Classification of hazelnut cultivars: comparison of DL4J and ensemble learning algorithms

Caner KOC[1], Dilara GERDAN[1], Maksut B. EMİNOĞLU[1], Uğur YEGÜL[1]*, Bulent KOC[2], Mustafa VATANDAŞ[1]

[1]*Ankara University, Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, Ankara, Turkey;* *ckoc@ankara.edu.tr; dilaragerdan22@gmail.com; eminoglu@agri.ankara.edu.tr: yegul@ankara.edu.tr (*corresponding author);* *vatandas@agri.ankara.edu.tr*
[2]*Clemson University, Department of Agricultural Sciences, Clemson, South Carolina, USA; bulent@clemson.edu*

Abstract

Classification of hazelnuts is one of the values adding processes that increase the marketability and profitability of its production. While traditional classification methods are used commonly, machine learning and deep learning can be implemented to enhance the hazelnut classification processes. This paper presents the results of a comparative study of machine learning frameworks to classify hazelnut (*Corylus avellana* L.) cultivars ('Sivri', 'Kara', 'Tombul') using DL4J and ensemble learning algorithms. For each cultivar, 50 samples were used for evaluations. Maximum length, width, compression strength, and weight of hazelnuts were measured using a caliper and a force transducer. Gradient boosting machine (Boosting), random forest (Bagging), and DL4J feedforward (Deep Learning) algorithms were applied in traditional machine learning algorithms. The data set was partitioned into a 10-fold-cross validation method. The classifier performance criteria of accuracy (%), error percentage (%), F-Measure, Cohen's Kappa, recall, precision, true positive (TP), false positive (FP), true negative (TN), false negative (FN) values are provided in the results section. The results showed classification accuracies of 94% for Gradient Boosting, 100% for Random Forest, and 94% for DL4J Feedforward algorithms.

Keywords: cultivar classification; DL4J; ensemble learning; hazelnut; machine learning

Introduction

Hazelnut (*Corylus avellana* L.) is an important agricultural crop produced by a small number of countries in the world. Shelled hazelnut production in the world in 2017 was 1,006,178 tonnes. Approximately 67% (675,000 tonnes) of this production was from Turkey, and 14% (132,281 tonnes) was from Italy. The remaining 18% share was from Azerbaijan, USA, China, Georgia, Chile, Iran, France, and Spain (Food and Agriculture Organization of the United Nations, 2019). Turkey, Georgia, Italy, and Azerbaijan are the leading countries in hazelnut trade (Ercisli *et al.*, 2011). Hazelnut is mostly produced in Ordu, Giresun, Samsun, Sakarya, Düzce, Trabzon, and Zonguldak provinces of Turkey (Bostan and Güler, 2016).

Knowing the physical properties of agricultural products is important for the design and development of machinery, equipment, and systems for various harvesting and post-harvest processes (Sayıncı *et al.*, 2015). Shape, geometric dimensions, mass, density, hardness, static and sliding coefficient, coefficient of friction, compression strength are the major technical properties of biological materials (Mohsenin, 1980). Among these properties, the shape is a crucial factor for the classification, grading, and packaging of agricultural products (Sayıncı *et al.*, 2015). Agricultural products are cleaned from foreign substances and classified based on their types, size, and quality. Biological materials are classified based on their physical, chemical, and biological properties. Mechanical, thermal, optical, and electrical properties of materials are used in the separation process (Mohsenin, 1980).

The emergence of machine learning, big data technologies, and high-performance computing have enabled new opportunities in multidisciplinary agricultural studies (Liakos *et al.*, 2018). Artificial intelligence and computer vision applications have been increased in recent years. As in every field, applications of these technologies in agriculture have also increased. Non-time-consuming, fast, non-destructive classification models are used for the post-harvest classification of agricultural products. Machine learning is widely used in solving classification problems. For example, Kavdir and Guyen (2004) used fuzzy logic for apple grading. Classification results obtained by the fuzzy logic expert system was 89%. Pronprasit and Natwichai (2013) predicted of mango fruit quality by using NIR spectroscopy and ensemble classification. The standard error of prediction values of the proposed ensemble classifier, naïve ensemble classifier, and single classifier were 0.95, 1.08, and 0.99, respectively. Recent studies underlined that ensemble classifiers could perform well in changing environments. Nandi *et al.* (2014) worked on a computer vision-based system for mango fruit grading with Support Vector Machine (SVM). The accuracy of the system was 90%. Teimouri *et al.* (2014) studied to discriminate normal, broken, split, wrinkled, doubles, or twins and shell of almond using artificial neural networks. Comparing Otsu dynamic thresholding and watershed versus 96%, 99%, and 98% accuracies were obtained from Artificial Neural Network (ANN) model. El-Bendary *et al.* (2015) studied tomato maturity with 250 images. They used principal components analysis (PCA) in addition to support vector machines (SVMs) and linear discriminant analysis (LDA) algorithms for feature extraction and classification. Their results showed 85% accuracy using SVMs algorithm with a linear kernel function and 84% accuracy using LDA algorithm for ripeness classification. Results were found to be 84% successful for LDA and 77% successful for ANN. Brahimi *et al.* (2017) tried to predict the diseases with convolutional neural network (CNN) by using 14828 tomato leaf images, which caught nine different diseases. According to the results, the accuracy of the model was 99%.

Manickavasagan *et al.* (2018) classified palm varieties grown in various regions of Saudi Arabia based on their hardness. Instead of the traditional expensive and time-consuming methods used to determine the quality of dried fruits, they developed an artificial neural network and linear discriminant analysis methods by subtracting histogram and tissue properties from monochrome images of 1800 samples to classify dried fruits as soft, semi-hard, and hard. Ishikawa *et al.* (2018) classified strawberries in their studies by taking advantage of the shape information obtained from digital images. Using SHAPE software (The University of Tokyo 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, JAPAN), they obtained fruit length, width, projection area, and fruit boundary from a total of 2969 images. The machine learning method was successful in diagnosing strawberry fruits with nine different shapes. Vaishnav and Rao (2018) classified 49 (including apple, mango, cherry, orange, and pineapple) fruit images that are obtained from Google. Logistic regression, neural network, K-nearest neighbour, decision tree, random forest, Naive Bayes were used for the classification process, and the data set was run ten-fold cross-validation method. Logistic regression predicted 85% for the first set of experiments and 91% accuracy for the second set of experiments.

Saranya *et al.* (2019) used machine learning and deep learning algorithms for apple, banana, orange, and pomegranate fruit to classify. R, G, B colour, size, height, and width of fruit were used as physical properties in the study. The comparative results of study were 49 for K-nearest neighbour, 61 for support vector machine. On the other hand, classification system accuracy was 97% for convolutional neural network. Shin *et al.* (2019)

used three image processing techniques and two supervised machine learning techniques for early detection of powdery mildew disease. ANN determined the highest accuracy, with a 90% success rate. Datta *et al.* (2019) used image processing and Faster R-CNN to detect cracked eggs. The performance of the damage detection in the test data was evaluated using the mean sensitivity (MAP) value. The network yielded a MAP value higher than 75% and was able to detect most of the damaged side. Caladcad *et al.* (2020) classified Philippine coconuts according to maturity levels using artificial neural network, random forest, and support vector machine. A total of 132,300 data were partitioned 10-fold-cross validation method to estimate maturity levels (pre-mature, mature, and over-mature). The study results that the total accuracy of system was 80%. Also, random forest model predicted 91% and 84% accuracies for training and testing, respectively. Miraei Ashtiani *et al.* (2020) estimated mass of almond kernels using from its shell features such as length, maximum width, and maximum thickness. For this purpose, a multi-linear regression model, multilayer perceptron neural network, radial basis function neural network, adaptive neuro-fuzzy inference system, and support vector machine algorithms were used. The high overall accuracy was 96% for the mass of almond kernels using the radial basis function neural network classifier. A similar study was conducted by Vidyarthi *et al.* (2020). In the study, size and mass of 100 pistachio kernels were predicted using random forest algorithm. The measured length of 18.002 mm for all pistachio kernels used in the study was approximated as 18.608 mm using the random forest classifier.

The objective of this research was to implement a comparative study of machine learning frameworks to classify native hazelnut (*Corylus avellana* L.) varieties using DL4J and ensemble learning algorithms.

Materials and Methods

Load measurement set-up

An inductive force transducer and NI 6009 (National Instruments, USA) data acquisition board were used for load measurements. A testing unit designed for this study was used to compress the hazelnut to measure the forces required to break hazelnuts shells. A 10 mm diameter hydraulic cylinder was mounted at the center of the testing unit for the ease of connection of the inductive force measuring device with the testing unit (Figure 1). The piston rod moved at a constant speed of 1.5 m.s^{-1} to compress the samples while breaking the hazelnut shells. The breaking force of each sample was determined from the force-displacement curve recorded on the computer during compression.
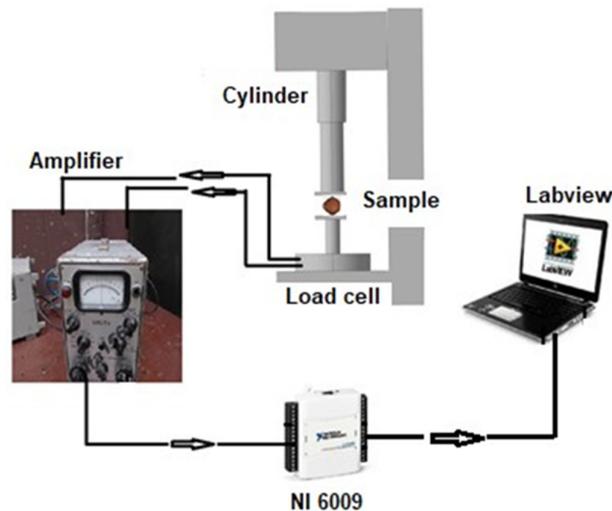

Figure 1. Experimental set-up

Data acquisition system

A graphical program developed in LABVIEW 8.2 (National Instruments Corporation, 11500 N Mopac Expwy Austin, TX 78759-3504, USA) was used for the data acquisition. The developed program facilitates the collection and recording of the inductive load data with the NI 6009 data acquisition board to a personal computer via USB cable. Force was applied until the nutshells were cracked. The force applied to the shell was recorded on the computer in a text format every 5 ms. The developed graphical program also displayed the measurements on the computer screen during the experiments (Figure 2).
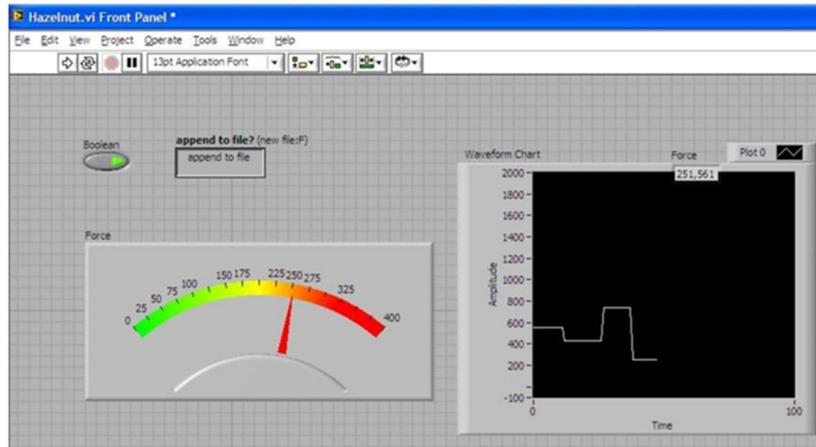

Figure 2. Display of the developed graphical program

Sensor calibration

For sensor calibration, constant weights ranging from 1 kg to 10 kg were used. The analog force value read from the sensor was transmitted to the NI 6009 data acquisition board. The analog data was then converted to digital data with the calibration equation, which was obtained in Excel in Figure 3.
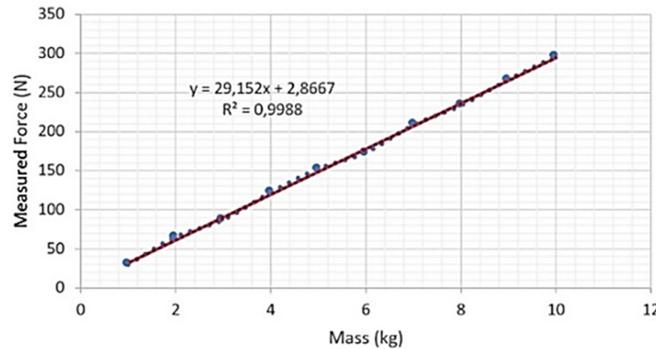

Figure 3. Calibration equation of sensor

Data collection

Three local hazelnut varieties commonly grown in Turkey were selected for the study. These hazelnut varieties are 'Sivri', 'Kara' and 'Tombul' varieties shown in Figure 4. A total of 150 fruits were measured, 50 of each fruit variety were randomly chosen. Basic dimensions of hazelnut maximum length, width, compression strength, weight, and shell thicknesses were measured with a caliper, a precision scale and a force transducer. The compression directions of hazelnut samples are given in Figure 5.
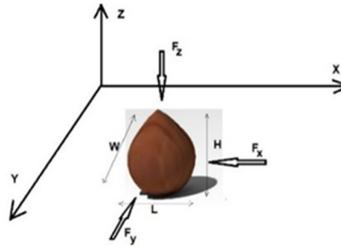
Figure 4. Hazelnut cultivars

Figure 5. Characteristic dimensions of hazelnuts and the compression directions: length (L), width (W) height (H) and forces applied in three directions

Deep learning

Deep learning is a subset of machine learning that uses multi-layered artificial neural networks to provide state-of-the-art accuracy in tasks such as artificial intelligence, object identification, speech recognition, and language translation. Deep learning is used in areas such as automated driving, aerospace and defense, medical research, industrial automation, electronics, and agriculture. Deep learning frameworks include Theano with Keras warper, Torch, Caffe, Tensorflow, and Deeplearning4J (DL4J) (Kovalev *et al.*, 2016). In this study, DL4J Feedforward algorithms were selected for data classification as deep learning frameworks. The main purpose of DL4J feedforward is mostly used for data classification. The algorithm was chosen for high-performance instead of traditional machine learning algorithms.

The first step in building network architecture is to create a network. Generally, neural networks are structured in layers, so the type and number of layers must determine to create a particular network architecture. In the KNIME program, each different layer is represented as its own node. Layer nodes that are intended to be used sequentially to form a network are linked together. The first node in each architecture will always work with a DL4J model initializer that will create an empty deep learning model. This model is used for storing the network and its configuration. Each layer node can have different parameters that can be configured in a particular layer node dialog box. Since this layer will be initiated by a learning node, it is not necessary to add an output layer to an architecture. When using KNIMA, the DL4J Model Initializer is selected and then dense layer is added and the necessary settings were made. Because of this reason, DL4J feedforward architecture, activation function as "ReLU" and weight initialization strategy as "XAVIER" were selected.

Ensemble learning

Ensemble methods aim to improve the performance of a particular statistical learning or model technique (Al-Barazanchi *et al.*, 2017). The general principle of community methods is to create a combination of some methods rather than using a single fit of the method (Bühlmann, 2012). A community (ensemble) is a collection of estimators (average of all estimates) that come together to give a final estimate. The reason we use the ensemble is that many different determinants used to predict the same target variable will do a better job than any single predictor alone. Ensemble techniques are classified as Boosting, Bagging (Random Forest), AdaBoost, and Stacking (blending, MAVL) (Seker and Erdogan, 2018).

Bagging

Bagging is a method that is generating multiple versions of a predictor and using these to get an aggregated predictor (Aoki *et al.*, 2015). Also, Bagging is a simple collection technique in which we build many independent determinants/models/learners and combine them using some average model techniques. Bagging that often considers homogeneous weak learners learns them independently from each other in parallel and combines them following deterministic averaging process (Shi *et al.*, 2018). Each observation is selected by modifying it to be used as input for each of the models. Thus, each model will have different observations based on the bootstrap process. This technique reduces error by reducing variance, as it takes many unrelated data to generate a final model. An example of a bagging community is a random forest model. In this study, the random forest algorithm was selected as the bagging method. The random forest algorithm developed by Leo Bieman generates multiple trees to solve a problem and creates different decision trees. Random forest algorithm can be used in both classification and regression problems. Random forest is an advanced version of the classification and regression trees (CART) algorithm, in which many trees are created based on subsets of data. At each node, branches are formed according to the criteria of the CART algorithm (GINI index). The GINI index measures class homogeneity and can be expressed by the following equation (Eq. 1, 2, and 3) (Salzberg, 1994).

$$\text{Gini impurity (p)} = 1 - \sum_{i=1}^{N} p_i^2 \tag{1}$$

In equation 1, '*p*' is the whole dataset, '*N*' is the number of classes, and '*pi*' is the frequency of class '*i*' in the same dataset.

$$\text{Gini index} = \sum_{j=1}^{K} w_j \text{ Gini impurity (j, after)} \tag{2}$$

$$w_j = \frac{\text{\# data in subset (j,after)}}{\text{\# data in dataset (before)}} \tag{3}$$

Here, K is the number of subsets generated by the split, and (j, after) is subset j after the split.

T training data set, *C_i* and *(f(C_i, T)* shows the probability that the selected sample belongs to the "*C_i class*" (Akar and Güngör, 2012).

Gradient boosting (Boosting method)

Boosting is a very successful technique for solving the two-class classification problems. Boosting is a community in which predictions are made sequentially, not independently. This technique uses the logic that subsequent estimators learn from the mistakes of previous estimators. Therefore, the likelihood of observations in subsequent models is uneven, and those with the highest errors appear most. The estimators can be selected from a range of models such as decision trees, regressors, and classifiers as the new estimators learn from the mistakes of the previous estimators. Gradient boost is an example of the acceleration algorithm. In this study, the gradient boosting algorithm was selected as a boosting method. The gradient boosting can be expressed by the following equation (Eq. 4, 5, and 6) (Natekin and Knoll, 2013).

Initialize $f_{k0} = 0, k = 1,2, \dots K$

For   m = 1 to M:

1. set $$p_k(x) = \frac{e^{f_k(x)}}{\sum_{l=1}^{K} e^{f_l(x)}}, k = 1,2, \dots, K \tag{4}$$

2. For   k = 1 to K;

    a.   compute $r_{ikm} = y_{ik} - p_k(x_i), i = 1,2, \dots, N$

    b.   Fit a regression tree to the targets $r_{ikm}, i = 1,2, \dots, N$

    c.   Giving terminal regions $R_{jim}, j = 1,2, \dots, J_m$

    d.   compute $\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{x_i \in R_{jkm}}(r_{ikm})}{\sum_{x_i \in R_{jkm}} |r_{ikm}|(1-|r_{ikm}|)}, j = 1,2, \dots, J_m$

    e.   Update $$f_{km}(x) = f_{k,m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jkm} I(x \in R_{jkm}) \tag{5}$$

Output $$\hat{f}_k(x) = f_{kM}(x) , k = 1,2, \dots , K \tag{6}$$

Data analysis

Before starting the data analysis, the ''statistic'' node was connected to the Knime workflow to find the missing values, and then the data set was partitioned 10-fold-cross validation method to prevent overfitting. The values in the dataset were labeled in Excel file as follows (Table 1). DL4J Feedforward, random forest, and gradient boosting machine were connected to workflow, respectively, in Figure 6.
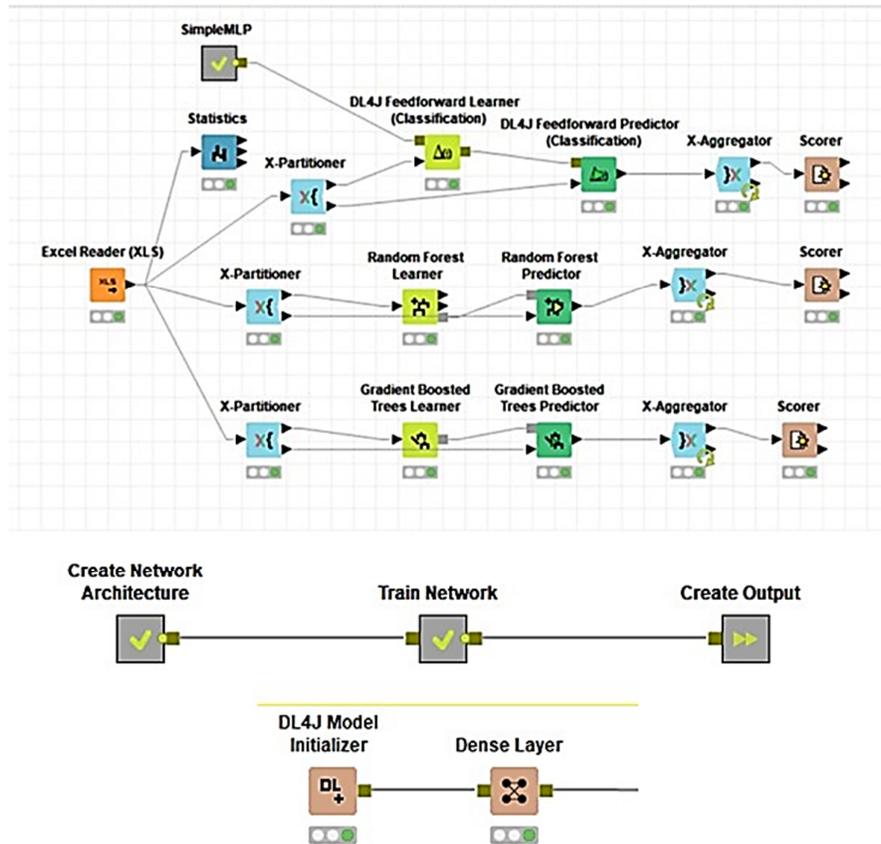

Figure 6. KNIME workflow

At the end of analysis, accuracy, Cohen's Kappa, TP, FP, TN, FN, P, R, and F-means values were given in the relevant part of the study. The "scatter plot" node was connected to provide a more visual representation of the results.

One of the classifier performances, accuracy, shows how accurate the system can predict. Recall or sensitivity or true positive rate (TPR) is the value that the system is predicting what rate is the passing assessment of the overall passing assessments. Precision or predicted position value (PPV) is the value that indicates how true or correct the system can predict. Also, F-measure is the overall efficiency assessment derived from the means of precision and recall (Takran *et al.*, 2017).

$$\text{Accuracy(1-Error)} = \frac{TP}{(TP+TN+FP+FN)} \tag{7}$$

$$\text{Recall} = \frac{TP}{(TP+FP)} \tag{8}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \qquad (9)$$

$$\text{F} - \text{Measure} = \frac{2\text{xPrecisionxRecall}}{\text{Precision+Recall}} \qquad (10)$$

Results and Discussions

The classifier performances, which are accuracy, error, F-measure, Cohen's Kappa, recall, precision, TP, FP, TN, FN values were given in Tables 1-2. As a result of the comparison, it was found that the Random Forest (Bagging method), was better than the other algorithm. When the shell compression strength was considered only, boosting method was found to be successful with 72%, Bagging method (Random forest) with 83%, and DL4J (Deep learning algorithm) with 71% accuracy. In cases where the compression strength value is higher than 431.212 in the Z-axis, the 'Sivri' hazelnut was classified as 100% also, in cases where the compression strength value is less than 103.22 'Tombul' hazelnut was classified as 100%. When hazelnut weights were added in addition to shell compression strength values, GBM method was found to be successful with 85%, Random Forest method with 87%, and DL4J with 73% accuracy. It was found that there is a parameter that increases the success rate in weight classification. On the other hand, in addition to the shell compression strength values and weight of hazelnut shell thickness, a Gradient Boosting method was found to be successful with 95%, Random Forest method with 100%, and DL4J with 95% accuracy. Comparison of deep learning and ensemble learning algorithm values are given in Figure 7.
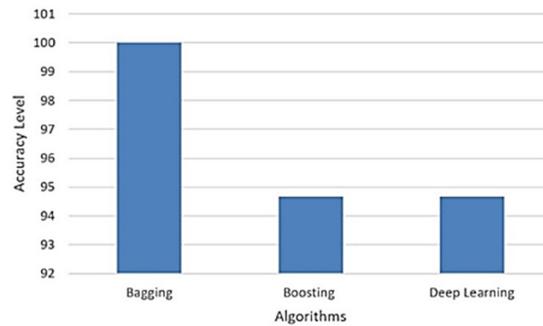

Figure 7. Comparison of deep learning and ensemble learning algorithm values

Menesatti *et al.* (2008) evaluated the potential use and effectiveness of shaped techniques to distinguish four traditional Italian cultivars. Approximately 400 hazelnut color RGB images were analyzed by a morphological method based on the elliptic Fourier approach to closed contours in a two-dimensional plane. The coefficients of the harmonic equations were analyzed by partial least square discrimination analysis (PLS-DA). The results showed that the percentage of correct classification for a lateral view is high (77% to 99% in the independent test). The hazelnut samples selected in the study have success rates varying between 94% and 100% according to the classification parameters considered. A higher success rate was obtained than the classification made based on the RGB image processing methods reported in the literature. At the same time, good positioning results (95 and 97 respectively) were found for random positioning appearance, especially for in-shell kernel between two round varieties and between two long varieties. Guvenc *et al.* (2015) studied on the classification of partly shell removed hazelnut kernel, shell removed, and rotten hazelnuts kernels by using a computer vision approach. In the proposed method, the processed hazelnut kernels were classified with 94% accuracy. This ratio is parallel to the classification results in the study. Giraudo *et al.* (2018) developed an automated identification of defective hazelnut kernels based on multivariate analysis of RGB images through the principal component analysis and subsequently used to build classification models. Their system success was approximately found as 97% of the test samples. The 97% success rate obtained in this study was a little more successful than the gradient boosting machine (94%) and DL4J Feedforward (9%) algorithms presented

in the literature. However, it was lower than the random forest (100%) algorithm. Solak and Altınışık (2018) worked on the classification of hazelnut cultivars by using image processing and mean-based, k-means clustering technique. Hazelnut fruits were classified with 100% accuracy using image processing techniques. A comparison of mean-based and K-means clustering methods was found that 90% to 100% accuracy. The classification results in this study are also close to the results of Solak and Altinisik (2018). Jalali and Mahmoudi (2013) worked on pistachio nut cultivars sorting by data mining and fuzzy logic classifier. A correct classification rate of 93% for training and 97% for the testing set was observed. Confusion matrix was selected as the model performance criteria of algorithms for compare to success rate. The performance evaluations and accuracy criteria of the algorithm were shown in Tables 1 and 2.

Table 1. The classifier performance of algorithms

Confusion Matrix	Algorithms	CC	WC	Accuracy (%)	Error (%)	FM	CK
'Sivri'	DL4J	71	4	95%	5.333	0.923	0.92
'Kara'						0.936	
'Tombul'						0.98	
'Sivri'	Random Forest	75	0	100%	0	1	1
'Kara'						1	
'Tombul'						1	
'Sivri'	Gradient Boosting	71	4	95 %	5.333	0.96	0.92
'Kara'						0.958	
'Tombul'						0.923	

*CC: Correct Class, WC: Wrong Class, FM: F-Measure, CK: Cohen's Kappa

Table 2. The classifier performance of algorithms

Algorithms	Cultivars	Recall	Precision	Sensitivity	TP	FP	TN	FN
	'Sivri'	0.96	0.889	0.96	24	3	47	1
	'Kara'	0.88	1	0.88	22	0	50	3
DL4J	'Tombul'	1	0.962	1	25	1	49	0
	'Sivri'	1	1	1	25	0	50	0
	'Kara'	1	1	1	25	0	50	0
Random Forest	'Tombul'	1	1	1	25	0	50	0
	'Sivri'	0.96	0.96	0.96	24	1	49	1
	'Kara'	0.92	0.958	0.92	23	1	49	2
Gradient Boosting	'Tombul'	0.96	0.923	0.923	24	2	48	0

*TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative

Conclusions

There are many studies on the post-harvest classification of various agricultural products. Researchers have focused on techniques and methods for developing prototypes capable of non-destructive classification. Among these methods, artificial intelligence and machine learning stand out as current issues. These methods are the primary issues used in developing solutions for the classification and automation of agricultural products. In this study, three different hazelnut cultivars ('Sivri', 'Kara', 'Tombul') were classified based on five attributes of hazelnuts using the most common classification and prediction algorithms. The best prediction was found by the Bagging method. The results showed 100 % accuracy for Random Forest, 95% accuracy for gradient boosting machine, and 95% accuracy for the DL4J Feedforward algorithm, respectively. With the integration of classification machines, machine learning has a new future for automated decision-making or support to provide practical tools for knowledge-based agricultural practices. The results of this research can be used in future studies by the researchers for classification issues.

Authors' Contributions

Conceptualization: MV, DG, Data curation: DG, CK, Formal analysis: DG, CK, UY, MBE, Funding acquisition: CK, UY, MBE, Investigation: CK, DG, UY, MBE, MV, BK, Methodology: DG, CK, Software: DG, Writing - original draft: DG, Writing-review and editing: BK, MBE, UY, MV. All authors read and approved the final manuscript.

Acknowledgements

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

Akar Ö, Güngör O (2012). Classification of multispectral images using random forest algorithm. Journal of Geodesy and Geoinformation 1(2):139-146. *https://doi.org/10.9733/jgg.241212.1*

Al-Barazanchi KK, Al-Neami AQ, Al-Timemy AH (2017). Ensemble of bagged tree classifier for the diagnosis of neuromuscular disorders. In: Proceedings of the 4th International Conference on Advances in Biomedical Engineering. Beirut, Lebanon pp 1-4.

Aoki K, Sakamoto M, Furutani H (2015). Analysis of genetic disease haemophilia a by using machine learning. Journal of Robotics Networking and Artificial Life 2(2):115-119. *https://doi.org/10.2991/jrnal.2015.2.2.11*

Bostan SZ, Güler SK (2016). Quality changes during in-shell storage of some hazelnut cultivars. Bahçe 45(2):41-53.

Brahimi M, Boukhalfa K, Moussaoui A (2017). Deep learning for tomato diseases: classification and symptoms visualization. Applied Artificial Intelligence 31(4):299-315. *https://doi.org/10.1080/08839514.2017.1315516.*

Bühlmann P (2012). Bagging, boosting and ensemble methods. In: Handbook of Computational Statistics. Springer, Berlin, Heidelberg.

Caladcad JA, Cabahug S, Catamco MR, Villaceran PE, Cosgafa L, Cabizares KN, ... Piedad EJ (2020). Determining Philippine coconut maturity level using machine learning algorithms based on acoustic signal. Computers and Electronics in Agriculture 172:105327. *https://doi.org/10.1016/j.compag.2020.105327*

Datta AK, Botta B, Gattam SSR (2019). Damage detection on chicken eggshells using Faster R-CNN. American Society of Agricultural and Biological Engineers 1:1

El-Bendary N, El Hariri E, Hassanien AE, Badr A (2015). Using machine learning techniques for evaluating tomato ripeness. Expert Systems with Applications 42(4):1892-1905. *https://doi.org/10.1016/j.eswa.2014.09.057*

Ercisli S, Ozturk I, Kara M, Kalkan F, Seker H, Duyar O, Erturk Y (2011). Physical properties of hazelnuts. International Agrophysics 25(2):115-121.

FAOSTAT (2019). Food and Agriculture Organization of the United Nations. Retrieved 2020 August 1 from *http://www.fao.org/faostat/en/#data/QC/visualize*

Giraudo A, Calvini R, Orlandi G, Ulrici A, Geobaldo F, Savorani F (2018). Development of an automated method for the identification of defective hazelnuts based on RGB image analysis and colour grams. Food Control 94:233-240. *https://doi.org/10.1016/j.foodcont.2018.07.018*

Guvenc SA, Senel FA, Cetisli B (2015). Classification of processed hazelnuts with computer vision. In: Proceedings of the 23th Signal Processing and Communications Applications Conference. Malatya, Turkey, pp 1362-1365.

Ishikawa T, Hayashi A, Nagamatsu S, Kyutoku Y, Dan I, Wada T, ... Kochi N (2018). Classification of strawberry fruit shape by machine learning. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences 42(2):463-470. *https://doi.org/10.5194/isprs-archives-XLII-2-463-2018*

Jalali A, Mahmoudi A (2013). Pistachio nut cultivars sorting by data mining and fuzzy logic classifier. International Journal of Agriculture and Crop Sciences 5(2):101-109.

Kavdir I, Guyer DE (2004). Apple grading using fuzzy logic. Turkish Journal of Agriculture and Forestry 27(6):375-382.

Kovalev V, Kalinovsky A, Kovalev S (2016). Deep learning with theano, torch, caffe, tensorflow, and deeplearning4J: Which one is the best in speed and accuracy? In: Proceedings of the XIII International Conference on Pattern Recognition and Information Processing. Minsk, Belarus, pp 99-103.

Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D (2018). Machine learning in agriculture: A review. Sensors 18(8):2674-2703. *https://doi.org/10.3390/s18082674*

Manickavasagan A, Al-Shekaili NH, Al-Mezeini NK, Rahman MS, Guizani N (2018). Computer vision technique to classify dates based on hardness. Journal of Agricultural and Marine Sciences 22(1):36-41.

Menesatti P, Costa C, Paglia G, Pallottino F, D'Andrea S, Rimatori V, Aguzzi J (2008). Shape-based methodology for multivariate discrimination among Italian hazelnut cultivars. Biosystems Engineering 101(4):417-424. *https://doi.org/10.1016/j.biosystemseng.2008.09.013*

Miraei Ashtiani SH, Rohani A, Aghkhani MH (2020). Soft computing-based method for estimation of almond kernel mass from its shell features. Scientia Horticulturae 262:109071. *https://doi.org/10.1016/j.scienta.2019.109071*

Mohsenin NN (1980). Physical properties of plant and animal materials. Gordon and Breach Science Publishers, One Park Avenue, New York, USA.

Nandi CS, Tudu B, Koley C (2014). Computer vision-based mango fruit grading system. In: Proceedings of the International Conference on Innovative Engineering Technologies. Srinagar, India, pp 28-29.

Natekin A, Knoll A (2013). Gradient boosting machines, a tutorial, Frontiers in neurorobotics 7(21):1-21. *https://doi.org/10.3389/fnbot.2013.00021*

Pronprasit R, Natwichai J (2013). Prediction of mango fruit quality from Nir spectroscopy using an ensemble classification. International Journal of Computer Application 83(14):25-30.

Salzberg SL (1994). C4.5: Programs for Machine Learning by J. Ross Quinlan. Mach Learn 16:235-240. *https://doi.org/10.1007/BF00993309*

Saranya N, Srinivasan K, Kumar SP, Rukkumani V, Ramya R (2019). Fruit classification using traditional machine learning and deep learning approach. In: Proceedings of the In International Conference on Computational Vision and Bio Inspired Computing. Springer, Cham pp 79-89.

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

Seker SE, Erdogan D (2018). End-to-end data science with KNIME. Demet Erdoğan Publishing, Turkey.

Shi L, Duan Q, Zhang J, Xi L, Ma X (2018). Rough set-based ensemble learning algorithm for agricultural data classification. Filomat 32(5):1917-1930. *https://doi.org/10.2298/fil1805917s.0*

Shin J, Chang YK, Nguyen-Quang T, Heung B, Ravichandran P (2019). Optimizing parameters for image processing techniques using machine learning to detect powdery mildew in strawberry leaves. American Society of Agricultural and Biological Engineers 1:1.

Solak S, Altınışık U (2018). Detection and classification of hazelnut fruit by using image processing techniques and clustering methods. Sakarya University Journal of Science 22(1):56-65. *https://doi.org/10.16984/saufenbilder.303850*

Takran T, Chartrungruang B, Tantranont N, Somhom S (2017). Constructing a Thai homestay standard assessment model by implementing a decision tree technique. International Journal of the Computer, the Internet and Management 25(2):106-112.

Teimouri N, Omid M, Mollazade K, Rajabipour A (2014). A novel artificial neural network assisted segmentation algorithm for discriminating almond nut and shell from background and shadow. Computers and Electronics in Agriculture 105:34-43. *https://doi.org/10.1016/j.compag.2014.04.008*

Vaishnav D, Rao BR (2018). Comparison of machine learning algorithms and fruit classification using orange data mining tool. In: Proceedings of the 3rd International Conference on Inventive Computation Technologies. Coimbatore, Tamilnadu, India pp 603-607.

Vidyarthi SK, Tiwari R, Singh SK, Xiao H (2020). Prediction of size and mass of pistachio kernels using random Forest machine learning. Journal of Food Process Engineering 43(9):e13473. *https://doi.org/10.1111/jfpe.13473*