European Food Research and Technology (2024) 250:409–423
https://doi.org/10.1007/s00217-023-04392-w

ORIGINAL PAPER

Binary classification of pumpkin (*Cucurbita pepo* L.) seeds based on quality features using machine learning algorithms

Necati Çetin[1] · Ewa Ropelewska[2] · Sali Fidan[3] · Şükrü Ülkücü[4] · Pembe Saban[5] · Seda Günaydın[5] · Ali Ünlükara[5]

Received: 15 July 2023 / Revised: 8 October 2023 / Accepted: 13 October 2023 / Published online: 7 November 2023

Abstract
Mass, size, and shape attributes are important for the design of planters, breeding studies, and quality assessment. In recent years, machinery design and system development studies have taken these factors into consideration. The aim of this study is to explore classification models for four pumpkin seed varieties according to their physical characteristics by machine learning. Binary classification is important because it ensures that the quality characteristics of the seeds are very similar to each other. The pumpkin seed varieties of Develi, Sena Hanım, Türkmen, and Mertbey were discriminated in pairs. Five machine learning algorithms (Naïve Bayes, NB; support vector machine, SVM; random forest, RF; multilayer perceptron, MLP; and kNN, k-nearest neighbors) were applied to assess the classification performance. In all pairs, the pumpkin seed varieties of Develi and Mertbey were discriminated with the highest accuracies of 85.00% for the MLP model and 84.50% for the SVM model and 83.50% for the RF. In the MLP algorithm, TP rate reached to 0.790 for Develi and 0.910 for Mertbey, Precision to 0.898 for Develi and 0.813 for Mertbey, F-measure to 0.840 for Develi and 0.858 for Mertbey, PRC area to 0.894 for Develi and 0.896 for Mertbey, and ROC area to 0.907 for both varieties. Variety of pairs was followed by Sena Hanım and Türkmen (84.50%, MLP) and Türkmen and Mertbey (82.50%, SVM). For the selected input attributes, the highest mass (0.23 g), length (22.08 for Mertbey, 21.43 for Sena Hanım), and geometric mean diameter (8.79 mm) values were obtained from Sena Hanım variety, while shape index (3.40) from Mertbey variety. Multivariate statistical results showed that differences in attributes were significant ($p < 0.01$). Wilks' lambda statistics found that the portion of the unexplained difference between groups was 46.60%. Develi and Sena Hanım varieties with the lowest Mahalanobis distance values had similar characteristics. Present results revealed that SVM and MLP may be used effectively and objectively for the classification of pumpkin seed varieties.

Keywords Pumpkin seed · Variety classification · Mass · Shape index · Machine learning

Abbreviations

ANN	Artificial neural network
AR	Aspect ratio
BPNN	Propagation neural network
CPU	Central process unit
CFS	Correlation-based feature selection
DD-SIMCA	Data-driven soft-independent modeling of class analogy
D_g	Geometric mean diameter
DT	Decision tree
E	Elongation
FAO	Food and agriculture organization
kNN	K-nearest neighbors
L	Length
LDA	Linear discriminant analysis
LR	Logistic regression
LS-SVM	Least squares support vector machine

✉ Necati Çetin
necati.cetin@ankara.edu.tr

[1] Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, Ankara University, Ankara, Turkey

[2] Fruit and Vegetable Storage and Processing Department, The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland

[3] Transitional Zone Agricultural Research Station, Eskişehir, Turkey

[4] Department of Plant Protection, Graduate School of Natural and Applied Science, Erciyes University, Kayseri, Turkey

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

ML	Machine learning
MLP	Multilayer perceptron
NB	Naïve Bayes
PA	Projected area
PA-3DCNN	Position attention embedded three-dimensional convolutional neural network
PC	Principal component
PRC	Precision–recall
R	Roundness
RBNN	Radial basis neural network
RF	Random forest
RMSE	Root mean square error
ROC	Receiver operating characteristic
S	Surface area
SI	Shape index
SVM	Support vector machine
T	Thickness
TP	True positive
V	Volume
W	Width
φ	Sphericity

Introduction

Pumpkin, a member of the Cucurbitaceae family, is a medically and economically golden plant species [1]. According to FAO data, 24 million tons of pumpkin are produced annually in the world. While China ranks first in world pumpkin production, it is followed by Ukraine, Russia, America, Spain, and Turkey [2]. Snack food, pumpkin, and squash varieties are grown for commercial purposes [3]. The meat part of the squash is used in soups, vegetable dishes, cakes, desserts, and confectionery [4]. While the seeds are consumed as snacks [5], the waste parts are used in animal nutrition [6]. In addition to being consumed fresh or roasted, pumpkin seeds are used as a food supplement, in salads, meals, and sauces, in the pharmacological field, in the production of cosmetic products, in the production of soap and candles, by obtaining oil from the seeds [3, 7].

The use of pumpkin-containing drugs rich in omega-3, fatty acids (linoleic acid, palmitic acid, oleic acid, and steric acid), zinc, and selenium draws attention worldwide [8–10]. β-carotene, which has an anti-aging effect, strengthens immunity and prevents the formation of tumors and cataracts, is abundant in pumpkin seeds [11, 12]. Thanks to these unsaturated fatty acids, it strengthens memory, prevents cancer, and plays an active role in reducing inflammation in the body [13, 14]. It is a rich source of protein, lutein, phenolic compounds, vitamins B1, B2, and C, α-tocopherol (vitamin E), and nutrients (Mg, K, Fe, Na, Se, P, Zn, and Mn) [15–17].

The physical properties of agricultural products (such as shape, size, sphericity, surface area, bulk weight, moisture content, porosity, specific gravity, color, and mass) are important in terms of gaining consumer appreciation and post-harvest technologies [18–20]. Customers prefer products that look healthy and regular in shape, color, and size [21]. As the moisture content of the seeds increases, the breaking strength decreases. Friction coefficient, porosity, and axial dimension increase [22]. The size and shape data of seeds provide convenience in the design and manufacture of standard packages [23]. In addition, the shape and size characteristics of seeds are considered in the design of sorting and grinding machines [24–26]. The physical attributes of pumpkin seeds should be known for the design of equipment that will help from planting seeds to post-harvest processing and marketing [27]. These measures take a lot of time and effort. To solve these issues, novel technologies have been created. Development technologies might be easily and quickly identified, classified, and sorted. To describe the features employed in the quality assessment of seeds, however, such pragmatic techniques are required.

Artificial intelligence is the approach that imitates the human brain and can make decisions and finalize the process in the new formation by transferring human characteristics [28]. Machine learning is the performance of a specific task through the acquisition and interpretation of extensive data by computer systems. With the advantage of machine learning, it is possible to efficiently categorize samples [29]. Machine learning uses multi-layered mathematical operations to learn and manipulate complex data. It is also modeled by mimicking the human brain [30]. Classification processes are carried out by processing data through machine learning algorithms. Machine learning is mostly implemented using neural networks, trees, and support vector machines [26, 31].

Many studies were performed to detect only the mass, size, and shape attributes of pumpkin seeds [18, 23, 32–34]. However, there are limited numbers of studies about shape and size-based classification of Cucurbitaceae. Generally, classification studies related to pumpkin seed [35, 36, 53–55] and watermelon seed [37–39]. However, literature reviews presented that there were no studies on the binary classification of pumpkin seeds using machine learning models. The novelty of this study is related to binary classification of the pumpkin seeds based on similar physical attributes by machine learning and analytical methods. The aim of the study was to develop binary classification models by five machine learning algorithms (NB, SVM, RF, MLP, and kNN) for the classification of four different pumpkin seed varieties based on mass, shape, and size.

Materials and methods

Plant material and sample preparation

In this study, seeds of four pumpkin varieties (Develi, Sena Hanım, Türkmen, and Mertbey) were used as the plant materials. Pumpkin seeds were harvested on 16 September 2021 from Develi District (38° 16′ 25.7″ N, 35° 25′ 03.1″ E) in Kayseri province of Turkey. Deformed, dirt and hollow seeds were removed before analysis and preserved at 4 ± 0.5 °C throughout the analysis.

Shape and dimension measurements

The mass of the products was measured by classical methods with the use of a precise electronic scale (±0.001 g), and principal physical properties such as length (L, mm), width (W, mm), and thickness (T, mm) were determined by instrumental methods using a digital caliper (±0.01 mm). For mass, shape, and size, 100 pumpkin seeds were handled from each variety. Size (geometric mean diameter, D_g, mm; volume, V, mm^3; projected area, PA, mm^2 and surface area, S, mm^2) and shape (aspect ratio, AR; elongation, E; roundness, R; shape index, SI; and sphericity, φ, %) attributes were found using equations given in Table 1. The flow chart of the binary classification of pumpkin seed varieties by machine learning is presented in Fig. 1. These stages consist of determining size, shape, and mass attributes, implementing feature selection, performing cross-validation, binary classifying by machine learning, and evaluating performance metrics.

Table 1 Size and shape equations used in the calculations

Attributes	Equations*	References
Shape index (SI)	$SI = (2 \cdot L)/(W + T)$	[48]
Geometric mean diameter (D_g, mm)	$D_g = (L \cdot W \cdot T)^{(1/3)}$	[49]
Volume (V, mm^3)	$V = (\pi/6) \cdot D_g^3$	[50]
Projected area (A_p)	$A_p = (\pi/4) \cdot D_g^2$	[49]
Surface area (S, mm^2)	$S = \pi \cdot D_g^2$	[50]
Roundness (R)	$R = A_p/A_c$	[49]
Sphericity (φ)	$\varphi = (D_g/L) \cdot 100$	[49]
Aspect ratio (R_a)	$R_a = W/L$	[51]
Elongation (E)	$E = L/T$	[52]

*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)

Multivariate analysis

Experimental data were evaluated in one-factor analyses, and Tukey's multiple comparison test was utilized to evaluate significant means ($p < 0.05$). Linear discriminant analysis was used to evaluate differences between the variations. The discriminant analysis variety group centroids was applied to create a scatter plot. The principal components (PCs) were evaluated for multivariate tests. Hotelling's pair-wise comparisons with Bonferroni correction and squared Mahalanobis distances were used to determine whether pumpkin seed varieties were similar or different from one another. Software versions PAST v3.20 [40] and SPSS v20.0 [41] were used to conduct statistical analyses.

Feature selection, validation methodology, and classification

Weka® v3.8 software was used to apply a classification strategy of machine learning models [42]. Five machine learning classifiers were run on a computer with an 8 GB memory and core i7 CPU running at 4.2 GHz. The primary physical characteristics served as the basis for the machine learning classification of variations. Machine learning algorithms used the primary physical characteristics to categorize different pumpkin seed varieties. The classification of pumpkin seed varieties using machine learning models was based on the main physical attributes. Mass, length, geometric mean diameter, and shape index were used as the criteria for classifying because these attributes have been selected by CFS attribute selection. 100 pumpkin seed samples were determined for each attribute. Total sample size was 5200, and a total of 1300 were used for each variety. The k-fold cross-validation method was applied for model performance evaluation. The k value was chosen as 10 since the current data set had 10 sub-sets. Training processes were utilized in

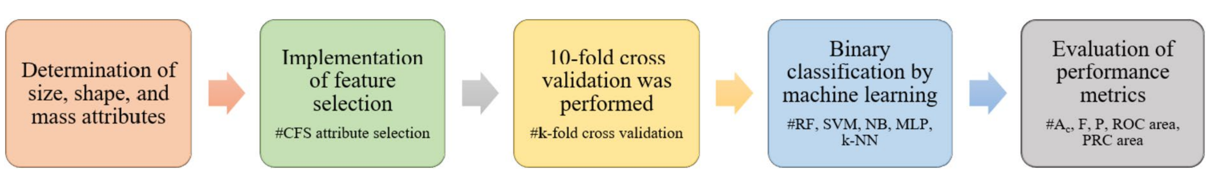


Fig. 1 The flow chart of the binary classification of pumpkin seed varieties by machine learning

the 10 iterations. One sub-set was used for testing and the other subsets (9 sub-sets) were used for training, in each iteration. Each k sub-sample was utilized once for testing, respectively [43]. The k-fold cross-validation procedure is presented in Fig. 2.

Machine learning approaches

The model development was performed on a variety of datasets (inputs), including physical attributes, such as mass, length, geometric mean diameter, and shape index. A total of 400 data, 100 from each attribute, were used for each binary analysis. The models were created using different algorithms from the groups of Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), Multilayer Perceptron (MLP), and k-Nearest Neighbors (kNN) in a test validation mode of ten fold cross-validation. In this study, the Chebyshev distance rule with the LinearNN Search was performed in the search process in the k-NN method, and the k values were 5. SVM was decided upon Pearson VII (PUK) kernel function. The numbers of neurons in input, hidden, and output layers were all 4-3-2 ANN structures in the binary classification of the pumpkin seed varieties. The numbers of epochs, learning ratio, momentum coefficient, and activation function were chosen as 500, 0.3, and 0.2, and the sigmoid function in all MLP classifications, respectively. The MLP model structure is given in Fig. 3, and detailed information about the ML models is provided in Fig. 4.

The outcomes include accuracies for each pair and confusion matrices for the pairs of four kinds of pumpkin seeds. In addition, accuracy (A_c), F-measure (F), precision (P), ROC (Receiver Operating Characteristic) area, and PRC (precision–recall) area. Performance indices were determined by Eqs. (1), (2), and (3) [44].

$$A_c = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \times 100 \tag{1}$$

$$F = \frac{2 \times P \times S_e}{P + S_e} \tag{2}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

Fig. 2 Ten fold cross-validation methodology

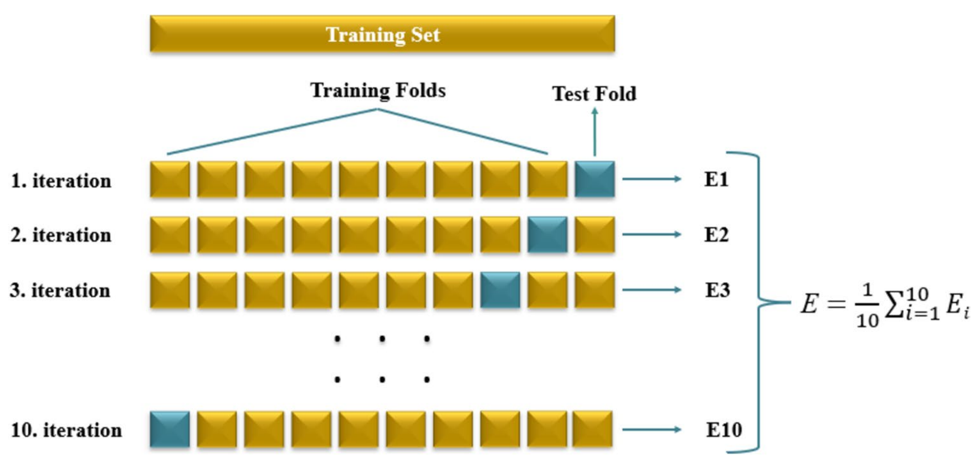
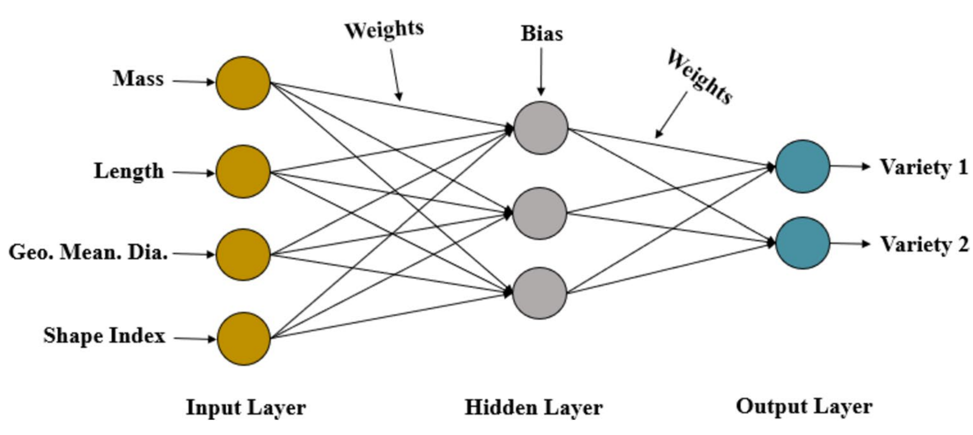


Fig. 3 MLP structure model



Algorithm	Main parameters
RF	batchSize: 100; breakTiesRandomly: False; doNotCheckCapabilities: False; debug: False; numExecutionSlots: 1; numIterations: 100; seed: 1
SVM	batchSize: 100; c:1.0; debug: False; doNotCheckCapabilities: False; filterType: Normalize training data; kernel: Pearson VII (PUK); regOptimizer: RegSMOImproved
NB	batchSize: 100; debug: False; displayModelInOldFormat: False doNotCheckCapabilities: False; useKernelEstimator: False; useSupervisedDiscretization: False
MLP	batchSize: 100; debug: False; doNotCheckCapabilities: False; decay: False; hiddenLayers: a ((attribs + classes) / 2); normalizeNumericClass: True; momentum: 0.2; learningRate: 0.3; normalizeAttributes: True; NominalToBinaryFilter: True; validationThreshold: 20; trainingTime: 500; seed: 0
k-NN	KNN:3; batchSize: 100; distanceWeighting: No distance Weighting; doNotCheckCapabilities: False; meanSquared: False; nearestNeigbourSearchAlgorithm: LinearNNSearch (distanceFunction: EuclideanDistance)

Fig. 4 Detailed information on ML models

TN = Number of true negatives, TP = Number of true positives, FN = Number of false negatives and FP = Number of false positives.

To compare the results of various categorization schemes using statistical metrics, the information provided by the ROC curve must be condensed into a single response variable [43, 45]. Because it falls between 0 and 1 and facilitates comparisons between classifiers, the region under the complete ROC curve was proposed as a suitable performance metric [46]. A threshold that perfectly separates exists when the ROC area value is close to 1, which indicates that most positive class samples have been given scores higher than any non-class samples.

Results and discussion

Seed physical attributes

Size, shape, and mass attributes of four pumpkin seed varieties were obtained, and binary classification was utilized for varieties. The results of the physical attributes are tabulated in Tables 2 and 3. In this study, all physical attributes were found to be significant ($p < 0.01$). The Sena Hanım variety had the greatest mass with the value of 0.23 g, while the Türkmen had the lowest mass (0.18 g). The highest volume and the length were determined from Sena Hanım (*V*: 363.61 mm^3 and *L*: 21.43 mm) and Mertbey (*V*: 357.44 mm^3 and *L*: 22.08 mm) varieties. The greatest and the lowest thickness

Table 2 Size and mass attributes for pumpkin seed varieties

Varieties	Mass (M, g)	Volume (V, mm^3)	Length (L, mm)	Width (W, mm)	Thickness (T, mm)	Geometric mean diam. (D_g, mm)	Projected area (PA, mm^2)	Surface area (SA, mm^2)
Develi	0.20 ± 0.05^{bc}	311.95 ± 74.51^{b}	19.83 ± 1.61^{b}	10.52 ± 1.06^{a}	2.83 ± 0.42^{c}	8.36 ± 0.66^{b}	55.27 ± 8.72^{b}	221.09 ± 34.89^{b}
Sena H	0.23 ± 0.06^{a}	363.61 ± 90.53^{a}	21.43 ± 1.82^{a}	10.73 ± 1.06^{a}	2.97 ± 0.40^{ab}	8.79 ± 0.76^{a}	61.16 ± 10.30^{a}	244.63 ± 41.21^{a}
Türkmen	0.18 ± 0.04^{d}	267.37 ± 67.02^{c}	19.03 ± 1.88^{c}	9.28 ± 1.31^{c}	2.87 ± 0.32^{bc}	7.94 ± 0.64^{c}	49.85 ± 8.20^{c}	199.39 ± 32.80^{c}
Mertbey	0.21 ± 0.05^{b}	357.44 ± 93.37^{a}	22.08 ± 2.11^{a}	10.04 ± 1.33^{b}	3.03 ± 0.37^{a}	8.74 ± 0.78^{a}	60.43 ± 10.62^{a}	241.72 ± 42.49^{a}
Mean	0.21 ± 0.06	325.09 ± 90.56	20.59 ± 2.22	10.14 ± 1.32	2.92 ± 0.39	8.46 ± 0.79	56.67 ± 10.52	226.71 ± 42.07
F values	16.394**	29.847**	56.743**	28.753**	6.126**	30.767**	30.465**	30.465**

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

**Significant at $p < 0.01$

Table 3 Shape attributes for pumpkin seed varieties

Varieties	Sphericity (*S*, %)	Shape Index (SI)	Roundness (*R*)	Aspect ratio (AR)	Elongation (*E*)
Develi	42.28 ± 2.96^{a}	2.99 ± 0.30^{c}	0.19 ± 0.03^{a}	0.14 ± 0.02^{b}	7.14 ± 1.09^{a}
Sena H	41.09 ± 2.53^{b}	3.14 ± 0.26^{b}	0.17 ± 0.02^{b}	0.14 ± 0.02^{b}	7.31 ± 0.98^{a}
Türkmen	41.90 ± 3.22^{ab}	3.16 ± 0.38^{b}	0.18 ± 0.03^{ab}	0.15 ± 0.02^{a}	6.72 ± 0.96^{b}
Mertbey	39.73 ± 3.02^{c}	3.40 ± 0.31^{a}	0.16 ± 0.03^{c}	0.14 ± 0.02^{b}	7.35 ± 0.93^{a}
Mean	41.25 ± 3.09	3.17 ± 0.35	0.17 ± 0.03	0.14 ± 0.02	7.13 ± 1.02
F values	14.775**	28.229**	13.454**	9.212**	8.543**

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

**Significant at $p < 0.01$

values were found as 3.03 and 2.83 mm from Mertbey and Develi varieties, respectively. The greatest geometric mean diameter was projected, and surface area values were determined from Sena Hanım (D_g: 8.79 mm and *SA*: 244.63 mm^2) and Mertbey (D_g: 8.74 mm and *SA*: 241.72 mm^2).

Develi variety had the highest sphericity (42.28%) and roundness (0.19). An almost spherical seed form is indicated by roundness values close to 1. However, the lowest sphericity and the roundness were obtained Mertbey with the values of 39.73 and 0.16, respectively. The greatest shape index was obtained from Mertbey (3.40) variety while the lowest one was obtained from Develi (2.99) variety. All varieties were defined as oval, as the shape index was above 1.25. The Türkmen variety had the highest aspect ratio value as 0.15. The greatest elongation values were found from Mertbey, Sena Hanım, and Develi with the values of 7.35, 7.31, and 7.14, respectively. Türkmen had the lowest elongation with the value of 6.72. With decreasing sphericity and roundness, increasing shape index values were seen in Çetin et al. [56] who obtained similar results as well.

Complying with the results, Devi et al. [33] indicated mean length, width, and thickness values of pumpkin seeds as 16.81, 8.87, and 2.75 mm, respectively. In addition, geometric mean diameter and single seed weight attributes were found as 7.42 mm, and 0.20 g, respectively. Khodabakhshian et al. [32] investigated main shape and size attributes of pumpkin seeds at different moisture contents (4%, 8%, 14%, and 20%) and varieties of Zaria and Gaboor. Authors reported length width, thickness, diameter, and sphericity attributes changed between 14.90 and 17.55 mm, 6.91 and 8.93 mm, 3.05 and 4.95 mm, 7.18 and 9.45 mm, 0.54 and 0.53 for Zaria variety, and 15.86 and 18.96 mm, 5.17 and 7.94 mm, 2.92 and 4.69 mm, 6.38 and 9.11 mm, and 0.45 and 0.53 for Gaboor variety, respectively. Contrary to the findings, Priyadarshini et al. [34] handled seed length, width, thickness, elongation (*L/T* ratio), and single seed weight of 12 different cucumber genotypes and reported grand mean values as 11.10 mm, 4.60 mm, 2.52 mm, 4.36, and 0.28 g, respectively. Results differences may be due to the product species differences. Paksoy and Aydin [23] found length, width, thickness, geometric mean diameter, volume, sphericity, and mass of pumpkins seeds to be 18.16 mm, 9.80 mm, 2.67 mm, 7.72 mm 43.0%, 0.73 cm^3, and 0.29 g, respectively. Similar findings were also reported by Altuntaş et al. [18] for pumpkin seed length, width, thickness, geometric mean diameter, sphericity, surface area, single volume seed, and unit seed mass with the values of 19.92 mm, 11.30 mm, 3.22 mm, 9.71 mm, 60.55%, 2.54 cm^2, 0.11 cm^3, and 0.21 g, respectively. These differences were primarily attributed to varieties, climate conditions, and moisture contents [47].

Discrimination analysis

Linear discriminant analysis for physical attributes of pumpkin seed varieties is shown in Table 4. The more dependent variables the function describes, the higher the eigenvalues. In the study, eigenvalues were determined as 0.490, 0.381, and 0.042 for functions 1, 2, and 3, respectively. The effect size of the functions is explained by the square of the correlation. The first two functions explained 95.4% of the total variation as 53.7% and 41.7%, respectively. The best estimation is explained by Wilks' lambda. Wilks' lambda ideal was significant for each estimative estimator, and in the case of the current investigation, it was significant for three outcomes. The unexplained portion of the differences between the groups was determined to be 46.6% in Wilks' lambda statistics. Eight estimators' relative relevance was determined by the discriminant function coefficients. The chi-square value was found as 299.742 for functions 1–3. Geometric mean diameter and length were discovered to have the highest loadings in function 1 according to the loadings. The shape index and the sphericity in function 2 showed the most significant loadings.

Group centroids of four different varieties based on their canonical discriminant functions are displayed in Fig. 5. Differences between components, geometric mean diameter, length, shape index, and sphericity attributes were taken into account as significant discriminate attributes. For the Sena Hanım and Türkmen varieties, length, and geometric mean diameter proved the discrimination analysis in the canonical function 1. The sphericity, shape index, and roundness

Table 4 Discriminant analysis results

Eigenvalue statistics	Function 1	Function 2	Function 3
Eigenvalues	0.490	0.381	0.042
% of variance	53.7	41.7	4.6
% of cumulative variance	53.7	95.4	100.0
Canonical correlation	0.574	0.525	0.200
Significance test of functions	1–3	2–3	3
Wilks' Lambda	0.466	0.695	0.960
Chi-square	299.742	142.897	16.073
Df	24	14	6
p (sigma)	0.000**	0.000**	0.013**
Standardized canonical discriminant function coefficients	Function 1	Function 2	Function 3
Mass	0.292	−0.404	−0.571
Volume	0.730	0.299	−4.989
Length	2.696	0.407	7.094
Width	0.028	0.685	5.734
Thickness	0.008	0.013	6.850
Geo. mean. dia	−2.372	−1.330	−9.791
Sphericity	1.551	2.782	0.315
Shape index	0.255	3.104	−1.334

**Highly significant ($p < 0.01$)

attributes for the Develi and Mertbey varieties confirmed the position in the canonical function 2 axis.

Pair-wise comparison and multivariate tests

Statistics using Hotelling Trace, Pillai Trace, and Wilks' Lambda revealed that all varieties of physical attributes were significant ($p < 0.01$). Table 5 provides MANOVA, Bonferroni corrected, and Mahalanobis distance values. The percentage of variance in dependent variables was represented using Wilks' Lambda statistics, which was then explained by variations in independent variables. The Wilks' Lambda statistic, which is smaller, reveals that the differences between the groups in the study increased and varied from 0 to 1. The sum of variances, which explains the most discrimination of independent factors in dependent variables, is considered by the Pillai trace statistics, which is regarded as the most reliable among multivariate analyses. In the study, Pillai's trace, Wilks' Lambda, and Hotelling trace values were obtained with the values of 0.752, 0.405, and 1.105, respectively. Cetin et al. [20] found that variations with a Mahalanobis distance of less than 3 exhibit remarkably similar characteristics. The Develi and the Sena Hanım varieties with the smallest Mahalanobis distances shared similar characteristics. The greatest value was found in the distance between the Sena Hanım and the Türkmen varieties, and the varieties showed different characteristics. Additionally, Bonferroni corrected *p* values supported these findings.

Performance results of binary classification

Binary variety classification of pumpkin seeds was performed for variety pairs. Five machine learning techniques (RF, SVM, NB, MLP, and kNN) were used to generate classification models for size, shape, area, and mass attributes in each pair scenario. All five classifiers were able to achieve classification accuracies that were only fairly adequate in the case of the model based on the physical attributes of pumpkin seeds for Develi and Sena Hanım (Table 6). The MLP gave a high accuracy of 73.00%, while the RF had the lowest accuracy of 70.00%. These findings were also validated by the values of other performance metrics. TP rate, Precision, F-measure, PRC area, and ROC area were 0.740 and 0.690, 0.705 and 0.726, 0.722 and 0.708, 0.652 and 0.656, and 0.715 for Develi and Sena H., respectively. For Develi and Türkmen pairs, the greatest accuracy value was obtained in the MLP algorithm (72.00%). kNN algorithm had the lowest accuracy value of 65.50%. In the case of pumpkin seeds of the Develi and the Türkmen varieties, classification accuracies for both classifiers were rated as slightly lower and yet still acceptable. In the study, Develi and Mertbey pairs had the greatest classification accuracies among the variety pairs.

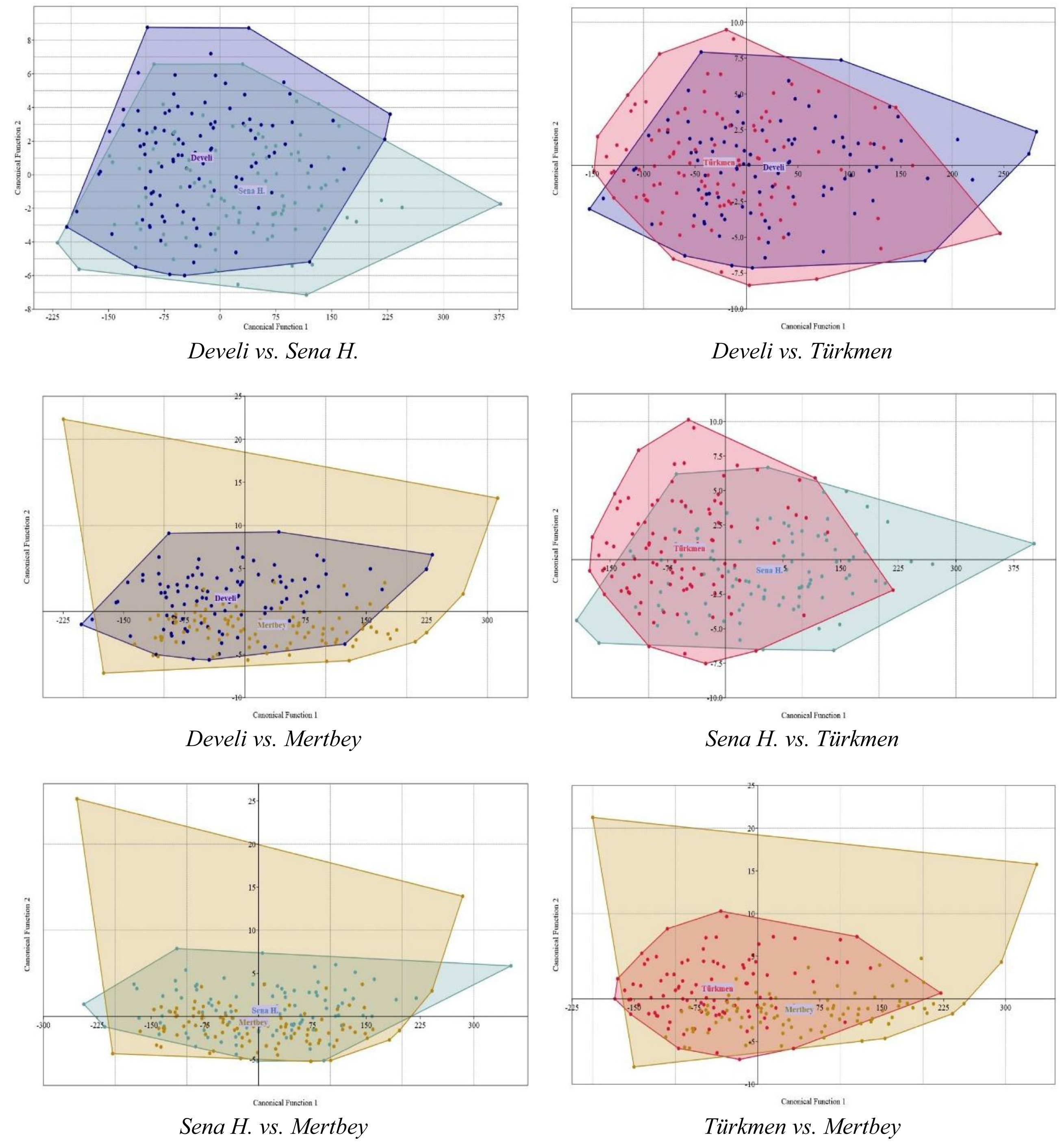


Fig. 5 Scatter plots of the pumpkin seed varieties from the point of the discriminant scores and group centroids function 1 and 2 (*Develi: Darkblue, Sena H.: Cadetblue, Türkmen: Crimson, Mertbey: Golden. **In the figure, the parts with the variety names are the group centroids.)

MLP algorithm had the greatest accuracy value of 85.00%. In the MLP algorithm with the highest accuracy, TP rate reached to 0.790 for Develi and 0.910 for Mertbey, Precision to 0.898 for Develi and 0.813 for Mertbey, *F*-measure to 0.840 for Develi and 0.858 for Mertbey, PRC Area to 0.894 for Develi and 0.896 for Mertbey, and ROC area to 0.907 for both varieties (Table 6). MLP was followed by SVM and RF with the values of 84.50% and 83.50%, respectively.

Seeds of Sena Hanım and Türkmen varieties were discriminated by five algorithms with accuracy values of between 77.50% and 84.50%. Herein, it was observed

Table 5 Differences among the pumpkin seed varieties

MANOVA results						
Effect	Statistics	Value	Hypothesis df	Error df	*F*	*p* (sigma)
Variables	Pillai's trace	0.752	36	1161	10.80	0.000**
	Wilks' Lambda	0.405	36	1138	11.33	0.000**
	Hotelling Trace	1.105	36	1151	11.78	0.000**

Hotelling's pairwise comparisons

Bonferroni corrected *p* values in upper triangle. Mahalanobis distances in lower triangle

Variables	Develi	Sena *H*	Türkmen	Mertbey
Develi		6.70E−08	5.89E−12	9.71E−20
Sena H	1.575		1.56E−21	9.43E−13
Türkmen	2.249	4.045		1.16E−20
Mertbey	3.677	2.529	3.837	

**Highly significant ($p<0.01$)

that seeds were classified with 84.50% percent accuracy in the confusion matrices despite the fact that MLP was the most successful algorithm (Table 7). The lowest accuracy (77.50%) was found in the RF algorithm. The ROC area value of the Sena Hanım and Türkmen varieties was obtained as 0.904. According to the classification performance, the next pair was Sena Hanım and Mertbey, and the highest accuracy values were observed in MLP (74.50%). In the MLP algorithm, for Sena Hanım, TP ratio, *F*-measure, Precision, PRC area, and ROC area reached the following values: 0.750, 0.743, 0.746, 0.803, and 0.776, respectively. These values were determined as 0.740, 0.747, 0.744, 0.803, and 0.804 for Mertbey, respectively (Table 7).

The pair of the Türkmen and Mertbey varieties were found to have a classification accuracy of more than 87.00%. The SVM model yielded an accuracy of 82.50% in the binary classification. The performance metrics for Türkmen and Mertbey were 0.810 and 0.840 (TP rate), 0.835 and 0.816 (Precision), 0.822 and 0.828 (*F*-measure), 0.825 (ROC area), and 0.771 and 0.765 (PRC area), respectively (Table 8).

Each pumpkin variety's separate ROC area curve was shown for all models created using all size and shape attributes (Fig. 5). The predictive model's efficacy is graphically represented by the receiver operating curve, which demonstrated that the classifier correctly classified the varieties. The MLP and the SVM algorithms produced the largest ROC area values, as was to be expected. Because the values obtained are larger, the ROC area values ensure very excellent performance for automatic identification of any understudy of the variety classification. As seen in Fig. 6, the ROC area curves, the best classified soybean variety pair, were Develi vs Mertbey, Türkmen vs Mertbey, and Sena Hanım vs Türkmen. Here, the worst classified pair was determined as Develi vs Sena Hanım, Develi vs Türkmen, and Sena Hanım vs Mertbey.

MLP and RF showed an excellent ability to classify among the variations in order to maximize the distance between groups and minimize the distance between classes. The MLP accuracy values for these varieties were very promising. In addition, it has been revealed that the SVM algorithm also comes to the fore in this study. Pumpkin seeds are very similar to each other due to their structure and physical attributes. For this reason, the fact that the accuracy values obtained are medium–high encourages future studies. Within the scope of findings, studies that are compatible and have similar or different aspects are clearly presented.

Similarly, Demir et al. [35] used the Radial Basis Neural Network (RBNN) and Propagation Neural Network (BPNN) to predict the physical attributes of the pumpkin seeds and reported RMSE values as 0.0025 and 0.6875 for RBNN and BPNN, respectively. The authors also mentioned its superiority in RBNN structure prediction and that these algorithms could be an alternative approach to the traditional methods. Koklu et al. [36] determined the physical attributes of two pumpkin seed varieties as "Ürgüp Sivrisi" and "Çerçevelik" and classified them using algorithms such as LR, MLP, SVM, RF, and k-NN, and authors indicated accuracy values of the models 87.92, 88.92, 88.64, 87.56, and 87.64, respectively. The reason why these results are higher than our findings is due to the structure of the selected varieties. So that the "Ürgüp Sivrisi" variety has a more oval shape, while the "Çerçevelik" variety has a round shape. Li et al. [53] classified pumpkin seeds by convolutional neural network and hyperspectral imaging technology. The authors indicated that PA-3DCNN had the greatest accuracy than the other algorithms with values of 99.14% and 95.24% for training and test sets, respectively. In addition, the accuracies were changed between 65.18% and 99.14% for eight different models. Prasad et al. [54] implemented

Table 6 Performance metrics and confusion matrices for Develi variety

Classifiers	Predicted class (%)		Actual class	Accuracy (%)	tp rate	Precision	*F*-measure	ROC area	PRC area
Develi vs Sena H									
RF	Develi	Sena H							
	75	25	Develi	70.00	0.740	0.679	0.708	0.748	0.719
	35	65	Sena H		0.650	0.714	0.681		0.760
SVM	Develi	Sena H							
	74	26	Develi	72.00	0.740	0.705	0.722	0.715	0.652
	30	70	Sena H		0.690	0.726	0.708		0.656
NB	Develi	Sena H							
	77	23	Develi	72.50	0.770	0.706	0.737	0.768	0.739
	32	68	Sena H		0.680	0.747	0.712		0.755
MLP	Develi	Sena H							
	76	24	Develi	73.00	0.760	0.717	0.738	0.755	0.711
	30	70	Sena H		0.700	0.745	0.722		0.723
kNN	Develi	Sena H							
	71	29	Develi	72.50	0.710	0.732	0.721	0.772	0.721
	26	74	Sena H		0.740	0.718	0.729		0.754
Develi vs Türkmen									
RF	Develi	Türkmen							
	70	30	Develi	68.00	0.700	0.673	0.686	0.706	0.659
	34	66	Türkmen		0.660	0.688	0.673		0.710
SVM	Develi	Türkmen							
	75	25	Develi	68.50	0.750	0.664	0.704	0.685	0.623
	38	62	Türkmen		0.620	0.713	0.663		0.632
NB	Develi	Türkmen							
	69	31	Develi	68.00	0.690	0.676	0.683	0.740	0.718
	33	67	Türkmen		0.670	0.684	0.677		0.735
MLP	Develi	Türkmen							
	75	25	Develi	72.00	0.750	0.708	0.728	0.737	0.713
	31	69	Türkmen		0.690	0.734	0.711		0.713
kNN	Develi	Türkmen							
	68	32	Develi	65.50	0.680	0.648	0.663	0.696	0.626
	37	63	Türkmen		0.630	0.663	0.646		0.698
Develi vs Mertbey									
RF	Develi	Mertbey							
	80	20	Develi	83.50	0.800	0.860	0.829	0.909	0.918
	13	87	Mertbey		0.870	0.813	0.841		0.876
SVM	Develi	Mertbey							
	80	20	Develi	84.50	0.800	0.879	0.838	0.845	0.803
	11	89	Mertbey		0.890	0.817	0.852		0.782
NB	Develi	Mertbey							
	82	18	Develi	80.50	0.820	0.796	0.808	0.878	0.846
	21	79	Mertbey		0.790	0.814	0.802		0.862
MLP	Develi	Mertbey							
	79	21	Develi	85.00	0.790	0.898	0.840	0.907	0.894
	9	91	Mertbey		0.910	0.813	0.858		0.896
kNN	Develi	Mertbey							
	75	25	Develi	82.00	0.750	0.872	0.806	0.889	0.858
	11	89	Mertbey		0.890	0.781	0.832		0.854

Table 7 Performance metrics and confusion matrices for Sena H. variety

Classifiers	Predicted class (%)		Actual class	Accuracy (%)	TP rate	Precision	*F*-measure	ROC area	PRC area
Sena H. vs Türkmen									
RF	Sena H	Türkmen							
	74	26	Sena H	77.50	0.740	0.796	0.767	0.836	0.862
	19	81	Türkmen		0.810	0.757	0.783		0.780
SVM	Sena H	Türkmen							
	74	26	Sena H	78.50	0.740	0.813	0.775	0.785	0.732
	17	83	Türkmen		0.830	0.761	0.794		0.717
NB	Sena H	Türkmen							
	76	24	Sena H	78.00	0.760	0.792	0.776	0.872	0.888
	20	80	Türkmen		0.800	0.769	0.784		0.888
MLP	Sena H	Türkmen							
	84	16	Sena H	84.50	0.840	0.848	0.844	0.904	0.909
	15	85	Türkmen		0.850	0.842	0.846		0.901
kNN	Sena H	Türkmen							
	79	21	Sena H	81.50	0.790	0.832	0.810	0.870	0.870
	16	84	Türkmen		0.840	0.800	0.820		0.807
Sena H. vs Mertbey									
RF	Sena H	Mertbey							
	73	27	Sena H	75.50	0.730	0.768	0.749	0.776	0.788
	22	78	Mertbey		0.780	0.743	0.761		0.788
SVM	Sena H	Mertbey							
	68	32	Sena H	71.50	0.680	0.731	0.705	0.715	0.657
	25	75	Mertbey		0.750	0.701	0.725		0.651
NB	Sena H	Mertbey							
	68	32	Sena H	70.00	0.680	0.708	0.694	0.756	0.741
	28	70	Mertbey		0.720	0.692	0.706		0.736
MLP	Sena H	Mertbey							
	75	25	Sena H	74.50	0.750	0.743	0.746	0.803	0.776
	26	74	Mertbey		0.740	0.747	0.744		0.804
kNN	Sena H	Mertbey							
	64	36	Sena H	69.00	0.640	0.711	0.674	0.736	0.699
	26	74	Mertbey		0.740	0.673	0.705		0.676

and designed machine learning models that included LR, SVM, DT, NB, ANN, and kNN for the classification of pumpkin seed varieties and obtained average accuracies of 99.81%, 52.20%, 100.00%, 52.00%, 95.80%, and 77.20%, respectively. They reported that DT had the best results and could be effectively used for the classification of pumpkin seeds. Gulzar et al. [55] proposed a system of classification of 14 different seeds (sunflower, onion, mustard, kidney beans, flax, fenugreek, black eyed peas, black pepper, chickpea, coriander, corn, cumin, fennel, and pumpkin) using machine learning and deep learning. The results showed that classification accuracy reached 99% for the test set. Since these seeds are of different types separated from each other, the results are quite high. However, the lowest results were obtained in pumpkin seeds. Liu et al. [37] applied LS-SVM, BPNN, and RF algorithms to discriminate watermelon seeds. According to spectral + morphology features for Julong variety, LS-SVM, BPNN, and RF results were found as 92%, 84%, and 87%, while these values for Xiali variety were found as 83%, 75%, and 91%, respectively. Mukasa et al. [39] classified triploid watermelon seeds from diploid and tetraploid seeds. Authors created a classification model with ML techniques by one-class classification using SVM quadratic and DD-SIMCA models. The SVM quadratic and the DD-SIMCA models yielded triploid accuracies of 84.3% and 69.5%, respectively. Ahmed et al. [38] evaluated deep learning and conventional machine learning methods for the classification of watermelon seeds by morphological patterns. The authors indicated accuracy values of 87.3% and 83.6% for ResNet-50 and LDA algorithms, respectively. The findings showed that classification based on

Table 8 Performance metrics and confusion matrices for the Türkmen and Mertbey varieties

Classifiers	Predicted class (%)		Actual class	Accuracy (%)	TP rate	Precision	*F*-measure	ROC area	Prc area
Türkmen vs Mertbey									
RF	Türkmen	Mertbey							
	75	25	Türkmen	79.00	0.750	0.815	0.781	0.868	0.870
	17	83	Mertbey		0.830	0.769	0.798		0.861
SVM	Türkmen	Mertbey							
	81	19	Türkmen	82.50	0.810	0.835	0.822	0.825	0.771
	16	84	Mertbey		0.840	0.816	0.828		0.765
NB	Türkmen	Mertbey							
	83	17	Türkmen	80.00	0.830	0.783	0.806	0.880	0.849
	23	77	Mertbey		0.770	0.819	0.794		0.849
MLP	Türkmen	Mertbey							
	77	23	Türkmen	80.50	0.770	0.828	0.798	0.888	0.884
	16	84	Mertbey		0.840	0.785	0.812		0.884
kNN	Türkmen	Mertbey							
	72	28	Türkmen	76.00	0.720	0.783	0.750	0.845	0.846
	20	80	Mertbey		0.800	0.741	0.769		0.805

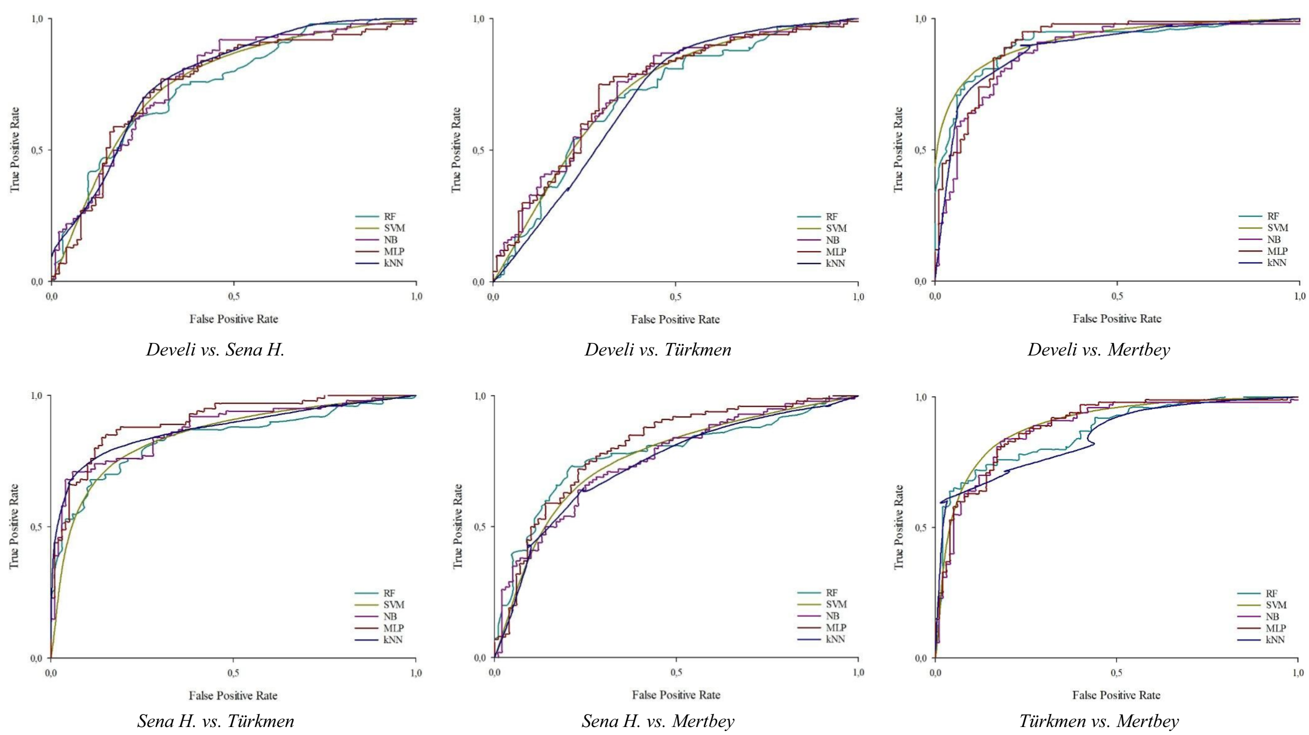


Fig. 6 ROC curves of classified pairs based on selected physical attributes

physical attributes could be performed using machine learning algorithms. The attributes and the algorithms studied have proven their usability by giving successful results in many similar studies.

Conclusion

The effectiveness of machine learning was demonstrated to discriminate pumpkin seeds in terms of physical characteristics. For classification models, the data were prepared through a series of preprocessing and then datasets and models were created with selected features (mass, length, geometric mean diameter, and shape index). Using these datasets, MLP and SVM from machine learning algorithms became the most successful models. In addition, the varieties with the highest accuracy values were Develi and Mertbey, while the less-accuracy values were Develi and Turkmen. The practical importance of the study is the classification of seeds with very similar characteristics correctly and quickly using the machine learning technique. In addition, accurately classifying pumpkin seeds that meet specific criteria is crucial for food and agricultural industries. Based on the present findings, a new approach could be suggested as a valuable control tool in development of planters for the agricultural machinery, breeding research, and the seed industry. In this study, we encountered some limitations and had suggestions for future research. One limitation was the time-consuming process of measuring shape, size, and mass attributes. To overcome this, we recommend using modern techniques like image processing with affordable, yet effective hardware, such as webcams, action cameras, or mobile phone cameras. Furthermore, future studies can expand by incorporating more data sets, attributes, and algorithms.

Author contributions NÇ: conceptualization, methodology, software, formal analysis, visualization, computation, and writing; ER: data curation and writing; SF: sources and investigation; ŞÜ: sources and formal analysis; PS: formal analysis and investigation; SG: investigation and writing; AÜ: conceptualization, investigation, and writing.

Funding No funding was received for this study.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing interest.

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

Informed consent Not applicable.

References

1. Montesano D, Rocchetti G, Putnik P, Lucini L (2018) Bioactive profile of pumpkin: an overview on terpenoids and their health-promoting properties. Food Sci Curr Opin 22:81–87. https://doi.org/10.1016/j.cofs.2018.02.003
2. FAO (2023) Production/yield quantities of pumpkins, squash and gourds in World + (Total). https://www.fao.org/faostat/en/#data/QCL/visualize. Accessed 24 April 2023
3. Hagos M, Yaya EE, Chandravanshi BS, Redi-Abshiro M (2023) Determination of fatty acids composition by GC-MS and physicochemical parameters of pumpkin (*Cucurbita maxima*) seed oil cultivated in Ethiopia. Bull Chem Soc Ethiop 37(3):565–577. https://doi.org/10.4314/bcse.v37i3.3
4. Setiawan B, Aulia SS, Sinaga T, Sulaeman A (2021) Nutritional content and characteristics of pumpkin cream soup with tempeh addition as supplementary food for elderly. Int J Food Sci 8:2021. https://doi.org/10.1155/2021/6976357
5. Igbozulike AO, Amamgbo N (2019) Effect of moisture content on physical properties of fluted pumpkin seeds. J Biosyst Eng 44:69–76. https://doi.org/10.1007/s42853-019-00015-z
6. Ponka R, Bouba AA, Fokou E, Tambe ST, Beaucher E, Piot M, Leonil J, Gaucheron F (2015) Protein, mineral and amino acid content of some Cameroonian traditional dishes prepared from pumpkin (*Cucurbita maxima* Duch.). J Food Compost Anal 43:169–174. https://doi.org/10.1016/j.jfca.2015.06.009
7. Liubych V, Novikov V, Pushka O, Pushka I, Cherchel V, Kyrpa M, Kolibabchuk T, Kirian V, Moskalets T (2023) Development of wheat bread recipe with pumpkin paste. East Eur J Enterp Technol 11:121. https://doi.org/10.15587/1729-4061.2023.274259
8. Hussain A, Kausar T, Sehar S, Sarwar A, Ashraf AH, Jamil MA, Noreen S, Rafique A, Iftikhar K, Aslam J, Quddoos MY, Majeed MA, Zerlash M (2022) Utilization of pumpkin, pumpkin powders, extracts, isolates, purified bioactives and pumpkin based functional food products; a key strategy to improve health in current post COVID 19 period; an updated review. Appl Food Res 2(2):100241. https://doi.org/10.1016/j.afres.2022.100241
9. Abdelnour SA, Metwally MG, Bahgat LB, Naiel MA (2023) Pumpkin seed oil–supplemented diets promoted the growth productivity, antioxidative capacity, and immune response in heat-stressed growing rabbits. Trop Anim Health Prod 55(1):55. https://doi.org/10.1007/s11250-023-03460-3
10. Joshi N, Chauhan K (2022) Nutritional properties of pumpkin seeds: emerging seed for a therapeutic benefit. World J Pharm Res 11(9):782–787. https://doi.org/10.26717/BJSTR.2019.21.003586
11. Ambi AA, Bashir M, Hafsat BM, Nura M, Abubakar H, Shatu M (2021) Effect of processing on vitamin A and some mineral contents of pumpkin (*Cucurbita maxima*) seed. Bayero J Pure Appl Sci 13(2):113–116. https://doi.org/10.4314/bajopas.v13i2.15
12. Chuwa C, Dhiman AK (2023) Nutrition and health benefits of ripe pumpkin fruit, pulp and powder. Recent Res Sci Technol 4:123–133. https://doi.org/10.9734/bpi/rpst/v4/9411F
13. Tau S, Miller TW (2023) The role of cancer cell bioenergetics in dormancy and drug resistance. Cancer Metastasis Rev 42(1):87–98. https://doi.org/10.1007/s10555-023-10081-7
14. Cansino S, Torres-Trejo F, Estrada-Manilla C, Flores-Mendoza A, Ramírez-Pérez G, Ruiz-Velasco S (2023) Nutrient effects on working memory across the adult lifespan. Nutr Neurosci 26(5):456–469. https://doi.org/10.1080/1028415x.2022.2055378
15. Rezig L, Harzalli Z, Gharsallah K, Mahfoudhi N, Chouaibi M, Majdoub H, Oueslati I (2022) Microwave and roasting impact on pumpkin seed oil and its application in full-fat mayonnaise formula. Foods 11(18):2732. https://doi.org/10.3390/foods11182732
16. Yu M, Peng M, Chen R, Chen J (2023) Effect of thermal pretreatment on the physiochemical properties and stability of pumpkin seed milk. Foods 12(5):1056. https://doi.org/10.3390/foods12051056
17. Syed QA, Akram M, Shukat R (2019) Nutritional and therapeutic importance of the pumpkin seeds. Seed 21(2):15798–15803. https://doi.org/10.26717/BJSTR.2019.21.003586
18. Altuntaş E (2008) Some physical properties of pumpkin (*Cucurbita pepo* L.) and watermelon *Citrullus lanatus* L. seeds. Int J

Agric Sci 14(1):62–69. https://doi.org/10.1501/Tarimbil_0000000519
19. Jahanbakhshi A, Kheiralipour K (2020) Evaluation of image processing technique and discriminant analysis methods in postharvest processing of carrot fruit. Food Sci Nutr 8(7):3346–3352. https://doi.org/10.1002/fsn3.1614
20. Cetin N, Yaman M, Karaman K, Demir B (2021) Determination of some physicomechanical and biochemical parameters of hazelnut (*Corylus avellana* L.) cultivars. Turkish J Agric Forest 44(5):439–450. https://doi.org/10.3906/tar-1905-115
21. Liu X, Le Bourvellec C, Yu J, Zhao L, Wang K, Tao Y, Renard CMGC, Hu Z (2022) Trends and challenges on fruit and vegetable processing: insights into sustainable, traceable, precise, healthy, intelligent, personalized and local innovative food products. Trends Food Sci Technol 125:12–25. https://doi.org/10.1016/j.tifs.2022.04.016
22. Saiedirad MH, Tabatabaeefar A, Borghei A, Mirsalehi M, Badii F, Varnamkhasti MG (2008) Effects of moisture content, seed size, loading rate and seed orientation on force and energy required for fracturing cumin seed (*Cuminum cyminum* Linn.) under quasi-static loading. J Food Eng 86(4):565–572. https://doi.org/10.1016/j.jfoodeng.2007.11.021
23. Paksoy M, Aydin C (2004) Some physical properties of edible squash (*Cucurbita pepo* L.) seeds. J Food Eng 65(2):225–231. https://doi.org/10.1016/j.jfoodeng.2004.01.019
24. Kumar S, Vishwakarma SSR, Singh BR (2019) Evaluating the effects of some relevant parameters on physico-mechanical and aerodynamic properties of sunflower seed. New Agr 30(2):1–8
25. Ali A, Qadri S, Mashwani WK, Brahim Belhaouari S, Naeem S, Rafique S, Jamal F, Chesneau C, Anam S (2020) Machine learning approach for the classification of corn seed using hybrid features. Int J Food Prop 23(1):1110–1124. https://doi.org/10.1080/10942912.2020.1778724
26. Çetin N (2022) Machine learning for varietal binary classification of soybean (*Glycine max* (L.) Merrill) seeds based on shape and size attributes. Food Anal Methods 15(8):2260–2273. https://doi.org/10.1007/s12161-022-02286-3
27. de Oliveira DEC, Sousa Bueno SG, Guimarães Carvalho JM, da Costa Júnior JR, Resende O, Soares Costa V (2022) Mechanical properties of pumpkin seeds. Commun Sci 13:e3586. https://doi.org/10.14295/CS.v13.3586
28. Verma S, Sharma R, Deb S, Maitra D (2021) Artificial intelligence in marketing: systematic review and future research direction. Int J Inf Manage 1(1):100002. https://doi.org/10.1016/j.jjimei.2020.100002
29. Srunitha K, Padmavathi S (2016) Performance of SVM classifier for image based soil classification. In: 2016 International conference on signal processing, communication, power and embedded system (SCOPES). IEEE, pp 411–415
30. Okwu MO, Tartibu LK, Okwu MO, Tartibu LK (2021) Artificial neural network metaheuristic optimization: nature-inspired algorithms swarm and computational intelligence, theory and applications. Springer Nat 927:133–145
31. Azizi A, Gilandeh YA, Mesri-Gundoshmian T, Saleh-Bigdeli AA, Moghaddam HA (2020) Classification of soil aggregates: a novel approach based on deep learning. Soil Tillage Res 199:104586. https://doi.org/10.1016/j.still.2020.104586
32. Khodabakhshian Kargar R (2012) Mechanical strength and physical behavior of pumpkin seed and its Kernel. Thai J Vet Med 45(1):37–43
33. Devi NM, Prasad R, Palmei G (2018) Physico-chemical characterisation of pumpkin seeds. Int J Chem Stud 6(5):828–831
34. Priyadarshini M, Das S, Muduli KC, Mohanty S, Sahoo S, Pradhan BR (2021) Characterisation of cucumber genotypes through seed morphological characters. J Pharmacogn Phytochem 10(1):2158–2161
35. Demir B, Eski I, Kuş ZA, Ercisli S (2017) Prediction of physical parameters of pumpkin seeds using neural network. Not Bot Horti Agrobo 45(1):22–27. https://doi.org/10.15835/nbha45110429
36. Koklu M, Sarigil S, Ozbek O (2021) The use of machine learning methods in classification of pumpkin seeds (*Cucurbita pepo* L.). Genet Resour Crop Evol 68(7):2713–2726. https://doi.org/10.1007/s10722-021-01226-0
37. Liu W, Xu X, Liu C, Zheng L (2019) Rapid discrimination of high-quality watermelon seeds by multispectral imaging combined with chemometric methods. J Appl Spectrosc 85:1044–1049. https://doi.org/10.1007/s10812-019-00757-w
38. Ahmed MR, Yasmin J, Park E, Kim G, Kim MS, Wakholi C, Mo C, Cho B-K (2020) Classification of watermelon seeds using morphological patterns of x-ray imaging: a comparison of conventional machine learning and deep learning. Sensors 20(23):6753. https://doi.org/10.3390/s20236753
39. Mukasa P, Wakholi C, Faqeerzada MA, Amanah HZ, Kim H, Joshi R, Suh HK, Kim G, Lee H, Kim MS, Baek I, Cho BK (2022) Nondestructive discrimination of seedless from seeded watermelon seeds by using multivariate and deep learning image analysis. Comput Electron Agric 194:106799. https://doi.org/10.1016/j.compag.2022.106799
40. Hammer O, Harper DA (2001) Past: paleontological statistics software package for educaton and data anlysis. Palaeontol Electron 4(1):9–10
41. IBM SPSS® (2010) Statistical software SSS Inc. IBM Company©, Version 200
42. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. ACM SIGKDD Explor Newslett 11(1):10–18
43. Stegmayer G, Milone DH, Garran S, Burdyn L (2013) Automatic recognition of quarantine citrus diseases. Exp Syst App 40(9):3512–3517. https://doi.org/10.1016/j.eswa.2012.12.059
44. Ropelewska E, Szwejda-Grzybowska J (2021) A comparative analysis of the discrimination of pepper (*Capsicum annuum* L.) based on the cross-section and seed textures determined using image processing. J Food Process Eng 44(6):e13694. https://doi.org/10.1111/jfpe.13694
45. Pietersma D, Lacroix R, Lefebvre D, Wade KM (2003) Performance analysis for machine-learning experiments using small data sets. Comput Electron Agric. https://doi.org/10.1016/S0168-1699(02)00104-7
46. Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 30(7):1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2
47. Taskin O (2020) Evaluation of freeze drying for whole, half cut and puree black chokeberry (*Aronia melanocarpa* L.). Heat Mass Transf 56:2503–2513. https://doi.org/10.1007/s00231-020-02867-0
48. Ozkan G, Koyuncu MA (2005) Physical and chemical composition of some walnut (*Juglans regia* L) genotypes grown in Turkey. Grasas Aceites 56(2):141–146. https://doi.org/10.3989/gya.2005.v56.i2.122
49. Mohsenin NN (1986) Physical properties of plant and animal materials: structure, physical characteristics and mechanical properties. Gordon and Breach Science Publishers, New York
50. Sayinci B, Ercisli S, Akbulut M, Şavşatli Y, Baykal H (2015) Determination of shape in fruits of cherry laurel (*Prunus laurocerasus*) accessions by using Elliptic Fourier analysis. Acta Sci Pol Hortorum Cultus 14(1):63–82
51. Omobuwajo TO, Akande EA, Sanni LA (1999) Selected physical, mechanical and aerodynamic properties of African breadfruit

(*Treculia africana*) seeds. J Food Eng 40(4):241–244. https://doi.org/10.1016/S0260-8774(99)00060-6

52. Fıratlıgil-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
53. Li X, Feng X, Fang H, Yang N, Yang G, Yu Z, Shen J, Geng W, He Y (2023) Classification of multi-year and multi-variety pumpkin seeds using hyperspectral imaging technology and three-dimensional convolutional neural network. Plant Methods 19(1):1–18. https://doi.org/10.1186/s13007-023-01057-3
54. Prasad KV, Hanumesh V, Kumar Swamy K, Renuka S (2023) Pumpkin seeds classification: artificial neural network and machine learning methods. J Int Acad Phys Sci 27(1):23–33
55. Gulzar Y, Hamid Y, Soomro AB, Alwan AA, Journaux L (2020) A convolution neural network-based seed classification system. Symmetry 12(12):1–18. https://doi.org/10.3390/sym12122018
56. 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 Meth 14(8):1666–1681. https://doi.org/10.1007/s12161-021-02002-7