Journal of Biodiversity and Environmental Sciences (JBES)
ISSN: 2220-6663 (Print) 2222-3045 (Online)
Vol. 11, No. 5, p. 258-267, 2017
http://www.innspub.net

INNSPUB

RESEARCH PAPER OPEN ACCESS

Comparison of some of the data mining Algorithms in classifying charleston gray watermelon variety using morphological properties

Amir Alipasandi, Asghar Mahmoudi*, Hossein Behfar, Hossein Ghaffari

Agriculture Department, Faculty of Biosystems Engineering, University of Tabriz, Iran

Article published on November 30, 2017

Key words: Data mining, Classification, Sensory evaluation, Watermelon, Morphological properties

Abstract

It is predicted that crops like watermelon will be decreased in coming years and price of this product in the market will increase. This highlights the necessity of measures to choose high-quality watermelons by the final consumer. Also according to the concept of virtual water, sorted and more desirable watermelons can be exported at higher prices in order to create more benefit. In general, the aim of this study is to provide a measure that is based on Charleston Gray watermelon morphological characteristics and evaluation of the classification ratio in unripe, ripped and overripe classes, by data mining algorithms. The results of the sensory evaluation showed that experts (human) were able to classify 52% of the samples correctly. The correct classification algorithm K Nearest Neighbor was significantly higher than the classification of LVQ Neural Networks and Discriminant Analysis but classification results of different distance metrics of this algorithm showed no significant difference using them. The highest correct classification with the amount of 67.3 percent belonged to Support Vector Machine algorithm with Gaussian kernel function. Although at first glance it may seem that, this amount is far from ideal but it should be noted that this amount is 15% higher than the classification made by humans. Incidentally, this classification was done based on morphological characteristics of samples which measuring them does not require sophisticated tools and methods.

***Corresponding Author:** Asghar Mahmoudi ✉ a_mahmoudi@tabrizu.ac.ir

Introduction

Watermelon quality at the time of consumption depends on its ripeness. Recognizing the stage of harvesting to pick ripped fruits is very important because the competitive market demands high-quality products which have equal or less price for similar products. Watermelons of a specific kitchen garden are usually harvested once or twice merely according to their weight. This can be due to lowering labor costs, increasing harvesting speed and other factors causes watermelons with varying degrees of ripeness available in the market. If it is possible to recognize watermelon ripeness stage or determine its harvest time, in addition to increasing consumer satisfaction, it can be possible to take advantage of the alterant industries alongside the fresh product sale. On the other hand, given that watermelon is a plant that needs plenty of water during its growth period and due to the incoming drought crisis, it's predicted that planting crops such as watermelon is projected to decline in the coming years and the price of this product will rise increasingly in the market. This highlights the necessities of existence criteria for selecting a high-quality watermelon for the end consumer.

There are several methods available to recognize ripped watermelon from unripe watermelon which shows different stages of growth. Some people use changing the color of ivy from green to brown near the stem connecting to fruit. Some recognize the amount of ripeness and harvest time from the fruit size, exterior properties such as the appearance of the skin and color, but it is very difficult to judge watermelon ripeness using these features. The most common method used traditionally by people to determine watermelon ripeness is tapping onto it and judging using the created sound. In addition to human factor error, this may be just a good way for people with a lot of experience and expertise (Stone *et al.*, 1996). In most of these methods, recognizing ripped from unripe watermelon is very difficult and requires a lot of experience and expertise. In addition, there is no guarantee for the correct identification and it's likely that the person choice may be wrong and the choice may be left to chance. Another method is vibrational stimulation which in this way after impulsing the generated vibration is measured by accelerometer. This method also has limitations which the most important of them is its necessity of attaching accelerometer on watermelon surface. Also, the mass of accelerometer can cause errors (Muramatsu *et al.*, 1997; Nourain *et al.*, 2004).

In addition, by impacting the excitation energy focuses on a small band of frequency and time this causes limitations in the precise determination of the parameters (Taniwaki *et al.*, 2009). Studies have also been conducted to assess the quality and classification of watermelon. Diezma *et al.*, (2004) Studied internal quality in a seedless watermelon by acoustic impulse response and spectral parameters and find a significant correlation between maximum frequency and watermelon ripeness. In this research, frequency domain features were used for this purpose. Abbaszadeh *et al.* (2011) have applied Laser Doppler Vibrometry technology (LDV) to evaluate the ripeness of watermelons. The sugar and firmness of the samples were measured as ripeness indices. The coefficients of determination and the mean square error for the estimation of the fruit sugar were 0.9 and 0.79 respectively. For the estimation of the fruit firmness, they were 0.89 and 0.035. In another study, they (Abbaszadeh *et al.*, 2012) predicted the customer satisfaction of watermelon on the basis of data obtained from sensory evaluation using a fuzzy logic model. In this study, 43 watermelons of the Crimson Sweet variety were subjected to sensory evaluation. Samples classified in ranges of ripeness based on common quality indices such as sweetness, flavor, color and texture and also customer satisfaction. The results showed high accuracy of the fuzzy model in implementation a model which correlate qualitative features and customer satisfaction.

Morphological features are widely used in automated grading, sorting and detection of objects in the industry. Shape features are physical dimensional measures that characterize the appearance of an object. Area, perimeter, major and minor axes lengths, as well as the aspect ratio are some of the most commonly measured morphological features. Many studies have been conducted on the wide range of agricultural products to achieve classification objectives using morphological features. Fruits like apple (Currie *et al.* 2000; Xiaobo *et al.* 2008), citrus (Ding *et al.* 2000; Blasco *et al.* 2009) and watermelon (Sadrnia *et al.*, 2007; Koc 2007) Vegetables like Eggplant (Nunome *et al.*, 2001), Pepper (Zygier *et al.* 2005), Bell pepper (Ngouajio *et al.* 2003), Tomato (Morimoto *et al.* 2000; Brewer *et al.*, 2007; Xiao *et al.*, 2008), Nuts like Pistachio (Ghazanfari *et al.*, 1997), Hazelnut (Menesatti *et al.* 2008), Almond (Antonucci *et al.*, 2011), Various Cereals (Choudhary *et al.* 2008), Rice (Yadav and Jindal 2001; Rabiei *et al.*, 2004; Zheng *et al.*, 2007) are subjects of the studies that morphological characteristics have been used for classification applications using various techniques.

Most studies that relate to the classification of agricultural products are considered as an application of supervised learning in data mining techniques. Supervised learning is essentially a synonym for classification. Classification is the process of finding a model in which classes are detected and concepts are described. These models are based on the analysis of a training data set and it's used to predict the class of instances whose class is unknown. This model may take different forms, such as Decision Trees, Neural Networks, Bayesian classifiers, Support Vector Machines, K Nearest Neighbor and mathematical formulas (Han, 2011).

In general, the aim of this study was to provide the criteria which based on the morphological characteristics of the Charleston Gray watermelon through evaluating classification rate of different data mining algorithms in unripe, ripped and over ripped classes. Morphological characteristics are criteria for non-destructive assessments and there is no need for complex technology to measure them.

Materials and methods

Sample preparation

Training a classifier model is conducted by providing the appropriate samples and after that, the model becomes ready to respond to the examples that did not participate in the training process and the model will face them in the future. Samples must be provided in a way that can describe perfectly the category which they are categorized. In this regard, in three stages, a total of about 2500kg of Charleston Gray watermelon was prepared from kitchen garden located in Parsabad city of Ardabil province.

In the first stage, which was coincided with the beginning of the harvest season, mature watermelons were selected with farmhand help. In fact, it can be said this category includes quite premature watermelons which were few days left to ripen but they have the appropriate size and weight. From consumer's perspective, these watermelons in terms of texture, color, and taste are considered unripe. It is clear that choosing watermelons of this category is very difficult even for the farmer who has a lot of experience and skill in this work in this stage picking watermelon is done often based on criteria such as shape, size, weight and most importantly elapsed days from the planting.

Two weeks after the first stage, the harvest of watermelons that was completely ripped from farmer's point of view started at intervals of two days in three steps. At this stage in the samples selection, many efforts were made to make no significant difference in samples size with first stage watermelons.

The third category is related to over ripped watermelons. Watermelon usually become over ripped when it left over in the kitchen garden for various reasons. At this time, the area of the stem that is attached to the fruit is gradually become dried and the connection between watermelon and soil is interrupted. After that, the food that is needed for plant survival supplied from the plant itself instead of soil. The occurrence of such situation along physiological changes makes sugar converts into starch which leads to spongy texture and finally fermentation that is by no means desirable to the consumer. In order to prepare the over ripped samples in the third stage, the farmer was asked to choose such watermelons based on the time component as well as the criteria he has gained with his personal experience. Also, numbers of second stage samples were kept in the laboratory at the room temperature for 15, 30 and 45 days at separate time intervals to ensure that after the end of the harvest season there will be some samples available for training model. A portion of these samples is shown in Fig. 1.

Fig. 1. Stored samples to provide over ripped class members.

Morphological measured parameters

Considering the oval shape of Charleston Gray variety, parameters such as perimeter in the thickest part, the maximum length of samples and their weight were measured. The ratios of the perimeter to length, weight to perimeter and weight to length were obtained from the above measurements and used in models formation. Table 1 presents the statistical characteristics of the above parameters.

Sensory evaluation for formation the categories

Because of a large number of quality parameters of food that is felt in the mouth and considering that chewing is a destructive process, performing destructive tests seems a logical way to evaluate food quality (Bourne, 2002). For this purpose, many panel tests formed to sensory evaluation and to identify the label of samples class. In this way, each panel was composed of five members and from the general public and its members changed in each test so that people do not find a subjective background to the type and category of watermelons. Before and after cutting watermelons, comments were recorded.

Table 1. Statistical characteristics of measured parameters.

	Length (cm)	Perimeter (cm)	Weight (kg)	$\frac{\text{Weight}}{\text{Perimeter}}$	$\frac{\text{Weight}}{\text{Length}}$	$\frac{\text{Perimeter}}{\text{Length}}$
Maximum	60	68	11.3	0.17121	0.2132	0.7647
Minimum	30	40	3.05	0.06489	0.0847	0.7119
Average	40.41	52.9	5.47	0.102	0.1337	0.3218

Comments before cutting watermelon used to compare the accuracy of classification algorithms with the accuracy of human categorization and comments that were recorded after cutting watermelon used to determine the class of each sample. At the end of the sensory evaluation, the number of samples in each class was determined. Accordingly, 98, 119 and 83 samples were placed in unripe, ripped and over ripped classes respectively. Fig. 2 shows one of the test panels have been formed.

Method of scoring

In this study, the method of scoring in the panel tests is slightly different. The assignment of scores is such that, before cutting watermelons, people were allowed to use their knowledge and experience to estimate the amount of watermelon ripeness. Score 3 was assigned to fully ripped watermelon. Score five was assigned to fermented watermelons and score 1 was assigned to fully unripe watermelons. Score 2 and 4 had also interstitial state. After cutting watermelons, the scores were assigned based on criteria such as color, texture, sweets, flavor, and aroma. Table 2 shows scores assignment for 5 sample watermelons.

Triple Classes Formation

To put samples in separate classes, there should be a kind of discrepancy based on the assigned scores.

In the present study, the mean of the scores was used as a criterion for discretization in order to form classes. So that if this value is within the range of [1, 2.5] then the sample is placed in the unripe class and if this value is within range of [2.6, 3.4] the sample is placed in the ripped class. The value of this quantity in the range of [3.5, 5] is a reflection of an over ripped watermelon. This discrepancy, however, creates crisp and fragile boundaries which may face border samples with serious challenge but to categorize the samples in the three classes we have to do similar works.

Fig. 2. Panel test formation and sensory evaluation.

Table 2. scores assignment and classification for 5 sample watermelons.

	Expert comments before cutting							Expert comments after cutting						
No.	Scores					Average	Class	Scores					Average	Class
1	5	3	4	3	2	3.4	ripped	2	2	2	2	2	2	unripe
10	3	3	3	3	4	3.2	ripped	2	2	2	2	2	2	unripe
20	5	3	4	3	3	3.6	over ripped	3	3	3	3	3	3	ripped
30	1	2	2	2	2	1.8	unripe	5	5	4	4	5	4.6	over ripped
40	5	4	4	3	2	3.6	over ripped	2	2	2	3	2	2.2	unripe

Comparison of classification algorithms

Evaluation criterion in this study is the classification which is obtained from scores that people have been assigned to the quality indices of samples after cutting watermelons. The results of classification studies can be summarized in the form of confusion Matrix. In the present study, the overall performance of each classification algorithms was evaluated based on this matrix and its algorithm correct classification. In an ideal classifier, most of the samples placed on the main diagonal of the confusion matrix and it is desirable that the rest of the elements except the main diagonal of this matrix have zero or near zero values.

Classification using discriminant analysis

This method is similar to Multiple Regression with this difference that in Multiple Regressions, the dependent variable is always a quantitative variable with normal distribution while in this analysis, the dependent variable does not only have any normal distribution but is a qualitative variable with finite levels. This method is for combining the $x_1, x_2, \ldots, x_n$ variables to obtain the new variable y. The purpose of the Discriminant Analysis is to find a linear function such as Equation 1

$$Y=\beta+\beta_1x_1+\beta_2x_2+\ldots+\beta_nx_n \quad 1$$

So that the probability of Equation 2 becomes maximum.

$$P(Y=y|(X_1,X_2 \ldots X_n)=(x_1,x_2 \ldots x_n)) \quad 2$$

In the case where the dependent variable contains k level (class), the aim is to attribute new observations (x1, x2,..., xn) to one of the k-groups based on y (discriminant function). In this study, the stepwise discriminant analysis was performed using SPSS 16 software.

Classification using LVQ Neural Network

Different types of neural networks can be used to pattern recognize applications and samples classification. The present study uses LVQ networks. This particular kind of neural network is the generalization of SOM (Self-Organizing Maps) idea to solve supervised learning problems.

The first layer of these networks is a competitive layer and their output layer is a linear layer. The main application of this kind of neural network is to solve the classification problems that cover a wide range of applications of intelligent systems.

To design these types of neural networks, several algorithms have so far been suggested that the LVQ1 and LVQ2 algorithms are the most popular types of these algorithms for training this type of neural network. These algorithms were implemented in Matlab 2015b software and used for samples classification.

Classification Using K Nearest Neighbor

The KNN classifiers do their work based on similarity. This is done by comparing a given test tuple with training tuples that are similar to it. When this algorithm faces an unknown tuple, a k-nearest-neighbor classifier searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k “nearest neighbors” of the unknown tuple. As in the present study, a good value for k can be determined experimentally. Starting with k = 1, we use a test set to estimate the error rate of the classifier. This process can be repeated each time by incrementing k to allow for one more neighbor. The k value that gives the minimum error rate may be selected. The KNN classifiers use distanced based metrics to classify samples. This algorithm, along with the Euclidean, Manhattan, Cosine and Chebyshev distances implemented in Matlab 2015b software, was used to classify the samples.

Classification Using Support Vector Machine

This algorithm is a method used to classify both linear and nonlinear data. Briefly, SVM uses a nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for the linear optimal separating hyperplane. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane. The SVM finds this hyperplane using support vectors (essential training tuples) and margins (defined by the support vectors). Because SVM algorithms are capable of modeling complicated nonlinear decision boundaries, even the fastest ones can have low speeds during training, but they generally have high accuracy. Unlike back propagation neural networks that reach many local optimizations, SVM training always finds a global solution. In this study, SVM algorithm with Linear, Quadratic, Cubic and Gaussian kernel functions implemented in Matlab 2015b software was used for classifying samples.

Comprehensiveness and generality of the models

Models built only on the basis of training data, generally have high correlation coefficients and high classification rate, due to their overfitting but these coefficients are valid only for test data. When using validation techniques, the correlation coefficients of the model, as well as the rate of the correct classification, is reduced, however, the comprehensiveness of the model is enhancing and during faces with samples that did not participate in the training process, it responds much better. This makes the proposed model valid. In this study, k-fold cross-validation was used for models validation and the value of the parameter k is considered 10 for cross-validation. In general, k-fold cross-validation method due to low variance is recommended for accuracy estimation of a classification model.

Results and discussion

Classification by Panel Test

Table 3 shows the confusion matrix resulting from experts classification. As can be seen, 45 samples of 98 samples are categorized correctly in their own class but 51 of these 98 samples which belong to the unripe class are mistakenly categorized by experts in the ripped class. Also, only two samples of the unripped class are classified incorrectly in the over ripped class.

Table 3. Panel test confusion matrix.

	Unripe	Ripped	Over ripped
Unripe	45 45.92 %	51 50.04 %	2 2.04 %
Ripped	24 20.18 %	84 70.58 %	11 9.24 %
Over ripped	21 25.30 %	35 42.17 %	27 32.53 %

In the ripped class, 84 samples of 119 samples are categorized correctly in their own class but 24 samples in unripe class and 11 samples in over ripped class were classified incorrectly. The sum of elements on the main diagonal of confusion matrix divided by the total number of samples represents the correct classification of each classifier. In this study, this amount was 52% for panel test classification. The results of this matrix indicate that the most common mistakes were made by people in the unripe class and 52.04% of this class samples are classified incorrectly in the ripped category. Looking at the third row of this matrix it is also seen that only 32.53% of the over ripped samples are correctly classified and rest of samples in this class categorized incorrectly. With a general look at this matrix, it can be seen 170 samples of all samples (56.66%) are classified in the ripped class by experts. In other words, it can be concluded that classification Interference in ripped class is more than the other classes and experts have categorized most of the samples in this class.

Classification Using Discriminant Analysis

Table 4. shows the mean comparison of morphological variables.

Table 4. Comparison means of morphological variables used in Discriminant analysis.

Variables	Wilks's lambda	F	df1	df2	Sig.
Length	0.799	37.354	2	297	0.000
Perimeter	0.742	51.749	2	297	0.000
Weight	0.685	68.295	2	297	0.000
Weight/Perimeter	0.688	67.339	2	297	0.000
Weight/Length	0.682	69.309	2	297	0.000
Perimeter/Length	0.933	1.066	2	297	0.346

As can be seen, for 5 variables, a significant level of 0.000 is obtained which indicates the difference in the mean of these variables in the triple groups. In table 5 it can be observed how variables enter into the equation according to Wilks's lambda in the stepwise method.

Table 5. Entering variables into the discrimination equation during different steps.

Step	Variables	Wilks's lambda
First	Weight/Length	
	Weight/Length	0.799
Secon d	Length	0.682
	Weight/Length	0.683
Third	Length	0.672
	Weight	0.649

According to Table 5 in the first step, only the variable ratios of weight to length, in the second step the variable length, and in the third step, the variable weight are entered into the equation and no variables have been removed after entering into the equation. Also, the values of Wilks's lambda, which are always between zero and one, are presented. Confusion matrix which is provided using Discriminant Analysis classification is presented in Table 6. Overly 57/6% of the samples were classified correctly.

Table 6. Discriminant Analysis confusion matrix.

	Unripe	Ripped	Over ripped
Unripe	71 72.4 %	22 22.4 %	5 5.2 %
Ripped	18 15.1 %	62 52.1 %	39 32.8 %
Over ripped	7 8.4 %	36 43.4 %	40 48.2 %

By comparing Table 6 and Table 3, it can be seen discriminant analysis has done a better classification for the unripe class in the event that panel experts have been a better classification for ripped class.

Classification using LVQ Neural Network

LVQ Neural Networks unlike multi-layer perceptron neural networks with backpropagation approach result the same exact answer each time they pass the training procedure for a specific number of neurons in the hidden layer and that's why they are more reliable in classification project but they usually have less capability in classification. In this study, the correct classification for hidden layer neurons greater than 6 began to decrease. That is why this number was chosen for the number of hidden layer neurons which resulted 57% of the correct classification.

Classification Using K Nearest Neighbor

In this algorithm, the correct classification for the Euclidean, Cosine, Manhattan, and Chebyshev distances was 63, 61.7, 64.3 and 64 percent, respectively and its accuracy was more than Discriminant Analysis and LVQ Neural Network accuracy.

Classification Using Support Vector Machine

The results of the classification using SVM algorithm with Linear, Quadratic, Cubic, and Gaussian kernel functions were 56, 58.3, 64.7, 67.3 percent, respectively. Table 7 shows the confusion matrix of the SVM algorithm with Gaussian kernel function.

Table 7. SVM confusion matrix.

	Unripe	Ripped	Over ripped
Unripe	69 70.4 %	23 23.47 %	6 6.13 %
Ripped	12 10.1 %	91 76.47 %	16 13.43 %
Over ripped	6 7.2 %	35 42.17 %	42 50.63 %

It was observed that with the complexity of the kernel function, the correct classification rate improved which shows the complexity of the relationship between the studied six morphological parameters and the amount of watermelon ripeness.

Finally based on this study, the sensory evaluation results showed that the experts were able to classify 52% of the samples correctly. The correct classification of the LVQ Neural Network and Discriminant Analysis did not differ significantly from expert's classification. The correct classification of the K Nearest Neighbor algorithm was considerably greater than the correct classification of the LVQ Neural Network and Discriminant Analysis but the results of the classification by different distance metrics for this algorithm did not show a significant difference in their use. Among the distance metrics used by this algorithm, the Manhattan distance, which is easier to calculate and suitable for implementation on small and portable hardware, has the highest correct classification rate. Discriminant Analysis can be used quickly and easily in early evaluations and classifications estimation. Also, the attributes which are more effective on classification that this method identifies among all features are very useful in data analysis and this method can be used as a feature selection or data reducing method. Other algorithms used in this study do not have this possibility.

Conclusion

The highest level of classification with 67.3% was related to the SVM algorithm with Gaussian kernel function. Although at first glance it may seem that this amount is far from ideal, hower it should be noted that this is 15% higher than the classification made by a human. In addition, the classification is based on the morphological characteristics of the samples, which does not require complex tools and sophisticated methods. It should be noted that the correct classification for this method without using k-fold cross validation was obtained of 97.3%. Which indicates that models which are based only on training data, despite providing a high and discrimination rate, are not valid for future data.

References

Abbaszadeh R, Rajabipour A, Delshad M, Mahjub M, Ahmadi H, Lague C. 2011. Application of vibration response for the nondestructive ripeness evaluation of watermelons, Australian Journal of Crop Science **5,** 920-925.

Abbaszadeh R, Rajabipour A, Labbafi R, Ahmadi H. 2012. Prediction of watermelon customer-friendly based on sensory evaluation data using expert fuzzy model. Proceedings of the 7th National Congress of Agricultural Engineering (Biosystems Mechanics) and Mechanization. Sep 4-6. Shiraz. Iran. (in Farsi) Antonucci, Francesca, Corrado Costa, Federico Pallottino, Graziella Paglia, Valentina Rimatori, Donato De Giorgio, and Paolo Menesatti. 2012. Quantitative method for shape description of almond cultivars (*Prunus amygdalus* Batsch), Food and Bioprocess Technology **5,** 768-785.

Anon. 2013. FAO Food and Nutrition Series. Statistical database http://faostat. fao.org.

Blasco JN, Aleixos S, Cubero J, Gómez-Sanchís E, Moltó. 2009. Automatic sorting of satsuma (*Citrus unshiu*) segments using computer vision and morphological features, Computers and Electronics in Agriculture **66,** 1-8.

Bourne, Malcolm. 2002. Food texture and viscosity: concept and measurement (Academic press).

Brewer, Marin Talbot, Jennifer B Moyseenko, Antonio J Monforte, Esther van der Knaap. 2007. Morphological variation in tomato: a comprehensive study of quantitative trait loci controlling fruit shape and development, Journal of Experimental Botany **58,** 1339-1349.

Choudhary R, Jayas Paliwal, DS, Jayas. 2008. Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images, Biosystems Engineering **99,** 330-337.

Currie AJ, Ganeshanandam S, Noiton DA, Garrick D, Shelbourne CJA, Oraguzie N. 2000. Quantitative evaluation of apple (*Malus*× *domestica* Borkh.) fruit shape by principal component analysis of Fourier descriptors, Euphytica **111,** 221-227.

Diezma-Iglesias B, Ruiz-Altisent M, Barreiro P. 2004. Detection of Internal Quality in Seedless Watermelon by Acoustic Impulse Response, Biosystems Engineering **88,** 221-30.

Ding, Wei, Hirohisa Nesumi, Yasushi Takano, Yasuo Ukai. 2000. Quantitative evaluation of the three-dimensional fruit shape and size of Citrus species based on spherical harmonic descriptors, Euphytica **114,** 103-115.

Ghazanfari A, Irudayaraj J, Kusalik A, Romaniuk M. 1997. Machine vision grading of pistachio nuts using Fourier descriptors, Journal of agricultural engineering research **68,** 247-252.

Han, Jiawei, Jian Pei, Micheline Kamber. 2011. Data mining: concepts and techniques (Elsevier). Menesatti, Paolo, Corrado Costa, Graziella Paglia, Federico Pallottino, Stefano D'Andrea, Valentina Rimatori, and Jacopo Aguzzi. 2008. Shape-based methodology for multivariate discrimination among Italian hazelnut cultivars, Biosystems Engineering **101,** 417-424.

Koc, Ali Bulent. 2007. Determination of watermelon volume using ellipsoid approximation and image processing, Postharvest Biology and Technology **45,** 366-371.

Morimoto T, Takeuchi T, Miyata H, Hashimoto Y. 2000. Pattern recognition of fruit shape based on the concept of chaos and neural networks, Computers and Electronics in Agriculture **26,** 171-186.

Muramatsu, Noboru, Keiichi Tanaka, Toshikazu Asakura, Yuko Ishikawa-Takano, Naoki Sakurai, Naoki Wada, Ryoichi Yamamoto, and Donald J Nevins. 1997. Critical comparison of an accelerometer and a laser Doppler vibrometer for measuring fruit firmness, HortTechnology **7,** 434-438.

Ngouajio, Mathieu, William Kirk, Ronald Goldy. 2003. A simple model for rapid and nondestructive estimation of bell pepper fruit volume, HortScience **38,** 509-511.

Nourain, Jamal, Yibin B, Ying, Jianping Wang, Xiuqin Rao. 2004. Determination of acoustic vibration in watermelon by finite element modeling. In Optics East 213-223. International Society for Optics and Photonics.

Nunome, Tsukasa, Keizo Ishiguro, Tatemi Yoshida, Masashi Hirai. 2001. Mapping of fruit shape and color development traits in eggplant (*Solanum melongena* L.) based on RAPD and AFLP markers, Breeding science **51,** 19-26.

Rabiei B, Valizadeh M, Ghareyazie B, Moghaddam M, Ali AJ. 2004. Identification of

QTLs for rice grain size and shape of Iranian cultivars using SSR markers, Euphytica **137,** 325-332.

Sadrnia, Hassan, Ali Rajabipour, Ali Jafary, Arzhang Javadi, Younes Mostofi. 2007. Classification and analysis of fruit shapes in long type watermelon using image processing, Int. J. Agric. Biol **1,** 68-70.

Stone ML, Armstrong PR, Zhang X, Brusewitz GH, Chen DD. 1996. Watermelon maturity determination in the field using acoustic impulse impedance techniques, Transactions of the ASAE **39,** 2325-2330.

Taniwaki, Mitsuru, Takanori Hanada, and Naoki Sakurai. 2009. Postharvest quality evaluation of "Fuyu" and "Taishuu" persimmons using a nondestructive vibrational method and an acoustic vibration technique, Postharvest Biology and Technology **51,** 80-85.

Xiao, Han, Ning Jiang, Erin Schaffner, Eric J Stockinger, and Esther van der Knaap. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit, science **319,** 1527-1530.

Xiaobo, Zou, Zhao Jiewen, Li Yanxiao, Shi Jiyong, Yin Xiaoping. 2008. Apples shape grading by Fourier expansion and genetic program algorithm. In Natural computation, 2008. ICNC'08. Fourth International Conference on 85-90.

Yadav, BK, Jindal VK. 2001. Monitoring milling quality of rice by image analysis, Computers and Electronics in Agriculture **33,** 19-33.

Yamamoto, Hiromichi, Mutsuo Iwamoto, and Shiko Haginuma. 1980. Acoustic impulse response method for measuring natural frequency of intact fruits and preliminary applications to internal quality evaluation of apples and watermelons, Journal of Texture Studies **11,** 117-136.

Zheng TQ, Xu JL, Li ZK, Zhai HQ, Wan JM. 2007. Genomic regions associated with milling quality and grain shape identified in a set of random introgression lines of rice (*Oryza sativa* L.), Plant Breeding **126,** 158-163.

Zygier S, Ben Chaim A, Efrati A, Kaluzky G, Borovsky Y, Paran I. 2005. QTLs mapping for fruit size and shape in chromosomes 2 and 4 in pepper and a comparison of the pepper QTL map with that of tomato, Theoretical and Applied Genetics **111,** 437-445.