



Comparative Evaluation of Some Quality Characteristics of Sunflower Oilseeds (*Helianthus annuus* L.) Through Machine Learning Classifiers

Necati Çetin¹ · Kevser Karaman² · Erman Beyzi³ · Cevdet Sağlam¹ · Bahadır Demirel¹

Received: 11 November 2020 / Accepted: 4 March 2021 / Published online: 15 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Sunflower seeds are rich in oil and oleic acid, thus having valuable nutritional properties. Sunflower is also resistant to dry conditions and can adapt easily to harsh environmental conditions. Physico-chemical properties play a great role in classification, grading, and quality assessment of sunflower seeds. In present study, six different machine learning algorithms (decision tree, DT; random forest, RF; support vector machine, SVM; multiple linear regression, MLR; Naïve Bayes, NB; and multilayer perceptron, MLP) were used to evaluate the classification performance for six different sunflower oilseed varieties. Additionally, characteristic properties of the oilseeds were evaluated by multivariate tests (MANOVA) and discriminant analysis. The best accuracy values were recorded as 80.16, 79.68, and 78.89 for RF, SVM, and MLP, respectively, and the lowest MAE value (0.088) was observed in NB. The MANOVA, Wilks' lambda, and Pillai Trace statistics revealed that differences in physical attributes of the sunflower varieties were significant ($p < 0.01$). Colombi and Transol varieties with the lowest Mahalanobis distances had the similar attributes.

Keywords Sunflower oilseed · Biochemical · Physical · Feature selection · Random forest

Introduction

Sunflower is native to North America and always attracted the attentions of scientist as an important industrial crop (Badouin et al. 2017). In evolutionary biology, the genus *Helianthus* is a long-standing model of hybrid traits (Rieseberg et al. 1995) and adaptive introgression (Vandenbrink et al. 2014). It is also seen as a model for tracking the sun and understanding the development of flowering in plant science (Tähtiharju et al. 2011). Sunflower seeds are rich in oil (40–50%) (Demir and Basalma 2018) and oleic acid (30%) (Avni et al. 2016). Sunflower seeds also have high polyunsaturated and unsaturated fatty acids and high linoleic acid contents (Kiriamiti et al.

2002; Salgin et al. 2006). The fact that sunflower was responsible for about 87% of vegetable oil production makes it more prominent than the other oilseeds. Sunflower seed is also rich in minerals, vitamins, flavonoids, glycosides, phenolic acids, alkaloids, tannins, caffeic acid and sterols etc. (Adeleke and Babalola 2020). The seeds constitute an important source of protein and offer a balanced source of amino acids (Kiriamiti et al. 2002; Salgin et al. 2006). Therefore, they play an important role human nutrition and health. In general, 90% of sunflower seeds is used for oil production and the rest for fresh consumption as an appetizer (Khodabakhshian et al. 2009).

There are many different engineering traits that distinguish agricultural products. Shape and size properties (dimensions, area, mass, elongation and spherical parameters) are important quality indicators of agricultural products. These features should be well-known while designing machinery and systems used in different processes such as classification and quality assessment of agricultural products (Costa et al. 2011). The physical properties of sunflower seeds should also be well-known while designing various equipment such as cleaner, grader, dehuller, separator, and oil expeller (Gupta et al. 2007).

Size, mass and shape attributes can be determined for a single seed. However, the frequency distributions of these properties should be defined for all sets of seeds when designing the

✉ Necati Çetin
necaticeitin@erciyes.edu.tr

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

² Department of Agricultural Biotechnology, Faculty of Agriculture, Erciyes University, Kayseri, Turkey

³ Department of Field Crops, Faculty of Agriculture, Erciyes University, Kayseri, Turkey

relevant equipment (Khazaei et al. 2008). In order to minimize the number of damaged and defective seeds, reduce the power consumption of threshing machines and increase the efficiency of the machines, the relevant machines and classifiers should be so designed and optimized as to comply with the physical properties of the seeds (Mirzabe et al. 2012). Shape, size, mass and biochemical attributes designate the quality of seeds for various purposes (Sorkheh et al. 2018). Product quality is characterized as a combination of the texture, flavor and visual appearance of the fruit, and consumers often shape their preferences based on these parameters (Kays 1999).

Machine learning-based product classification applications are effective tools used in design of accurate and reliable classifiers. Such applications include various algorithms such as decision tree, artificial neural network, genetic algorithm, regressions and fuzzy logic. Additionally, there are powerful algorithms for training of various machine learning models and for adapting difficult input-output mapping strategies as well as selecting and removing useful features. These models are generally used for correct selection of descriptive features in the quality assessment of agricultural products (Omid et al. 2010; Mollazade et al. 2012).

Several researchers studied size and shape attributes of sunflower seeds (Gupta and Das 1997; Santalla and Mascheroni 2003; Khodabakhshian et al. 2010; Mirzabe et al. 2012; Jafari et al. 2011; Malik and Saini 2016; Munder et al. 2017; Ortiz-Hernandez et al. 2020), mineral composition (Ingale and Shrivastava 2011; Kirbaslar et al. 2012), and fatty acid composition of sunflower seeds (Gupta and Shrivastava 2004; Ekin et al. 2005; Izquierdo and Aguirrezábal 2008; Werteker et al. 2010; Ingale and Shrivastava 2011; Alberio et al. 2016; Li et al. 2017). Some other researchers used different machine learning algorithms for quality traits of sunflower seeds (Kavdir and Guyer 2008; Omid 2011; Mollazade et al. 2012; Zielinska et al. 2012; Barbosa et al. 2015; Soltani and Omid 2015; Siedliska et al. 2017; Niaziyan et al. 2018; Abdipour et al. 2019; Gumus et al. 2018; Kurtulmuş 2020). However, present literature reviews revealed that there were any studies about discrimination and classification of sunflower seeds based on physical and chemical properties with the use of machine learning algorithms. Therefore, this study was conducted to (i) determine the size, shape, mass attributes, fatty acid, and mineral composition of sunflower seeds; (ii) distinguish the sunflower oilseed varieties; and (iii) classify and compare the samples by using the different machine learning algorithms.

Materials and Methods

Plant Material and Sample Preparation

Seeds of six sunflower cultivars (Transol, Colombi, Tunca, 63MM54, LG5582, and P64LC53) were used as the plant

material of the present study. Sunflower seeds were harvested from a local farm located about 48 km away from Konya (Turkey) in October 2018 (38° 18' 54.5" N 32° 46' 35.6" E). Foreign materials such as dust, dirt, bark and deformed and immature seeds were manually eliminated from the raw materials. The seeds were preserved in a fridge at 5 ± 0.5 °C until the time of analysis. Before the analyses, sufficient quantity of seeds was removed from the fridge and kept under room conditions for 2 hours. Seed mass was determined with a precise balance (± 0.001).

Determination of Proximate Composition

The methods specified in AOAC (2000) were used to determine crude ash and oil content of the sunflower seeds. Samples were dried in an oven at 105 °C for 4 hours to get dry matter content. All proximate analyses were performed in three replications.

Determination of Fatty Acid Composition

Fatty acid composition of the sunflower seeds was determined with the use of fatty acid methyl esters (Sukhija and Palmquist 1988) in a gas chromatography device equipped with a flame ionization detector and a capillary column. Helium (H₂) was used as the carrier gas. Resultant peaks were compared to authentic standards to get fatty acid quantities.

Determination of Element Composition

Initially, 0.5 g ground sample was supplemented with 10 ml nitric + perchloric acid. Following the wet-digestion of the resultant mixture, diluted samples were subjected to element (B, Ca, Cd, Cu, Fe, K, Mg, Mn, Na, Ni, P, S, and Zn) analyses in an ICP-OES device (Mertens 2005a, 2005b).

Shape and Dimensions

A digital caliper (± 0.01 mm) was used to measure seed length (L, mm), width (W, mm) and thickness (T, mm). Then, the volume (mm³), sphericity (φ , %), shape index (SI), roundness (R), elongation (E), aspect ratio (AR), geometric mean diameter (D_g, mm), projected area (PA, mm²) and surface area (S, mm²) of the samples were calculated with the equations given in Table 1.

Feature Selection and Validation Methodology

Many different techniques are used for the selection of the features. In this study, correlation-based feature selection (CFS) was used. A search algorithm and Pearson correlations are used in CFS to assess feature subsets. It allows individual characteristics to be considered with the level of interrelationship between them

Table 1 Equations used to calculate the size and shape properties of the sunflower oilseed varieties

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)	$S = \pi \cdot D_g^2$	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)	$R = A_p/A_c$	Mohsenin (1986)
Elongation (E)	$E = L/T$	Firathgil-Durmus et al. (2010)
Aspect ratio (R_a)	$R_a = W/L$	Omobuwajo et al. (1999)
Projected area (A_p)	$A_p = (\pi/4) \cdot D_g^2$	Mohsenin (1986)

* L , length (mm); W , width (mm); T , thickness (mm); D_g , geometric mean diameter (mm); A_p , projected area (mm²); A_c , the biggest circular area (mm²)

to estimate the CFS class label (Hall 1999). Best first search algorithm was used to select the best subset (Witten and Frank 2005). CFS (CfsSubsetEval) algorithm was applied on physical properties of sunflower seeds with the use of Weka® v3.8 software (Hall et al. 2009) and 7 features (length, width, thickness, mass, projected area, sphericity, and aspect ratio) were selected. In the study, 105 samples were measured for each feature. A total of 4410 values were used for classification after CFS attribute selection.

Cross-validation evaluates the generalization ability of each model by comparing its performance in a dataset not used during training to fit the parameters of different models. This method is applied practically and effectively in classification and prediction (Stegmayer et al. 2013). In the present study, in order to validate the classification model generated, the k-fold cross-validation method was utilized. In machine learning applications, the k value is usually chosen as 5 or 10 (Ataş et al. 2012) which was selected as 10 in the present study. The dataset was divided into 10 subsets by using 10-fold cross-validation technique. In this way, every subset had an equal proportion of each class example. Training and testing were carried out with 10 iterations. In each iteration, 1 subset was used for testing and the rest of the subsets (9 subsets) were used for training and also with each of the k sub-samples used exactly once as the testing respectively. After that, the average of all errors in each iteration was taken as the error estimate of the classifier. Based on the confusion matrix, the values of evaluation metrics were calculated for algorithms. After that, the performance of classification algorithms was compared by using specified evaluation metrics (Witten and Frank 2005; Stegmayer et al. 2013).

Classification Techniques

A classification technique of machine learning was utilized in the present study by using the Weka® v3.8 software (Hall et al. 2009). Six different machine learning classifiers were

performed in a laptop having a core i5 central processing unit (CPU), 2.5 GHz and 8 GB memory. The classification of varieties using machine learning was based on the main physical properties. Shape (sphericity, aspect ratio), size (length, width, thickness, projected area) and mass were used as the criteria for classifying, because these properties have been chosen by CFS attribute selection.

Decision Tree (DT)

Decision tree (DT), a supervised learning algorithm, is able to learn specific data, develop a set of rules to classify or predict an attribute (Drazin and Montag 2012). The method depends on “divide and conquer” strategy (Safavian and Landgrebe 1991). DTs are arranged so that a class is rejected at each layer of the tree. The last class under the tree is considered as the specified class. There are two different DT induction algorithms, namely J48, C4.5, and logistic model tree (LMT) used in classifications (Quinlan 1993; Landwehr et al. 2005). In this study, the decision tree so called as J48, C4.5 was used, and number of leaves was 28 and the size of tree was 55. The schematic illustration of applied DT is presented in Fig. 1.

Random Forest (RF)

Random forest algorithms generate more than one decision tree to train each tree in 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 ensemble of trees constructed by RF in assigned class of data sets (Berhane et al. 2018). Since bootstrap and ensemble scheme could overcome over fitting problem inherited from DT, there is no pruning step in RF (Ali et al. 2012). In addition, RF generally has a high predictive accuracy and is robust against noise (Breiman 2001; Rodriguez-Galiano et al. 2012).

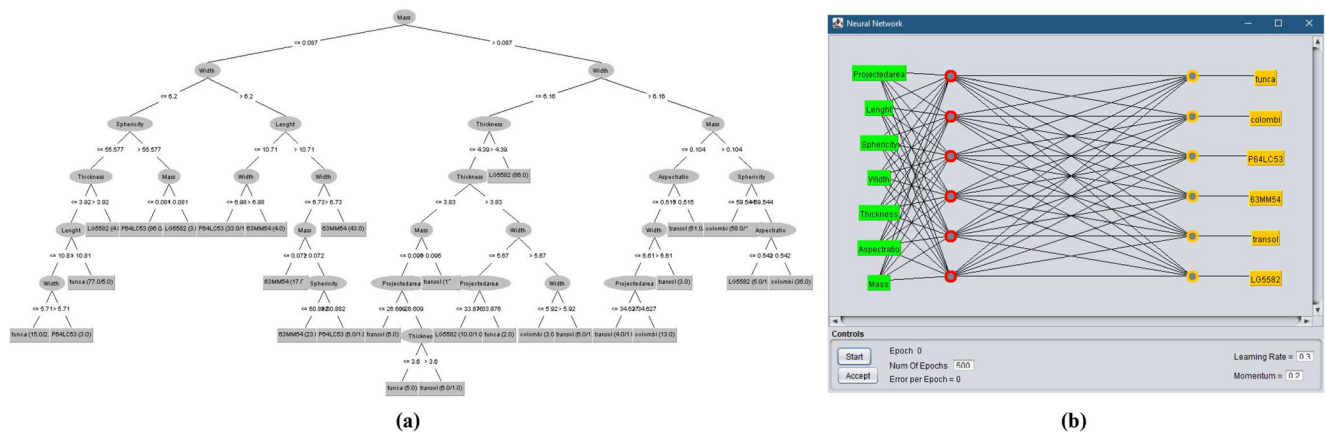


Fig. 1 Structure models of the decision tree (a) and multilayer perceptron (b) algorithms

Support Vector Machines (SVMs)

SVM has different learning algorithms used in classification and regression models. In SVM method, the optimal limit is achieved separately based on probability distribution of the training vectors in the data set. The vectors located close by the hyper plane are defined as support vectors. If the space cannot be separated linearly, a kernel function is used to examine the relationships between the investigated traits (Vapnik 2000).

Multiple Linear Regression (MLR)

MLR elucidate the linear relationships of a dependent variable with the independent variables of the interest. MLR, which is one of the prediction methods, was applied to the data sets in the study. With the prediction techniques used in present analyses (Witten and Frank 2005; Wu et al. 2008), a response variable (y) from a range of predictor variables was estimated with the use of the following Eq. 1 (Vereecken and Herbst 2004; Hastie et al. 2009):

$$y = a + \sum_{i=1}^n b_i x_i + \varepsilon \tag{1}$$

where *a* is an intercept, *x_i* is a predictor variable, *b_i* is a regression coefficient, and ε represents the error.

Naïve Bayes (NB)

This algorithm computes the previous possibilities of each attribute in each class. It is assumed that the possibilities are independent from each other. Classification is done using known probabilities of each class and known probabilities of each attribute. The subsequent probability of class *C_i* can be written as (Eq. 2):

$$P(C_i) \times P(a_1 = x_1 \text{ and } a_2 = x_2 \dots \text{ and } a_n = x_n | C_i) \tag{2}$$

Assuming that the attributes are independent, the value of this expression can be reached by the following Eq. 3:

$$P(C_i) \times P(a_1 = x_1 | C_i) \times P(a_2 = x_2 | C_i) \times \dots \times P(a_n = x_n | C_i) \tag{3}$$

Thus, for equation *i* = 1, 2, the product is calculated and the classification with the highest value is selected (Bramer 2007).

Multilayer Perceptron (MLP)

ANN models are composed of interconnected processing elements similar to biological neurons and weighted connections similar to brain snaps (Karray and Silva 2004). Multilayer perceptron (MLP) is commonly used ANN for classification purposes (Omid et al. 2010). MLPs consist of input layers, hidden layers, and output layers. MLP is a feed-forward neural network (FFNN), namely, data flow through input layers toward to output layers in a single direction. In classification, MLP model is performed with the use of back propagation algorithm (Karray and Silva 2004). MLP is commonly used in classification of agricultural products (Marini et al. 2004; Mollazade et al. 2012). In this study, 7-6-6 MLP structure model with a learning rate of 0.3, momentum of 0.2, number of epochs of 500 and sigmoid function was used for classification (Fig. 1).

Statistical Analyses and Variety Classification

One-factor variance analysis was applied the experimental data with the use of SPSS v20.0 (IBM SPSS® 2010) software and significant means were compared with the use of Tukey’s test (*p*<0.05). Canonical correlation was applied to examine the relationships among the investigated parameters. A scatter

plot with canonical discrimination functions presented the group centroids of varieties. The principal components were assessed for multivariate tests (MANOVA). Hotelling's pairwise comparisons were conducted with the use of PAST v3.20 software to see the differences in sunflower oilseed varieties (Hammer et al. 2001).

Model Performance Evaluation

Model performance was assessed through statistical indicators including classification accuracy (A_c), sensitivity (S_e), specificity (S_p), F -measure (F), precision (P), kappa statistic (κ), mean absolute error (MAE), and root mean square error (RMSE) and was calculated using Eqs. 4, 5, 6, 7, 8, 9, 10, and 11, respectively. Kappa statistics measure interrater reliability among the classifiers. Kappa statistics of -1 indicates "total disagreement," 0 indicates "random classification," and 1 indicates "perfect agreement" (Viera and Garrett 2005). The following equations are used for calculations (Parker 2001):

$$A_c = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (4)$$

$$S_e = \frac{TP}{TP + FN} \times 100 \quad (5)$$

$$S_p = \frac{TN}{TN + FP} \times 100 \quad (6)$$

$$F = \frac{2 \times P \times S_e}{P + S_e} \quad (7)$$

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$\kappa = \frac{\text{Prob}(\alpha) - \text{Prob}(e)}{1 - \text{Prob}(e)} \quad (9)$$

$$\text{MAE} = \sum_{i=1}^n \frac{|E_i - M_i|}{n} \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (E_i - M_i)^2}{n}} \quad (11)$$

where n = number of data instances, M_i = measured target value, E_i = predicted target value, $\text{Prob}(\alpha)$ = total agreement probability, $\text{Prob}(e)$ = hypothetical probability of chance agreement, TP = number of true positives, FP = number of false positive, TN = number of true negatives, and FN = number of false negatives. The accuracy was assessed through goodness of the prediction (Colton 1974). The correlation coefficients of between 0 and 0.25 indicate little or no relationship, 0.25 and 0.50 indicate fair degree relationship, 0.50 and 0.75 indicate moderate to good relationship, and 0.75 and 1.0 indicate very good to excellent relationship.

It is possible to use relative operating characteristic (ROC) curves to visualize the tradeoffs between the correctly

classification of positive and negative rates. A ROC curve is a graphical plot of the true positive rate, also known as sensitivity, versus the false positive rate or one minus the specificity, as its threshold of discrimination is varied. A classifier with a particular tradeoff between sensitivity and specificity represents each point on the ROC curve. The information expressed by the ROC curve must be collapsed into a single response variable to compare the output of multiple classification schemes with statistical instruments (Pietersma et al. 2003; Stegmayer et al. 2013). The region under the entire ROC curve (AUC) was suggested as an acceptable performance index for this reason (Bradley 1997) because it is between 0 and 1 and makes it easier to compare classifiers with each other. Additionally, higher AUC value indicates a better predictive capability of a model. When AUC is close to 1 , it means that a score higher than any non-class sample has been allocated to most of the positive class samples, which means that there is a threshold that perfectly distinguishes.

Results and Discussion

Six sunflower varieties (105 single seed) were evaluated in terms of fatty acid and mineral composition, biochemical characteristics, and physical attributes and experimental findings were provided in subsequent sections.

Biochemical Attributes

The proximate compositions of sunflower seeds are given in Table 2. The ash content of sunflower seeds ranged between 3.20 and 4.80% . The greatest ash content was obtained from Transol variety and the least was obtained from Tunca variety. Ash content of sunflower seeds was reported as 4.82 – 4.87% (Ingale and Shrivastava 2011). Present findings on ash content comply with the results of Gupta and Shrivastava (2004). Seed oil contents ranged from 39.9 to 48.1% with the greatest value in Tunca and the smallest value in Colombi varieties. Oil content of sunflower seeds was reported as between 38.13 and 50.20% (Ekin et al. 2005). Differences in oil contents were mostly attributed to environmental conditions, genotypes and cultural practices.

The fatty acid composition of sunflower seeds is given in Table 3. Significant differences were observed in crude oil content and fatty acid composition of sunflower varieties. Among 10 fatty acids identified in sunflower seeds, oleic acid (C18:1 n9c) was identified as the major fatty acid followed by linoleic acid (C18:2 n6c) and palmitic acid (C16:0). Saturated fatty acids are responsible for the cholesterol level in humans; however, it depends on the carbon-chain length and bioavailability (Webb and Sanders 1991; Cater and Garg 1997). Related to this information, behenic acid (C22:0) and arachidic acid (C20:0) acids are less responsible as compared

Table 2 Proximate composition of sunflower oilseed varieties

Varieties	Ash (%)	Crude oil (%)
LG5582	4.35±0.05 ^b	46.25±0.69 ^a
Transol	4.80±0.20 ^a	38.45±0.42 ^d
63MM54	3.85±0.15 ^c	41.74±1.02 ^{bc}
P64LC53	4.30±0.10 ^b	43.60±1.46 ^b
Colombi	3.80±0.00 ^c	39.89±0.61 ^{cd}
Tunca	3.20±0.10 ^d	48.07±1.02 ^a
Mean	4.05±0.53	43.00±3.57
<i>F</i> values	65.22 ^{**}	47.19 ^{**}

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

^{**}Significant at $p < 0.01$

to the myristic (C14:0), palmitic (C16:0) and stearic acids (C18:0). In present study, behenic (C22:0) and arachidic (C20:0) acids were the minor fatty acids in all sunflower varieties. The highest myristic acid (C14:0) was determined in P64LC53 and Tunca varieties (0.07%). The highest behenic (C22:0) and arachidic (C20:0) acids were observed in 63MM54 variety. Similar findings were reported by Izquierdo and Aguirrezábal (2008). Palmitic and stearic acids ranged between 3.58–5.91% and 2.66–3.49 %, respectively. In a previous study, similar palmitic acid (C16:0) and higher stearic acid (C18:0) contents were reported (Alberio et al. 2016). For dietary purposes, higher mono-unsaturated fatty acids, especially higher oleic acid is an important factor for human health (Esmaili et al. 2012; Salimon and Farhan 2012). Also, in recent years, many studies have been carried out to get more oleic acid in seed with genetic improvements. Present oleic acid (18:1 n-9) contents varied between 22.3 and 26.5%. Oleic acid content of the studied materials is very low as compared to high oleic acid varieties. Present findings on oleic acid contents are consistent with results of Li et al. (2017). Angeloni et al. (2017) reported varying oleic acid contents with the genotypes and growing seasons. Werteker et al. (2010) reported highly variable oleic acid contents in sunflower seeds as compared to rapeseed and soybean seeds. Present linoleic acid (C18:2n6) contents ranged between 66.00 and 67.43%. Such findings comply with the results of Li et al. (2017). Linoleic acid is an essential fatty acid and plays an important role in various physiological functions of human body, especially in prevention of cardiovascular diseases (Connor 2000; Arts et al. 2001). Present linolenic acid (C18:3n3) contents ranged between 0.04 and 0.10%. In a previous study, linolenic acid (C18: 3n3) contents were reported as between 0 and 3.9% (Werteker et al. 2010). Konuskan et al. (2017) indicated significant effects of variety, ecological and growing conditions and agronomic practices on fatty acid composition of oilseeds.

Table 3 Fatty acid composition (%) of different sunflower oilseed varieties

Varieties	C14:0	C16:0	C16:1	C18:0	C18:1n9	C18:2n6	C18:3n3	C20:0	C20:1n9	C22:0
LG5582	0.06±0.00 ^{ab}	5.39±0.00 ^d	0.05±0.00 ^c	2.95±0.05 ^{ab}	23.06±0.16 ^{cd}	66.78±0.20 ^{ab}	0.04±0.01 ^d	0.44±0.00 ^b	0.11±0.00 ^d	1.12±0.01 ^b
Transol	0.07±0.01 ^{ab}	5.81±0.06 ^c	0.05±0.00 ^c	2.70±0.31 ^b	23.76±0.60 ^{bc}	66.30±0.78 ^b	0.05±0.00 ^{cd}	0.32±0.01 ^c	0.12±0.01 ^c	0.83±0.05 ^c
63MM54	0.04±0.00 ^b	4.06±0.01 ^e	0.09±0.00 ^a	3.49±0.01 ^a	24.13±0.08 ^b	66.00±0.09 ^b	0.06±0.01 ^c	0.45±0.01 ^a	0.14±0.01 ^b	1.53±0.00 ^a
P64LC53	0.07±0.00 ^a	6.52±0.01 ^a	0.08±0.00 ^b	3.40±0.47 ^a	22.35±0.39 ^d	66.03±0.08 ^b	0.10±0.00 ^a	0.36±0.01 ^c	0.11±0.00 ^d	0.99±0.01 ^c
Colombi	0.06±0.03 ^{ab}	3.58±0.01 ^f	0.09±0.01 ^a	3.14±0.00 ^{ab}	26.50±0.06 ^a	64.94±0.04 ^c	0.07±0.00 ^b	0.34±0.00 ^d	0.19±0.00 ^a	1.09±0.01 ^b
Tunca	0.07±0.00 ^a	5.91±0.01 ^b	0.05±0.00 ^c	2.66±0.21 ^b	22.49±0.39 ^d	67.43±0.17 ^a	0.05±0.00 ^{cd}	0.33±0.00 ^{de}	0.13±0.00 ^c	0.88±0.00 ^d
Mean	0.06±0.01	5.21±1.08	0.07±0.02	3.06±0.39	23.71±1.47	66.25±0.83	0.06±0.02	0.38±0.05	0.13±0.03	1.07±0.24
<i>F</i> values	3.53 [*]	6205.87 ^{**}	210.40 ^{**}	6.15 ^{**}	18.39 ^{**}	61.37 ^{**}	117.30 ^{**}	591.00 ^{**}	243.20 ^{**}	532.99 ^{**}

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

^{*}Significant at $p < 0.05$

^{**}Significant at $p < 0.01$

Table 4 Mineral composition of different sunflower oilseed varieties (mg/kg)

Varieties	B	Ca	Cd	Cu	Fe	K	Mg
LG5582	18.25±0.13 ^a	647.33±5.86 ^c	0.25±0.02 ^e	22.09±0.65 ^b	42.35±0.86 ^b	5517.00±44.54 ^c	1939.33±27.50 ^b
Transol	15.13±0.69 ^{bc}	193.33±4.16 ^e	1.60±0.04 ^d	21.27±0.27 ^b	38.07±0.62 ^c	3675.67±16.92 ^f	1287.00±7.94 ^d
63MM54	18.89±0.23 ^a	972.33±54.22 ^a	1.63±0.04 ^a	23.67±0.57 ^a	45.14±0.55 ^a	5634.67±10.69 ^b	1436.67±17.16 ^c
P64LC53	14.29±0.21 ^c	855.00±28.00 ^b	1.50±0.04 ^b	21.87±0.39 ^b	35.89±0.39 ^d	8507.33±53.63 ^a	2514.00±51.97 ^a
Colombi	15.35±0.11 ^b	554.33±9.07 ^d	0.65±0.03 ^d	17.46±0.39 ^c	39.46±0.38 ^c	5077.00±31.58 ^d	982.00±5.29 ^e
Tunca	8.77±0.13 ^d	224.00±3.00 ^e	0.97±0.04 ^c	18.67±0.50 ^c	32.80±0.38 ^e	4514.33±16.26 ^e	1271.33±14.22 ^d
Mean	15.11±3.40	574.39±300.99	1.10±0.54	20.84±2.22	38.95±4.18	5487.67±1546.03	1571.72±25.19
<i>F</i> values	380.47**	475.49***	805.05**	70.77**	188.80**	7473.52***	1388.68***

Varieties	Mn	Na	Ni	P	S	Zn
LG5582	22.35±0.26 ^b	254.67±6.51 ^c	6.90±0.02 ^b	8685.00±55.05 ^a	2948.67±47.23 ^b	78.05±0.32 ^a
Transol	13.27±0.16 ^c	156.33±3.21 ^d	8.62±0.16 ^a	4578.33±12.01 ^e	1251.00±19.31 ^e	35.46±0.23 ^e
63MM54	21.49±0.56 ^b	307.00±3.61 ^b	5.73±0.06 ^c	6651.67±24.13 ^c	2406.67±23.25 ^c	55.93±0.43 ^c
P64LC53	28.28±0.56 ^a	447.67±11.68 ^a	5.85±0.06 ^c	7835.67±47.18 ^b	3952.67±38.28 ^a	58.28±0.54 ^b
Colombi	21.46±0.12 ^b	236.67±7.23 ^c	3.50±0.06 ^c	4663.67±37.29 ^e	2109.67±7.64 ^d	39.21±0.52 ^d
Tunca	10.95±0.35 ^d	162.00±3.00 ^d	5.46±0.09 ^d	4938.33±39.88 ^d	2056.33±47.90 ^d	38.32±0.58 ^d
Mean	19.63±6.02	260.72±101.63	6.01±1.59	6225.44±1661.25	2454.17±863.12	50.87±15.44
<i>F</i> values	859.17**	797.67**	1180.77**	6277.72**	2180.61**	3921.18**

Means indicated with different letters in the same column are significantly different ($p < 0.05$)***Significant at $p < 0.01$

Element composition of sunflower seed samples was tabulated in Table 4. As can be seen from Table, 13 different elements were detected in seed samples and differences in micro element composition of the varieties were found to be significant ($p < 0.05$). Phosphorus (P), potassium (K), and sulfur (S) were the major elements with average concentrations of 6225.4, 5487.7, and 2454.2.6 mg/kg, respectively. The greatest K content was obtained from P64LC53 (8507 mg/kg) variety and the lowest from Transol (3675.7 mg/kg) variety. The greatest magnesium content was detected in P64LC53 and the least in Colombi variety. The lowest element concentrations of all sunflower varieties were observed for cadmium which ranged between 0.04 and 1.63 mg/kg. Following this mineral, trace quantities of Ni, Ba, Cu, Fe, Mn and Zn were observed (ranging between 6.01 and 50.87 mg/kg). Kirbaslar et al. (2012) reported the mineral composition of some seeds including sunflower and stated that potassium was the major element (3094.1 mg/kg) and Cu, Zn and Fe were observed in trace quantities (10.13, 28.30 and 38.65 mg/kg, respectively). Ingale and Shrivastava (2011) reported Ca and P of two sunflower genotypes respectively as between 1070 and 1500 mg/kg and between 4000 and 3900 mg/kg. Among the present samples, 63MM54 and P64LC53 were found to be prominent for mineral composition. These two genotypes showed the significantly higher concentrations for majority of the major and minor elements (Table 4).

Shape and Dimensional Attributes

The gravitational, dimensional and area attributes of the varieties are provided in Table 5. There were significant differences in physical attributes of the sunflower varieties ($p < 0.01$). Seed mass is an appropriate physical property to determine the quality of seed. The greatest seed weights were obtained from Colombi and LG5582 varieties and the least from 63MM54, P64LC53 and Tunca varieties. Similar with the present findings, Munder et al. (2017) reported seed weights of hybrid

sunflower variety (PR65H22) as between 0.024 and 0.108. Demir et al. (2018) indicated projected area as an important classification criterion. Projected areas of the present sunflower seed samples varied between 27.37 and 41.46 mm² with the greatest value from Colombi and the least from Tunca varieties. Mirzabe et al. (2012) reported projected area of Sirena sunflower variety as between 27.38 and 69.82 mm². The greatest surface areas were obtained from LG5582 and Colombi varieties (165.82 and 174.03 mm², respectively). However, Ortiz-Hernandez et al. (2020) reported greater surface areas for P64H41 precocious hybrid sunflower seeds than the present values. The greatest volume and geometric mean diameter values were determined in Colombi variety as 201.50 mm³ and 7.26 mm, respectively ($p < 0.05$). Malik and Saini (2016) reported volume and equivalent (geometric mean) diameter values of PSH-996 variety at five different moisture contents (07.60–25.00% w.b.) respectively as between 192.61 and 262.77 mm³ and between 7.15 and 7.93 mm and those values were greater than the present ones. Gupta and Das (1997) reported geometric mean diameter of sunflower seeds as 4.72 mm. The greatest length and widths were obtained from Colombi and 63MM54 varieties. Santalla and Mascheroni (2003) reported seed length and width as 11.526 and 5.008 mm, respectively. Khodabakhshian et al. (2010) investigated the physical properties of Shahroodi variety and reported the thickness values as between 3.88 and 4.94 mm.

The dimension and shape attributes are important for sunflower dehuller and separator. Sphericity, shape and aspect ratio of the sunflower varieties are provided in Table 6. All sunflower varieties were not close to sphere. The highest sphericity values were determined in 63MM54 and P64LC53 as 60.82 and 60.36% varieties, respectively. Gupta and Das (1997) reported average sphericity of Morden variety as 57%. Increasing sphericity values were observed with decreasing shape index values. All varieties with a shape index greater than 1.25 were described as oval. Decreasing roundness and sphericity values were observed with increasing

Table 5 Gravitational, dimensional and area attributes for sunflower oilseed varieties

Varieties	Mass (<i>M</i> , g)	Volume (<i>V</i> , mm ³)	Length (<i>L</i> , mm)	Width (<i>W</i> , mm)	Thickness (<i>T</i> , mm)	Geometric mean diam. (<i>D_g</i> , mm)	Projected area (<i>PA</i> , mm ²)	Surface area (<i>SA</i> , mm ²)
LG5582	0.11±0.01 ^a	168.31±26.22 ^b	11.66±0.48 ^c	5.47±0.39 ^d	5.02±0.47 ^a	6.83±0.36 ^b	36.76±3.82 ^b	147.03±15.29 ^b
Transol	0.10±0.01 ^b	158.63±40.10 ^b	12.05±0.90 ^b	6.36±0.63 ^b	3.89±0.56 ^d	6.66±0.60 ^c	35.16±6.13 ^c	140.64±24.51 ^c
63MM54	0.07±0.01 ^c	159.45±21.72 ^b	11.05±0.43 ^d	6.76±0.30 ^a	4.07±0.43 ^c	6.71±0.30 ^{bc}	35.48±3.20 ^{bc}	141.92±12.79 ^{bc}
P64LC53	0.07±0.01 ^c	122.81±19.84 ^c	10.21±0.59 ^e	5.96±0.43 ^c	3.85±0.39 ^d	6.15±0.33 ^d	29.79±3.20 ^d	119.15±12.80 ^d
Colombi	0.11±0.01 ^a	201.50±29.58 ^a	12.39±0.79 ^a	6.80±0.45 ^a	4.56±0.40 ^b	7.26±0.35 ^a	41.46±4.05 ^a	165.82±16.20 ^a
Tunca	0.08±0.01 ^c	108.16±17.39 ^d	11.46±0.76 ^c	5.48±0.37 ^d	3.28±0.32 ^e	5.89±0.32 ^c	27.37±2.94 ^c	109.47±11.75 ^e
Mean	0.09±0.02	153.14±40.61	11.47±0.98	6.14±0.70	4.11±0.70	6.59±0.59	34.33±6.12	137.34±24.49
F values	363.00**	162.10**	135.98**	193.75**	203.17**	165.58**	164.90**	164.90**

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

**Significant at $p < 0.01$

Table 6 Sphericity, shape and aspect ratio attributes for sunflower oilseed varieties

Varieties	Sphericity (<i>S</i> , %)	Shape index (<i>SI</i>)	Roundness (<i>R</i>)	Aspect ratio (<i>AR</i>)	Elongation (<i>E</i>)
LG5582	58.61±2.57 ^b	2.23±0.15 ^c	0.34±0.03 ^b	0.47±0.03 ^c	2.14±0.14 ^a
Transol	55.38±3.90 ^c	2.37±0.24 ^b	0.31±0.04 ^c	0.53±0.05 ^d	1.91±0.18 ^b
63MM54	60.82±2.90 ^a	2.05±0.13 ^d	0.37±0.04 ^a	0.61±0.03 ^a	1.64±0.09 ^e
P64LC53	60.36±3.23 ^a	2.09±0.16 ^d	0.37±0.04 ^a	0.58±0.05 ^b	1.72±0.14 ^d
Colombi	58.75±3.85 ^b	2.19±0.21 ^c	0.35±0.05 ^b	0.55±0.05 ^c	1.83±0.15 ^c
Tunca	51.59±3.24 ^d	2.63±0.23 ^a	0.27±0.03 ^d	0.48±0.04 ^e	2.10±0.18 ^a
Mean	57.58±4.60	2.26±0.27	0.33±0.05	0.54±0.07	1.89±0.24
<i>F</i> values	117.06 ^{**}	130.83 ^{**}	110.32 ^{**}	195.33 ^{**}	186.99 ^{**}

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

^{**}Significant at $p < 0.01$

shape index values. Roundness values close to unity indicate an almost circular seed shape. All average roundness values of sunflower seeds were close to each other and varied between 0.27 and 0.37. According to Table 6, the highest aspect ratio was observed in 63MM54 variety and the greatest elongation values were found in Tunca (2.10) and LG5582 (2.14) varieties. Jafari et al. (2011) reported average elongation value of Shamshiri variety as 3.26. Elongation and aspect ratio were negatively correlated. Increasing aspect ratios were observed with decreasing sphericity.

Discrimination of the Sunflower Varieties

Results of canonical discriminant functions are provided in Table 7. The higher the eigenvalues, the greater the function explains dependent variable. Functions effect size depends on square of the canonical correlation values. First three functions explained 93.3% of the total variation in physical attributes of the varieties. Since the results of biochemical attributes are insufficient in number, it is not handled by this method. Wilks' lambda generally gives the best prediction model and is important for each predictor variables those are ideal. In this work, canonical correlations for 5 results were significant ($p < 0.01$). In Wilks' lambda statistics, only 13% of the difference was not explained.

The standardized canonical discriminant function coefficients present the relative importance of 12 predictors for both physical and chemical attributes. In function 1, the highest loadings were observed for thickness and geometric mean diameter. In function 2, the greatest loadings were observed for thickness and aspect ratio. Additionally, for function 1, palmitic and oleic acid, for the function 2, arachidic and eicosenoic acids were the strongest predictors for the biochemical attributes.

Figure 2 shows the centroids of 6 sunflower oilseed varieties based on their canonical discriminant functions. With respect to the differences between principal components, thickness, geometric mean diameter and aspect ratio were

considered as an important distinguishing feature. Thickness was identified as the most important feature that enables discrimination of varieties. The 63MM54, P64LC53 and LG5582 varieties which are similar in many attributes located on the right of the canonical function 1 axis. Since Colombi and Transol are particularly similar in shape, they were located close to each other. Tunca, which has the lowest values in terms of most attributes, is clearly distinguished from the other varieties and located on the bottom of the canonical function 2 axis. The aspect ratio for 63MM54 and P64LC53 varieties constituted a separate group.

Multivariate Tests (MANOVA) and Pairwise Comparison

The results of the MANOVA test, Mahalanobis distances and Bonferroni corrected are given in Table 8. According to Wilks' lambda and Pillai Trace statistics, all varieties were found to be significant in terms of the physical attributes ($p < 0.01$). Pillai Trace statistics, considered to be the most reliable among multivariate evaluations, consider the sum of the variance that explains the greatest discrimination of independent variables in dependent variables. Wilks' lambda clarified the variance percentage in dependent variables and expressed them with differences in independent variables. The smaller the "Wilks' lambda" statistics, the greater the differences among the groups to be analyzed (Sayıncı et al. 2015). Mostly, the varieties which Mahalanobis distance is lower than 3 showed statistically similar physical attributes ($p > 0.05$). According to Mahalanobis distances among the sunflower varieties, the closest distances were observed between Colombi and Transol varieties, and between 63MM54 and P64LC53 varieties, while the distance between the Tunca and all other varieties had the greatest values. Confusion matrix results are presented in Table 8. Discriminant analysis was able to classify sunflower varieties with 77.94% accuracy. The greatest correctly classification was obtained from Tunca and

Table 7 Discriminant analysis results for the studied parameters of sunflower oilseed varieties

<i>Eigenvalue statistics of discriminant functions</i> Statistics	Physical attributes				
	Function 1	Function 2	Function 3	Function 4	Function 5
Eigenvalues	5.331	2.095	1.330	0.526	0.099
% of variance	56.8	22.3	14.2	5.6	1.1
% of cum. variance	56.8	79.2	93.3	98.9	100.0
Canonical correlation	0.918	0.823	0.755	0.587	0.300
Significance test of canonical functions	1–5	2–5	3–5	4–5	5
Wilks' lambda	0.013	0.083	0.256	0.596	0.910
Chi-square	2694.057	1547.981	846.376	321.154	58.497
df	50	36	24	14	6
p (sigma)	0.000**	0.000**	0.000**	0.000**	0.000**

<i>Standardized canonical discriminant function coefficients</i>											
Variables	Function 1	Function 2	Function 3	Function 4	Function 5	Variables	Function 1	Function 2	Function 3	Function 4	Function 5
Mass	.631	-.558	-.005	-.312	-.389	Ash	-1.333	5.093	.454	.091	.523
Length	.921	2.042	.459	-6.599	-4.639	Crude oil	-8.006	3.265	3.663	2.093	.558
Width	1.533	-.612	.633	4.838	-3.860	Palmitic	18.578	-.572	-.377	-.157	-.193
Thickness	4.112	3.970	-.418	6.034	-1.765	Stearic	5.165	-.869	1.869	.933	.608
Geo. mean diameter	-2.511	-2.099	.886	9.799	3.191	Oleic	14.828	1.255	-.587	-.559	.535
Projected area	-2.104	-2.030	-1.150	-9.865	6.070	Linolenic	2.496	-1.619	1.513	-3.06	.681
Shape index	-1.405	-1.502	-2.761	4.811	3.992	Arachidic	2.158	11.014	-1.406	-8.64	-.143
Roundness	-2.419	-2.997	-.806	-5.869	-1.703	Eicosenoic	.955	-8.818	-2.147	-3.10	-.083
Aspect ratio	.570	3.668	1.229	1.910	2.621	B	1.321	.643	-1.316	.758	-.308
Elongation	.742	1.031	2.231	1.539	-.280	Ca	-1.167	-2.744	2.430	-.008	.659

**Highly significant ($p < 0.01$)

Fig. 2 Scatter plots of the sunflower oilseed varieties in terms of the discriminant scores and group centroids function 1 and 2

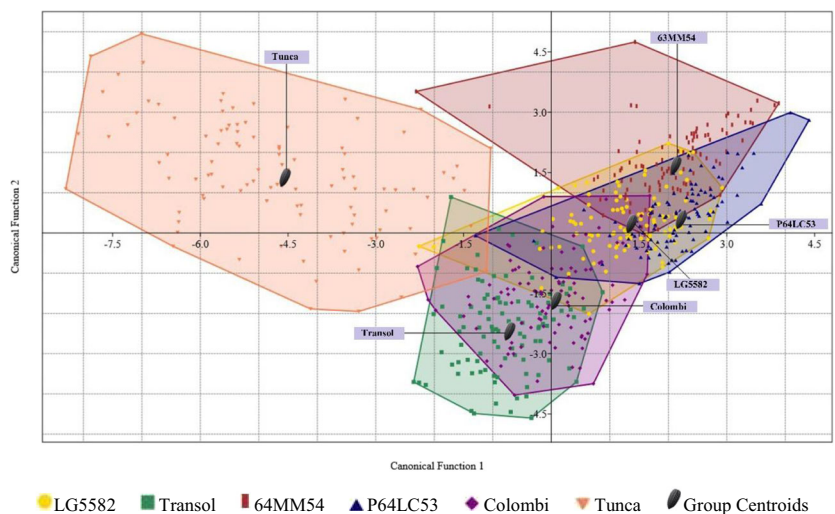


Table 8 Differences among the sunflower oilseed based on seed outlines for physical attributes

The results of MANOVA							
Effect	Statistics	Value	Hypothesis df	Error df	F	p (sigma)	
Variables	Pillai's trace	2.601	65	3080	51.39	0.000**	
	Wilks' lambda	0.011	65	2896	71.13	0.000**	
Hotelling's pairwise comparisons. Bonferroni corrected <i>p</i> values in lower triangle, Mahalanobis distances in upper triangle							
Varieties	LG5582	Transol	63MM54	P64LC53	Colombi	Tunca	
LG5582		20.56	13.26	16.73	10.45	42.81	
Transol	5.58E-69		25.69	20.06	3.39	30.08	
63MM54	3.83E-54	5.43E-77		9.04	17.62	46.97	
P64LC53	7.68E-62	4.18E-68	1.99E-42		14.00	49.98	
Colombi	1.11E-46	4.135E-19	1.33E-63	6.62E-56		31.28	
Tunca	2.95E-96	8.22E-83	7.27E-100	2.71E-102	2.77E-84		
Discriminant analysis confusion matrix (% correctly classified: 77.94)							
Varieties	LG5582	Transol	63MM54	P64LC53	Colombi	Tunca	Total
LG5582	87	0	11	0	7	0	105
Transol	0	79	0	1	25	0	105
63MM54	10	0	75	19	0	1	105
P64LC53	0	3	8	94	0	0	105
Colombi	15	27	2	2	59	0	105
Tunca	0	4	1	0	3	97	105
Total	112	113	97	116	94	98	630

**Highly significant ($p < 0.01$)

P64LC53 varieties. However, among all varieties, Colombi and 63MM54 had the worst correct classification.

Classification Results

Classification matrix for size, shape, area and mass attributes and performance comparisons made with the use of 6 classification algorithms (DT, RF, SVM, MLR, NB and MLP) are given in Table 9. The confusion matrix of classifiers is shown in Table 10. Classification accuracies ranged from 74.44 to 80.16% and RF algorithm had the greatest A_c and the lowest RMSE (0.218) as compared to the other algorithms. DT which has 55 tree and 28 leaves yielded busy and complicated tree.

Mass was identified as an important classification node in the structure of trees. For this reason, if an automatic system for sorting is developed based on the classifiers, mass attribute can be used as a main criterion for product quality assessment (Banakar et al. 2017). DT classification accuracy was 75.71% which was the third lowest accuracy after NB (74.44%) and MLR (75.24%). Similar to present findings, Zielinska et al. (2012) reported that engineering properties-based classification accuracy of four different red clover seeds ranged from 41.0 to 78.5% for Naïve Bayes and from 52.8 to 85.4% for decision tree algorithms. Omid (2011) determined reported the correct classification rates and RMSEs respectively as

Table 9 Performance results of machine learning classifiers

Classifiers	Accuracy (%)	Kappa statistic	MAE	RMSE	Sensitivity (%)	Specificity (%)	Precision	F-measure	ROC area
DT	75.71	0.71	0.095	0.260	0.757	0.951	0.762	0.758	0.910
RF	80.16	0.76	0.095	0.218	0.802	0.960	0.802	0.801	0.967
SVM	79.68	0.76	0.227	0.319	0.797	0.959	0.793	0.793	0.942
MLR	75.24	0.70	0.193	0.286	0.752	0.950	0.748	0.732	0.943
NB	74.44	0.69	0.088	0.259	0.744	0.949	0.745	0.739	0.957
MLP	78.89	0.75	0.092	0.228	0.789	0.958	0.790	0.789	0.954

Table 10 Confusion matrix of classifiers for sunflower oilseed varieties (10-fold cross-validation)

DT	Tunca	Colombi	P64LC53	63MM54	Transol	LG5582	RF	Tunca	Colombi	P64LC53	63MM54	Transol	LG5582
Tunca	83	1	11	0	8	2	Tunca	81	0	11	1	12	0
Colombi	1	81	0	0	21	2	Colombi	0	78	0	1	22	4
P64LC53	6	0	81	16	0	2	P64LC53	5	0	84	15	0	1
63MM54	0	1	23	78	3	0	63MM54	0	1	8	93	3	0
Transol	10	30	2	1	62	0	Transol	8	28	2	0	67	0
LG5582	2	6	3	1	1	92	LG5582	1	2	0	0	0	102
SVM	Tunca	Colombi	P64LC53	63MM54	Transol	LG5582	MLR	Tunca	Colombi	P64LC53	63MM54	Transol	LG5582
Tunca	89	0	11	0	5	2	Tunca	91	1	9	0	2	2
Colombi	0	81	0	1	19	4	Colombi	0	80	1	3	11	10
P64LC53	8	0	78	17	0	2	P64LC53	12	0	74	17	0	2
63MM54	0	3	9	93	0	0	63MM54	0	0	7	96	1	1
Transol	13	29	2	3	58	0	Transol	23	46	1	4	29	2
LG5582	1	1	0	0	0	103	LG5582	0	1	0	0	0	104
NB	Tunca	Colombi	P64LC53	63MM54	Transol	LG5582	MLP	Tunca	Colombi	P64LC53	63MM54	Transol	LG5582
Tunca	90	0	9	0	6	0	Tunca	80	0	13	1	9	2
Colombi	0	80	0	2	21	2	Colombi	0	71	0	0	31	3
P64LC53	11	0	75	19	0	0	P64LC53	3	0	85	16	0	1
63MM54	0	5	14	86	0	0	63MM54	0	2	8	92	2	1
Transol	20	37	1	4	43	0	Transol	12	20	2	1	69	1
LG5582	3	7	0	0	0	95	LG5582	1	1	1	0	2	100

*Colored rows display the true positive values of each model (total number of samples: 105)

99.52% and 0.07 for training set and respectively as 95.56% and 0.21 for the test set (Table 9).

There are mainly four types of kernel functions (polynomial, normalized polynomial, radial basis and Pearson VII function) to select the best SVM. In this study, the appropriate kernel function for SVM was Pearson VII function. The classification matrix of SVM showed that among 630 sunflower seeds, 502 seeds were categorized precisely (Table 10). The accuracy calculated was 79.68% for SVM. However, SVM had the highest MAE and RMSE values. Soltani and Omid

(2015) reported the accuracy of SVM Pearson VII function-based function as 0.891. Mollazade et al. (2012) obtained the greatest *r* and lowest RMSE from SVM polynomial and Pearson VII function-based kernel functions.

A multilayer perceptron (MLP) neural network was applied for analysis of size, shape, and mass of the sunflower seeds. Seven physical attributes were used as inputs, there were six hidden layers and 6 outputs (i.e., 7-6-6 structure). Accuracy, Kappa, sensitivity, specificity, precision and F-measure values of MLP classifier were respectively identified as 78.89%, 0.75,

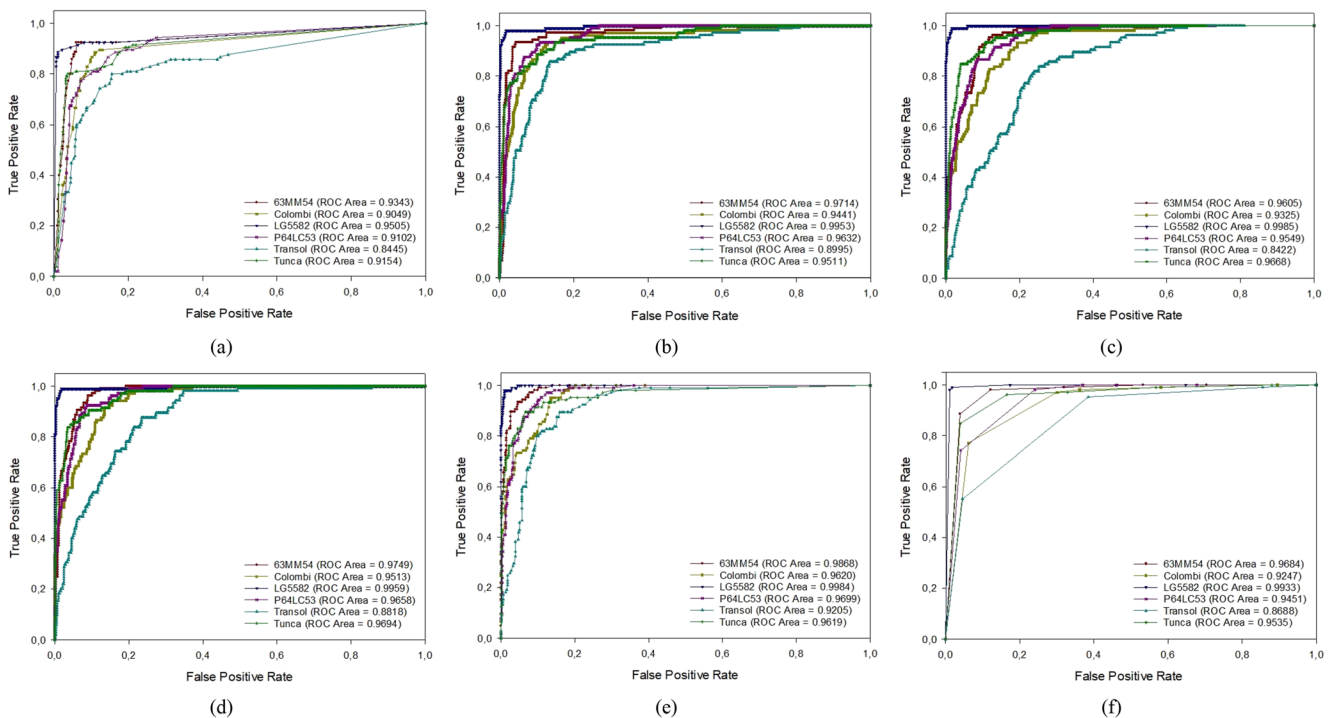


Fig. 3 ROC Curves and area values for machine learning algorithms (a DT, b MLP, c MLR, d NB, e RF, f SVM)

0.789, 0.958, 0.790, and 0.789. Kavdır and Guyer (2008) used artificial neural networks, MLP and DT classifiers to apples and reported the greatest classification accuracy (83.33%) for MLP, then (75.56%) for DT. Gumus et al. (2018) informed that the accuracy of six olive oil classifications based on biochemical attributes ranged from 81.63 to 85.71%. The authors expressed that ANN results exhibited a well-correlation with PCA results. Zielinska et al. (2012) used multilayer perceptron and reported the accuracy of varietal classification as between 57.6 and 89.6%. These findings support the present ones.

NB classifiers had the lowest accuracy, kappa statistic, sensitivity, specificity and precision as 74.44%, 0.69, 0.744, 0.949 and 0.745, respectively. In fact, NB had the lowest MAE values. Present findings comply with the results of Gumus et al. (2019) reporting accuracy as 79.59% for Naïve Bayes, 79.59% for decision tree, and 83.67% for support vector machine. In contrast, accuracy and kappa statistic values of Naïve Bayes (74.44%—0.690) reported by Siedliska et al. (2017) were greater than those of the present work (87.6%—0.724) for fresh cherries, but similar with the accuracy values of random forest (80.3%). Barbosa et al. (2015) also reported greater accuracy (95.4%) for Nb and Rf than the present study.

MLR is a simple model used for estimation in agricultural practices (Abdipour et al. 2019). In this study, the minimum value of F-measure was obtained for MLR and the calculated values of A_c , S_e , and S_p were 75.24%, 0.752, and 0.950, respectively. Among all classifiers, MLR was the best classifiers for LG 5582, 63MM54 and Tunca varieties. Similarly, Abdipour et al. (2019) used ANN and MLR and indicated that MLR was less capable to predict seed yield in both training as $R^2=0.710$, RMSE=0.064, and MAE=0.052 and testing $R^2=0.686$, RMSE=0.071, and MAE=0.054. Present findings comply with the results of Niazian et al. (2018) reporting the ability of MLR model to predict seed yield for training ($R^2=0.81$) and testing ($R^2=0.79$) stages. Author also indicated that the performance of ANN was better than MLR.

The accuracy of the predictive model was presented in the ROC curve. As seen in Fig. 3, the receiver operating curve is a graphical representation of the effectiveness of the predictive model confirmed that the classifier accurately identified varieties. As expected, the highest ROC area values were obtained from RF, MLP, and SVM. In particular, the ROC area values ensure quite a high performance for automatic identification of any understudy of the variety classification, as the values obtained are higher than 0.84 in all cases.

Conclusion

The present study focused on fatty acid and mineral profile and physical attributes of six different sunflower oilseed varieties and detailed discrimination and classifications were

performed based on physical attributes. In general, the greatest shape and dimension values were observed in Colombi and LG5582 varieties, while the highest mineral and fatty acid composition values were observed in 63MM54 and P64LC53 varieties. In terms of physical attributes, excluding shape index and elongation, Tunca variety had the lowest values. The first three functions explained 93.3% of the total variation in physical attributes of the varieties. In the Wilks' lambda statistics, unexplained part of the similarities or dissimilarities among the groups was 13%. Colombi and Transol varieties with the lowest Mahalanobis distances had the closest attributes. Before classification, feature selection was applied to physical attributes, and projected area, length, sphericity, width, thickness, aspect ratio and mass were identified as the best features by using CFS-Best First method. Among six classifiers (DT, RF, SVM, MLR, NB and MLP), the RF and SVM (with Pearson VII kernel function) yielded better outcomes. MLP with a 7-6-6 topology also yielded quite a well classification for sunflower varieties.

Correct classification of sunflower oilseeds that meet specific requirements is important in the food and agricultural industries. By providing a correctly classified final product, the proposed algorithms could help food and seed companies (Kurtulmuş 2020). By using the algorithms and techniques suggested in the study, a separator or sorter could be designed, and the actual classification accuracy of the system could be determined during the process. Cleaning, dehulling, packaging, and oil expelling apparatuses are generally designed based on physical features of the seeds. Such features are also utilized in the design of conveyor systems used in product transportation. In addition, the specified features can be helpful in the selection, breeding, quality assessment, breeding, chemical, and pharmaceutical industries (Çetin et al. 2020). This research showed that using machine learning algorithms, the classification of sunflower seeds was feasible. The results observed from the tests were encouraging in that the approach proposed here could classify related sunflower oilseed varieties. Future studies related to seed classification with machine learning should also be inspire more dataset, features, and algorithms by these results.

Declarations

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

Informed Consent Informed consent is not applicable in this study.

Conflict of Interest Necati Çetin declares that he has no conflict of interest. Kevser Karaman declares that she has no conflict of interest. Erman Beyzi declares that he has no conflict of interest. Cevdet Sağlam declares that he has no conflict of interest. Bahadır Demirel declares that he has no conflict of interest.

References

- Abdipour M, Younessi-Hmazekhanlu M, Ramazani SHR (2019) Artificial neural networks and multiple linear regression as potential methods for modeling seed yield of safflower (*Carthamus tinctorius* L.). *Ind Crop Prod* 127:185–194. <https://doi.org/10.1016/j.indcrop.2018.10.050>
- Adeleke BS, Babalola OO (2020) Oilseed crop sunflower (*Helianthus annuus*) as a source of food: nutritional and health benefits. *Food Sci Nutr* 8:4666–4684. <https://doi.org/10.1002/fsn3.1783>
- Alberio C, Izquierdo NG, Galella T, Zuil S, Reid R, Zambelli A, Aguirrezábal LA (2016) A new sunflower high oleic mutation confers stable oil grain fatty acid composition across environments. *Eur J Agron* 73:25–33. <https://doi.org/10.1016/j.eja.2015.10.003>
- Ali J, Khan R, Ahmad N, Maqsood I (2012) Random forests and decision trees. *Int J Comput Sci* 9:272–278
- Angeloni P, Echarte MM, Irujo GP, Izquierdo N, Aguirrezábal L (2017) Fatty acid composition of high oleic sunflower hybrids in a changing environment. *Field Crop Res* 202:146–157. <https://doi.org/10.1016/j.fcr.2016.04.005>
- AOAC (2000) Official methods of analysis of AOAC international, 17th edn. AOAC International, Gaithersburg
- Arts MT, Ackman RG, Holub BJ (2001) Essential fatty acids in aquatic ecosystems: a crucial link between diet and human health and evolution. *Can J Fish Aquat Sci* 58(1):122–137. <https://doi.org/10.1139/f00-224>
- 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
- Avni T, Anupriya S, Rai P, Maan K, Naryansamy CCN (2016) Effects of heating and storage on nutritional value of sunflower oil. *DU J Undergr Res Innov* 2(1):196–202
- Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L et al (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546(7656):148–152. <https://doi.org/10.1038/nature22380>
- Banakar A, Zareiforoush H, Baigvand M, Montazeri M, Khodaei J, Behroozi-Khazaei N (2017) Combined application of decision tree and fuzzy logic techniques for intelligent grading of dried figs. *J Food Process Eng* 40(3):1–12. <https://doi.org/10.1111/jfpe.12456>
- Barbosa RM, Batista BL, Barião CV, Varrique RM, Coelho VA, Campiglia AD, Barbosa F Jr (2015) A simple and practical control of the authenticity of organic sugarcane samples based on the use of machine-learning algorithms and trace elements determination by inductively coupled plasma mass spectrometry. *Food Chem* 184:154–159. <https://doi.org/10.1016/j.foodchem.2015.02.146>
- 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>
- Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn* 30(7):1145–1159
- Bramer M (2007) Principles of data mining, vol 180. Springer, London. <https://doi.org/10.1007/978-1-4471-7307-6>
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Cater NB, Garg A (1997) Serum low-density lipoprotein cholesterol response to modification of saturated fat intake: recent insights. *Curr Opin Lipidol* 8(6):332–336. <https://doi.org/10.1097/00041433-199712000-00002>
- Çetin N, Yaman M, Karaman K, Demir B (2020) Determination of some physicochemical and biochemical parameters of hazelnut (*Corylus avellana* L.) cultivars. *Turk J Agric For* 44(5):439–450
- Colton T (1974) Statistics in Medicine. Little Brown and Co, New York, p 179
- Connor WE (2000) Importance of n-3 fatty acids in health and disease. *Am J Clin Nutr* 71(1):171–175. <https://doi.org/10.1093/ajcn/71.1.171S>
- Costa C, Antonucci F, Pallottino F, Aguzzi J, Sun DW, Menesatti P (2011) Shape analysis of agricultural products: a review of recent research advances and potential application to computer vision. *Food Bioprocess Technol* 4:673–692. <https://doi.org/10.1007/s11947-011-0556-0>
- Demir İ, Basalma D (2018) Response of different level of nitrogen and sulphur doses on oil yield and seed nutrients content of sunflower (*Helianthus annuus* L.). *Fresenius Environ Bull* 27(9):6337–6342
- Demir B, Sayıncı B, Çetin N, Yaman M, Çömlek R, Aydın 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):e271. <https://doi.org/10.3989/gya.0104181>
- Drazin S, Montag M (2012) Decision tree analysis using weka. Machine Learning-Project II, University of Miami, 1–3.
- Ekin Z, Tunçtürk M, Yılmaz I (2005) Evaluation of seed, oil yields and yield properties of different sunflower (*Helianthus annuus* L.) hybrid varieties in Van, Turkey. *Pak J Biol Sci* 8:683–686. <https://doi.org/10.3923/pjbs.2005.683.686>
- Esmaili A, Shaykhmoradi F, Naseri R (2012) Comparison of oil content and fatty acid composition of native olive genotypes in different region of Lian. *Iran Intl J Agric Crop Sci* 4(8):434–438
- Fıratlıgil-Durmus E, Sárka 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>
- Gumus ZP, Ertas H, Yasar E, Gumus O (2018) Classification of olive oils using chromatography, principal component analysis and artificial neural network modelling. *J Food Meas Character* 12(2):1325–1333. <https://doi.org/10.1007/s11694-018-9746-z>
- Gumus O, Yasar E, Gumus ZP, Ertas H (2019) Comparison of different classification algorithms to identify geographic origins of olive oils. *J Food Sci Technol* 57:1–9. <https://doi.org/10.1007/s13197-019-04189-4>
- Gupta RK, Das SK (1997) Physical properties of sunflower seed. *J Agric Eng Res* 66(1):1–8. <https://doi.org/10.1006/jaer.1996.0111>
- Gupta M, Shrivastava SK (2004) Proximate composition of some new varieties of oil seeds. *Int J Chem Sci* 2(3):375–378
- Gupta RK, Arora G, Sharma R (2007) Aerodynamic properties of sunflower seed (*Helianthus annuus* L.). *J Food Eng* 79(3):899–904. <https://doi.org/10.1016/j.jfoodeng.2006.03.010>
- Hall MA (1999) Correlation-based feature selection for machine learning. PhD Thesis, University of Waikato, Hamilton
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H. 2009. The WEKA data mining software: an update. SIGKDD Explorations. Available online: <http://www.cs.waikato.ac.nz>. (Accessed 12 April 2019).
- Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontol Electron* 4(1):9
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference and prediction, 2nd edn. Springer-Verlag, New York. <https://doi.org/10.1007/978-0-387-84858-7>
- IBM SPSS® (2010). Statistical software. SSS Inc., IBM Company®, Version 20.0.
- Ingale S, Shrivastava SK (2011) Chemical studies of new varieties of sunflower (*Helianthus annuus*) LSF-11 and LSF-8 seeds. *Agric Biol JN Am* 2(8):1171–1181
- Izquierdo NG, Aguirrezábal LAN (2008) Genetic variability in the response of fatty acid composition to minimum night temperature

- during grain filling in sunflower. *Field Crop Res* 106(2):116–125. <https://doi.org/10.1016/j.fcr.2007.10.016>
- Jafari S, Khazaei J, Arabhosseini A, Massah J, Khoshtaghaza MH (2011) Study on mechanical properties of sunflower seeds. *Food Sci Technol* 14(1):6
- Karray FO, Silva CD (2004) *Soft computing and intelligent systems design: theory, tools and applications*. Addison Wesley Pearson Press, New York
- Kavdir I, Guyer DE (2008) Evaluation of different pattern recognition techniques for apple sorting. *Biosyst Eng* 99(2):211–219. <https://doi.org/10.1016/j.biosystemseng.2007.09.019>
- Kays SJ (1999) Preharvest factors affecting appearance. *Postharvest Biol Technol* 15:233–247. [https://doi.org/10.1016/S0925-5214\(98\)00088-X](https://doi.org/10.1016/S0925-5214(98)00088-X)
- Khazaei J, Jafari, S., Noorolah. S. 2008. Lognormal vs. Normal and Weibull distributions for modeling the mass and size distributions of sunflower seeds and kernels. *World Conference on Agricultural Information and IT*.
- Khodabakhshian R, Emadi B, Abbaspour Fard MH (2009) Aerodynamic properties of sunflower seed, kernel and its hull affected by moisture content and size, azargol variety as a case study. *International Agricultural Engineering Conference, Bangkok*
- Khodabakhshian R, Emadi B, Fard MA (2010) Some engineering properties of sunflower seed and its kernel. *Nong Ye Ke Xue Yu Ji Shu* 4(4):37
- Kirbaslar FG, Türker G, Özsoy-Günes Z, Ünal M, Dülger B, Ertas E, Kizilkaya B (2012) Evaluation of fatty acid composition, antioxidant and antimicrobial activity, mineral composition and calorie values of some nuts and seeds from Turkey. *Records Nat Prod* 6(4):339
- Kirihamiti HK, Rascol E, Marty A, Condoret JS (2002) Condoret, Extraction rate of oil from high oleic sunflower seeds with supercritical carbon dioxide. *Chem Eng Process* 41:711–718. [https://doi.org/10.1016/S0255-2701\(01\)00191-X](https://doi.org/10.1016/S0255-2701(01)00191-X)
- Konuskan OMER, Konuskan DB, Levai CM (2017) Effect of foliar boron fertilization on chemical properties and fatty acid compositions of corn (*Zea mays* L.). *Rev Chim (Bucharest)* 68(9):2073–2075. <https://doi.org/10.3965/j.ijabe.20160901.1790>
- Kurtulmuş, F. 2020. Identification of sunflower seeds with deep convolutional neural networks. *J Food Meas Characterization*, 1–10. <https://doi.org/10.1007/s11694-020-00707-7>
- Landwehr N, Hall M, Frank E (2005) Logistic model trees. *Mach Learn* 59:161–205. <https://doi.org/10.1007/s10994-005-0466-3>
- Li WP, Shi HB, Zhu K, Zheng Q, Xu Z (2017) The quality of sunflower seed oil changes in response to nitrogen fertilizer. *Agron J* 109(6):2499–2507. <https://doi.org/10.2134/agronj2017.01.0046>
- Malik MA, Saini CS (2016) Engineering properties of sunflower seed: effect of dehulling and moisture content. *Cogent Food Agric* 2(1):1145783. <https://doi.org/10.1080/23311932.2016.1145783>
- Marini F, Magn 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>
- Mertens D (2005a) Plants preparation of laboratory sample. In: Horwitz W, Latimer GW (eds) *Official Methods of Analysis of AOAC International*, 18th edn. AOAC International, Gaithersburg, pp 1–2
- Mertens D (2005b) Metal in plants and pet foods. In: Horwitz W, Latimer GW (eds) *Official Methods of Analysis of AOAC International*, 18th edn. AOAC International, Gaithersburg, pp 3–4
- Mirzabe AH, Khazaei J, Chegini GR (2012) Physical properties and modeling for sunflower seeds. *Agric Eng Int CIGR J* 14(3):190–202
- Mohsenin NN (1986) *Physical properties of plant and animal materials: structure, physical characteristics and mechanical properties*. Gordon and Breach Science Publishers, New York
- 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>
- Munder S, Argyropoulos D, Mueller J (2017) Class-based physical properties of air-classified sunflower seeds and kernels. *Biosyst Eng* 164:124–134. <https://doi.org/10.1016/j.biosystemseng.2017.10.005>
- Niazian M, Sadat-Noori SA, Abdipour M (2018) Modeling the seed yield of Ajowan (*Trachyspermum ammi* L.) using artificial neural network and multiple linear regression models. *Ind Crop Prod* 117:224–234. <https://doi.org/10.1016/j.indcrop.2018.03.013>
- Omid M (2011) Design of an expert system for sorting pistachio nuts through decision tree and fuzzy logic classifier. *Expert Syst Appl* 38(4):4339–4347. <https://doi.org/10.1016/j.eswa.2010.09.103>
- Omid M, Khojastehnazhand M, Tabatabaefar 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>
- 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](https://doi.org/10.1016/S0260-8774(99)00060-6)
- Ortiz-Hernandez AA, Araiza-Esquivel M, Delgadillo-Ruiz L, Ortega-Sigala JJ, Durán-Muñoz HA, Mendez-García VH, Yacamán MJ, Vega-Carrillo HR (2020) Physical characterization of sunflower seeds dehydrated by using electromagnetic induction and low-pressure system. *Innov Food Sci Emerg Technol* 60:102285. <https://doi.org/10.1016/j.ifset.2019.102285>
- 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>
- Parker J (2001) Rank and response combination from confusion matrix data. *Inf Fusion* 2:113–120. [https://doi.org/10.1016/S1566-2535\(01\)00030-6](https://doi.org/10.1016/S1566-2535(01)00030-6)
- Pietersma D, Lacroix R, Lefebvre D, Wade KM (2003) Performance analysis for machine-learning experiments using small data sets. *Comput Electron Agric* 38(1):1–17
- Quinlan JR (1993) Combining instance-based and model-based learning. In *Proceedings of the tenth international conference on machine learning*, pp 236–243
- Rieseberg LH, Van Fossen C, Desrochers AM (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375:313–316. <https://doi.org/10.1038/375313a0>
- 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>
- Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. *IEEE Trans Syst, Man, Cybernet* 21(3):660–674. <https://doi.org/10.1109/21.97458>
- Salgin U, Doker O, Calimli A (2006) Extraction of sunflower oil with supercritical CO₂: experiments and modeling. *J Supercrit Fluids* 38(3):326–331. <https://doi.org/10.1016/j.supflu.2005.11.015>
- Salimon J, Farhan N (2012) Physicochemical properties of Saudi extra virgin olive oil. *Int J Chem Environ Eng* 3(3):205–208
- Santalla EM, Mascheroni RH (2003) Equilibrium moisture characteristics of high oleic sunflower seeds and kernels. *Dry Technol* 21(1):147–163. <https://doi.org/10.1081/DRT-120017288>
- Sayıncı B, Ercişli S, Akbulut M, Şavşatlı Y, Baykal H (2015) Determination of shape in fruits of cherry laurel (*Prunus laurocerasus*) accessions by using Elliptic Fourier analysis. *Acta Sci Polonorum, Hortorum Cultus* 14(1):63–82
- Siedliska A, Baranowski P, Zubik M, Mazurek W (2017) Detection of pits in fresh and frozen cherries using a hyperspectral system in transmittance mode. *J Food Eng* 215:61–71. <https://doi.org/10.1016/j.jfoodeng.2017.07.028>

- Soltani M, Omid M (2015) Detection of poultry egg freshness by dielectric spectroscopy and machine learning techniques. *LWT-Food Sci Technol* 62(2):1034–1042. <https://doi.org/10.1016/j.lwt.2015.02.019>
- Sorkheh K, Kazemifard A, Rajabpoor S (2018) A comparative study of fuzzy linear regression and multiple linear regression in agricultural studies: a case study of lentil yield management. *Turk J Agric For* 42:402–411. <https://doi.org/10.3906/tar-1709-57>
- Sukhija PS, Palmquist DL (1988) Rapid method for determination of total fatty acid content and composition of feedstuffs and feces. *J Agric Food Chem* 36(6):1202–1206
- Stegmayer G, Milone DH, Garran S, Burdyn L (2013) Automatic recognition of quarantine citrus diseases. *Expert Syst Appl* 40(9):3512–3517
- Tähtiharju S, Rijpkema AS, Vetterli A, Albert VA, Teeri TH, Elomaa P (2011) Evolution and diversification of the CYC/TB1 gene family in Asteraceae—a comparative study in *Gerbera* (Mutisieae) and sunflower (Heliantheae). *Mol Biol Evol* 29(4):1155–1166. <https://doi.org/10.1093/molbev/msr283>
- Vandenbrink JP, Brown EA, Harmer SL, Blackman BK (2014) Turning heads: the biology of solar tracking in sunflower. *Plant Sci* 224:20–26. <https://doi.org/10.1016/j.plantsci.2014.04.006>
- Vapnik VN (2000) Methods of pattern recognition. In: *In The nature of statistical learning theory*. Springer, New York, pp 123–180. https://doi.org/10.1007/978-1-4757-3264-1_6
- Vereecken H, Herbst M (2004) Statistical regression. In: Pachepsky YA, Rawls WJ (eds) *Development of pedotransfer functions in soil hydrology*. Elsevier, Amsterdam, pp 3–18. [https://doi.org/10.1016/S0166-2481\(04\)30001-2](https://doi.org/10.1016/S0166-2481(04)30001-2)
- Viera AJ, Garrett JM (2005) Understanding interobserver agreement: the kappa statistic. *Fam Med* 37(5):360–363
- Webb DR, Sanders RA (1991) Caprenin 1. Digestion, absorption, and rearrangement in thoracic duct-cannulated rats. *J Am Coll Toxicol* 10(3):325–340. <https://doi.org/10.3109/10915819109079813>
- Werteker M, Lorenz A, Johannes H, Berghofer E, Findlay CS (2010) Environmental and varietal influences on the fatty acid composition of rapeseed, soybeans and sunflowers. *J Agron Crop Sci* 196(1):20–27. <https://doi.org/10.1111/j.1439-037X.2009.00393.x>
- Witten IH, Frank E (2005) *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann Press, San Francisco
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ..., Zhou, Z. H. 2008. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14 (1), 1–37. <https://doi.org/10.1007/s10115-007-0114-2>
- Zielinska M, Zapotoczny P, Białobrzewski I, Zuk-Golaszewska K, Markowski M (2012) Engineering properties of red clover (*Trifolium pratense* L.) seeds. *Ind Crop Prod* 37(1):69–75. <https://doi.org/10.1016/j.indcrop.2011.12.002>

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