Spanish Journal of Agricultural Research
19 (4), e0211, 11 pages (2021)
eISSN: 2171-9292
https://doi.org/10.5424/sjar/2021194-18056
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)

RESEARCH ARTICLE

OPEN ACCESS

Classification of hazelnut varieties by using artificial neural network and discriminant analysis

Omer Keles and Alper Taner
Ondokuz Mayıs University, Faculty of Agriculture, Department of Agricultural Machinery, Samsun, Turkey.

Abstract

Aim of study: This study was conducted to classify hazelnut (*Corylus avellana* L.) varieties by using artificial neural network and discriminant analysis.

Area of study: Samsun Province, Turkey.

Material and methods: The physical, mechanical and optical properties of 11 hazelnut varieties were determined for three major axes. The parameters of physical, mechanical and optical properties were included as independent variables, while hazelnut varieties were included as dependent variables. Models were created for each of the three axes to classify hazelnut varieties.

Main results: Classification success rates with Artificial Neural Networks (ANN) and Discriminant Analysis (DA) were found as 89.1% and 92.7% for X axis, as 92.7% and 92.7% for Y axis and as 86.8% and 88.7% for Z axis, respectively. The classification results of ANN and DA models were found to be very close to each other. Both models can be used in the classification of hazelnut varieties.

Research highlights: The results obtained for the identification and classification of hazelnut varieties show the feasibility and effectiveness of the proposed models.

Additional key words: *Corylus avellana*; artificial intelligence; multivariate statistical methods.

Authors' contributions: AT analyzed the data and wrote the paper. Both authors designed and conducted the experiments, read and approved the final article.

Citation: Keles, O; Taner, A (2021). Classification of hazelnut varieties by using artificial neural network and discriminant analysis. Spanish Journal of Agricultural Research, Volume 19, Issue 4, e0211. https://doi.org/10.5424/sjar/2021194-18056.

Received: 04 Mar 2021. **Accepted:** 23 Nov 2021.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Correspondence should be addressed to Alper Taner: alper.taner@omu.edu.tr

Introduction

Hazelnut (*Corylus avellana* L.) is a very healthy nutrient for humans and animals since it contains fatty acids, vitamins, proteins and minerals. It has a considerable economic contribution to the agriculture sector in Turkey (Alasavar *et al.*, 2003). Hazelnut is one of the most important raw materials for the pastry and chocolate industry (Fallico *et al.*, 2003). It is widely used to add distinct and favorable flavour and texture to pastries, confectioneries, cornflakes, dairy products and sauces (Parcerisa *et al.*, 1998).

The production area and amount of hazelnut were 966,196 ha and 863,888 tons in the world, respectively. In the same year, 515,000 tons of hazelnut were produced in Turkey in the total area of 728,381 ha (FAOSTAT, 2018). Turkey is ranked as the first in the world with approximately 70% of the hazelnut production.

The physical properties of agricultural materials are important parameters used to design agricultural and food processing machines such as crop processing, handling, sieving, storing and drying machines and equipment. In addition, their physical properties, mechanical properties and optical properties are used for classification of varieties (Mohsenin, 1970; Tabatabaeefar, 2003). Therefore, the physical properties of agricultural materials must be determined accurately and the correlations among them must be mathematically modelled.

Classification of hazelnut varieties is very important for the development of agricultural industry and profitable farming of hazelnut. Varieties are extremely important factors in terms of productivity and processing hazelnut. Each variety has its specific physical and mechanical properties. The varieties can be distinguished from each other based on their physical properties. In this way, variety standardization can also be achieved.

It is very important for producers, industrialists, traders and consumers to know the varieties of that agricultural products they are interested in. Producers want to know the variety of the agricultural product they grow to make a correct breeding; industrialists and dealers want to know the variety of the agricultural product they trade and process to set standards; and consumers want to know the variety of the agricultural product they buy. For these reasons, reliable methods are needed to identify varieties (Chen *et al.*, 2010; Pourreza *et al.*, 2012).

The recent increase of commercial interest in hazelnut has caused the development of new analytical methods for the standardization and traceability of its varieties. The insufficiency of univariate statistical methods for examining possible associations among more than two features creates a problem. This problem occurs because all features may affect each other. This situation causes complex associations among those features. Therefore, the application of multivariate statistical methods has gained value (Oddone *et al.*, 2009). Multivariate statistical methods have been developed to explain multiple associations (Özdamar, 2004). Artificial neural networks and discriminant analysis are also used in classification studies (Yang *et al.*, 2003).

Artificial neural networks (ANN) is one of the most important artificial intelligence methods. ANN is frequently used in biological practices for classification. ANN is extremely efficient and successful in studies with non-linear data. For this reason, ANN has a very important potential in the classification of agricultural products (Visen *et al.*, 2002; Dubey *et al.*, 2006; Guiné *et al.*, 2015).

Discriminant analysis is a multivariate statistical method that determines the relationships between categorical dependent variables and metric independent variables aiming to group objects into two or more categories. Discriminant analysis finds the linear combinations of independent variables with the largest difference between group means. In the analysis, the groups are determined in advance as special research groups. A set of distinguishing variables related to the characteristics that measure the expected difference in the groups are chosen to distinguish among the groups. Linear combinations are made from these variables. Then, new observations are classified into the groups by using linear combinations (Alpar, 2017).

Image processing technique is one of the technologies used for classification. Image processing technique consists of the analysis of images taken with a camera or a scanner and transferred to computers by using image processing programs (Demirbaş & Dursun, 2007). Image processing is frequently used in classification, sorting, quality control and automation processes of agricultural products such as fruit, vegetable and cereals. Thus, productivity increases and production costs decrease. In addition, higher quality and more healthy products are provided to consumers (Chen *et al.*, 2010).

The study area of our work is defined as the identification of fruit classification tasks to determine class according to the specific type. In addition, the study is an example of practical application to be used in different samples by presenting approaches to researchers. This study aims to classify hazelnut varieties with ANNs and discriminant analysis using physical, mechanical, and optical properties.

Material and methods

'İncekara', 'Kalınkara', 'Kan', 'Kuş', 'Okay28', 'Palaz', 'Sivri', 'Tombul', 'Uzunmusa', 'Yassı Badem', and 'Yuvarlak Badem' hazelnut varieties were used in the study.

Biological material test device (Lloyd Instrument LRX Plus, Lloyd Instruments Ltd, An AMATEK Company and NEXYGEN Plus software) was used to find out hazelnut shells' mechanical properties of under compression load. In order to measure the mechanical properties of hazelnut shells, the device was equipped with a load cell of 500 N and the load cell had a measurement accuracy of 0.5% (Fig. 1).

Colour parameters (L, a, b) were measured by using colorimeter (Minolta, model CR-400). A camera with a resolution of 12 mega pixels was used to take the pictures.

Adobe Photoshop and MATLAB program were used for image processing. MATLAB was used for modelling with artificial neural networks and SPSS 21.0 program was used for the statistical evaluation of discriminant analysis and the data.

In the study, physical, mechanical and optical properties of hazelnut varieties were determined for each of three major axes (X, Y, Z) (Fig. 2). Physical properties include geometric mean diameter, sphericity, grain volume, surface area, shell thickness and grain weight; mechanical

Figure 1. Biological material test device

properties include rupture force, rupture energy and deformation, optical properties include colour parameters (L, a, b) and image processing parameters.

In order to determine the physical, mechanical and optical properties, 50 hazelnuts randomly were selected for each variety to determine their physical properties. The length (L), width (W) and thickness (T) and shell thickness values of each hazelnut were measured. Grain weights were measured according to Aydın (2002). Measurements were taken at 8% humidity.

The physical properties of hazelnut cultivars used in the study are presented (Table 1). The following equations were used to calculate physical properties include geometric mean diameter (D_g), sphericity (Ø), surface area (S) and grain volume (V) of hazelnut (Mohsenin, 1970; Jain & Bal, 1997; Taner *et al.*, 2018).

$$D_g = (LWT)^{1/3} \tag{1}$$

$$\emptyset = \frac{(LWT)^{1/3}}{L} \tag{2}$$

$$S = \left(\frac{\pi B L^2}{2L - B}\right) \tag{3}$$

$$V = \left(\frac{\pi B^2 L^2}{6(2L - B)}\right) \tag{4}$$

$$B = (WT)^{1/2} \tag{5}$$

Mechanical properties were found by performing rupture tests on three major axes with biological material test device. Rupture force and deformation were determined from the obtained force-deformation graph, while rupture energy was determined from the area under this graph. The mechanical properties of hazelnut cultivars used in the study are presented (Table 2).

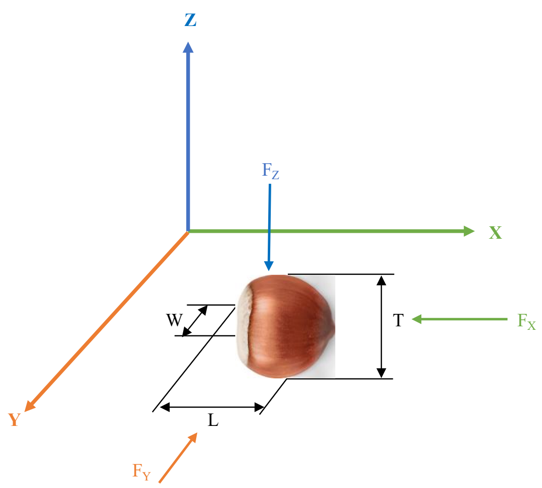

Figure 2. Three major axes and three major perpendicular dimensions of hazelnut

Colour parameters (L, a, b) were measured with a colorimeter. "L" value means white when it is 100 and black when it is 0; "a" value means red when it is positive and green when it is negative; and "b" value means yellow when it is positive and blue when it is negative (Francisco *et al.*, 2021). The optical properties of hazelnut cultivars used in the study are presented (Table 2).

Twelve independent and 11 dependent variables were used in the study. Independent variables were geometric mean diameter, sphericity, grain volume, shell thickness, rupture force, rupture energy, deformation, L, a, b and image processing parameters. Dependent variables were the 11 hazelnut varieties.

A total of six models were developed with independent variables: ANN_X, ANN_Y and ANN_Z for ANN, and DA_X, DA_Y and DA_Z models for discriminant analysis.

Four hundred and eighty-one data were used for the X axis, while 476 data were used for the Y axis and 472 data were used for the Z axis. The data used in the study were normalized between 0 and 1 (Purushothaman & Srinivasa, 1994); 80% of the data were used for training, while 10% were used for test and 10% were used for validation to develop ANN and discriminant models.

Feed Forward Back Propagation network structure was used for ANN models (Jacobs, 1988; Minai & Williams, 1990). Levenberg-Marquardt and Scaled Conjugate Gradient algorithms were used as training algorithms (Levenberg, 1944; Marquardt, 1963). As a transfer function, Hyperbolic Sigmoid Tangent function was used in the hidden layer, while softmax function was used in the output layer (Bishop, 1995). Termination cycle limit of the training algorithm was chosen as 1000. Verification was conducted in all trainings in the study. Training was terminated successfully when six verification errors stayed sequentially below the chosen error value or one verification error was equal to 0.

In order to find out the optimum number of neurons, the number of neurons in the hidden layer was increased sequentially in selected amounts and their performance was evaluated (Taner *et al.*, 2010). As a result of these tests, the network with the lowest error was chosen. For the single layer network, firstly, the number of neurons was increased by 10 from 10 to 100 and the networks were compared. Secondly, ten more trials were conducted by increasing or decreasing the number of neurons by one ranging from five more or five less neurons for the network that passed the test with the least error during the initial training.

In the discriminant analysis, univariate normal distribution analysis and then multivariate normal distribution analysis were conducted for each variable by using Mahalanobis distance (Sharma, 1996; Alpar, 2017). Equality of variance-covariance matrices was tested with Box-M test. In order to test if there was a multi linear connection problem, correlation matrix consisting of correlation coefficients between variables was calculated (Tabachnick & Fidell, 2007). Discriminant function is a

Table 1. Physical properties of the hazelnut cultivars used in the study

Variety	Axes	Geometric mean diameter (mm)	Sphericity (%)	Grain volume (mm^3)	Surface area (mm^2)	Shell thickness (mm)	Grain weight (g)
İncekara	X	17.99±0.75	0.87±0.03	2592±335	921±78	0.85±0.12	2.09±0.40
	Y	18.32±0.74	0.84±0.03	2622±345	935±80	0.85±0.10	2.21±0.31
	Z	17.79±0.65	0.83±0.04	2397±295	882±69	0.84±0.09	1.98±0.35
Kalınkara	X	17.46±0.81	0.88±0.04	2408±366	875±86	0.92±0.12	1.92±0.40
	Y	17.48±0.64	0.83±0.02	2266±261	850±64	0.87±0.09	1.94±0.37
	Z	17.32±0.73	0.85±0.03	2261±314	844±76	0.93±0.09	1.86±0.30
Kan	X	16.82±0.62	0.97±0.06	2436±366	875±87	0.83±0.13	1.81±0.25
	Y	16.76±0.52	0.93±0.04	2253±229	833±55	0.81±0.11	1.74±0.22
	Z	16.63±0.61	0.95±0.05	2266±313	835±75	0.88±0.11	1.73±0.27
Kuş	X	17.56±0.80	0.87±0.03	2428±346	881±83	0.91±0.11	2.26±0.34
	Y	17.40±0.60	0.85±0.02	2270±245	848±60	0.86±0.09	2.10±0.26
	Z	17.48±0.59	0.85±0.03	2302±260	856±62	0.94±0.13	2.12±0.25
Okay28	X	19.87±0.72	0.98±0.03	4046±462	1227±94	0.89±0.08	2.75±0.34
	Y	19.61±0.65	0.98±0.04	3871±454	1192±94	0.90±0.12	2.58±0.34
	Z	19.67±0.77	0.98±0.04	3928±518	1203±105	0.97±0.13	2.64±0.33
Palaz	X	16.45±0.72	1.05±0.04	2575±371	909±89	1.04±0.22	2.04±0.35
	Y	17.52±0.65	1.03±0.03	2962±362	997±82	0.89±0.08	2.13±0.30
	Z	16.85±0.66	1.00±0.04	2539±349	899±82	0.96±0.12	1.85±0.32
Sivri	X	16.05±0.88	0.81±0.03	1727±308	711±81	0.86±0.12	1.70±0.30
	Y	16.60±0.72	0.78±0.03	1844±253	750±67	0.84±0.08	1.79±0.27
	Z	17.11±0.64	0.79±0.03	2024±265	798±65	1.00±0.13	1.97±0.23
Tombul	X	16.48±0.94	0.94±0.06	2194±396	817±99	0.89±0.13	1.67±0.33
	Y	15.98±0.84	0.92±0.04	1925±318	750±83	0.83±0.11	1.71±0.33
	Z	16.35±0.61	0.94±0.05	2129±314	801±77	0.89±0.10	1.69±0.28
Uzunmusa	X	17.10±0.96	0.96±0.04	2539±475	898±111	0.73±0.12	1.82±0.40
	Y	16.89±0.82	0.93±0.04	2320±401	848±96	0.71±0.11	1.69±0.31
	Z	17.09±0.82	0.93±0.04	2390±399	866±95	0.74±0.09	1.71±0.30
Yassı Badem	X	17.14±0.70	0.68±0.02	1855±235	780±64	1.15±0.13	2.28±0.35
	Y	16.79±0.81	0.68±0.03	1732±273	748±73	1.11±0.16	2.17±0.35
	Z	16.96±0.64	0.68±0.03	1783±223	763±59	1.25±0.10	2.25±0.30
Yuvarlak Badem	X	16.53±0.79	0.69±0.04	1684±273	727±72	0.76±0.10	1.78±0.34
	Y	17.10±0.72	0.68±0.03	1831±262	776±67	0.82±0.10	2.05±0.30
	Z	16.77±0.81	0.68±0.03	1729±261	746±71	0.84±0.08	1.92±0.31

linear combination of independent variables and it was expressed as the following equation:

$$y = a + \sum b_i x_i \tag{6}$$

Here, y is the discriminant index, a is the constant, b is the discriminant coefficient and x is the independent variables (Tabachnick & Fidell, 2007). Eigenvalues, canonical correlation and Wilks' Lambda values were calculated to evaluate discriminant functions (Kalaycı, 2018). Image processing was carried out by capturing the image, filtering, converting the image to grey image, thresholding and determining the dimensions (Gonzalez & Woods, 2008; Ağın & Taner, 2015) (Fig. 3).

Results and discussion

Classification with discriminant analysis

In discriminant analyses which were conducted separately for the three major axes, the existences of the variables

Table 2. Mechanical and optical properties of the hazelnut cultivars used in the study

Variety	Axes	Mechanical properties			Optical properties			
		Rupture force (N)	Deformation (mm)	Rupture energy (J)	L	a	b	Image processing
İncekara	X	177.59±68.11	1.04±0.31	0.10±0.06	59.40±2.63	13.07±1.69	23.35±3.81	95379±10715
	Y	169.86±51.08	1.39±0.33	0.13±0.07	58.67±3.07	13.42±2.15	23.19±3.40	98672±12703
	Z	140.30±38.90	1.09±0.25	0.08±0.04	59.52±3.09	13.57±2.10	24.17±3.50	95829±12441
Kalınkara	X	185.80±42.42	0.99±0.21	0.10±0.04	63.57±2.54	11.92±1.70	25.05±3.31	93932±12935
	Y	222.67±46.07	1.65±0.28	0.19±0.06	63.48±2.19	11.40±1.47	23.04±2.87	87103±9280
	Z	169.69±44.89	1.09±0.26	0.10±0.05	62.69±3.06	12.24±1.54	24.85±3.12	86972±13737
Kan	X	178.23±45.47	0.99±0.25	0.09±0.04	59.21±2.29	15.11±1.69	26.06±3.18	94077±9039
	Y	127.32±28.77	1.10±0.18	0.08±0.03	57.09±2.78	16.10±1.89	26.76±4.03	91407±9035
	Z	135.54±50.55	0.88±0.19	0.06±0.04	58.88±3.25	15.13±2.03	25.80±3.16	90695±11310
Kuş	X	234.71±58.53	1.00±0.21	0.11±0.05	65.71±2.86	13.66±1.50	36.34±3.72	95568±10417
	Y	165.64±41.77	1.22±0.23	0.11±0.05	66.48±2.88	13.84±1.54	37.81±2.54	87688±8886
	Z	139.94±34.42	0.78±0.18	0.06±0.03	66.22±2.73	13.55±1.35	36.21±2.62	90225±7645
Okay28	X	189.97±66.22	0.89±0.31	0.10±0.06	60.38±1.92	14.92±1.63	32.36±5.00	133121±10019
	Y	160.97±46.92	1.08±0.29	0.10±0.05	61.59±2.03	14.88±1.37	32.98±5.06	129866±10353
	Z	170.35±48.96	0.94±0.27	0.09±0.05	61.31±1.78	14.57±1.15	31.33±4.50	131254±11317
Palaz	X	164.97±47.39	0.96±0.33	0.08±0.04	60.09±1.48	15.02±1.22	29.10±3.23	108636±10025
	Y	142.67±33.17	0.93±0.21	0.07±0.03	61.01±1.64	14.62±1.37	29.20±3.05	108206±8611
	Z	165.73±42.10	0.92±0.25	0.08±0.04	61.13±1.82	14.64±1.31	29.11±3.26	102093±9733
Sivri	X	185.17±58.94	0.99±0.28	0.10±0.06	61.73±2.36	15.23±1.08	35.29±2.88	74257±9527
	Y	173.78±39.60	1.21±0.22	0.11±0.04	60.63±2.53	15.36±1.40	34.35±3.26	75058±8452
	Z	186.19±56.62	0.91±0.22	0.09±0.05	60.19±2.37	15.52±1.02	33.62±2.69	80324±8416
Tombul	X	164.94±49.45	0.80±0.27	0.07±0.04	56.14±2.07	16.41±0.99	29.62±2.76	87728±10621
	Y	146.85±37.14	0.85±0.17	0.07±0.03	55.20±2.27	16.50±1.27	29.16±3.26	86626±10394
	Z	155.55±40.19	0.83±0.19	0.07±0.03	56.02±2.89	16.29±1.65	30.27±3.24	88181±10065
Uzunmusa	X	144.92±34.44	0.87±0.22	0.07±0.03	61.76±2.71	15.11±1.75	31.46±4.39	97168±12064
	Y	98.74±33.54	0.96±0.24	0.05±0.03	60.98±2.34	15.61±1.28	32.51±4.44	90949±10747
	Z	109.87±26.91	0.89±0.20	0.05±0.02	60.90±2.51	15.45±1.34	31.60±3.81	93177±10759
Yassı Badem	X	255.03±58.44	1.02±0.24	0.14±0.06	55.88±2.33	16.29±0.89	30.37±2.85	73393±7267
	Y	228.88±74.44	1.23±0.32	0.16±0.08	56.05±1.97	16.09±1.14	29.51±3.30	69201±8616
	Z	294.16±55.35	1.15±0.22	0.16±0.06	56.26±2.20	15.99±0.88	29.99±2.95	72411±7980
Yuvarlak Badem	X	138.42±46.24	0.76±0.24	0.06±0.03	55.16±2.86	17.37±0.97	31.76±3.59	67312±8939
	Y	97.86±31.71	0.85±0.28	0.05±0.03	54.59±2.77	17.26±1.07	31.39±3.71	72616±10265
	Z	142.92±36.37	0.93±0.24	0.07±0.03	55.45±2.90	17.03±0.83	31.06±3.45	68461±8865

fitting into multiple normal distribution, the variance-covariance matrices being equal and not having multiple connection problems were tested. The data used in the study showed multivariate normal distribution. The variance-covariance matrices were tested with Box-M test. Quadratic discriminant function was used since the variance-covariance matrices were not equal (Tabachnick & Fidell, 2007). In order to test whether there were multi linear connection problems, correlation matrix consisting of the correlation coefficients between variables were calculated (Alpar, 2017). The correlation coefficient of hazelnut surface area was found to be well above 0.90; therefore, this variable was excluded from the independent variables (Alpar, 2017). In discriminant analysis, the most effective variables for separating the varieties were found to be sphericity, grain volume, geometric mean diameter and colour parameters, while the least effective ones were found to be deformation, rupture energy, rupture force and grain weight variables.

For all three main axes, the significance level of independent variable groups was checked with equality of

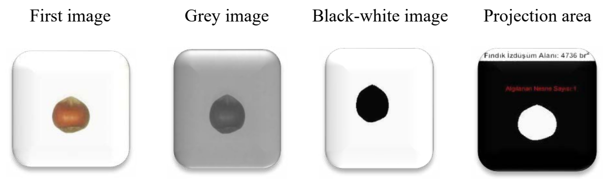

Figure 3. Image processing application

group means test. Wilks' Lambda and F values were considered for these tests. Smaller Wilks' Lambda values and greater F values indicate that the contribution of the independent variables to the dependent variables is higher. All of the independent variables were found to be statistically significant (p<0.05) (Güzeller, 2016).

Classification with discriminant analysis for X axis (DA_X)

Eigenvalues of discriminant functions, canonical correlation and Wilks' Lambda values were calculated. Eigenvalues of the functions were found to range between 0.014 and 11.474. Eigenvalues greater than 0.40 can be considered to be good (Kalaycı, 2018). Greater eigenvalue expresses that more of the variance in the dependent variable can be explained by the function. The first function explained 57.5% of the total variance. Canonical correlation values of the functions were between 0.118 and 0.959 and the square of these values ranged between 0.014 and 0.92. The square of canonical correlation value shows the variance that the function can explain in the dependent variable. Wilks' Lambda values, which show the part of the functions that cannot be explained by the differences between groups of the total variance, were between 0.001 and 0.986. These values indicate the significance of the eigenvalue statistics for each discriminant function (Kalaycı, 2018).

In order to evaluate the accurate classification of discriminant analysis, that is, the success of the analysis, the classification table of training and test sets was obtained (Table 3). According to the classification results obtained with DA_X model, the discriminant function that was chosen for the training set classified 397 of the 426 samples accurately into their varieties, while 29 samples were classified inaccurately. The general accurate classification rate of DA_X model was found as 93.2%. On the other hand, DA_X model classified 51 of the 55 samples in test set to their varieties accurately, while 4 samples were classified inaccurately. Its general accurate classification rate was found to be 92.7%.

Classification with discriminant analysis on Y axis (DA_Y)

Eigenvalues of the discriminant functions were found to range between 0.005 and 12.397. The first function explained 54.6% of the total variance. Canonical correlation values of the functions were between 0.071 and 0.962 and the square of these values ranged between 0.005 and 0.93. Wilks' Lambda values were between 0.0003 and 0.995 (Kalaycı, 2018).

According to the classification results obtained with DA_Y model, the discriminant function that was chosen for the training set classified 403 of the 421 samples accurately into their varieties, while 18 samples were classified inaccurately. The general accurate classification rate of DA_Y model was found as 95.7%. On the other hand, DA_Y model classified 51 of the 55 samples in test set to their varieties accurately, while 4 samples were classified inaccurately. Its general accurate classification rate was found to be 92.7%. (Table 3).

Classification with discriminant analysis on Z axis (DA_Z)

Eigenvalues of the discriminant functions were found to range between 0.007 and 11.910. The first function explained 57.7% of the total variance. Canonical correlation values of the functions were between 0.081 and 0.960 and the square of these values ranged between 0.006 and 0.92. Wilks' Lambda values, which show the part of the functions that cannot be explained by the differences between groups of the total variance, were between 0.001 and 0.993 (Kalaycı, 2018).

According to the classification results obtained with DA_Z model, the discriminant function that was chosen for the training set classified 399 of the 419 samples accurately into their varieties, while 20 samples were classified inaccurately. The general accurate classification rate of DA_Z model was found as 95.2%. On the other hand, DA_Z model classified 47 of the 53 samples in test set to their varieties accurately, while 6 samples were classified inaccurately. Its general accurate classification rate was found to be 88.7% (Table 3).

Classification with artificial neural networks

Since the network structure used in ANN applications is similar for three main axes, they were evaluated together. In the study, the best results were obtained by using feed forward back propagation network structure and Levenberg-Marquardt training algorithm. In addition, hyperbolic sigmoid tangent function was chosen for the hidden layer and softmax function was chosen for the output layer as transfer functions. While the networks were trained, it was ensured that the training was terminated before overfitting occurred by controlling it with the verification set. The success of the network was controlled with the test set. The best training results were obtained

Table 3. Classification results obtained with DA_X, DA_Y, and DA_Z models

Varieties	y1		y2		y3		y4		y5		y6		y7		y8		y9		y10		y11		Success (%)	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
y1	35	5	3		1																		89.7	100
y2	2		32	5	1		1								1		1						84.2	100
y3					37	5							1		2								92.5	100
y4			1				33	5					2		1								89.2	100
y5									35	5							1						97.2	100
y6					1	1					37	4											97.4	80
y7													40	5									100	100
y8					2										37	5							94.9	100
y9					3	2			2		1				1		33	3					82.5	60
y10																			38	4	1	1	97.4	80
y11																					40	5	100	100
DA_X													General classification success (%)										93.2	92.7

Varieties	y1		y2		y3		y4		y5		y6		y7		y8		y9		y10		y11		Success (%)	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
y1	36	4	4	1																			90	80
y2			36	3		2																	100	60
y3					34	5					2						3						87.2	100
y4							38	5					1										97.4	100
y5									35	5													100	100
y6					2						37	4											94.9	100
y7													37	5					1				97.4	100
y8															40	5							100	100
y9					3								1				36	5					90	100
y10																			37	4	1	1	97.4	80
y11																					37	5	100	100
DA_Y													General classification success (%)										95.7	92.7

Varieties	y1		y2		y3		y4		y5		y6		y7		y8		y9		y10		y11		Success (%)	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
y1	37	4	2	1																			94.9	80
y2	3	1	34	4					1														89.5	80
y3					33	5					5						1						84.6	100
y4							40	5															100	100
y5									32	5	1						1						94.1	100
y6					3	2					33	3			2	1							86.8	50
y7													39	3					1				98	100
y8															37	5							100	100
y9																1	39	5					100	83
y10																			36	3			100	100
y11																					39	5	100	100
DA_Z													General classification success (%)										95.2	88.7

y1:'İncekara', y2:'Kalınkara', y3:'Kan', y4:'Kuş', y5:'Okay28', y6:'Palaz', y7:'Sivri', y8:'Tombul', y9:'Uzunmusa', 10:'Yassı Badem', y11:'Yuvarlak Badem'.

Table 4. Classification results obtained with ANN_X, ANN_Y, and ANN_Z models

Varieties	y1		y2		y3		y4		y5		y6		y7		y8		y9		y10		y11		Success (%)	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
y1	39	5																					100	100
y2		1	38	4																			100	80
y3					38	4									2			1					95	80
y4			2				35	5															94.6	100
y5									36	5													100	100
y6						1					38	4											100	80
y7													40	5									100	100
y8			1												38	5							97.4	100
y9	1					1				1							39	3					97.5	60
y10																			38	4	1	1	97.4	80
y11																					40	5	100	100
ANN_X													General classification success (%)										98.4	89.1

Varieties	y1		y2		y3		y4		y5		y6		y7		y8		y9		y10		y11		Success (%)	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
y1	40	5																					100	100
y2		3	36	2																			100	40
y3					38	5											1						97.4	100
y4							39	4						1									100	80
y5									35	5													100	100
y6											39	4											100	100
y7													38	5									100	100
y8															40	5							100	100
y9													1				39	5					97.5	100
y10																			37	5	1		97.4	100
y11																					37	5	100	100
ANN_Y													General classification success (%)										99.3	92.7

Varieties	y1		y2		y3		y4		y5		y6		y7		y8		y9		y10		y11		Success (%)	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
y1	39	5																					100	100
y2			37	5											1								97.4	100
y3					36	4					1	1					2						92.3	80
y4							40	4										1					100	80
y5									34	5													100	100
y6					2						36	4				1							94.7	83
y7								2					40	1									100	33
y8												1			37	4							100	80
y9						1											39	5					100	83
y10																			36	3			100	100
y11																					39	5	100	100
ANN_Z													General classification success (%)										98.6	86.8

y1:'İncekara', y2:'Kalınkara', y3:'Kan', y4:'Kuş', y5:'Okay28', y6:'Palaz', y7:'Sivri', y8:'Tombul', y9:'Uzunmusa', 10:'Yassı Badem', y11:'Yuvarlak Badem'.

from the networks having 60, 60 and 40 neuron numbers, respectively for X, Y and Z axes.

Classification with artificial neural networks on X axis (ANN_X)

According to the classification results obtained with ANN_X model, 419 of the 426 hazelnut samples were classified accurately into their varieties and 7 samples were classified inaccurately. Its general accurate classification rate was found to be 98.4%. By using the test set, 49 of 55 hazelnut samples were classified accurately into their varieties, while 6 samples were classified inaccurately. Its general accurate classification rate was found to be 89.1%. The samples belonging to 'İncekara', 'Kuş', 'Okay28', 'Sivri', 'Tombul' and 'Yuvarlak Badem' hazelnut varieties were classified accurately (Table 4).

Classification with artificial neural networks on Y axis (ANN_Y)

According to the classification results obtained with ANN_Y model, 418 of the 421 hazelnut samples were classified accurately into their varieties and 3 samples were classified inaccurately. Its general accurate classification rate was found to be 99.3%. By using the test set, 51 of 55 hazelnut samples were classified accurately into their varieties, while 4 samples were classified inaccurately. Its general accurate classification rate was found to be 92.7%. While one sample belonging to 'Kuş' variety was classified inaccurately, three samples belonging to 'Kalınkara' variety was classified inaccurately. All the other samples were classified accurately into their varieties (Table 4).

Classification with artificial neural networks on Z axis (ANN_Z)

According to the classification results obtained with ANN_Z model, 413 of the 419 hazelnut samples were classified accurately into their varieties and 6 samples were classified inaccurately. Its general accurate classification rate was found to be 98.6%. By using the test set, 46 of 53 hazelnut samples were classified accurately into their varieties, while 7 samples were classified inaccurately. Its general accurate classification rate was found to be 86.8%. While all samples belonging to 'İncekara', 'Kalınkara', 'Okay28', 'Yassı Badem' and 'Yuvarlak Badem' varieties were classified accurately, one sample belonging to 'Kan', 'Kuş', 'Palaz', 'Tombul' and 'Uzunmusa' was classified inaccurately and two samples belonging to 'Sivri' were classified inaccurately (Table 4).

Variety classification is required for many different purposes such as machine design, evaluation of consumer preference, cultivar identification for cultivar registrations, investigation of heritability of fruit shape traits, analysis of shape abnormalities (Kays, 1999; Cannon & Manos, 2001; Beyer *et al.*, 2002; Brewer *et al.*, 2007; Taner *et al.*, 2021).

Similar results were obtained with both ANN and discriminant analysis models for the variety classification of hazelnuts. The reason for obtaining these similar results with both models can be due to processing data used in this study with multivariate normal distribution analyses before fitting the models. In fact, ANN can be more successful in nonlinear data according to the literature (Ford *et al.*, 2004).

In previous studies, hazelnut cultivars have not been classified by using both ANN and discriminant analysis based on their physical, mechanical, and optical properties. However, one study using discriminant analysis were available in the literature (Menesatti *et al.*, 2008). In that study, four traditional Italian hazelnut cultivars were classified by discriminant analysis. They found the classification success of 77.5%-98.8%. ANN and discriminant analysis methods were used separately or together to classify different agricultural products (Chen *et al.*, 2010; Azizi *et al.*, 2016; Taner *et al.*, 2018).

In conclusion, the most effective variables for the classification of hazelnut varieties were sphericity, grain volume, geometric mean diameter and image processing variables, while the least effective ones were deformation, rupture energy, rupture force and grain weight variables. ANN and discriminant analysis gave close results to each other probably because of that nonlinear data used in this study were normally distributed. ANN and discriminant analysis models can be used successfully for the classification of hazelnut varieties. The classification performance of ANN and discriminant analysis models for the same hazelnut varieties grown in different parts of Turkey and in different climatic conditions can be determined in future studies. In addition, smartphone applications using the models proposed in this study can be developed.

References

Ağın O, Taner A, 2015. Determination of weed intensity in wheat production using image processing techniques. Anadolu J Agr Sci 30: 110-117. https://doi.org/10.7161/anajas.2015.30.2.110-117

Alasavar C, Shahidi F, Liyanapathirana CM, Oshima T, 2003. Turkish Tombul hazelnut (*Corylus avellana* L.) 1. Compositional characteristics. J Agr Food Chem 51: 3790-3796. https://doi.org/10.1021/jf0212385

Alpar R, 2017. Applied multivariate statistical. Detay Yayıncılık. Ankara, Turkey, 820 pp.

Aydın C, 2002. Physical properties of hazelnuts. Biosyst Eng 82(3): 297-303. https://doi.org/10.1006/bioe.2002.0065

Azizi A, Abbaspour-Gilandeh Y, Nooshyar M, Afkari-Sayah A, 2016. Identifying potato varieties using machine vision and artificial neural networks. Int J Food Prop 19: 618-635. https://doi.org/10.1080/10942912.2015.1038834

Beyer M, Hahn R, Peschel S, Harz M, Knoche M, 2002. Analysing fruit shape in sweet cherry (*Prunus avium* L.). Sci Hortic 96: 139-150. https://doi.org/10.1016/S0304-4238(02)00123-1

Bishop CM, 1995. Neural network for pattern recognition. Clarendon, Oxford. https://doi.org/10.1201/9781420050646.ptb6

Brewer MT, Moyseenko JB, Monforte AJ, van der Knaap E, 2007. Morphological variation in tomato: a comprehensive study of quantitative trait loci controlling fruit shape and development. J Exp Bot 58(6): 1339-1349. https://doi.org/10.1093/jxb/erl301

Cannon CH, Manos PS, 2001. Combining and comparing morphometric shape descriptors with a molecular phylogeny: the case of fruit type evolution in Bornean Lithocarpus (Fagaceae). Syst Biol 50(6): 860-880. https://doi.org/10.1080/106351501753462849

Chen X, Xun Y, Li W, Zhang J, 2010. Combining discriminant analysis and neural networks for corn variety identification, Comput Electron Agr 71: 48-53. https://doi.org/10.1016/j.compag.2009.09.003

Demirbaş HY, Dursun İ, 2007. Determination of some physical properties of wheat grains by using image analysis. J Agr Sci 13(3): 176-185.

Dubey BP, Bhagwat SG, Shouche SP, Sainis JK, 2006. Potential of artificial neural networks in varietal identification using morphometry of wheat grains. Biosyst Eng 95: 61-67. https://doi.org/10.1016/j.biosystemseng.2006.06.001

Fallico B, Arena E, Zappala M, 2003. Roasting of hazelnuts. Role of oil in colour development and hydroxy methyl furfural formation. Food Chem 81: 569-573. https://doi.org/10.1016/S0308-8146(02)00497-1

FAOSTAT, 2018. Classifications and standards. Food and Agriculture Organization of the United Nations, Rome. http://www.fao.org/faostat/en/#data.

Ford MG, Pitt WR, Whitley DC, 2004. Selecting compounds for focused screening using linear discriminant analysis and artificial neural networks. J Mol Graph Model 22: 467-472. https://doi.org/10.1016/j.jmgm.2004.03.006

Francisco EL, Jeffrey KB, Amarat HS, Anne P, Elizabeth AB, Jinhe B, Elena L, 2021. Color biogenesis data of tomatoes treated with hot-water and high temperature ethylene treatments. Data in Brief 36: 107123. https://doi.org/10.1016/j.dib.2021.107123

Gonzalez RC, Woods RE, 2008, Digital image processing, Pearson Int Ed, Pearson Prentice Hall, USA. ISBN: 0-13-168728-x 978-0-13-168728-8.

Guiné RPF, Almeida CFF, Correia PMR, Mendes M, 2015. Modelling the influence of origin, packing and storage on water activity, colour and texture of almonds, hazelnuts and walnuts using artificial neural networks. Food Bioprocess Technol 8(5): 1113-1125. https://doi.org/10.1007/s11947-015-1474-3

Güzeller CO, 2016. Multivariate statistics for everyone. Maya Akademi Yayıncılık, Ankara, Turkey, 346 pp.

Jacobs RA, 1988. Increased rate of convergence through learning rate adaptation. Neural Networks 1(4): 295-307. https://doi.org/10.1016/0893-6080(88)90003-2

Jain RK, Bal S, 1997. Properties of pearl millet. J Agr Eng Res 66(2): 85-91. https://doi.org/10.1006/jaer.1996.0119

Kalaycı Ş, 2018. SPSS Applied multivariate statistics techniques, 9th ed, Dinamik Akademi Yayıncılık, Turkey, 426 pp.

Kays SJ, 1999. Preharvest factors affecting appearance. Postharv Biol Technol 15: 233-247. https://doi.org/10.1016/S0925-5214(98)00088-X

Levenberg K, 1944. A method for the solution of certain nonlinear problems in least squares. Quart Appl Math 2: 164-168. https://doi.org/10.1090/qam/10666

Marquardt DW, 1963. An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 11: 431-441. https://doi.org/10.1137/0111030

Menesatti P, Costa C, Paglia G, Pallottino F, D'Andrea S, Rimatori V, Aguzzic J, 2008. Shape-based methodology for multivariate discrimination among Italian hazelnut cultivars. Biosyst Eng 101: 417-424. https://doi.org/10.1016/j.biosystemseng.2008.09.013

Minai AA, Williams RD, 1990. Back-propagation heuristics: a study of the extended Delta-bar-delta algorithm. Int Joint Conf on Neural Networks, vol.1, pp: 595-600, San Diego, CA, USA. https://doi.org/10.1109/IJCNN.1990.137634

Mohsenin NN, 1970. Physical properties of plant and animal materials. Gordon & Breach Sci Publ Inc, NY.

Oddone M, Aceto M, Baldizzone M, Musso D, Osella D, 2009. Authentication and traceability study of hazelnuts from Piedmont, Italy. J Agr Food Chem 57(9): 3404-3408. https://doi.org/10.1021/jf900312p

Özdamar K, 2004. Statistical data analysis with package programs (Multivariate analysis), Kaan Kitabevi, Eskişehir, Turkey, 502 pp.

Parcerisa J, Richardson DG, Rafecas M, Codoni R, Boatella S, 1998. Fatty acid, to cophero land sterol content of some hazelnut varieties (*Corylus avellana* L.)

harvested in Oregon (USA). J Chromat 805: 259-268. https://doi.org/10.1016/S0021-9673(98)00049-1

Pourreza A, Pourreza H, Abbaspour-Fard MH, Sadrnia H, 2012. Identification of nine Iranian wheat seed varieties by textural analysis with image processing. Comput Electron Agr 83: 102-108. https://doi.org/10.1016/j.compag.2012.02.005

Purushothaman S, Srinivasa YG, 1994. A back-propagation algorithm applied to tool wear monitoring. Int J Machin Tools Manufact 34(5): 625-631. https://doi.org/10.1016/0890-6955(94)90047-7

Sharma S, 1996. Applied multivariate techniques. John Wiley & Sons Inc., Canada.

Tabachnick BG, Fidell LS, 2007. Using multivariate statistics, 5th ed. Harber Collins Pub., London, 980 pp.

Tabatabaeefar A, 2003. Moisture-dependent physical properties of wheat. Int Agrophys 17: 207-211.

Taner A, Gültekin SS, Çarman K, 2010. Prediction of the parameters radial centrifugal pumps with artificial neural networks. Selcuk J Agr Food Sci 24(1): 28-38.

Taner A, Öztekin YB, Tekgüler A, Sauk H, Duran H, 2018. Classification of varieties of grain species by artificial neural networks. Agronomy 8: 123. https://doi.org/10.3390/agronomy8070123

Taner A. Öztekin YB, Duran H, 2021. Performance analysis of deep learning CNN models for variety classification in hazelnut. Sustainability 13: 6527. https://doi.org/10.3390/su13126527

Visen NS, Paliwal J, Jayas DS, White NDG, 2002. Specialist neural networks for cereal grain classification. Biosyst Eng 82: 151-159. https://doi.org/10.1006/bioe.2002.0064

Yang CC, Prasher SO, Landry JA, Ramaswamy HS, 2003. Development of a herbicide application map using artificial neural networks and fuzzy logic. Agr Syst 76(2): 561-574. https://doi.org/10.1016/S0308-521X(01)00106-8