Discrimination Between Closed and Open Shell (Turkish) Pistachio Nuts Using Undecimated Wavelet Packet Transform

6 authors, including:

Nuri F Ince
University of Houston
97 PUBLICATIONS   **840** CITATIONS

SEE PROFILE

Ibrahim Onaran
Mevlana University
30 PUBLICATIONS   **243** CITATIONS

SEE PROFILE

Ahmed Tewfik
University of Texas at Austin
577 PUBLICATIONS   **9,288** CITATIONS

SEE PROFILE

A. Enis Cetin
Bilkent University
290 PUBLICATIONS   **4,231** CITATIONS

SEE PROFILE

Discrimination Between Closed and Open Shell (Turkish) Pistachio Nuts Using Undecimated Wavelet Packet Transform

N. F. Ince, F. Goksu, A. H. Tewfik,
I. Onaran, A. E. Cetin, T. C. Pearson

ABSTRACT. *Due to low consumer acceptance and the possibility of immature kernels, closed-shell pistachio nuts should be separated from open-shell nuts before reaching the consumer. A system using impact acoustics as a means of classifying closed–shell nuts from open–shell nuts has already been shown to be feasible and have better discrimination performance than a mechanical system. The accuracy of an impact acoustics based system is determined by the signal processing and feature extraction procedures. In this article, a new time-frequency plain feature extraction and classification algorithm was developed to discriminate between open- and closed-shell pistachio nuts produced in the Gaziantep region of Turkey. The proposed approach relies on the analysis of the impact acoustics signal of pistachio nuts, which are emitted from their impact with a steel plate after dropping from a certain height. Features are extracted by decomposing the acoustic signals into time and frequency components, using double-tree undecimated wavelet packet transform. The most discriminative features from the dual tree nodes are selected by a wrapper strategy that includes the structural pruning of the double-tree feature dictionary. The proposed approach requires no prior knowledge of the relevant time or frequency content of the acoustic signals. The algorithm used a small number of features and achieved a classification accuracy of 91.7% on the validation data set, while separating the closed shells from the open ones. A previously implemented algorithm, which uses maximum signal amplitude, absolute integration, and gradient features, achieved 82% classification accuracy on the same dataset. The results show that the time-frequency features extracted from impact acoustics can be used successfully for classification of open- and closed-shell Turkish pistachios.*

Keywords. Classification, Impact acoustic, Turkish pistachios, Undecimated wavelet packet transform.

Closed-shell pistachio nuts could be rejected by consumers because they are difficult to open or can contain immature kernels. Therefore, their separation from open-shell nuts is crucial. Closed-shell pistachio nuts are currently separated from open-shell nuts by mechanical devices called "pinpickers." These machines can inadvertently damage the kernel of open-shell nuts by inserting a needle into the kernel meat. The hole created by the needle can give the appearance of an insect tunnel, which causes rejection by the consumer. In addition, according to Pearson (2001),

approximately 5% to 10% of all open-shell pistachio nuts in the U.S. are incorrectly classified by mechanical devices as having a closed shell, costing the industry $3.75 to $7.5 million per year in lost revenue. Therefore, high-accuracy classification systems are needed in the industry.

A number of classification devices have been designed for the separation of open- and closed-shell pistachio nuts. Prior versions of these systems were based on the processing of 2D images of pistachio nuts. Due to the high cost and slight improvements in the final classification accuracy over existing methods, these systems did not find widespread application in the industry. For a description of these computerized vision-based techniques and a short comparison of their performances, refer to Pearson (2001) and Ghazanfari et al. (1997). Recently, a new non-contact system based on impact acoustic emission has been proposed for food kernel inspection, which overcomes several limitations of the approaches summarized above. This system was designed to separate pistachio nuts with closed shells from those with open shells by analyzing their impact acoustics (Pearson, 2001). Acoustic recordings were obtained by impacting the pistachio nuts onto a steel plate. During the off-line training (learning) phase, a total of 359 features were extracted from the recorded signals. This rich feature set was formed from several properties of the signal, such as the integration of the absolute value of the signal, all frequency spectra magnitudes computed using a 256-point FFT with a Hanning window, and the gradient and magnitude thresholds from several time points. Because it is difficult to use all features in real-time processing, due to high computational cost and dimensionality, among those 359 features, a subset of three features were selected through an exhaustive search. The real-time constraint is that the nut under examination be classified before the next nut comes into the system. Using these three selected features on the validation set, the classification accuracy was approximately 97% with a throughput of 40 nuts per second, proving better performance than traditional mechanical devices on pistachio nuts produced in California.

Pearson's (2001) study using impact acoustics emphasized the importance of the signal processing stage. In particular, it has been shown that having *a priori* information about the relevant time and frequency content of the signals has important effects on classification accuracy. However, the adjustment of these parameters is demanding. Furthermore, the difference in size between nuts from region to region results in sound signals with different characteristics. For instance, pistachio nuts in the Gaziantep region of Turkey are much smaller than those nuts produced in California. Furthermore, the open-shell Gaziantep nuts have an ellipsoid shape and a thin split in the shell, which cause similar impact acoustics to that of closed shells. Sample images of these nuts are shown in figure 1. The physical differences make it necessary to develop an adaptive classification system that can adjust its parameters for a given signal (nut type).

In this article, we tackle these problems by developing an adaptive time-frequency plain feature extraction and classification algorithm that utilizes the same impact acoustic system. Based on the observations from the algorithms mentioned above, our algorithm does not depend on *a priori* knowledge of the time-frequency content of the signals under examination. Hence, it is more universally adaptable to other types of applications and uses features extracted from both the time and frequency content of signals, concurrently. In this respect, the proposed algorithm extracts adaptive time-frequency features from impact acoustics by using a dual-tree undecimated wavelet packet transform. It locates most discriminant features among classes by a wrapper strategy and classifies them with linear discriminant analysis (LDA). In particular, we tested our proposed system to discriminate between open- and closed-shell pistachio nuts collected from the Gaziantep region of Turkey.
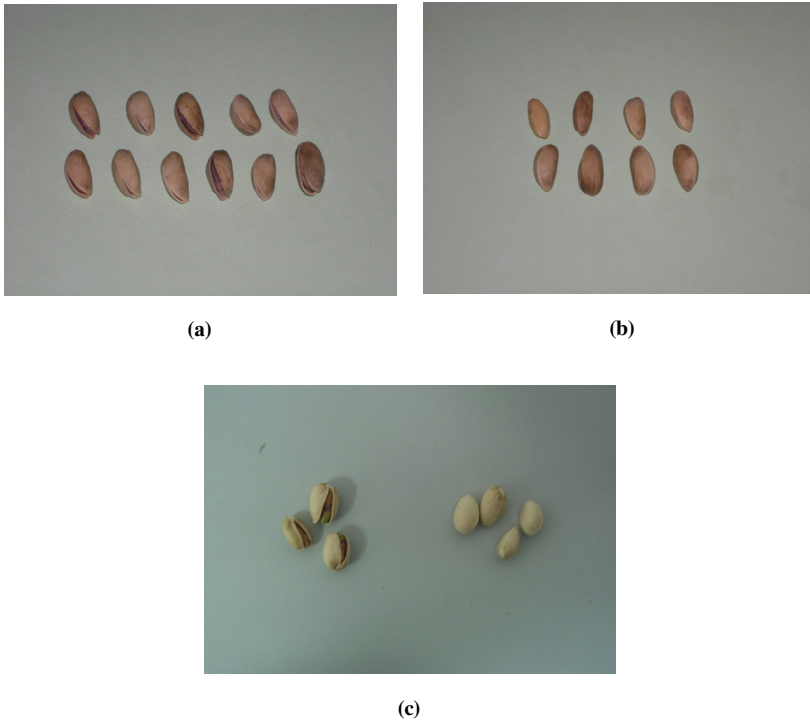
(a)

(b)

(c)

Figure 1. Sample pistachio images: (a) open-shell and (b) closed-shell pistachios from the Gaziantep region in Turkey, and (c) open and closed shell California pistachios.

This article is organized as follows: in the next section, we describe the data acquisition system to record the impact acoustic signals. Then we describe our feature extraction and subset selection procedures. In the Results section, we provide experimental results on the classification of open- and closed-shell pistachio nuts. Finally, we discuss our findings and challenges and the future steps of our research.

Materials and Systems

The apparatus used to record sound signals and sort the pistachio nuts is based on the system developed by Pearson (2001) and shown in figure 2. A nut is fed down a declining slide, at the end of which it drops onto a steel impact plate. The resulting acoustic signal from the impact is recorded with a highly directional microphone. The thickness of the impact plate and the direction of microphone are selected in such a way that the unwanted noise during the sound signal recordings is minimized. Output of the microphone is digitized with 100 kHz sampling frequency by a sound card attached to a PC and stored for further analysis. Each record is aligned according to an anchor time point, which was obtained by examining the absolute value of the impact acoustics. This point is detected when the absolute amplitude exceeds a threshold (≥ 1500). In particular, 40 samples before and 216 samples after the anchor point were used. The sound signals available this way are analyzed in an off-line manner for feature extraction. After features are selected and the decision rules are set, the system is run for validation. In real-time implemen–tation, a decision is given and either the nut is diverted by an air valve to one stream or
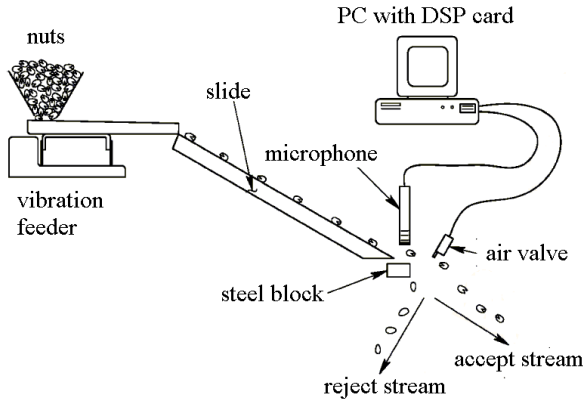
Figure 2. Impact sound classification system (modified from Pearson, 2001).

no action is taken and the nut continues in the direction of its momentum. Obviously, a timing constraint is present, and this is one reason to give a good decision as fast as possible. Although we do not validate the performance of our data in real-time, these time constraints have to be taken into consideration for practice.

For each type (closed and open) of pistachio, 200 recordings were obtained using this setup. Each acoustic recording was 256 samples long. Examples of sound signals are given in figure 3. We note that 256 sample provide enough time duration to cover the impact and its tail. After 256 samples, the impact signals vanished out.

The variations between open- and closed-shell pistachio sound signals can be observed in these time domain signals. However, as will be shown in the Results section, a method that is based only on time domain signal variations yields lower classification accuracy than time and frequency domain information; consequently, it is a better strategy to use information from both the time and frequency domains by a joint adaptation. To give a flavor of the time and frequency content of the signals, a short-time Fourier transform (STFT) is computed for each sound signal. A 256-point FFT is computed in 16 sample windows to obtain STFT maps. This window is shifted with
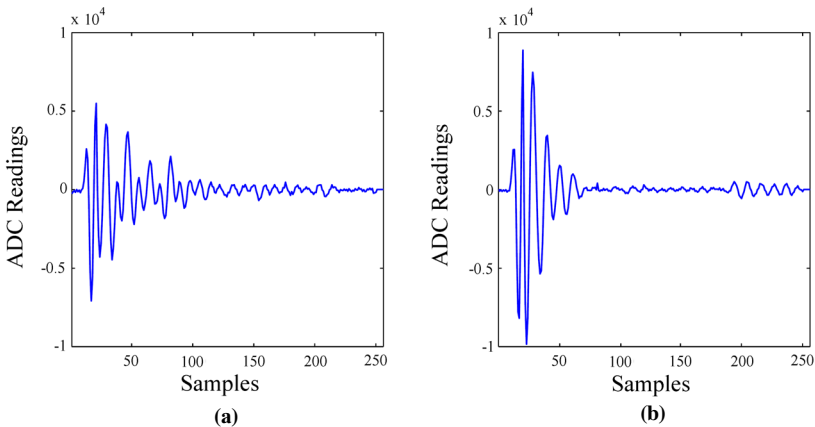

Figure 3. Sample impact sound signals of (a) an open-shell pistachio and (b) a closed-shell pistachio.
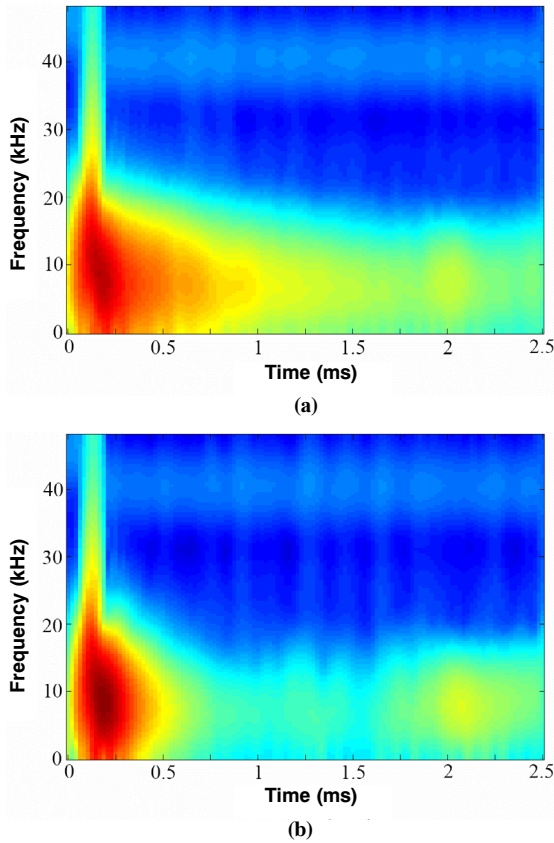
(a)

(b)

Figure 4. Averaged STFT images of (a) open-shell and (b) closed-shell pistachios.

87.5% overlap along the time axis. The absolute values of these STFTs are averaged over 200 signals for each class. Logarithms of these averages for each class are depicted in figure 4. Not only do we have the time variations but also the frequency variations for both classes of signals as well.

Feature Extraction

In the last several years, there has been a growing interest in exploring the time-frequency plane for classification by using adaptive strategies. The local discriminant bases (LDB), an algorithm of Saito et al. (2002), has been proposed to achieve this task. This approach first represents a given signal for each class in a redundant manner in a single pyramidal tree structure by wavelet packets (WP) or cosine packets (CP). The pyramidal tree structure is pruned from bottom to top such that the discrimination power between expansion coefficients in the nodes of the tree is maximized. Once the tree is pruned, a complete representation of the signal is obtained. This is followed by sorting the expansion coefficients according to their discrimination power and inputting them to a classifier for the final decision. This powerful method has been used in several applications, such as electromyogram and electroencephalogram classification, and successful results have been obtained (Englehart, 1999; Ince et al., 2006, Ince et al., 2007). However, the LDB algorithm has many drawbacks. First, the

WP/CPs, which are used to represent the signal in the nodes of the tree structure, do not satisfy the shift invariance property. Briefly, a shift in the signal is reflected by unpredictable changes in the expansion coefficients. This behavior is not appropriate for pattern recognition applications (Mallat, 2000). Furthermore, since a single tree is used, this algorithm can only adapt in a single axis, either time or frequency. Several studies have shown that the adaptation in both axes is crucial (Vetterli, 2001; Ince et al., 2006). Finally, the pruning and feature sorting stages do not account for the interactions/relations between different time-frequency cells. The obtained complete representation may not be the best subset for the classifier.

Here, as a first step, we use undecimated wavelet packet transform (UDWPT) to achieve a shift invariant signal representation. Unser (1995) has shown that the classification results obtained with undecimated wavelet transform (UDWT) is superior to discrete wavelet transform. We note that in our previous publications we used the term "wavelet transform" although we expanded the nodes in the high band. Therefore, in order to prevent confusion, we will use the term UDWPT, which expands the low and high frequency bands of the decomposition tree as shown in figure 5. Furthermore, we use a dual tree to adapt both in time and frequency axes. Finally, we use a greedy approach to select features from the redundant representation. These techniques are explained in detail in the following sections.

Undecimated Wavelet Packet Transform

The discrete wavelet transform (DWT) and its variants have been extensively used in 1D and 2D signal analysis due to their good localization properties both in time and frequency domains (Vetterli, 2001). However, the down-sampling operator at the outputs of each filter results in a shift variant decomposition. In practice, a shift in the signal is reflected by abrupt changes in the extracted expansion coefficients or related features. Unser (1995) proposed using the undecimated wavelet transform to extract subband energy features, which are shift invariant. This is achieved by removing the down-sampling operation. The output at any level of pyramidal filter bank (fig. 6) is computed by using an appropriate filter, which is derived by up-sampling the basic filter.
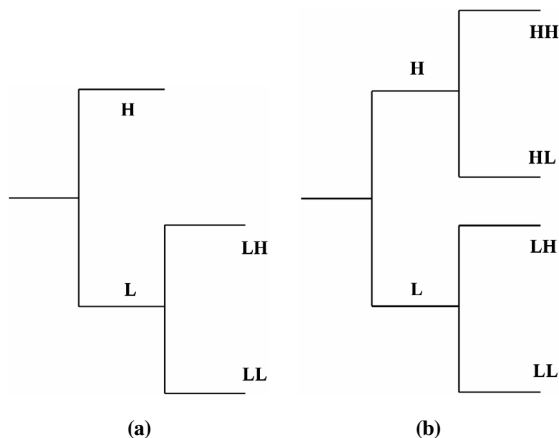

(a) (b)

Figure 5. The wavelet tree (a) and the wavelet packet tree (b). Note that in the wavelet packet tree, the nodes in the high band are expanded, which is a generalized form of wavelet transform and provides a richer decomposition. L and H stand for low and high frequency bands.
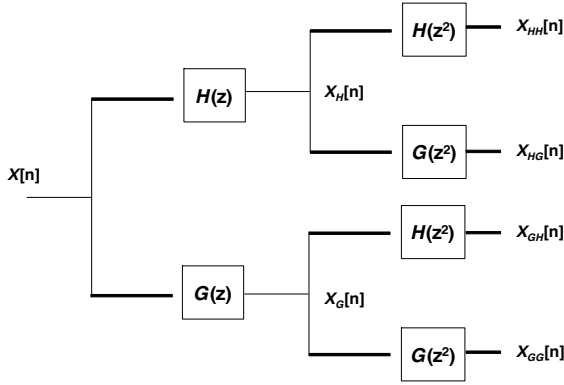
Figure 6. Pyramidal undecimated wavelet packet tree. The input *x*[*n*] is successively decomposed in low and high frequency bands.

A filter $g(n)$ with a z-transform $G(z)$ that satisfies the quadrature mirror filter condition is used to construct the pyramidal filter bank (fig. 6):

$$G(z)G(z^{-1}) + G(-z)G(-z^{-1}) = 1 \qquad (1)$$

The high-pass filter $h(n)$ is obtained by shifting and modulating $g(n)$ (Unser 1995). Specifically, the z transform of $h(n)$ is chosen as:

$$H(z) = zG(-z^{-1}) \qquad (2)$$

The subsequent filters in the filter bank are then generated by increasing the width of $h(n)$ and $g(n)$ at every step, e.g.:

$$G_{i+1}(z) = G(z^{2^i}), \ H_{i+1}(z) = H(z^{2^i}), \ (i = 0, 1, ..., N) \qquad (3)$$

In the signal domain, the filter generation can be expressed as:

$$g_{i+1}(k) = [g]_{\uparrow 2^i}, \ h_{i+1}(k) = [h]_{\uparrow 2^i} \qquad (4)$$

where the notation $[]_{\uparrow m}$ denotes the up-sampling operation by a factor of m. The subbands computed by UDWPT and the original signal constitute the frequency branch of the double-tree undecimated DWT.

Time Segmentation

Now in order to extract temporal information in each subband, as in the frequency decomposition tree, every subband is segmented into non-overlapping time segments at each level with a pyramidal tree structure successively. In each time segment, the sum of the squares of the samples, energy, is computed as one feature to be used in off-line training. The time segmentation explained above forms the second branch of the double tree. From then on, we keep the index information of the dual tree structure to be used in the later stage for dimension reduction via pruning.

To summarize this section, refer to the double tree structure in figure 7. This double tree uses 1-level in each plane. The vertical middle boxes are the frequency subbands of UDWPT. Box 1 represents the unfiltered original signal, box 2 represents the low-pass filtered signal, and box 3 represents the high-pass filtered signal. Each of these subbands
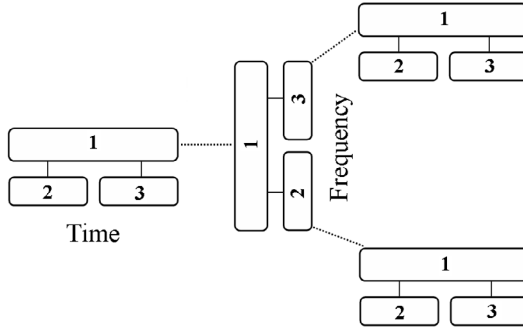
Figure 7. Double tree structure consisting of time and frequency planes.

is segmented in time into three segments as shown. Segment 1 covers the whole subband, segment 2 covers the first half of it, and segment 3 the second half of it. The parameters for deciding the number of features are the number of levels for frequency (F) and time segmentation (T). Let T be the number of levels in time and F be the number levels in frequency. There will be $2^{(F+1)} - 1$ subbands (including the original signal) and $2^{(T+1)} - 1$ time segments for each subband. This will produce the total number of features NF = $(2^{(F+1)} - 1)(2^{(T+1)} - 1)$.

Feature Subset Selection

As explained in the previous section, the dual tree has a total number of features NF = $(2^{(F+1)} - 1)(2^{(T+1)} - 1)$ for each sound signal, where F is the frequency level and T is the time level. For a typical value of $F = 3$ and $T = 3$, the dual tree produces 225 (NF = 225) features. Obviously, using a high-dimensional and correlated feature dictionary for classification is not an efficient approach. At this point, selection of a subset of features from this redundant dictionary is critical. Now let us shortly summarize some of the existing methods such as filtering and wrapper strategies that are widely used for subset selection in classification (Kohavi and John, 1997; Hall, 1998). The filtering approach uses a cost measure to select a subset of features and feeds them to the classifier for final decision. In general, the discrimination power of each individual feature is first estimated with a cost measure. Then the features are sorted according to their discrimination power in descending order. As a final step, the top subset is fed to the classifier. Since the subset is selected without implementing a search procedure, the filtering approaches have very low computational complexity. The subset selection procedure in the LDB algorithm, which is obtained by pruning and getting a rank of surviving features, is a filtering approach as well. Note also that the filtering approach does not evaluate the actual classification performance of the selected subset of features.

As an alternative strategy, a "wrapper" method is widely used to select a subset of features by inputting them to a classifier and measuring their combinatorial performance in classification. This method is very powerful in estimating the combinatorial performance of a set of features by implementing a search in the entire feature space. Sequential forward selection, sequential backward elimination, and floating search methods are some of popular strategies used in subset selection (Pudil et al., 1994). Typically, a cross-validation process is implemented on the training set to evaluate the classification performance of the inspected subset of features in classification. Since each set is processed by the classifier, the wrapper strategies have high computational complexity. Another drawback of the wrapper strategies is that they are very susceptible to overlearning. In a very rich feature dictionary, the wrapper methods can easily find a

subset and even use noise and/or correlated features, which in actuality do not carry the real discriminant information. Therefore, while selecting a subset, decorrelating the feature dictionary can be a crucial step. For this particular purpose, we will utilize the structure of the feature dictionary to select a subset with both filtering and wrapper strategies. In particular, we will prune the double-tree structure to select those nodes that do not overlap in time and/or frequency. What is meant with overlap is as follows. A selected node on the dual tree is a parent for the finer levels (for the nodes below it) or a child for the nodes above it. Therefore, each subspace of a particular node overlaps with its mother and child spaces. From now on, we will evaluate the efficiency of pruning by combining it with filtering and wrapper strategies in classification.

Three different types of methods are considered for feature selection (Type I, Type II, and Type III). The general structure of the algorithm for all three methods is given in figure 8. The leftmost box is the dictionary of features. LDA on the right is used both for classification and extracting the relationship among combinations of features. Rather than using a cross-validation procedure to asses the efficiency of features, we use a cost measure. For this reason, the output of the classifier is fed to a cost function to measure the discrimination power for that combination of features between classes. This measure will be used to select the best feature combination among other feature combinations. In this study, the Fisher discrimination (FD) (Duda et al., 2001) criterion:

$$FD = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{5}$$

and misclassification (MC) rate are used as cost functions to quantify the discrimination between two classes.

Type I

Type I is a sequential forward feature selection method. All of the features from the dictionary (from each class) enter the LDA one by one, and corresponding classification efficiency is measured for each feature using the cost function. After this search is done over all NF features, the best feature is selected by comparing the cost values of each one. In the next step, the second best feature, which will do the best in combination with the first selected one, is searched over the remaining features set one by one. This procedure is run until the desired number of features is reached. Type I uses all the boxes and connections in figure 8 except the feedback from the cost function to the dictionary. Since no dimension reduction is implemented, this approach has high computational complexity.

Type II

Type II is a modified version of Type I with an additional pruning module for dimension reduction. As in Type I, a sequential forward selection procedure is
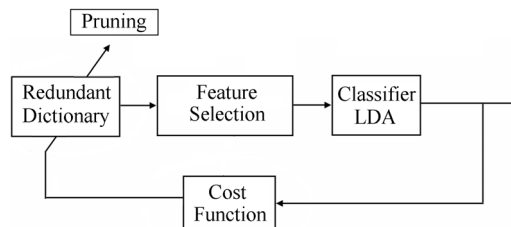

Figure 8. Block diagram of the proposed classification technique.

implemented. After selection of each feature, we use the feedback path from the cost function to the dictionary as in figure 8. The index of the selected feature corresponds to a node on the double tree, which has a frequency tree index and a time tree index in that subband. In the frequency tree, the nodes (subbands) that overlap with the selected frequency index are removed. Similarly in the time tree, the nodes that overlap with the selected time index are removed as well. In this way, only "good" potential features are kept in the dictionary; hence, the dictionary is pruned based on the last selected feature. Now the next feature, which will do the best in combination with the first selected one, is searched on the pruned dictionary. This procedure is run until the desired number of features is reached. Therefore, the only difference between Type II and Type I is that pruning is done on the dictionary based on the selected features.

Type III

Type III is the filtering approach with pruning. It does not use the LDA or a feedback path as in figure 8. Instead of using the classifier feedback, a cost value is computed for each node on the double tree individually. Then a pruning algorithm, as described by Saito et al. (2002), is run on the double tree from bottom to top to find the nodes with maximum discrimination power measured by the FD cost function. Once a node is selected, all nodes overlapping with the selected one are removed. After pruning the tree, the resulting feature set is sorted according to the features' corresponding discrimination power in descending order, and the top subset is input to the classifier. In this way, the most predictive features are input to the classification system. Since no feedback is used from the classifier, Type III has lower computational complexity compared to the other two methods.

The sound signals from each class are analyzed in an off-line manner for feature extraction by one of the three methods explained in the preceding sections. After features are selected and the decision rules are set, the system is run for validation.

Results

We tested the proposed approach on pistachio acoustic signals. We used a 2 times 2-fold cross-validation method to estimate the classification performance. Basically, half of the data set is used for training and the rest for testing. Then the test and training sets are swapped. This experiment is repeated twice. We use a frequency level $F = 3$ and time level $T = 3$ for the dual tree. After calculating the energy features in the nodes of the dual tree, they are converted to log scale. The log scale operation approximates the distribution of features to a Gaussian distribution where this assumption holds in the final classification step by LDA classifier.

Table 1 shows the classification accuracies obtained with the proposed methods with the Fisher discrimination (FD) cost measure. The classification error obtained with the base line algorithm of Pearson (2001) is given as BA. The comparison of all three types with their minimum errors and the number of features (NF) used to reach them are given. The Type I and Type II approaches, which use the feedback from the classifier, outperformed the Type III and BA approaches. As indicated before, the Type III only uses the same tree pruning method of the original LDB algorithm and does not account for the interactions between features. The classification accuracies we obtained strongly indicate that the evaluation of the classification performance of the combined features in the training stage is important and should be preferred to evaluating the individual discrimination power of the features. Furthermore, we note that the Type II approach always used a smaller number of features than the other approaches to achieve the minimal error.

Table 1. Open-shell and closed-shell pistachio nut minimum classification errors for the proposed types with FD criterion. NF stands for the number of features used to reach minimum error.

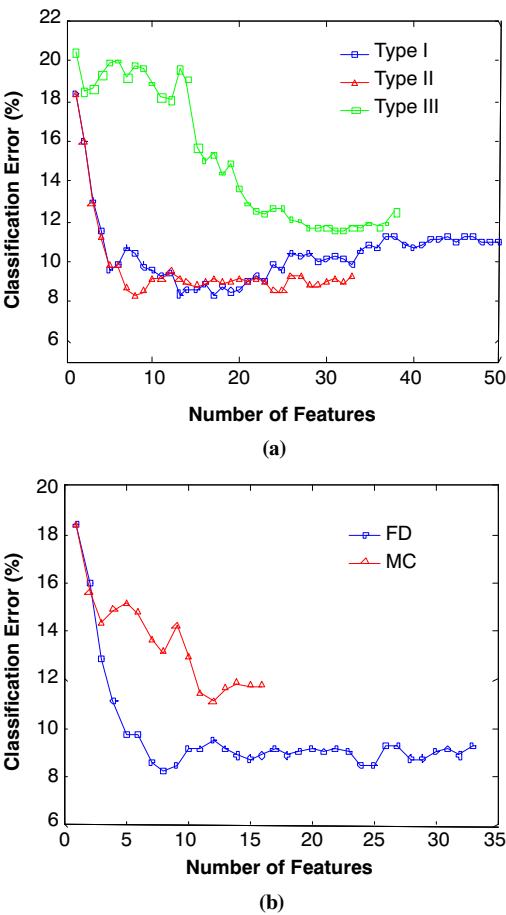| Type	Error (%)	NF
Type I	8.25	17
Type II	8.25	8
Type III	11.5	31
BA	18	3

(a)

(b)

Figure 9. Classification error curves for (a) all three types obtained with FD criterion, and (b) effect of cost functions (FD, MC) in classification for Type II.

The classification error curves versus the number of features for the results in table 1 are given in figure 9a. Since each selected feature index is used to prune the dual tree structure, the decorrelation of the redundant feature dictionary helps the Type II classifier to use a small number of features to achieve minimal error rate. Although the obtained classification accuracies of Type I and Type II are same, the number of features used carries significant importance for real-time applications. Since Type II uses a small number of features, it has lower computational complexity. It can be realized with a

Table 2. Effect of cost functions (FD and MC) for Type II.

Cost Function	Error (%)	NF
FD	8.25	8
MC	11.25	12

low-cost embedded system with a high throughput rate. Furthermore, using a small number of features can be a great advantage in generalization during classification.

To see the effects of the cost functions on the classification accuracy the results for Type II with the FD and MC criteria are given in table 2. We note that the FD criterion was superior to the MC cost. Interestingly, not only was the classification accuracy decreased, but also the number of features used to achieve minimum error was lower for the FD case. We believe that the evaluation of discrimination power with a measure that takes real values provides more accurate quantification of separability, as in the case of FD. On the other hand, for MC cost, the effectiveness of the combination of features is quantified with a discrete scale, indicating if a feature vector belongs to class A or B. We think that with MC cost, the classifier has more bias in overlearning, which is also indicated in figure 9b. For the FD case, the classifier tries to improve the distance between the class means and reduce their standard deviation, which is more effective.

In order to make an additional connection to the time-frequency images of each class given in figure 4, we prepared a time-frequency map of the first eight features of Type II. It reached the minimal error with this set. Selected features correspond to non-overlapping nodes on the dual tree selected by the wrapper strategy. These features may come from the same frequency level but different time segments or from the same time interval but different frequency subbands. The intensity of each segment on the map is related with that feature's individual discriminative power measured by the FD cost function. The resulting map is shown in figure 10. The most discriminant feature is selected from the tail of the impact. We note that the open shells produce an impact acoustic with a larger tail than closed shells.
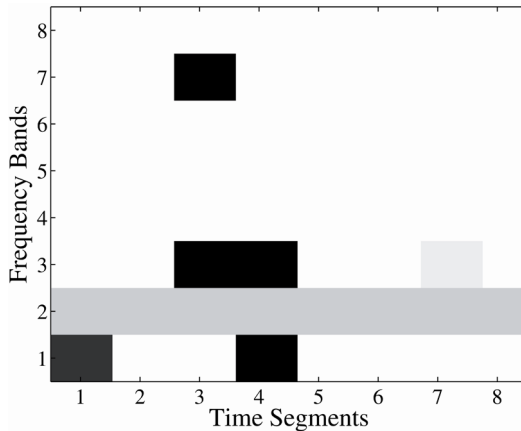

Figure 10. Discriminant time-frequency map of the first eight features of Type II. The darker regions have more discrimination power.

Conclusion

In this article, we described a new adaptive time-frequency plane feature extraction and classification algorithm to discriminate between open- and closed-shell pistachio nuts using their impact acoustics. The algorithm selects most discriminant features from a redundant structural dictionary by using a classifier feedback and a pruning procedure. In particular, the classification system uses the Fisher discrimination criterion or miss-classification cost to measure the effectiveness of the combination of features. The selected feature subset is fed to a linear discriminant analysis for final decision. We applied the algorithm to particular pistachio nuts from the Gaziantep region in Turkey, which are more difficult to classify than those from California. The best classification accuracy was 91.7%. This accuracy was obtained using only eight time-frequency features that are selected with the combination of dictionary pruning and classifier feedback with the Fisher criterion. When the pruning stage was removed, the same classification accuracy was obtained with 17 features, which will increase the computational complexity in real-time implementation. When the classifier feedback was removed, the system achieved a classification accuracy of 88.5%. This indicates that the evaluation of the combination of features by using a classifier feedback and a pruning procedure provides better results than evaluating individual discrimination power of the candidate features. In addition to different feature selection procedures, we also compared the effectiveness of the cost function at the end of the classifier. We observed that when the Fisher criterion is replaced with misclassification cost, the best results we obtained dropped to 88.7%.

A previously implemented algorithm, which uses maximum signal amplitude, absolute integration, and gradient features, achieved 82% classification accuracy on the same data set. The results we obtained show that our algorithm is superior to previous algorithms applied in this area. Our proposed algorithm does not depend on *a priori* knowledge of the time-frequency content of the signals under examination. Furthermore, its adaptation capability to both the time and frequency content of signals, concurrently, makes the algorithm a universal method for food kernel inspection, one that can resist the variability between nuts from region to region with respect to size and weight. One of the critical aspects to consider is the environmental noise that can disturb the impact signal. In real-life conditions, a mechanical device and several other sorting machines are expected to be running in the factory. In this case, the acoustic disturbances originating from other sorting machines and mechanical components may interfere with the impact acoustic recordings. Currently, we are exploring other impact signals, such as vibration, that may be more robust against interference from neighboring sorters than an acoustic signal.

References

Duda, R., P. Hart, and D. Stork. 2001. *Pattern Classification*. 2nd ed. New York, N.Y.: John Wiley and Sons.

Englehart, K., B. Hudgins, P. A. Parker, and M. Stevenson. 1999. Classification of the myoelectric signal using time-frequency based representations. *Medical Engineering and Physics on Intelligent Data Analysis in Electromyography and Electroneurography* 21: 431-438.

Ghazanfari, A., J. Irudayaraj, A. Kusalik, and M. Romaniuk. 1997. Machine vision grading of pistachio nuts using Fourier descriptors. *J. Agric. Eng. Res*. 68(3): 247-252.

Hall, M. A. 1998. Correlation-based feature selection for machine learning. PhD diss. Hamilton, New Zealand: Waikato University, Department of Computer Science.

Ince, N. F., S. Arica, and A. H. Tewfik. 2006. Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time-frequency tilings. *J. Neural Eng*. 3(3): 235-244.

Ince, N. F., A. H. Tewfik, and S. Arica. 2007. Extraction subject-specific motor imagery time-frequency patterns for single trial EEG classification. Computers in Biology and Med. 37(4): 499-508.

Kohavi, R., and G. H. John. 1997. Wrappers for feature subset selection. Artif. Intell. 97(1-2): 273-324.

Mallat, S. 2000. *A Wavelet Tour of Signal Processing*. 2nd ed. New York, N.Y.: Academic Press.

Pearson, T. C. 2001. Detection of pistachio nuts with closed shells using impact acoustics. *Applied Eng. in Agric*. 17(2): 249-253.

Pudil, P., J. Novovicova, and J. Kittler. 1994. Floating search methods in feature selection. *Pattern Recog. Letters* 15(11): 1119-1125.

Saito, N., R. R. Coifman, F. B. Geshwind, and F. Warner. 2002. Discriminant feature extraction using empirical probability density estimation and a local basis library. *Pattern Recog*. 35(12): 1842-1852.

Unser, M. 1995. Texture classification and segmentation using wavelet frames. *IEEE Trans. Image Proc*. 4(11): 1549-1560.

Vetterli, M. 2001. Wavelets, approximation, and compression. *IEEE Signal Proc. Magazine* 18(5): 59-73.