Australian Journal of Crop Science
AJCS
AJCS 4(5):353-358 (2010) ISSN:1835-2707

Integrating fuzzy data mining and impulse acoustic techniques for almond nuts sorting

Ebrahim Ebrahimi*[1], Kaveh Mollazade[2]

[1]Department of Mechanical Engineering of Agricultural Machinery, Faculty of Engineering, Islamic Azad University, Kermanshah Branch, Kermanshah, Iran
[2]Department of Agricultural Machinery Engineering, Faculty of Agricultural Engineering and Technology, University of Tehran, P.O. Box 4111, Karaj 31587-77871, Iran

***Corresponding author: ebrahimi.kiu@gmail.com**

Abstract

Sorting of agricultural products is one of the core research areas in the field of agricultural postharvest engineering. This paper presents an algorithm based method to sort four varieties of almond (Yalda, France, Shokofeh, and Shahrood 15) using impulse acoustic signals, selected features and rules generated from a decision tree, and fuzzy inference system. Two impact plates (Plywood and Stainless steel) and two fall heights (14 cm and 24 cm) were considered as test conditions. Results showed that the best decision tree will be obtained when impact plate is Stainless steel and fall height is 24 cm. In this condition, the fuzzy inference engine performance was found to be encouraging and its accuracy in the classification of almond varieties was 84.16%.

Keywords: Artificial intelligence; Decision tree; Impact; Postharvest; Sound signal

Introduction

Almond is one of the most important nut crops worldwide which produces fruits of high commercial value. According to the FAO (2007), Iran produces 110000 MT of almond every year. Several varieties of almond can be found in the traditional and modern gardens of Iran. In the harvest stage and after picking the almonds from the trees or in the silo stage, it may that these varieties be mixed together. In the agricultural product's marketing and processing industry, it is necessary that products be uniform in the case of variety. Because different variety of a product has different nutritional and taste content and this influences on its marketing price. So, it seems that a system which could sort the different varieties of almond be necessary before doing any processing or marketing tasks. The development of science and technology has made available many methods for the issue of fruit sorting and grading. These methods can be classified according to the principle used to detect fruit properties: mechanical, acoustic, optical, etc (Zerbini, 2006). There are few researches which used acoustic technique for sorting of nuts. Pearson (2001) developed an acoustic sorting system to separate pistachio nuts with open shells from those with closed shells. Also, an algorithm using speech recognition technology to distinguish pistachio nuts with closed shells from those with open shells was developed by Cetin et al. (2004). Omid et al. (2009) developed an intelligent pistachio nut sorting system based on acoustic emission technique. Their system combined acoustic emissions analysis, Principal Component Analysis (PCA) and Multilayer Feedforward Neural Network (MFNN) classifier. Other than nuts, acoustic sorting was used for other agricultural products. Pearson et al. (2007) developed a non-destructive, real time device to detect insect damage, sprout damage, and scab damage in kernels of wheat. Kernels are impacted onto a steel plate and the resulting acoustic signal analyzed to detect damage. In order to detect hollow heart in potato tubers an acoustic sorting system was developed by Elbatawi (2008). The classification accuracy of this system based on linear discriminate analysis was approximately 98%. Expert system applications have been increasingly applied to agricultural problems. Morimoto et al. (2000) developed an intelligent system for fruit shape recognition. In their system a one-dimensional profile data was made to characterize the fruit shape. Then a three-layer neural network was used for identifying and tracing the one-dimensional profile data and then evaluating their irregularities. Xu and Zhao (2009) combined the image processing, neural network, and genetic algorithm techniques for maturity detection of strawberries. Similar works are reported for the inspection of agricultural products (meats, fruits, vegetables and nuts) and processed foods such as cookies, breads, snacks, etc (Shaw, 1990; Day, 1993; Henrikson, 1996). Traditionally, these artificial intelligence technologies have been applied independently for the solution of problems. However, when these technologies are used together, the resulting hybrid system can perform better than the individual technologies. For example, one of the problems in developing an expert system is the extraction of knowledge from a human expert. In many cases, no human expert is available or it may be necessary to gather information from more than one expert. A data mining technique, on the other hand, can be developed using samples of data from the problem domain in the form of inputs and associated outputs. The decision trees can automatically extract (and encode) the knowledge from the data. The objective of the work reported

Table 1. Some physical properties of almond nuts

varieties	Mass (gr)	Major diameter (mm)	Intermediate diameter (mm)	Minor diameter (mm)	Arithmetic mean diameter (mm)	Geometric mean diameter (mm)
Yalda	5.10	36.42	23.59	17.75	25.92	24.44
France	3.47	30.67	22.48	16.60	23.25	22.52
Shokofeh	2.07	29.99	20.44	12.52	20.99	19.72
Shahrood 15	1.49	28.16	15.87	11.43	18.48	17.16

here was to use fuzzy logic inference system combined with data mining technique to develop a hybrid decision support system for sorting of four varieties of almond namely: Yalda, France, Shokofeh, and Shahrood 15 using impulse acoustic signals.

Material and methods

The research follows these steps: (1) For different almond varieties, raw acoustic signals were recorded in four conditions including two different impact surfaces and tow different fall heights. (2) Several features of raw acoustic signals were extracted. (3) Adequate features were selected by data mining technique. (4) Output of data mining was fed to the Fuzzy system. (5) An automatic system is presented for almond nut variety separation.

Almond varieties

Four almond varieties were collected from the Sanandaj Agriculture Researches Institute (SARI), Iran. These varieties were namely; Yalda, France, Shokofeh, and Shahrood 15 (see Fig. 1). The number of almond nuts, for each variety, used for each test in this study was 105. Table 1 shows some physical properties of almond nuts used in this study including average value of mass, major diameter, intermediate diameter, minor diameter, arithmetic mean diameter, and geometric mean diameter (Mollazade et al., 2009). The almonds were stored in a cold storage at a temperature of 12 °C and 82% relative humidity. They were taken out of storage 24 h prior to testing and allowed to reach room temperature. The room temperature was recorded at the start of the experiment and found to be in the range 20–23 °C (Elbatawi, 2008). Almonds were selected, numbered and measured. All varieties were tested daily using an acoustic system to acquire their acoustic spectrum.

Experimental apparatus

A schematic of the experimental apparatus for dropping of almonds onto the impact plate and collecting the acoustic emissions from the impact is shown in Fig. 2. The apparatus include a chute with adjustable height which was inclined 35° above the horizontal, on which almond nuts slide down and are projected onto the impact plate. Two vertical distances (14 and 24 cm) between the underneath section of chute and the surface of impact plate (h in Fig. 2) were considered as fall height. Also, experiments were done on two impact plates namely, Stainless steel and Plywood. So, we had four test conditions include two fall heights and two impact plates. A highly directional microphone (Dm-400 Hi-fidelity, Enping Huana Sound Instrument Manufactory, China) was used to minimize the effect of ambient noise. The microphone was installed

Fig 1. Four different varieties of almond

inside an isolated acoustic chamber to eliminate environmental noise effects. Output from the microphone was sent to a personal computer (Pentium 4 processor, 2.53 GHz, Windows 7 operating system) based data acquisition system. Acquired signals were digitized at a sampling frequency of 44.1 kHz, with 16 bit resolution. Matlab data acquisition toolbox was used for saving the signals. When almond nuts hit the impact plate, the amplitude of the microphone output signal ranged from 0 to 1.0 V. Data acquisition began when the microphone output was higher than 0.1 V. This threshold level was sufficient to trigger acquisition for virtually all nuts, whilst preventing false triggers from ambient sound. Since there were four test conditions, totally 1680 acoustic spectra were acquired. Each acquired signal was consist of 500 samples. Fig. 3 shows an example of acoustic signals of almond nuts recorded with the data acquisition system in different test conditions.

Feature extraction

In intelligent systems, feature extraction is very important. So, in this research 12 features of acoustic signals were extracted as described in the Table 2 (Lebold et al., 2000; Holcomb, 1997). Table. 2, List of features

Data mining process for feature selection

Data mining is the process of selecting, exploring and modeling large amounts of data in order to discover unknown patterns or relationships which provide a clear and useful result to the data analyst. Waikato Environment for Knowledge Analysis or WEKA, Version 3.4.3 was used for data mining. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization (Witten and Frank 2005). In this study, we made use of the J48 classifier

Table 2. List of features

Feature	Formula
Average	$Aver = \frac{\sum_{n=1}^{N} x(n)}{N}$
Standard deviation	$Stdv = \sqrt{\frac{n\sum_{n=1}^{N} x^2(n) - (\sum_{n=1}^{N} x(n))^2}{n(n-1)}}$
Variance	$Var = (Stdv)^2$
Skewness	$Skew = \frac{n}{(n-1)(n-2)} \sum_{n=1}^{N} (\frac{x(n) - Aver}{Stdev})^3$
Kurtosis	$Kurt = \frac{\sum_{n=1}^{N} (x(n) - \mu)^4}{N \times (\sigma^2)^2}$
Sum	$Sum = \sum_{n=1}^{N} x(n)$
Root Mean Square (RMS)	$RMS = \sqrt{\frac{\sum_{n=1}^{N} x^2(n)}{N}}$
Crest factor	$CF = \frac{x_{\max}}{RMS}$
Dispersion	$Disp = \sum_{n=1}^{N} \lvert x(n) - Aver \rvert$
Entropy	$E = \sum_{n=1}^{N} x(n).\log_2(x(n))$
Range	$R = x_{\max} - x_{\min}$
Coefficient variation	$CV = \frac{Stdv}{Aver}$

algorithm which is an implementation of the C4.5 decision tree learner. A standard tree induced with c4.5 consists of a number of branches, one root, a number of nodes and a number of leaves. One branch is a chain of nodes from root to a leaf; and each node involves one attribute. The occurrence of an attribute in a tree provides the information about the importance of the associated attribute. A decision tree is a tree based knowledge representation methodology used to represent classification rules. In these trees the top node is the best node for classification. The other features in the nodes of decision tree appear in descending order of importance. Only features that contribute to the classification appear in the decision tree and others do not. Features, which have less discriminating capability, can be consciously discarded by deciding on the threshold. The algorithm identifies the good features for the purpose of classification from the given training data set, and thus reduces the domain knowledge required to select good features for pattern classification problem (Saravanan, 2010).

Fuzzy logic

Fuzzy logic is considered a superset of conventional logic extended to handle the concept of partial truth. This idea was proposed and developed by Zadeh (1965). The main idea is that

Table 3. Results of decision trees for different test conditions

Test condition	Number of nodes	Number of leaves	Accuracy (%)
Plywood and 14 cm	73	37	82.33
Plywood and 24 cm	27	14	67.62
Steel and 14 cm	61	31	59.66
Steel and 24 cm	69	35	81.76

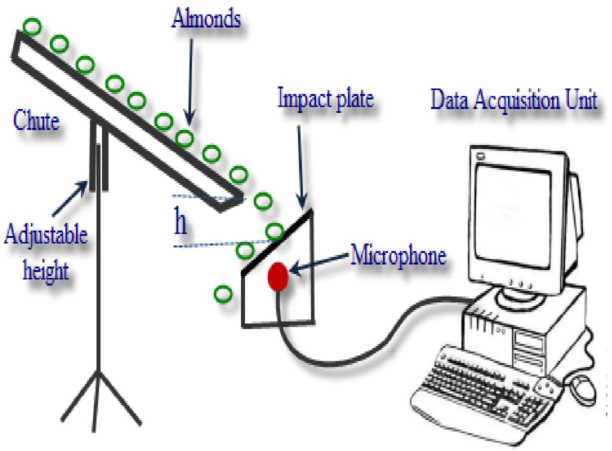

Fig 2. Schematic view of experimental apparatus

the decision is no longer a binary 0 or 1, but it can lie somewhere in between. This then leads to the concept of a fuzzy variable and membership functions. Using this representation, objects (in this research almond variety) are members of classes to some level or degree. The point of fuzzy logic is to map an input space to an output space, and the primary mechanism for doing this is a list of "if-then" statements called rules. Rules are the inputs for building a fuzzy inference engine. All rules are evaluated in parallel, and the order of the rules is unimportant. The real world data do not have sharply defined boundaries where information is often incomplete or sometimes unreliable. In quest for precision, scientists have generally tempted to manipulate the real world into artificial mathematical models that make no provision for gradation. Because fuzzy classifier provides the tools to classify information into broad, coarse categorisations or groupings, it has infinite possibilities for application which have proven to be much cheaper, simpler, and more effective than other systems in handling complex information (Sugumaran and Ramachandran, 2007). In the current study, fuzzy rules and threshold values of membership functions were made by using decision tree produced by WEKA. The shape of membership functions was used as trapezoidal. The selection of this membership function is to some extent arbitrary. The fuzzy toolbox available in MATLAB 7.2 was used for building fuzzy inference engine. The data sets of the features for each condition had 420 samples. In each test condition, two-thirds of samples were employed for training process (training the J45 algorithm) and the remaining samples for testing purposes (testing the fuzzy model).

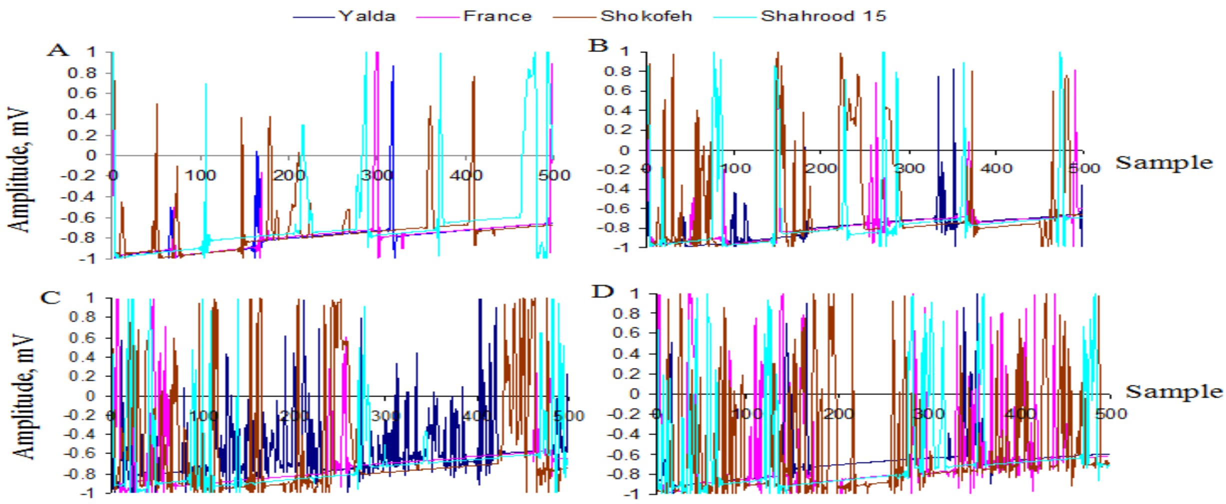

Fig 3. Typical impact acoustic signals of almond nuts varieties in different test conditions: A. Plywood and 14 cm, B. Plywood and 24 cm, C. Steel and 14 cm, and D. Steel and 24 cm

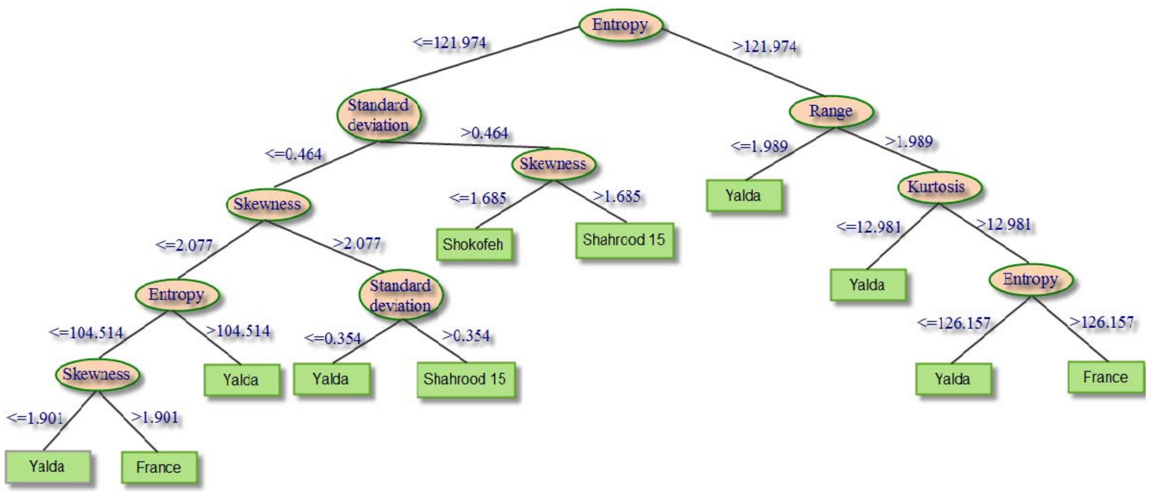

Fig 4. Decision tree with selected features for "Steel and 24 cm" condition

Table 4. Confusion matrix

Varieties	Yalda	France	Shokofeh	Shahrood 15
Yalda	26	2	2	0
France	4	24	1	1
Shokofeh	1	1	27	1
Shahrood 15	1	2	3	24

Results and discussion

Summary of decision tree results for different test conditions is shown in Table 3. According to these results, because of very low accuracy, "Plywood and 24 cm" and "Steel and 14 cm" conditions were not suitable for almond sorting. On the other hand, accuracy of decision trees produced for "Plywood and 14 cm" and "Steel and 24 cm" were suitable. But another problem was exist and that was the huge number of nodes and leaves of decision tree for these two selected conditions. The number of membership functions and rules in the fuzzy inference system is closely equal with the number of nodes and leaves in the consequent decision tree, respectively. If we want to have an online sorting system, this is very important that processing time be short. For this purpose we must have small number for fuzzy membership functions and fuzzy rules. According to the references (Witten and Frank, 2005; Saravanan et al., 2009), in the decision trees, features that have less discriminating capability can be consciously discarded by deciding on the threshold. This concept is made using of in selecting good features. The algorithm identifies the good features for the purpose of classification from the given training data set and thus reduces the domain knowledge required to select good features for pattern classification problem. So, pruning was

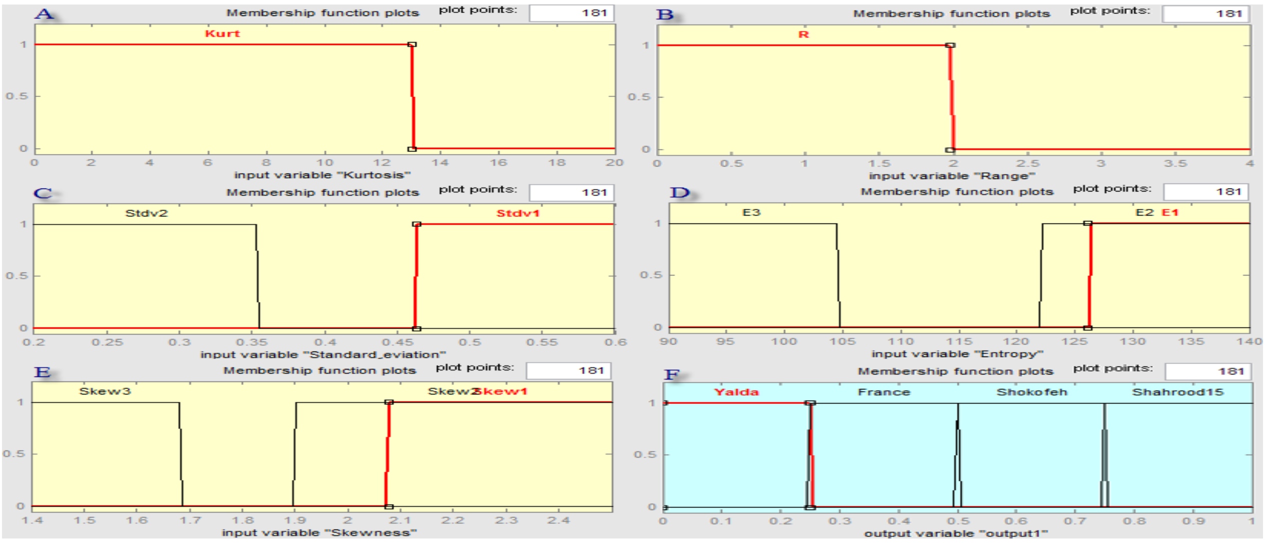

Fig 5. Membership function for: A. Kurtosis, B. Range, C. Standard deviation, D. Entropy, E. Skewness, and F. Almond variety (output)

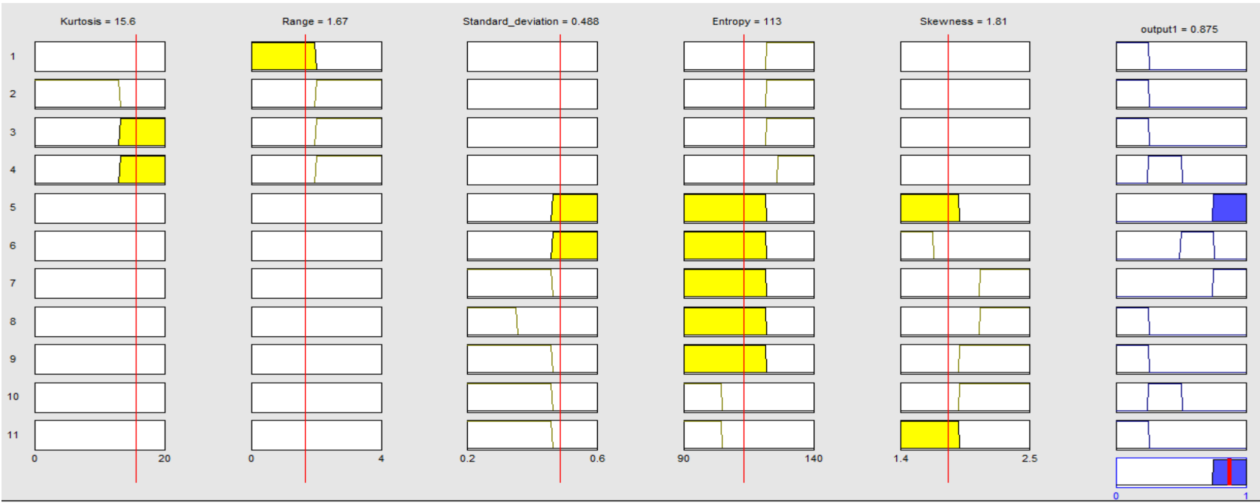

Fig 6. Rule viewer editor for a sample of data

done in order to reduce the size of trees by using “attribute selection” filter in the WEKA software. Finally, the size of decision tree for “Plywood and 14 cm” and “Steel and 24 cm” conditions was obtained respectively as 16 nodes and 17 leaves, and 10 nodes and 11 leaves, respectively. Since the size of decision tree for “Steel and 24 cm” was lower than that of the “Plywood and 14 cm”, this was concluded that Stainless steel as impact plate and 24 cm as fall height is suitable condition for construction of an online almond sorting system. Decision tree of this condition is shown in Fig. 4. In the tree the ovals represent the nodes and rectangles represent leaves. This tree shows that only five features (Entropy, Standard deviation, Skewness, Range, and Kurtosis) are significant for classification and other features must be ignored. Observing the values of the feature, based on which the branches of the decision tree are created, the membership functions for all five features are defined and presented in Fig. 5 (A-E) for Kurtosis, Range, Standard deviation, Entropy, and Skewness, respectively. There are four possible outcomes from a fuzzy classifier, namely: Yalda, France, Shokofeh, and Shahrood 15. Hence, four membership functions are defined with equal range and shown in Fig. 5 (F). Tracing a branch from the root node leads to a variety of almond (Refer Fig. 4) and decoding the information available in a branch in the form of ‘if - then’ statement gives the following rules : 1- If Entropy is E2 and Range is R then output is Yalda. 2- If Entropy is E2 and Range is not R and Kurtosis is Kurt then output is Yalda. 3- If Entropy is not E1 and Range is not R and Kurtosis is not Kurt then output is Yalda. 4- If Entropy is E1 and Range is not R and Kurtosis is not Kurt then output is France. 5- If Entropy is not E2 and Standard deviation is Stdv1 and Skewness is not Skew2 then output is Shahrood 15. 6- If Entropy is not E2 and Standard deviation is Stdv1 and Skewness is Skew2 then output is Shokofeh. 7- If Entropy is not E2 and Standard deviation is not Stdv1 and Skewness is Skew1 then output is Shahrood 15. 8- If Entropy is not E2 and Standard deviation is Stdv2 and

Skewness is Skew1 then output is Yalda. 9- If Entropy is not E2 and Standard deviation is not Stdv1 and Skewness is Skew2 then output is Yalda. 10- If Entropy is E3 and Standard deviation is not Stdv1 and Skewness is Skew2 then output is France. 11- If Entropy is E3 and Standard deviation is not Stdv1 and Skewness is not Skew2 then output is Yalda. Fig. 6 illustrates the application of the rules designed. Here each row corresponds to each rule as discussed above. The first five blocks in rows represents the fuzzy input membership functions. The sixth block corresponds to the membership functions for almond variety as shown in Fig. 5. With the help of sample inputs, rules were tested as follows, for a sample input of Kurtosis as 15.6, Range as 1.67, Standard deviation as 0.488, Entropy as 113, and Skewness as 1.81, which satisfies the fifth rule completely and the corresponding output condition is Shahrood 15, which is shown in the output block of the fifth row in the rule viewer shown in Fig. 6. Using the testing data, the fuzzy inference engine was evaluated and its performance was presented as confusion matrix in Table 4. The diagonal elements in the confusion matrix show the number of correctly classified instances. In the first row, the first element shows the number of data points belonging to “Yalda” class and classified by fuzzy logic as “Yalda”. The second element shows the number of data points belonging to “Yalda” class but misclassified as “France”. The third element shows the number of data points misclassified as “Shokofeh” and so on. According to the confusion matrix, total accuracy of fuzzy system was found to be 84.16%. In order to increase the accuracy of fuzzy logic inference system two procedures are suggested: using other features of acoustic signal different from features used in this study and using other feature selection techniques like Neural Networks.

Conclusion

This work has investigated the use of impulse acoustic and fuzzy data mining techniques as a nuts sorting technique. The work conducted has displayed the potential of fuzzy logic to classify acoustic spectra according to the almond variety. Its ability to classify shows considerable potential. Using membership function domain limits that are linked to the variability of a group of spectra that represent a particular almond variety allows the technique to be truly objective. This work outlines the procedure for arriving at this objective technique. Also the effect of almond fall height and impact plate type on the accuracy of fuzzy inference system was investigated. It was concluded that the optimum fuzzy system will be obtained when impact plate is Stainless steel and fall height is 24 cm.

Acknowledgement

This research was funded by Islamic Azad University, Kermanshah Branch. This assistance is gratefully appreciated.

References

Cetin AE, Pearson TC, Tewfik AH (2004) Classification of closed and open shell pistachio nuts using voice recognition technology. Trans ASAE 47(2): 659–664

Day W (1993) Computer opportunities in agriculture and horticulture. In: Hashimoto Y, Bot GPA, Day W, Tantau HJ, Nonami H (Eds.), The computerized greenhouse. Academic Press, San Diego, CA

Elbatawi IE (2008) An acoustic impact method to detect hollow heart of potato tubers. Bios Eng 100: 206 – 213

FAO (2007) FAOSTAT database. http://faostat.fao.org/faostat/ (accessed: 10 Feb., 2010)

Henrikson FW (1996) Food processing automation conference IV: Conference report. Trends Food Sci Tech 7 (3), 96–98.

Holcomb ZC (1997) Fundamentals of descriptive statistics. Pyrczak Press

Lebold M et al (2000) Review of vibration analysis methods for gearbox diagnostics and prognostics. In: Proceedings of the 54th meeting of the society for machinery failure prevention technology, Virginia Beach, VA, May 2000, 623-634

Mollazade K, Ahmadi H, Khorshidi J, Mohtasebi SS, Rajabipour A (2009) Moisture dependent physical and mechanical properties of cumin (*Cuminum cyminum* L.) seed. Int J Agric Biol Eng 2(2): 1-7.

Morimoto T, Takeuchi T, Miyata H, Hashimoto Y (2000) Pattern recognition of fruit shape based on the concept of chaos and neural networks. Comp Elec Agr 26: 171–186

Omid M, Mahmoudi A, Omid MH (2009) An intelligent system for sorting pistachio nut varieties. Exp Sys App 36: 11528–11535

Pearson TC (2001) Detecting of pistachio nuts with closed shells using impact acoustics. App Eng Agr 17(2): 249–253

Pearson TC, Cetin AE, Tewfik AH, Haff RP (2007) Feasibility of impact-acoustic emissions for detection of damaged wheat kernels. Digit Sig Proc 17: 617–633

Saravanan N, Cholairajan S, Ramachandran KI (2009). Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique. Exp Sys App 36: 3119–3135

Saravanan N, Kumar Siddabattuni VNS, Ramachandran KI (2010) Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vctor machine (PSVM). App Soft Comp 10: 344–360

Shaw WE (1990) Machine vision for detecting defects in fruits and vegetables. In: Food Processing Automation: Proceedings of the 1990 Conference, May 1990, Lexington, Kentucky. American Society of Agricultural Engineers, St. Joseph, MI.

Sugumaran V, Ramachandran KI (2007) Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. Mech Sys Sign Proc 21: 2237 – 2247

Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques. Elsevier, London

Xu L, Zhao Y (2009) Strawberry maturity neural network detecting system based on genetic algorithm. In IFIP International Federation for Information Processing, Volume 294, Computer and Computing Technologies in Agriculture II, Volume 2, eds Li D, Chunjiang Z (Boston: Springer) 1201–1208

Zadeh LA (1965) Fuzzy sets. Inform Cont 8(3): 338- 353

Zerbini PE (2006) Emerging technologies for non- destructive quality evaluation of fruit. J Fruit Ornam Plant Res 14(2): 13-23