0.919, 0.927, 0.939, and 0.948 for MLP, 1-NN, 3-NN, 5-NN, RF, and GP, respectively. Similar with the present findings, Gurbuz et al. (2018) reported a correlation coefficient of 0.8603 for almond mass prediction. Additionally, Rad et al. (2017) reported correlation coefficients for eggplant mass prediction of 4 different ANN structures as between 0.67 and 0.93. Soares et al. (2013) investigated the prediction of the bunch weight of banana plants and was founded the R^2 values as between 0.69 and 0.91 for 6 different ANN structures. Rad et al. (2015) determined predicted mass of melon with MLP models and coefficient of determination was lower than the present value.

According to the results, Gaussian processes had the highest correlation coefficient as 0.976 and 0.948 for nut and kernel mass prediction, respectively. Generally, machine learning algorithms use RMSE because it is easy to perform mathematical operations. The lowest RMSE was also observed in Gaussian processes as compared to the other algorithms. Demir et al. (2020) reported similar RMSE values for walnut. Among the models developed, the second greatest correlation coefficients were observed in MLP with a 18–9-1 structure as 0.973 for nuts and 0.931 for kernels. RF yielded similar correlation coefficients with MLP for mass prediction. Network topology is the main parameter in designing MLPs because the topology type has an important effect on prediction. The number of hidden layers, neurons, and epochs are significant parameters for MLP design. It is preferable that the number of neurons in the hidden layer was low since it leads to a decrease in the networks size and an increase in the network learning speed. The MAE values of selected learning algorithms for kernel and nut were as follows: MLP (0.041; 0.034), 1-NN (0.056; 0.043), 3-NN (0.047; 0.037), 5-NN (0.045; 0.035), RF (0.041; 0.032), and GP (0.038; 0.029). In all algorithms, kernel MAE values were found to be lower than nuts. Among the kNN's predictors, the best performance was achieved by 5-NN. In this case, correlation coefficients increased with increasing "k" values. Overall, each base learner carried out all evaluation parameters very well with all achieving an R value of >0.90. These results point out that all machine learning methods had sufficient success in mass prediction of pistachio cultivars.

Table 8 Comparison of the performance of all machine learning algorithms (nut values in upper, kernel values in lower)

Algorithms	R	MAE	RMSE	RAE	RRSE
MLP	0.973	0.041	0.054	21.04%	23.34%
1-NN	0.950	0.056	0.078	29.08%	31.28%
3-NN	0.963	0.047	0.062	24.37%	26.88%
5-NN	0.966	0.045	0.060	23.29%	25.80%
RF	0.972	0.041	0.054	21.07%	23.29%
GP	0.976	0.038	0.050	19.39%	21.68%
MLP	0.931	0.034	0.047	33.84%	36.98%
1-NN	0.890	0.043	0.058	42.62%	46.44%
3-NN	0.919	0.038	0.050	36.09%	39.43%
5-NN	0.927	0.035	0.047	34.18%	37.66%
RF	0.939	0.032	0.043	31.45%	34.35%
GP	0.948	0.029	0.040	28.48%	31.72%

Conclusion

In the present study, classification and discrimination were carried out for 6 different pistachio cultivars and machine learning algorithms were tested for mass prediction of pistachio nuts and kernels from the physical attributes. Multilayer Perception (MLP), kNN, Random Forest (RF), and Gaussian process (GP) algorithms were implemented. The Gaussian processes (GP) had yielded better outcomes as compared to the other machine learning algorithms. GP, MLP, and RF also yielded quite well predictions for nut and kernel mass. In the structure of the MLP, the optimized number of layers was used to improve the correlation coefficient of the

prediction model. Classifying correct pistachio is crucial in the food process industries. By providing a correctly classified final product, the proposed algorithms could help for mass-based separator or sorter design (Çetin et al. 2021). The present finding should also be taken into consideration in post-harvest technologies for these pistachio cultivars.

Declarations

Ethics Approval This study does not contain any studies with human participants or animals performed by any of the authors.

Consent to Participate Informed consent is not applicable in this study.

Conflict of Interest Cevdet Saglam declares that he has no conflict of interest. Necati Cetin declares that he has no conflict of interest.

References

Abidi W (2016) Pomological and physical attributes of pistachio (*Pistacia vera* L.) varieties grown in west-central Tunisia. J New Sci 28(4):1582–1588

Afonso Junior PC, Correa PC, Pinto FAC, Queiroz DM (2007) Aerodynamic properties of coffee cherries and beans. Biosyst Eng 98:39–46. https://doi.org/10.1016/j.biosystemseng.2007.03.010

Ali J, Khan R, Ahmad N, Maqsood I (2012) Random Forests and decision trees. Int J Comput Sci 9:272–278

Altuntas E, Mutlu A (2007) Determination of some physical properties of pistacia (*Pistacia vera* L.) nut and its kernel (In English). J Agric Fac Gaziosmanpasa Univ 24(1):19–25

Ashtiani SHM, Motie JB, Emadi B, Aghkhani MH (2014) Models for predicting the mass of lime fruits by some engineering properties. J Food Sci Technol 51(11):3411–3417

Ataş M, Yardimci Y, Temizel A (2012) A new approach to aflatoxin detection in chili pepper by machine vision. Comput Electron Agric 87:129–141

Berhane T, Lane C, Wu Q, Autrey B, Anenkhonov O, Chepinoga V, Liu H (2018) Decision-tree, rule-based, and Random Forest classification of high-resolution multispectral imagery for wetland mapping and inventory. Remote Sens 10:580. https://doi.org/10.3390/rs10040580

Breiman L (2001) Random Forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Colton T (1974) Statistics in medicine. Little Brown and Co, New York, p 179

Çetin N, Yaman M, Karaman K, Demir B (2020) Determination of some physicomechanical and biochemical parameters of hazelnut (Corylus avellana L.) cultivars. Turk J Agric For 44(3):1–12. https://doi.org/10.3906/tar-1905-115

Çetin N, Karaman K, Beyzi E, Sağlam C, Demirel B (2021) Comparative evaluation of some quality characteristics of sunflower oilseeds (*Helianthus annuus* l.) through machine learning classifiers. Food Anal Methods 14:1666–1681

Demir B, Eski I, Kus ZA, Ercisli S (2017) Prediction of physical parameters of pumpkin seeds using neural network. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 45(1):22–27

Demir B, Sayinci B, Çetin N, Yaman M, Çömlek R, Aydin Y, Sütyemez M (2018) Elliptic Fourier based analysis and multivariate approaches for size and shape distinctions of walnut (Juglans regia L.) cultivars. Grasas Aceites 69(4):271. https://doi.org/10.3989/gya.0104181

Demir B, Sayinci B, Çetin N, Yaman M, Çömlek R (2019) Shape discrimination of almond cultivars by elliptic fourier descriptors. Erwerbs-Obstbau 61(3):245–256. https://doi.org/10.1007/s10341-019-00423-7

Demir B, Eski İ, Gürbüz F, Kuş ZA, Sesli Y, Ercişli S (2020) Prediction of walnut mass based on physical attributes by Artificial Neural Network (ANN). Erwerbs-Obstbau 62(1):47–56. https://doi.org/10.1007/s10341-019-00468-8

Ercisli S, Sayinci B, Kara M, Yildiz C, Ozturk I (2012) Determination of size and shape features of walnut (Juglans regia L.) cultivars using image processing. Sci Hortic 133:47–55

Eski İ, Demir B, Gürbüz F, Kuş ZA, Yilmaz KU, Uzun M, Ercişli S (2018) Design of neural network predictor for the physical properties of almond nuts. Erwerbs-Obstbau 60(2):153–160. https://doi.org/10.1007/s10341-017-0349-3

FAOSTAT (2019). http://faostat.fao.org/ (Access date 12.11.2019)

Firatligil-Durmuş E, Šárka E, Bubník Z, Schejbal M, Kadlec P (2010) Size properties of legume seeds of different varieties using image analysis. J Food Eng 99(4):445–451. https://doi.org/10.1016/j.jfoodeng.2009.08.005

Galedar MN, Jafari A, Tabatabaeefar A (2008) Some physical properties of wild pistachio (*Pistacia vera* L) nut and kernel as a function of moisture content. Int Agrophys 22(2):117-124

Ghabel R, Rajabipour A, Ghasemi-Varnamkhasti M, Oveisi M (2010) Modeling the mass of Iranian export onion (Allium cepa L.) varieties using some physical characteristics. Res Agric Eng 56(21):33–40

Gonzalez JPB, Ortiz FAP, Rodriguez CPP (2017) Mass and volume estimation of passion fruit using digital images. IEEE Latin America Transactions 15(2):275–282. https://doi.org/10.1109/TLA.2017.7854623

Gurbuz F, Demïr B, Eski İ, Ku ZA, Yilmaz KU, Ilikçioğlu E, Ercişli S (2018) Estimation of the weights of almond nuts based on physical properties through data mining. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 46 (2): 579–584. https://doi.org/10.15835/nbha46210631

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining Software: an update. SIGKDD Explorations. Available online: http://www.cs.waikato.ac.nz. Accessed 15 February 2017

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron 4(1):9

Hassan-Beygi SR, Ghanbarian D, Farahmand M (2010) Prediction of saffron crocus corm mass by geometrical attributes. Sci Hortic 124(1):109–115

IBM SPSS® (2010) Statistical software. SSS Inc., IBM Company©, Version 20.0

Karray FO, Silva CD (2004) Soft computing and intelligent systems design: theory, tools and applications. Addison Wesley Pearson Press, New York

Kashaninejad M, Tabil LG, Mortazavi A, Safekordi A (2003) Effect of drying methods on quality of pistachio nuts. Dry Technol 21(5):821–838. https://doi.org/10.1081/DRT-120021688

Kashaninejad M, Mortazavi A, Safekordi A, Tabil LG (2006) Some physical properties of pistachio (*Pistacia vera* L.) nut and its kernel. J Food Eng 72(1):30–38. https://doi.org/10.1016/j.jfoodeng.2004.11.016

Kola O, Hayoğlu İ, Türkoğlu H, Parildi E, Ak BE, Akkaya MR (2018) Physical and chemical properties of some pistachio varieties (*Pistacia vera* L.) and oils grown under irrigated and non-irrigated conditions in Turkey. Qual Assur Saf Crops Foods 10(4):383–388. https://doi.org/10.3920/QAS2017.1152

Khoshnam F, Tabatabaeefar A, Ghasemi Varnamkhasti M, Borghei A (2007) Mass modeling of pomegranate (Punica granatum L.) fruit with some physical characteristics. Sci Hortic 114(1):21–26

Kus ZA, Demir B, Eski I, Gurbuz F, Ercisli S (2017) Estimation of the colour properties of apples varieties using neural network. Erwerbs-Obstbau 59(4):291–299

MacKay DJC (1998) Introduction to Gaussian processes. NATO ASI Ser F Comput Syst Sci 168:133–166

Mahawar MK, Jalgaonkar K, Kumar M, Meena VS, Bhushan B (2017) Determination of some physical properties of date palm fruits (cv. Khadrawy and medjool). Acta Agroph 24:217–223

Marini F, Magri AL, Balestrieri F, Fabretti F, Marini D (2004) Supervised pattern recognition applied to the discrimination of the floral origin of six types of Italian honey samples. Anal Chim Acta 515(1):117–125. https://doi.org/10.1016/j.aca.2004.01.013

Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. Int J Remote Sens 39:2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Moosavi AA, Sepaskha A (2012) Artificial neural networks for predicting unsaturated soil hydraulic characteristics at different applied tensions. Arch Agron Soil Sci 58(2):125–153

Mohammadi Moghaddam T, Razavi SMA (2019) Application of Artificial Neural Networks (ANN) and image processing for prediction of the geometrical properties of roasted pistachio nuts and kernels. J Nuts 10(1):79–93. https://doi.org/10.22034/jon.2019.554956.1038

Mohsenin NN (1986) Physical properties of plant and animal materials. Gordon and Breach Science Publisher, New York, p 891

Mollazade K, Omid M, Arefi A (2012) Comparing data mining classifiers for grading raisins based on visual features. Comput Electron Agric 84:124–131. https://doi.org/10.1016/j.compag.2012.03.004

Naderi-Boldaji M, Fattahi R, Ghasemi-Varnamkhasti M, Tabatabaeefar A, Jannatizadeh A (2008) Models for predicting the mass of apricot fruits by geometrical attributes (cv. Shams, Nakhjavan, and Jahangiri). Sci Hortic 118(4):293–298

Nettleton DF, Orriols-Puig A, Fornells A (2010) A study of the effect of different types of noise on the precision of supervised learning techniques. Artif Intell Rev 33:275–306. https://doi.org/10.1007/s10462-010-9156-z

Omobuwajo TO, Akande AE, Sanni LA (1999) Selected physical, mechanical and aerodynamic properties African Breadfruit (Treculia africana) seeds. J Food Eng 40:241–244. https://doi.org/10.1016/S0260-8774(99)00060-6

Omid M, Khojastehnazhand M, Tabatabaeefar A (2010) Estimating volume and mass of citrus fruits by image processing technique. J Food Eng 100(2):315–321. https://doi.org/10.1016/j.jfoodeng.2010.04.015

Ozkan G, Koyuncu MA (2005) Physical and chemical composition of some walnut (Juglans regia L.) genotypes grown in Turkey. Grasas Aceites 56:141–146. https://doi.org/10.3989/gya.2005.v56.i2.122

Polat R, Aydin C, Ak B (2007) Some physical and mechanical properties of pistachio nut. Bulg J Agric Sci 13(2):237

Rad MRN, Fanaei HR, Rad MRP (2015) Application of Artificial Neural Networks to predict the final fruit weight and Random Forest to select important variables in native population of melon (Cucumis melo L.). Sci Hortic 181:108–112. https://doi.org/10.1016/j.scienta.2014.10.025

Rad MRN, Ghalandarzehi A, Koohpaygani JA (2017) Predicting eggplant ındividual fruit weight using an Artificial Neural Network. Int Jo Veg Sci 23(4):331–339. https://doi.org/10.1080/19315260.2017.1290001

Rashidi M, Keshavarzpour F (2012) Modeling of tangerine mass based on geometrical properties. World Appl Sci J 16(5):740–743

Rasmussen CE, Williams CK (2006) Gaussian process for machine learning. MIT Press

Razavi SM, Emadzadeh B, Rafe A, Amini AM (2007) The physical properties of pistachio nut and its kernel as a function of moisture content and variety: Part I. Geometrical properties. J Food Eng 81(1):209–217. https://doi.org/10.1016/j.jfoodeng.2006.11.003

Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2012) An assessment of the effectiveness of a Random Forest classifier for land-cover classification. ISPRS J Photogramm Remote Sens 67:93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Romero JR, Roncallo PF, Akkiraju PC, Ponzoni I, Echenique VC, Carballido JA (2013) Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. Comput Electron Agric 96:173–179. https://doi.org/10.1016/j.compag.2013.05.006

Saglam C, Kup F (2013) Determination of some important physical and mechanical properties of different varieties of pistachio nut (*Pistacia Vera* L.) grown in Şanliurfa (In English). Harran J Agric Food Sci 17(3):18–24

Sayinci B, Kara M, Ercişli S, Duyar Ö, Ertürk Y (2015) Elliptic Fourier analysis for shape distinction of Turkish hazel- nut cultivars. Erwerbs-Obstbau 57:1–11. https://doi.org/10.1007/s10341-014-0221-7

Seferoglu S, Seferoglu HG, Tekintas FE, Balta F (2006) Biochemical composition influenced by different locations in Uzun pistachio cv. (*Pistacia vera* L.) grown in Turkey. J Food Compos Anal 19(5):461–465. https://doi.org/10.1016/j.jfca.2006.01.009

Shabani A, Ghaffary KA, Sepaskhah AR, Kamgar-Haghighi AA (2017) Using the artificial neural network to estimate leaf area. Sci Hortic 216:103–110

Soares JDR, Pasqual M, Lacerda WS, Silva SO, Donato SLR (2013) Utilization of artificial neural networks in the prediction of the bunches' weight in banana plants. Sci Hortic 155:24–29. https://doi.org/10.1016/j.scienta.2013.01.026

Stegmayer G, Milone DH, Garran S, Burdyn L (2013) Automatic recognition of quarantine citrus diseases. Expert Syst Appl 40(9):3512–3517

Stroshine R, Hamann DD (1994) Physical properties of agricultural materials and food products. Course manual. Purdue Univ

Tabatabaeefar A, Rajabipour A (2005) Modeling the mass of apples by geometrical attributes. Sci Hortic 105(3):373–382

Tomaino A, Martorana M, Arcoraci T, Monteleone D, Giovinazzo C, Saija A (2010) Antioxidant activity and phenolic profile of pistachio (*Pistacia vera* L., variety Bronte) seeds and skins. Biochimie 92:1115–1122. https://doi.org/10.1016/j.biochi.2010.03.027

Unver H, Sakar E, Sülüşoğlu M (2016) Determination of pomological and morphological characteristics with fatty acid composition of high kernel ratio walnut genotypes. Erwerbs-Obstbau 58(1):11–18. https://doi.org/10.1007/s10341-015-0261-7

Warmund MR (2008) Kernel color of three black walnut cultivars after delayed hulling at five successive harvest dates. HortScience 43:2256–2258. https://doi.org/10.21273/HORTSCI.43.7.2256

Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques. Morgan Kaufmann Press, San Francisco

Zhang H, Song T, Wang K, Wang G, Hu H, Zeng F (2012) Prediction of crude protein content in rice grain with canopy spectral reflectance. Plant Soil Environ 58:514–520

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