Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 12(4): 2034 - 2044, 2022
Journal of the Institute of Science and Technology, 12(4): 2034 - 2044, 2022
ISSN: 2146-0574, eISSN: 2536-4618
Biosystems Engineering **DOI: 10.21597/jist.1115119**
Research Article
Received: 10.05.2022 **Accepted: 26.08.2022**

To Cite: Çetin N, 2022. Multivariate Machine Learning Approach for Size and Shape Prediction of Sunflower Seeds. Journal of the Institute of Science and Technology, 12(4): 2034 - 2044.

Multivariate Machine Learning Approach for Size and Shape Prediction of Sunflower Seeds

Necati ÇETİN[1*]

ABSTRACT: Sunflower constitutes an important source of protein, mineral, vitamin, fatty acid, and offer a balanced source of amino acids. Machine learning is mostly performed for the prediction of descriptive attributes in the quality evaluation of foods. In this study physical attributes of two different sunflower varieties (Metinbey and İnegöl Alası) were determined and algorithms were applied for size and shape prediction of these varieties. In addition, five different machine learning predictors were used as Multilayer Perceptron (MLP), Gaussian Processes (GP), Random Forest (RF), k-Nearest Neighbors (kNN), and Support Vector Regression (SVR). The prediction of surface area, volume, geometric mean diameter, aspect ratio, elongation, and shape index were based on the main physical attributes. İnegöl Alası variety had the greatest physical attributes. The seed length, width and thickness were obtained from İnegöl Alası variety as 23.89, 8.80 and 4.15 mm and from Metinbey as 17.88, 6.20 and 4.01 mm. All varieties were determined as significant in terms of the selected attributes as reported by Pillai Trace and Wilks' Lambda ($p<0.01$). In the Wilks' Lambda statistics, unexplained of the similarities or differences among the groups was 12.30%. Present findings revealed that MLP and SVR algorithms had the greatest correlation coefficients for all predicted attributes. In the study, the best predicted attributes were geometric mean diameter with an R value of 0.9989 (SVR), followed by volume and elongation with an R value of 0.9988 (MLP). Present findings revealed that MLP and SVR algorithms could potentially be used for size and shape prediction of sunflower varieties.

Keywords: Sunflower, volume, area, elongation, machine learning

[1]Necati ÇETİN (**Orcid ID:** 0000-0001-8524-8272) Erciyes University, Faculty of Agriculture, Department of Biosystems Engineering, Kayseri, Türkiye
*** Corresponding Author:** Necati ÇETİN, e-mail: necaticetin@erciyes.edu.tr

INTRODUCTION

Sunflower is largely used to supply the edible oil requirement. Besides the grains of these varieties are rich in nutrients, are mixed with salt, butter, and honey and used in confectionery, added as a seasoning on vegetables, fish, meat, and salads, and consumed as snacks, either roasted or unroasted (Ergen and Sağlam, 2005). Sunflower (*Helianthus annus* L.) belongs to the Asteraceae family and has long been cultivated and consumed in North America. It is also an important industrial crop and attracted the attention of researchers and consumers (Bodouin et al., 2017). Sunflower seeds are rich in nutrition content such as fat (50%), oil (39% to 49% as linoleic acid %80), and protein (20%) (Seiler, 2007). Globally, annual production of sunflower seeds was 52 million tons in 26.7 million ha. It was cultivated on 104 000 ha and 68 000 tons with oil production and seed production, respectively (FAOSTAT, 2018).

Physical attributes like volume, surface area, geometric mean diameter, sphericity, elongation, mass, porosity, and color are used to design for product processing, cleaning, oil expeller, transportation, storage, and dehuller (Tabatabaeefar and Rajabipour, 2005). Seed physical attributes are also important appearance trait for variety and quality identification. In addition, size and shape are used in the classification and breeding process (Demir et al., 2018; Cetin et al., 2020). It is important to know the physical properties for the classification and quality assessment of different varieties in the final product. In addition, it allows the control of these processes and allows the selection of parameters related to the operation of the machines (Demir et al., 2017). Knowledge about the size and shape is necessary to plan discrimination, sizing, planting, harvesting, and handling devices (Sahay and Singh, 2004). Area (projected and surface) properties are considered when designing pneumatic separation devices (Bwade and Aliyu, 2012). Quality properties of the agricultural product are identified as the visual, texture, and flavor of the food. Additionally, consumers mostly were preferred in terms of these attributes (Kays, 1999; Cetin et al., 2020).

Machine learning is effective and practice approach used in the design of rapid, accurate, and reliable predictors. These approaches include a lot of algorithms like decision trees, artificial neural network, k-nearest neighbors, and regressions. Such algorithms are mostly performed for the accurate choice of descriptive attributes in the quantity and quality evaluation of products (Omid et al. 2010; Mollazade et al. 2012). Machine learning algorithms also offer non-linear models able to estimate food quality attributes values within an input layer and output layer linkage (Zhang et al., 2012). Artificial neural network (ANN), gaussian processes (GP), k-nearest neighbors (kNN), random forest (RF), and support vector regression (SVR) are commonly used in prediction studies. It has been reported that ANN was useful in the classification and prediction in agricultural commodities. In the ANN, the impacts of synapses are offered by link weights that modulate the impact of the respective input signals (Çetin and Sağlam, 2022). GP is of great significance in prediction problems since all attributes inherited from a normal distribution could clearly be obtained (Sağlam and Çetin, 2022a). kNN is an easy-to-implement machine learning algorithm that can be used to solve prediction problems (Xu et al., 2017). RF algorithm produces more than one decision tree with the use of bootstrap samples from the original training data to train each tree and is a successful prediction (Sağlam and Çetin, 2022b). SVR is highly recognized for its superior performance of regression data due to its excellent performance capability when working with multidimensional data (Zhang and Ma, 2009; Çetin and Sağlam, 2022).

In the literature, different studies were carried out to evaluate the physical attributes of sunflower seeds (Santalla and Mascheroni, 2003; Khodabakhshian et al., 2010; Jafari et al., 2011; Malik and

Saini, 2016; Ortiz-Hernandez et al., 2020; Çetin et al., 2021). However, there are not any studies about the prediction of size and shape attributes. In the current study, principal physical attributes of two different sunflower varieties were measured for the purpose of discrimination and prediction of size and shape. In addition, five machine learning algorithms were applied for size and prediction from principal dimension attributes and the performance of algorithms was compared.

MATERIALS AND METHODS

Plant material

In the present study, Metinbey and İnegöl Alası sunflower varieties were used as material. These varieties, commonly used by local farmers in Kayseri Province (38°50'21.0"N 35°39'13.3"E), were used in the present experiments. Sunflowers were harvested from the Kayseri province of Turkey in the 2021 harvest season. Foreign materials with irregular shapes were separated. After that, 100 randomly selected seeds were studied. Sunflower seeds were kept in a refrigerator at 4±0.5°C constant moisture level throughout the study.

Principal dimension properties

Length (L, mm) width (W, mm), and thickness (T, mm) of each variety as principal dimension were measured with a digital caliper. For the dimension analysis, 100 sunflower seeds were sampled from each variety. The surface area (S, mm^2), volume (V, mm^3), geometric mean diameter (Dg, mm), shape index (SI), elongation (E), and aspect ratio (AR) were used for input-output parameters.

Discriminant analysis

Differences between the varieties and group centroids of the varieties were evaluated by using linear discriminant analysis. The principal components were assessed with the use of MANOVA. Similarities or dissimilarities of sunflower varieties were tested by Hotelling's pair-wise comparisons with two different approaches as squared Mahalanobis distances and Bonferroni correction (Çetin, 2022). Experimental data were subjected to one-factor analysis and significant means were compared by Tukey's multiple comparison test ($p<0.05$). Statistical analyses were performed by using SPSS v20.0 (IBM SPSS® 2010) and PAST v3.20 software (Hammer et al. 2001).

Machine learning approaches and validation

Five different machine learning predictors were used Artificial Neural Network, Gaussian Processes, Random Forest, k-Nearest Neighbors, and Support Vector Regression. The prediction of surface area, volume, geometric mean diameter, shape index, aspect ratio, and elongation were based on the main physical attributes. A total of 600 values were used for shape and size prediction for each attribute. In the present study, the k-fold cross-validation technique was performed for validation, and the k value was preferred as 10. K value is mostly selected as 10 or 5 in the prediction (Ataş et al. 2012). In the k-fold cross validation, the dataset was separated to 10 subsets. Training folds and test folds were performed with the 10 iterations. In each iteration, one subset was used for test folds and 9 subsets were used for training folds and with each of the k sub-samples used exactly once as the testing, respectively (Stegmayer et al. 2013). The 10-fold cross validation methodology is given in Figure 1.

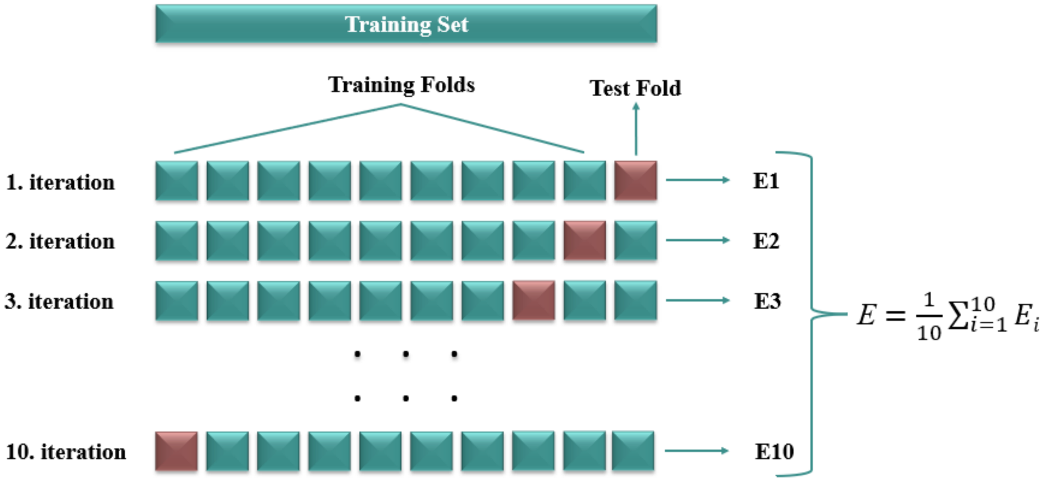

Figure 1. 10-fold cross validation methodology

Artificial neural networks (ANNs)

ANNs are machine learning with algorithms inspired by the nerve cell structure of the human cell. Multilayer Perceptron is widely applied in ANN structures. Multilayer perceptron was used as a feedforward artificial neural network. The neural network parameters of the MLP structure were selected as the number of epochs 500, learning rate 0.3, and momentum 0.2. Additionally, 3-10-6 MLP structure consisting of neurons in 3 input, 10 hidden, and 6 output layers for shape and size prediction. The applied MLP model structure is presented in Figure 2.

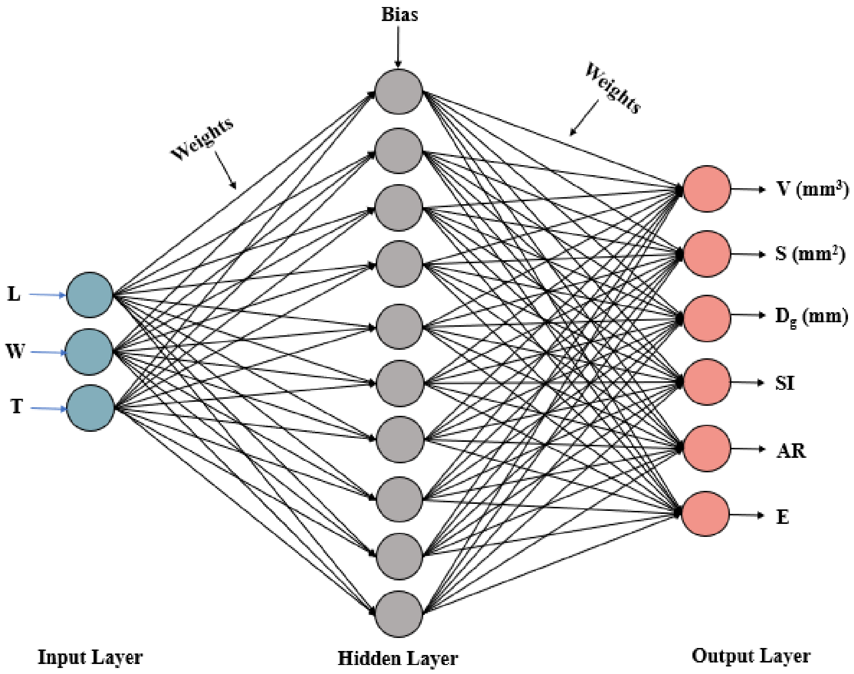

Figure 2. Structure of the MLP model for prediction of the sunflower shape and size

Gaussian processes (GP)

The Bayesian Gaussian technique in the GP is performed for non-linear regression. It also requires defining a kernel function with a noise editing indicator. Additionally, before the learning methods allow to choose the training data to be normalized (MacKay, 1998). GP is crucial in modeling because all attributes are derived from the normal distribution. The normal distribution of error, covariance, and variance could be obviously determined (Rasmussen and Williams, 2006). In the current study, as kernel function was determined as Pearson VII (PUK).

k-nearest neighbors (k-NN)

The k nearest neighbor (k-NN) is a useful and effective prediction model. In this model, all observation values as a cluster are considered. After that these clusters are combined into new clusters. Generally, Euclidean distance calculation was used (Nettleton et al., 2010; Romero et al., 2013). High k-values could result in overgeneralized outputs, while low k values could generate a very complex decision boundary. Here in, since a trained model is not generated, it is expected that the k-NN prediction will provide proper results as the number of training folds increases. For this reason, selecting the proper value in prediction gains significance (Maxwell et al., 2018). In the present study, the Euclidean distance rule was performed, and k values were selected as 3.

Random forest (RF)

Random Forest algorithm was performed for size and shape prediction. Random Forest, a decision is utilized with majority of the ensemble of trees built by RF in data sets assigned class (Berhane et al. 2018). Afterward, the ensemble scheme and bootstrap can overcome overfitting problems inherited from Decision Tree, there is no pruning step in Random Forest. Additionally, Random Forest has a high estimative correlation coefficient and is robust against noise (Breiman, 2001).

Support vector regression (SVR)

SVR has core-based functionality and is used to reveal prediction problems. The selection of kernel function has an important effect on the SVR performance. SVR presents accurate and successful findings in revealing problems. In Support Vector Regression, nonlinear kernel transformation formula is utilized to map the inputs in feature space (Vapnik, 2000). In this procedure, in the transformed space the correlation between outputs and inputs is linearized. In the study, the polynomial kernel function was chosen. The general SVR equation is presented in Equation (1) below:

$$* \ y = w\Phi(x) + b \quad (\Phi : R_n \rightarrow R_N) = \pi r^2 \tag{1}$$

Where; w ∈ RN is the coefficient factor, b is the bias term, x ∈ Rn is the input, y ∈ RN is the output, and Φ is the mapping function whose input is transformed into a high-dimensional vector.

Model performance evaluation

Evaluation of performance was calculated by the following Equation (2-4): correlation coefficient (R), mean absolute error (MAE) root mean square error (RMSE), (Parker, 2001).

$$* \ R = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(M_i - \dot{M})\,(E_i - \dot{E})}{S_M \ S_E} \tag{2}$$

$$* \ MAE = \sum_{i=1}^{n}\frac{|\ E_i - M_i\ |}{n} \tag{3}$$

$$* \ RMSE = \sqrt{\frac{\sum_{i=1}^{n}(E_i - M_i)^2}{n}} \tag{4}$$

Where; M_i: Determined target value, Ṁ: Mean of determined target values, Ė: Mean of predicted target values, E_i: Predicted target value, SE: Sum of predicted target values, SM: Sum of determined target values and n: Number of data.

R values were obtained to evaluate the prediction success in relation to the rule specified in Colton (1974). R values of between 0 and 0.25 indicate lower, 0.25 and 0.50 moderate, 0.50 and 0.75 moderate or high; 0.75 and 1.00 report perfect correlation.

RESULTS AND DISCUSSION

Physical attributes of sunflower seeds

Two sunflower varieties were assessed in terms of size and shape and findings were provided in followed. The physical attributes of the sunflowers are presented in Table 1 ($p<0.01$). The highest seed length, width and thickness were obtained from İnegöl Alası variety as 23.89, 8.80 and 4.15 mm, respectively. Length (17.88 mm), width (6.20 mm), and thickness (4.01 mm) attributes of Metinbey variety were slightly less than the İnegöl Alası. Santalla and Mascheroni (2003) indicated width and length attributes as 5.008 mm and 11.526 mm. Khodabakhshian et al. (2010) reported the thickness of Shahroodi variety values as between 3.88 - 4.94 mm. Volume of the sunflower varieties varied between 111.00 - 835.83 mm^3 with the greatest value from İnegöl Alası (462.41 mm^3) variety and the least from Metinbey (235.88 mm^3) variety. The highest surface area value was determined from İnegöl Alası as 286.78 mm^2. Surface area of the Metinbey variety was calculated as 183.43 mm^2. The geometric mean diameter values of Metinbey and İnegöl Alası varieties were determined as 7.62 and 9.51 mm, respectively. Comply with the present findings, Malik and Saini (2016) indicated PSH-996 variety volume at the different moisture content as between 192.61 and 262.77 mm^3. Ortiz-Hernandez et al. (2020) determined surface area values between 150.01 and 159.71 mm^2 for P64H41 sunflower seed variety and those values were slightly lower than the present ones. Çetin et al. (2021) determined geometric mean diameter values of six different sunflower varieties as between 5.89 and 7.26 mm.

The highest shape index values were determined in İnegöl Alası (3.74) variety while the lowest was obtained from Metinbey (3.52). All sunflower varieties with a shape index greater than 1.25 were described as the oval. The greatest aspect ratio value was determined in Metinbey with the values of 0.23. The lower aspect ratio was observed in İnegöl Alası (0.18). Elongation values were determined as 4.49 and 5.91 for Metinbey and İnegöl Alası, respectively. In the current study, elongation and aspect ratio were negatively correlated. In contrast to the current results, Çetin et al. (2021) indicated average shape index and aspect ratio as 2.26 and 0.54. The reason for these findings is that the varieties used in the study were small sunflower oilseeds. Jafari et al. (2011) indicated Shamshiri variety elongation value as 3.26.

Table 1. Dimension, shape and size attributes for sunflower varieties

Varieties	Metinbey	İnegöl Alası	Mean	Min-max	F value
Length (*L*. mm)	17.88±1.48^b	23.89±2.32^a	20.88±3.58	8.16-30.06	477.72**
Width (*W*. mm)	6.20±0.70^b	8.80±1.08^a	7.50±1.59	4.13-11.29	407.48**
Thickness (*T*. mm)	4.01±0.40	4.15±0.68	4.08±0.56	2.88-5.86	3.22
Volume (*V*. mm^3)	235.88±57.70^b	462.41±126.90^a	349.14±150.20	111.00-835.83	264.08**
Surface area (*SA*. mm^2)	183.43±29.77^b	286.78±52.78^a	235.11±67.16	111.69-429.11	290.82**
Geo. mean diam. (D_g. mm)	7.62±0.62^b	9.51±0.88^a	8.57±1.22	5.96-11.69	310.51**
Shape Index (*SI*)	3.52±0.34^b	3.74±0.58^a	3.63±0.48	1.57-6.31	10.54**
Aspect ratio (*AR*)	0.23±0.04^a	0.18±0.03^b	0.20±0.04	0.11-0.51	107.26**
Elongation (E)	4.49±0.50^b	5.91±1.14^a	5.20±1.13	1.96-8.89	130.03**

Means indicated with diferent letters in the same column are signifcantly diferent ($p<0.05$)
**: signifcant at $p<0.01$

Multivariate tests (MANOVA) and discrimination analysis

Metinbey and İnegöl Alası were determined to be significant with regard to the size and shape attributes as reported by Pillai Trace and Wilks' Lambda ($p<0.01$). The MANOVA results, Bonferroni corrected and Mahalanobis distances are presented in Table 2. Wilks' Lambda indicated the percentage of variance in dependent and expressed them with dissimilarity in independent variables. The lower "Wilks' Lambda" values show that differences between groups to be analyzed increase and vary between 0 and 1 (Çetin et al., 2021). Pillai Trace is considered the most reliable among multivariate assessments. In addition, these statistical methods consider the sum of the variance that expresses the highest discrimination of independent variables dependent ones. Pillai's trace and Wilks' Lambda values were obtained as 0.882 and 0.118, respectively. Mostly, Mahalanobis distance of lower than three indicate significantly similar size and shape attributes (P>0.05). In this study, Mahalanobis distance value was determined as 28.063. It concluded that the Metinbey and İnegöl Alası varieties with the highest Mahalanobis distances had the contrast attributes. Additionally, these results are supported by Bonferroni corrected p values. Discriminant analysis results are presented in Table 2. The highest eigenvalues and higher functions present dependent variables. The correlation square explains the function's effect size. Wilks' lambda value presents best estimation. Wilks' lambda is important for each estimator variable that is proper. Unexplained part of the differences between the groups was 12.30% in the Wilks' lambda statistics, The discriminant analysis coefficients give a relative importance to 7 predictors. According to the loadings, function 1 had the greatest loading for the thickness, shape index, and geometric mean diameter.

Table 2. Discriminant analysis results

The results of MANOVA						
Effect	Statistics	Value	Hypothesis df	Error df	F	*p* (sigma)
Variables	Pillai's trace	0.882	12	187	116.21	0.000**
	Wilks' Lambda	0.118	12	187	116.21	0.000**
Eigenvalue statistics of discriminant functions		**Eigenvalues**	**% of variance**	**% of cumulative variance**		**Canonical correlation**
Function 1		6.560	100	100		0.932
Significance test of canonical functions		**Wilks' Lambda**	**Chi-square**	**df**		***p* (sigma)**
Function 1		0.123	391.44	9		0.000**
Canonical Discriminant Function Coefficients		**Function 1**	***Standardized Canonical Discriminant Function Coefficients***			**Function 1**
Volume		-0.004	Volume			-0.426
Length		-1.534	Length			-2.983
Width		-1.545	Width			-1.407
Thickness		-4.195	Thickness			-2.349
Geo. mean dia.		8.657	Geo. mean dia.			6.592
Shape index		4.886	Shape index			2.308
Elongation		1.650	Elongation			1.452

**Highly significant (p<0.01)

Figure 3 shows the centroids of two sunflower varieties relying on their canonical discriminant functions. In the dissimilarity between components, geometric mean diameter was considered as an important distinguishing trait. The length, thickness, and geometric mean diameter for İnegöl Alası variety confirmed the location on the right and bottom of the canonical function 1 axis. In addition, shape index and elongation for Metinbey variety were located on the left and above of canonical function 2 axis.

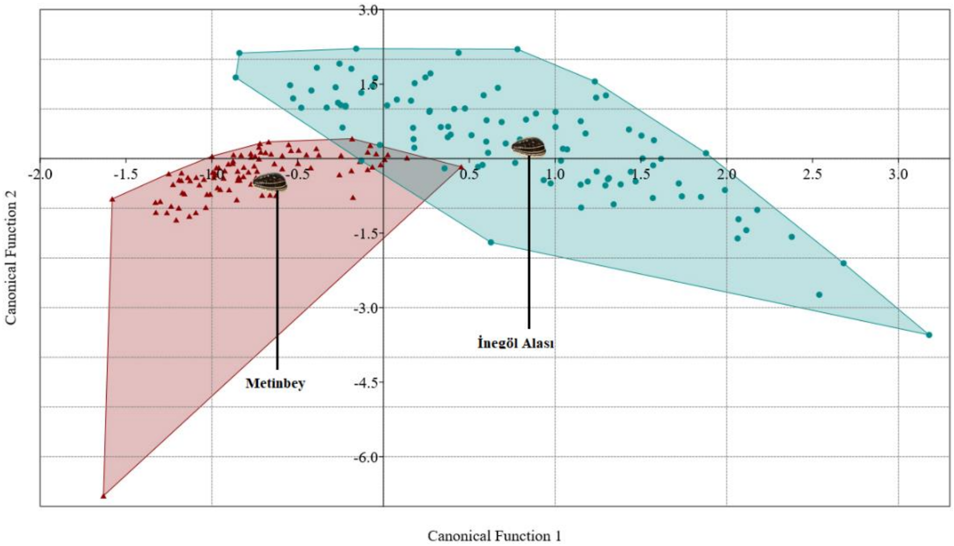

Figure 3. Scatter plots of the sunflower varieties

Performance results of machine learning algorithms

Machine learning models were separately created in terms of physical attributes. The findings were assessed based on the R, MAE, and RMSE of the prediction of size and shape attributes. Prediction results by MLP, GP, k-NN, RF, and SVR were tabulated in Table 3. The best estimation criteria were higher R and lower MAE and RMSE. Overall, each base learner performed all evaluation parameters very well with all achieving an R value of >0.8432. The greatest R for volume prediction was 0.9988, 0.9892, and 0.9886 for MLP, SVR and k-NN, respectively. The lowest MAE and RMSE values obtained from MLP as 0.0175 and 0.0372. The highest R for surface area prediction was obtained from MLP and SVR as 0.9982 and 0.9965. The lowest MAE (0.0119) and RMSE (0.0256) determined in MLP algorithm. SVR and MLP algorithms had the highest R for geometric mean diameter prediction with 0.9989 and 0.9979. SVR had the lowest MAE and RMSE values as 0.0039 and 0.0080. In the present study, the greatest R for shape index was 0.9970 (MLP) while the lowest was 0.8432 (k-NN). The highest MAE and RMSE values found in k-NN as 0.0398 and 0.0962. The highest correlation coefficient value for aspect ratio prediction obtained from MLP with the value of 0.9777. The lowest R value determined as 0.8739 from k-NN algorithm. Here in, the lowest MAE and RMSE found as 0.0064 and 0.0198 in MLP algorithm. In the elongation prediction successful results was obtained from MLP with the highest R (0.9988) and the lowest MAE (0.0048) and RMSE (0.258).

Table 3. Comparison of the model performance

Algorithms	Metrices	Volume	Surface Area	Geometric Mean Dia.	Shape Index	Aspect Ratio	Elongation
Multilayer Perceptron	R	0.9988	0.9982	0.9979	0.9970	0.9777	0.9988
	MAE	0.0175	0.0119	0.0053	0.0051	0.0064	0.0048
	RMSE	0.0372	0.0256	0.0118	0.0213	0.0198	0.0258
Gaussian Processes	R	0.9864	0.9864	0.9850	0.9132	0.8753	0.9398
	MAE	0.0359	0.0216	0.0100	0.0221	0.0199	0.0625
	RMSE	0.1358	0.0740	0.0314	0.0889	0.0527	0.0780
k-Nearest Neighbors	R	0.9886	0.9897	0.9897	0.8432	0.8739	0.9612
	MAE	0.0410	0.0271	0.0135	0.0398	0.0366	0.0381
	RMSE	0.0731	0.0467	0.0225	0.0962	0.0643	0.0908
Random Forest	R	0.9851	0.9855	0.9847	0.8665	0.8910	0.9769
	MAE	0.0499	0.0313	0.0154	0.0393	0.0315	0.0318
	RMSE	0.1027	0.0619	0.0289	0.0940	0.0566	0.0839
Support Vector Regression	R	0.9892	0.9965	0.9989	0.9823	0.9393	0.9882
	MAE	0.0553	0.0197	0.0039	0.0125	0.0214	0.0187
	RMSE	0.0887	0.0288	0.0080	0.0247	0.0410	0.0297

Similar to the present study, Demir et al. (2017) indicated that in some physical attribute estimations such as volume, area, diameter, and sphericity of pumpkin seeds using Radial Basis Neural Network and Back Propagation Neural Network, the RMSE was found as 0.6875 and 0.0025, respectively. Eski et al. (2017) presented the best RMSE value as 0.0001 for adaptive neuro fuzzy interface system (ANFIS) estimation of almond physical attributes (arithmetic and geometric mean diameter, projected and surface area, geometric mean diameter, sphericity, aspect ratio shape index and volume). Omid et al. (2010) predicted volume of citrus and, the authors found that the R^2 values for lime, lemon, tangerine, and orange were 0.970, 0.962, 0.959, and 0.985, respectively. Comply with the present study, Singh et al. (2020) applied RF, SVM, ANN, kNN, and KRR to estimate size of rice kernels. The authors indicated R^2 values between 0.898 (ANN) and 0.975 (RF-KRR). Concha-Meyer et al. (2020) to estimate the volume of the mushrooms, strawberries and tomatoes were performed simple linear regression and R^2 values were 0.92, 0.96, and 0.99, respectively. Saha et al. (2017) estimated cardamom capsule size and surface area and reported the lowest MAE and RMSE findings as 0.330 and 0.456 for minor diameter. In contrast to the current study, Ponce et al. (2018), reported RMSE results of olive-fruit size estimation (major axis and minor axis) between 0.4163 and 0.8036.

CONCLUSION

In this study, the present findings proved the availability of machine learning approaches for size and shape prediction of sunflower seeds based on principal dimension attributes. In the study, Multilayer Perceptron and Support Vector Regression algorithms yielded the greatest results in all attributes. ANN topology is a critical parameter because it has a significant effect on the prediction. In this case, prediction with the use of ANN of the physical attributes in advance could make it helpful for designers and classifiers. In addition, the selected attributes could be helpful in quality assessment industries.

For the seed industry, it is important to know the physical attributes of sunflower seeds and to make correct discrimination between varieties. Hulling, cleaning, drying, and packaging parts are generally designed according to the physical attributes of the seeds. In addition, these features can be used in selection, breeding, and quality assessment studies. Machine learning algorithms could help seed companies by providing an accurately predicted and discriminated of product. With the algorithms and structures proposed in the current study, a classifier and separator could be designed (Çetin, 2020). This study showed that some physical attributes of sunflower seeds can be predicted using machine learning algorithms and it is possible to discrimination of seeds with the use of discriminant analysis. Limiting situations in the present study, and recommendations for future studies, measurements of physical attributes of sunflower seeds took a lot of time. Instead, more up-to-date approaches such as image processing could be applied. Also, more data sets, attributes, and algorithms may be included in similar studies in future research.

Conflict of Interest

The author declares that there is no conflict of interest.

Author's Contributions

Author has contributed in experimental study and writing of the manuscript himself.

REFERENCES

Ataş M, Yardimci Y, Temizel A, 2012. A New Approach to Aflatoxin Detection in Chili Pepper by Machine Vision. Computers and Electronics in Agriculture 87:129-141.

Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Legrand L, 2017. The Sunflower Genome Provides Insights into Oil Metabolism, Flowering and Asterid Evolution. Nature 546:148-152.

Berhane T, Lane C, Wu Q, Autrey B, Anenkhonov O, Chepinoga V, Liu H, 2018. Decision-Tree, Rule-Based, and Random Forest Classification of High-Resolution Multispectral Imagery for Wetland Mapping and Inventory. Remote Sensing 10:580.

Breiman L, 2001. Random Forests. Machine Learning 45(1):5-32.

Bwade KE, Aliyu B, 2012. Investigations on the Effect of Moisture Content and Variety Factors on Some Physical Properties of Pumpkin Seed (*Cucurbitaceae* spp). International Journal of Engineering, Business and Enterprise Applications (IJEBEA) 3(1):20-24.

Çetin N, 2022. Machine Learning for Varietal Binary Classification of Soybean (*Glycine max* (L.) Merrill) Seeds Based on Shape and Size Attributes. Food Analytical Methods 15: 2260-2273.

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 Analytical Methods 14(8):1666-1681.

Çetin N, Sağlam C, 2022. Rapid Detection of Total Phenolics, Antioxidant Activity and Ascorbic Acid of Dried Apples by Chemometric Algorithms. Food Bioscience 47: 101670.

Cetin N, Yaman M, Karaman K, Demir B, 2020. Determination of Some Physicomechanical and Biochemical Parameters of Hazelnut (*Corylus avellana* L.) Cultivars. Turkish Journal of Agriculture and Forestry 44(5): 439-450.

Colton T, 1974. Statistics in Medicine. Little Brown and Co, New York.

Concha-Meyer A, Eifert J, Wang H, Sanglay G, 2018. Volume Estimation of Strawberries, Mushrooms, and Tomatoes with a Machine Vision System. International Journal of Food Properties 21(1):1867-1874.

Demir B, Eski I, Kus ZA, Ercisli S, 2017. Prediction of Physical Parameters of Pumpkin Seeds using Neural Network. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 45(1):22-27.

Ergen Y, Sağlam C, 2005. Bazı Çerezlik Ayçiçeği (*Helianthus annuus* L.) Çeşitlerinin Tekirdağ Koşullarinda Verim ve Verim Unsurlari. Tekirdağ Ziraat Fakültesi Dergisi 2(3):221-227.

Eski İ, Demir B, Gürbüz F, Kuş ZA, Uğurtan Yilmaz K, Uzun M, Ercişli S, 2018. Design of Neural Network Predictor for the Physical Properties of Almond Nuts. Erwerbs-Obstbau 60(2): 153-160.

FAOSTAT, 2018. Food and Agriculture Organization of the United Nations (FAO), http://www.fao.org/site, (accessed date: 22.03.2022).

Hammer Ø, Harper DAT, Ryan PD, 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4(1): 1-9.

IBM SPSS®, 2010. Statistical Software. SSS Inc., IBM Company©, Version 20.0.

Jafari S, Khazaei J, Arabhosseini A, Massah J, Khoshtaghaza MH, 2011. Study on Mechanical Properties of Sunflower Seeds. Food Science and Technology 14(1):1-12.

Kays SJ, 1999. Preharvest Factors Affecting Appearance. Postharvest Biology and Techonolgy 15: 233–247.

Khodabakhshian R, Emadi B, Fard MA, 2010. Some Engineering Properties of Sunflower Seed and its Kernel. Nong Ye Ke Xue Yu Ji Shu, 4(4):37-46.

MacKay DJC, 1998. Introduction to Gaussian processes. NATO ASI Series F Computer and Systems Sciences 168, 133-166.

Malik MA, Saini CS, 2016. Engineering Properties of Sunflower Seed: Effect of Dehulling and Moisture Content. Cogent Food & Agriculture 2(1):1-11.

Maxwell AE, Warner TA, Fang F, 2018. Implementation of Machine-learning Classification in Remote Sensing: An Applied Review. International Journal of Remote Sensing 39:2784-2817.

Mollazade K, Omid M, Arefi A, 2012. Comparing Data Mining Classifiers for Grading Raisins based on Visual Features. Computers and Electronics in Agriculture 84:124-131.

Nettleton DF, Orriols-Puig A, Fornells A, 2010. A Study of the Effect of Different Types of Noise on the Precision of Supervised Learning Techniques. Artificial Intelligence Review 33:275-306.

Omid M, Khojastehnazhand M, Tabatabaeefar A, 2010. Estimating Volume and Mass of Citrus Fruits by Image Processing Technique. Journal of Food Engineering 100(2):315-321.

Ortiz-Hernandez AA, Araiza-Esquivel M, Delgadillo-Ruiz L, Ortega-Sigala JJ, Durán-Muñoz HA, Mendez-Garcia VH, Vega-Carrillo HR, 2020. Physical Characterization of Sunflower Seeds Dehydrated by Using Electromagnetic Induction and Low-Pressure System. Innovative Food Science & Emerging Technologies 60:102285.

Parker JR, 2001. Rank and Response Combination from Confusion Matrix Data. Information Fusion 2(2), 113-120.

Ponce JM, Aquino A, Millán B, Andújar JM, 2018. Olive-Fruit Mass and Size Estimation using Image Analysis and Feature Modeling. Sensors 18(9):2930.

Rasmussen CE, Williams CK, 2006. Gaussian Process for Machine Learning. MIT Press, Cambridge, Massachusetts.

Romero JR, Roncallo PF, Akkiraju PC, Ponzoni I, Echenique VC, Carballido JA, 2013. Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. Computers and Electronics in Agriculture 96:173-179.

Sağlam C, Çetin N, 2022a. Machine Learning Algorithms to Estimate Drying Characteristics of Apples Slices Dried with Different Methods. Journal of Food Processing and Preservation e16496.

Saglam C, Cetin N, 2022b. Prediction of Pistachio (*Pistacia vera* L.) Mass Based on Shape and Size Attributes by using Machine Learning Algorithms. Food Analytical Methods 15(3): 739-750.

Saha KK, Uddin MZ, Rahman MM, Moniruzzaman M, Ali MA, Oliver MMH, 2021. Estimation of Cardamom Capsule Size and Surface Area Using Digital Image Processing Technique. Journal of the Bangladesh Agricultural University 19(3):398-405.

Sahay KM, Singh KK, 2004. Unit Operations of Agricultural Processing. Vikas Publishing House PVT LTD.

Santalla EM, Mascheroni RH, 2003. Equilibrium Moisture Characteristics of High Oleic Sunflower Seeds and Kernels. Drying Technology 21(1):147-163.

Seiler GJ, 2007. Wild Annual Helianthus Anomalus and H. Deserticola for Improving Oil Content and Quality in Sunflower. Industrial Crops and Products 25:95–100.

Singh SK, Vidyarthi SK, Tiwari R, 2020. Machine Learnt Image Processing to Predict Weight and Size of Rice Kernels. Journal of Food Engineering 274:109828.

Stegmayer G, Milone DH, Garran S, Burdyn L, 2013. Automatic Recognition of Quarantine Citrus Diseases. Expert Systems with Applications 40(9):3512-3517.

Tabatabaeefar A, Rajabipour A, 2005. Modeling the Mass of Apples by Geometrical Attributes. Scientia Horticulturae 105(3):373-382.

Vapnik VN, 2000. Methods of Pattern Recognition. The Nature of Statistical Learning Theory. Springer, New York.

Xu G, Shen C, Liu M, Zhang F, Shen W, 2017. A User Behavior Prediction Model Based on Parallel Neural Network and k-Nearest Neighbor Algorithms. Cluster Computing 20(2): 1703-1715.

Zhang H, Song T, Wang K, Wang G, Hu H, Zeng F, 2012. Prediction of Crude Protein Content in Rice Grain with Canopy Spectral Reflectance. Plant, Soil and Environment 58:514-520.

Zhang, R., Ma, J. (2009). Feature Selection for Hyperspectral Data Based on Recursive Support Vector Machines. International Journal of Remote Sensing 30: 3669-3677.