

FEATURE EXTRACTION FROM ACOUSTIC AND HYPERSPECTRAL
DATA BY 2D LOCAL DISCRIMINANT BASES SEARCH

HABİL KALKAN

DECEMBER 2008

FEATURE EXTRACTION FROM ACOUSTIC AND HYPERSPECTRAL DATA
BY 2D LOCAL DISCRIMINANT BASES SEARCH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HABİL KALKAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

DECEMBER 2008

Approval of the Graduate School of Informatics:

Prof. Dr. Nazife BAYKAL
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Yasemin YARDIMCI
Head of Department

This is to certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Yasemin YARDIMCI
Supervisor

Examining Committee Members

Prof. Dr. Fatoş T. YARMAN VURAL (METU, CENG) _____

Prof. Dr. Yasemin YARDIMCI (METU, II) _____

Prof. Dr. Enis ÇETİN (Bilkent Univ, EE) _____

Assoc. Prof. Dr. Erkan MUMCUOĞLU (METU, II) _____

Asst. Prof. Dr. Erhan EREN (METU, II) _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name, Last name : Habil Kalkan

Signature : _____

ABSTRACT

FEATURE EXTRACTION FROM ACOUSTIC AND HYPERSPETRAL DATA BY 2D LOCAL DISCRIMINANT BASES SEARCH

Kalkan, Habil

Ph.D., Department of Information Systems

Supervisor: Prof. Dr. Yasemin YARDIMCI

December 2008, 89 pages

In this thesis, a feature extraction algorithm based on 2D Local Discriminant Bases (LDB) search is developed for acoustic and hyperspectral data. The developed algorithm extracts the relevant features by both eliminating the irrelevant ones and/or by merging the ones that do not provide extra information on their own. It is implemented on real world data to separate aflatoxin contaminated or high risk hazelnuts from the sound ones by using impact acoustic and hyperspectral data. Impact acoustics data is used to sort cracked and intact shell hazelnuts with high classification accuracy. Hyperspectral images of the shelled and roasted (SRT) hazelnuts are used for classification and the algorithm extracted the spectral and spatial-frequency features for that classification. Aflatoxin concentration of the SRT category hazelnuts is automatically decreased to 0.7 ppb from 608 ppb by eliminating the detected contaminated ones.

Keywords: LDB, feature extraction, acoustic, hyperspectral, food safety.

ÖZ

2B YEREL AYIRTAÇ TABANLARI ARAŞTIRMASI İLE AKUSTİK VE HYPERSPEKTRAL VERİDEN ÖZİNİTELİK ÇIKARIMI

Kalkan, Habil

Doktora, Enformatik Enstitüsü

Tez Yöneticisi: Prof. Dr. Yasemin YARDIMCI

Aralık 2008, 89 sayfa

Bu tezde, akustik ve hyperspektral veriden 2B Yerel Ayırtaç Tabanları (YAT) araştırmasına dayalı bir öznelik çıkarımı algoritması geliştirilmiştir. Geliştirilen algoritma gereksiz özneliklerin atılması ve/veya tek başlarına fazladan bilgi taşımayanların birleştirilmesi sureti ile gerekli öznelikleri çıkarır. Çarpma akustik verisi kullanılarak çatlak kabuklu fındıklar sağlam kabuklu olanlardan yüksek bir sınıflandırma oranı ile ayıklanmıştır. Kavrulmuş iç fındıkların (SRT) sınıflandırılmasında hyperspektral görüntüler kullanılmış ve algoritma bu sınıflandırma için spektral ve uzamsal-frekans özneliklerini çıkarmıştır. Tespit edilen bozulmuş fındıkların ayıklanması ile SRT fındıklarındaki aflatoksin yoğunluğu otomatik olarak 608 ppb den 0.7 ppb'ye düşmüştür.

Anahtar Kelimeler: YAT, öznelik çıkarımı, akustik, hyperspektral, gıda güvenliği.

To my parents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following people for their contribution to the completion of my Ph.D. program.

First, I thank my supervisor Prof. Dr. Yasemin Yardımcı for her patient, valuable guidance, continuous support and friendly attitude throughout my research. Under her supervision, I have a chance to develop myself in both academic and personal subjects.

I have conducted some of my studies in University of Minnesota under the guidance of Dr. Nuri Firat Ince and Prof. Dr. Ahmet Tewfik and Dr. Tom Pearson from USDA-ARS-GMPRC. I would like to thank to them for their support and valuable comments.

I would also like to thank Prof. Dr. Enis Çetin and Assoc. Prof. Dr. Erkan Mumcuoğlu for their valuable suggestions. I also thank to Prof. Dr. Fatoş Tünay Yarman Vural and Asst. Prof. Dr. Erhan Eren for reviewing my work.

This thesis work is supported by The Scientific and Technical Research Council of Turkey (TÜBİTAK) by the project of 106E057:”Food Safety with Noninvasive Methods”.

Finally, I would like to express my deepest gratitude to my family for their unconditional support during my life. Special thanks goes to my wife, Nida, for her continuous support and understanding.

TABLE OF CONTENTS

ABSTRACT	iv
ÖZ	v
DEDICATION	vi
ACKNOWLEDGMENTS	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xvi

CHAPTER

1 INTRODUCTION	1
1.1 Motivation	2
1.2 Thesis statement	3
1.3 Thesis overview	4
2 LITERATURE REVIEW AND BACKGROUND	6
2.1 Feature Extraction and Selection	6
2.2 Best Bases Algorithm	10
2.3 Local Discriminant Basis Algorithm	11
2.4 Shift Invariance in Signal Classification	12

2.5	Dissimilarity Measure	14
2.6	Classification	15
3	REDUCTION OF AFLATOXIN IN HAZELNUTS BY IMPACT ACOUSTICS AND HYPERSPECTRAL IMAGING	16
3.1	Aflatoxin in Food Stuff	16
3.2	Detection of Aflatoxin Contaminated Food Items by Non-invasive Methods	16
3.3	Aflatoxin Problem in Hazelnut	18
3.4	Impact Acoustic Sorter System	20
3.5	Multi/Hyperspectral Imaging System for Food Safety	22
4	FEATURE EXTRACTION ALGORITHM BASED ON 2D LDB SEARCH	24
4.1	Feature Extraction from Acoustic Data by 2D Local Discriminant Bases Search	25
4.1.1	Adaptive segmentation and pruning in time axis	25
4.1.2	Adaptive segmentation and pruning in frequency axis	28
4.1.3	Selection of extracted time-frequency features	32
4.2	Feature Extraction from Hyperspectral Data by 2D Local Discriminant Bases Search	33
4.2.1	Feature Tree Generation	33
4.2.2	Adaptive Pruning in Spectral Axis	34
4.2.3	Adaptive Pruning in Spatial-Frequency Axis	35
4.2.4	Selection of extracted spectral spatial-frequency features	36
5	DATASETS USED IN EXPERIMENT	37
5.1	Acquisition of Impact Acoustics Data	37
5.2	Acquisition of Hyperspectral Data	41
5.3	Preprocessing the Hyperspectral Images	44

6	RESULTS AND DISCUSSIONS ON IMPACT ACOUSTIC DATA	
	CLASSIFICATION	45
6.1	Classification	48
6.2	Filter Selection	51
6.3	Effect of noise on classification	52
6.4	Effect of shift-invariance on classification	53
6.5	Effect of feature extraction algorithm on classification	54
6.6	Computational Complexity	57
7	RESULTS AND DISCUSSIONS ON HYPERSPECTRAL DATA CLASSIFICATION	58
7.1	Problem 1: Classification of <i>SRT_Cont</i> and <i>SRT_UnCont</i> Hazelnuts	59
7.2	Problem 2: Classification of <i>SRT_Afla+</i> and <i>SRT_Afla-</i> Hazelnuts	66
7.3	Problem 3: Classification of <i>SHD_Cont</i> and <i>SHD_UnCont</i> Hazelnuts	70
7.4	Problem 4: Classification of <i>SHD_Afla+</i> and <i>SHD_Afla-</i> Hazelnuts	74
8	CONCLUSIONS	79
8.1	Future work	82
	REFERENCES	83
	VITA	89

LIST OF TABLES

3.1	The mean contamination levels of the hazelnuts from four groups	19
5.1	Number of aflatoxin contaminated kernels and the mean aflatoxin level (ppb) of the three group SRT hazelnuts	42
5.2	Number of aflatoxin contaminated kernels and the mean aflatoxin level (ppb) of the three group SDT hazelnuts	42
6.1	Classification rate comparison of proposed LDB based method against the other methods	50
7.1	Minimum classification error obtained by four feature selection algorithms. The number of features of methods giving the error is shown in brackets	63
7.2	Minimum classification error obtained by four feature selection algorithms.	68
7.3	Minimum classification error of <i>SHD_Cont</i> and <i>SHD_UnCont</i> data obtained by four feature selection algorithms. The number of features of methods giving the error is shown in brackets	73
7.4	Minimum classification error of <i>SHD_Afla+</i> and <i>SHD_Afla-</i> data obtained by four feature selection algorithms. The number of features of methods giving the error is shown in brackets	76

LIST OF FIGURES

2.1	a) Filter and b) wrapper feature subset selection model	9
2.2	Two impact acoustic signal of hazelnut kernels from the same class.	13
3.1	A sample image from a) Regular shell hazelnuts (ReH), b) Cracked shell hazelnuts (CrH), c) Broken shell hazelnuts (BrH) and d) In-shell hazelnuts (InH)	19
3.2	The forms of exported hazelnut in TURKEY	20
3.3	Schematic of experimental apparatus for collecting acoustic emissions from hazelnut kernels	21
3.4	Hyperspectral imaging system with a rotating filter wheel	23
4.1	The block diagram of the 2D LDB based feature extraction for one dimensional signal	25
4.2	Time segmentation with LCP at 4 levels in dyadic tree structure .	26
4.3	The smooth windows obtained by LCP	27
4.4	Pyramidal filter tree up to second level. L and H stands for Low and High band, respectively	29
4.5	Filter bank	30
4.6	Undecimated filter banks up to third level decomposition	31
4.7	The 3 level full wavelet sub-band tree	32
4.8	The block diagram of the 2D LDB based feature extraction for hyperspectral data	33
4.9	k=4 levels binary spectral band tree	34
4.10	Full wavelet decomposition quad tree up to h=2 levels	34
5.1	Weights of 'EmH', 'CrH' and 'ReH' hazelnuts	37
5.2	Typical impact acoustic signals of ReH, CrH and EmH hazelnuts .	38
5.3	Signal energy of hazelnuts at various weights	39
5.4	The averaged spectrogram of a) CrH and b) ReH hazelnuts	40

5.5	Categories of hazelnut used in the experiment	41
5.6	A few spectral band images for <i>FlavusCont</i> , <i>WaterCont</i> and (<i>Un-Cont</i>) group of a) SRT and b) SHD hazelnuts	43
5.7	A raw spectral band image with unskinned region at the surface, b) masked image of 430 nm spectral band	44
6.1	The adaptive time segmentation grids (dotted lines) of a-) of set1 and b-) of set2	46
6.2	The location of most discriminative features in time-frequency axis	47
6.3	The time-frequency discrimination map of impact acoustic data. Darker regions indicate higher discrimination power	48
6.4	The classification error rates (%) with various numbers of features	49
6.5	Receiver Operating Characteristics (ROC) curves	50
6.6	The effect of selected wavelets and feature dimension on classification accuracy. a) Daubechies, b) Coiflet	51
6.7	The classification error curves for noise disturbed impact acoustic signals	52
6.8	The classification error curves for evaluating the efficiency of shift invariance property. The spin-cycle curve stands for the results obtained from DWT supported 1-Spin-Cycle procedure	54
6.9	The classification error curves of a) FFS and b) CFS method with extracted and candidate feature set	55
6.10	The classification error curves of CFS and FFS method with extracted feature set	56
7.1	The schematic assignments of hazelnuts <i>Afla+</i> and <i>Afla-</i> groups . .	59
7.2	Entropy map of the spectral spatial-frequency features	60
7.3	Spectral band pruning. The spectral bands (440-470) and (480-510) are pruned. The null bands in the tree were ignored at pruning	60
7.4	The location of the most discriminative features in spectral spatial-frequency axis	61
7.5	Spectral Spatial-Frequency feature map of SRT hazelnut data ranked by a) by FFS, b) by CFS algorithm	62

7.6	The classification error curves on FFS, CFS, PCA and Wrapper based selected ranked features	63
7.7	The classification error curves with the features selected from candidate and extracted features by FFS algorithm	65
7.8	The classification error curves with the features selected from candidate and extracted features by CFS algorithm	65
7.9	The classification error curves with the features selected from candidate and extracted features by PCA algorithm	66
7.10	Spectral Spatial-Frequency feature map of <i>SRT_Afla+</i> and 189 <i>SRT_Afla-</i> data which were ranked by a) by FFS, b) by CFS algorithm.	67
7.11	The classification error curves of <i>SRT_Afla+</i> and <i>SRT_Afla-</i> data by FFS, CFS, PCA and Wrapper based selected features . .	68
7.12	The classification error curves with the features selected from candidate and extracted features by FFS algorithm	69
7.13	The classification error curves with the features selected from candidate and extracted features by CFS algorithm	69
7.14	The classification error curves with the features selected from candidate and extracted features by PCA algorithm	70
7.15	Spectral Spatial-Frequency feature map of <i>SHD_Cont</i> and <i>SHD_UnCont</i> data which were ranked by a) by FFS algorithm	71
7.16	The classification error curves of <i>SHD_Cont</i> and <i>SHD_UnCont</i> data by FFS, CFS and PCA based selected features	72
7.17	The classification error curves with the features selected from candidate and extracted features of SHD category hazelnuts by FFS algorithm	73
7.18	The classification error curves with the features selected from candidate and extracted features by of SHD category hazelnuts CFS algorithm	74
7.19	The classification error curves with the features selected from candidate and extracted features by of SHD category hazelnuts PCA algorithm	74

7.20	Spectral Spatial-Frequency feature map of <i>SHD_Afla+</i> and <i>SHD_Afla-</i> data which were ranked by a) by FFS algorithm	75
7.21	The classification error curves of <i>SHD_Afla+</i> and <i>SHD_Afla-</i> data by FFS, CFS and PCA based selected features	76
7.22	The classification error curves with the features selected from candidate and extracted features of SHD category hazelnuts by FFS algorithm	77
7.23	The classification error curves with the features selected from candidate and extracted features of SHD category hazelnuts by CFS algorithm	77
7.24	The classification error curves with the features selected from candidate and extracted features of SHD category hazelnuts by PCA algorithm	78

LIST OF ABBREVIATIONS

μ	: mean
σ	: standart deviation
H	: Entropy
LDA	: Linear discriminant analysis
PCA	: Principal component analysis
LDB	: Local discriminant bases
LCP	: Local cosine packets
CFS	: Correlation based feature selection
FFS	: Fisher discriminant based feature selection
$SPCT$: Segmented principal component transformation
$STFT$: Short time fourier transform
SNR	: Signal to noise ratio
DWT	: Discrete wavelet transform
$UDWT$: Undecimated discrete wavelet transform
TF	: Time-frequency
$NATF$: Non-adaptive time-frequency features
$NASB$: Non-adaptive subband features
$BGYF$: bright greenish-yellow fluorescence
ppb	: part per billion
$HPLC$: High performance liquid chromatography
kNN	: k-Nearest neighbor
SVM	: Support vector machine
MLP	: Multi layer perceptron
HMM	: Hidden markov model

<i>FuH</i>	:	Fully developed hazelnut
<i>EmH</i>	:	Empty or undeveloped hazelnut
<i>ReH</i>	:	Fully developed hazelnut with regular shell
<i>CrH</i>	:	Fully developed hazelnut with cracked shell
<i>BrH</i>	:	Fully developed hazelnut with broken shell
<i>InH</i>	:	In-shell hazelnut
<i>SRT</i>	:	Shelled and roasted hazelnut
<i>SHD</i>	:	Shelled hazelnut
<i>UnCont</i>	:	Uncontaminated hazelnut
<i>FlavusCont</i>	:	Flavus incubated hazelnut
<i>WaterCont</i>	:	Pure water incubated hazelnut

CHAPTER 1

INTRODUCTION

The recent developments in electronics and computer technologies have made capturing and storing of images, signals or video in real-time easy and inexpensive. Machine learning algorithms analyze these data automatically. Feature extraction algorithms in machines learning focus on getting the useful and necessary content of the information. In this thesis a feature extraction technique based on Local Discriminant Bases (LDB) is developed to explore the location of the discriminative features for acoustic and hyperspectral data. The algorithm is tested on a real world problem in food safety field. The developed algorithm is robust and adaptive to other type of data and it is able to extract the relevant features which increase the performance of learning algorithms.

The feasibility of real-time food sorting and grading systems with satisfactory performance depends on how well the acquired data is represented with fewer number of features for classification. The data taken from food item may be high dimensional and low quality with low signal to noise ratio. High dimensional data is expected to increase the accuracy and effectiveness of classification algorithms. However, the increase in data dimension not only increases the computational complexity, but also decreases the classification accuracy, when a limited number of training data is available [1, 2]. Moreover, the high dimensional data increases the processing time which is inversely proportional to throughput of the system. Therefore, the data dimension should be reduced by either eliminating the irrelevant ones or by combining the features that do not provide extra information on their own.

The current dimensionality reduction algorithms are usually classified into two categories: feature extraction and feature selection. Feature selection is the process of moving to low dimensional feature space by eliminating the irrele-

vant features as much as possible and this process does not distort the original data by transformation. Unlike feature selection, feature extraction is defined as transforming to lower dimensional feature space by using various transformation methods. Although the reduction in dimension allows learning algorithms to operate faster, this may remove the significant interpretation of original data. Feature extraction is extensively studied by various researchers [3, 4, 5, 1].

Local Discriminant Bases (LDB) algorithm, which originated from Best Bases algorithm [6] is developed to extract local information from data by using a divide and conquer approach [7]. The original LDB algorithm searches discriminant bases, which are regarded as feature, of the 1D signals either in time or frequency axis. However, it has been shown that bases search in both axes is crucial for classification[8, 9, 10]. LDB algorithm is also adapted to hyperspectral data to get the relevant features by accepting the hyperspectral curve of pixels as a one dimensional signal [11, 12, 1]. In this thesis, LDB algorithm is applied to a practical nut sorting method to get the location of discriminative features in data.

1.1 Motivation

Hazelnuts are extensively used in the chocolate and confectionary industries. They are produced in Iran, USA and Mediterranean Countries such as Turkey, Italy, Spain, France, Greece and Portugal. However, Turkey is the major producer of hazelnuts in that 70% percent of the hazelnuts in worldwide is produced by Turkey and about 80% of the Turkish Hazelnut export is consumed by the European Union (EU) countries.

Unfortunately hazelnuts are prone to aflatoxin formation like many other natural products among which are corn, red pepper, almonds, and figs. The aflatoxins are a group of structurally related toxic compounds produced by certain strains of the fungi , *Aspergillus Flavus* and *Aspergillus Parasiticus*. Under favorable conditions of temperature and humidity, these fungi grow on certain foods and feeds, resulting in the production of aflatoxins. Aflatoxins produce acute necrosis, cirrhosis, and carcinoma of the liver on laboratory animals at low quan-

tities [13]. Due to the health risks they pose on humans; it is desirable to lower the aflatoxin levels in consumed food items. The present allowed level for aflatoxin contamination in seed is 20 ppb (ng/g) and 4 ppb in USA and EU, respectively.

The hazelnuts are subject to aflatoxin analysis before exporting. A total of 40 incremental samples of 300 gr each are taken from the lot of 2 tonnes and analyzed via chemical methods. A few highly aflatoxin contaminated kernels in the incremental sample may increase the aflatoxin concentration of the sample and finally may result the cancellation of export of the examined lot. In addition, the sample which has undergone chemical testing can no longer be consumed. Therefore, real-time passive screening procedures which do not destroy samples are highly desirable and can be used to detect and remove such contaminated samples.

For real time sorting systems, a classifier with maximum accuracy but with minimum feature vector dimension is desirable. In addition, the feature vector should be based on least amount of sensor data so that a simple and fast data acquisition system can be employed. Unlike the traditional feature extraction algorithm which use all the candidate features, LDB algorithm combines these features to obtain a more compact feature set. These features are fed to the feature selection step to identify the most relevant part of the data.

In this thesis, we developed 2D "Local Discriminant Bases" based feature extraction algorithms to detect contaminated or potentially contaminated hazelnuts by non-invasive and rapid methods. We used impact acoustics and hyperspectral imaging data depending on the hazelnut kernels type and the developed algorithm is tested on both types of data. The algorithm s developed to

1.2 Thesis statement

In this thesis, we modified the original LDB algorithm to two dimensions to get the exact location of the discriminative features in data space. The bases search is performed in time and frequency axis of impact acoustic data, spectral and spatial-frequency axis for hyperspectral data. The developed algorithm decreases the feature dimension by either eliminating the irrelevant ones and/or combining

the ones that do not provide extra information on their own.

This algorithm is implemented on selecting the contaminated or potentially contaminated hazelnut kernels by using impact acoustic and hyperspectral data. It will be able to sort the hazelnut kernels by fewer number of features with high classification accuracy. It is also observed that prior to feature selection, feature extraction is a critical step on getting the discriminative features.

1.3 Thesis overview

Chapter 2 reviews the principle of feature extraction and selection algorithms and provides a brief overview on Best Bases algorithm and Local Discriminant Bases algorithm which the proposed feature extraction algorithm originates from. It also gives the shift invariance concept that is used for discussion of proposed algorithm. Some dissimilarity metrics and the classifier used in the study are given in Section 2.

Chapter 3 explains the aflatoxin problem in foods and especially in hazelnuts which is among the major export goods of Turkey. It also presents some non-invasive aflatoxin detection algorithm developed for specific food items. This section also describes impact acoustics sorter system and hyperspectral imaging systems which are also used to acquire data for our algorithm.

Chapter 4 presents the proposed feature extraction algorithm by 2D Local Discriminant Bases search for impact acoustic and hyperspectral data. Section 4.1 focuses on the algorithm for acoustic data whereas Section 4.2 focuses on hyperspectral data. Both algorithms are same in principle but have some minor differences depending on the data type.

Chapter 5 presents the data acquisition procedures for impact acoustic data and the signal acquired from hazelnuts. The weight of the hazelnut for different classes and the weight versus signal energy correlation are given in this Section 5.1. Section 5.2 gives the hyperspectral image acquisition procedures and the hyperspectral images of shelled hazelnuts. The aflatoxin concentration of the classes and the preprocessing of the hyperspectral images is also given in Section 5.2.

Chapter 6 begins by presenting the output of the steps of feature extraction algorithm developed for impact acoustic data. The classification results obtained with extracted features are compared to other type of features extracted from the same dataset in Section 6.1. The filter selection criteria and the effect of noise and shift invariance on classification are given in succeeding three sections. The relevance of extracted features for classification is compared to candidate feature set by two different feature selection algorithms.

Chapter 7 presents the findings of hyperspectral data classification on four different classification problems defined on separating the mold contaminated or aflatoxin contaminated kernels of two different hazelnut forms. The outputs of the spectral spatial-frequency feature extraction steps are given in detail for the mold detection of shelled&roasted hazelnut kernels. The summary of classification results and the relevance of extracted features are given for each problem separately.

Chapter 8 presents conclusions and suggests future works.

CHAPTER 2

LITERATURE REVIEW AND BACKGROUND

2.1 Feature Extraction and Selection

In classification problems, it is expected that the high dimensional data should increase the accuracy and effectiveness of the classification. However, it is observed that the increase in the dimension decreases the classification accuracy when a limited number of training data is available [1]. When the dimension of the data increases the number of training samples should be increased exponentially in order to retain the accuracy of classifier. This is a common problem in hyperspectral data. In order to overcome this limitation, the number of dimension is reduced by eliminating the irrelevant ones or by combining features that do not provide extra information on their own [14, 5, 11].

The current dimensionality reduction algorithms are usually classified into two groups; feature extraction and selection. Feature extraction is defined as transforming to M dimensional feature space from N dimensional ($M < N$) measured data space [15] by using various transformation methods. Although the reduction in dimension allows learning algorithms to operate faster, this may remove the significant interpretation of original data.

Feature extraction algorithms are mostly developed depending on the problem and the data set. One of the common feature extraction algorithm is Principle Component Analysis (PCA) [4]. In PCA, new sequences of uncorrelated variables are generated by using Karhunen-Loeve transformation. The first M number of components are selected and used for classification or representation. This enables us to represent the high dimensional data by a few principle component.

One drawback of PCA is that it works with whole data set for transformation. Jia and Richards [5] segmented the PCA to remove the need for whole hyperspectral data set. They developed a PCA based feature extraction algorithm: Segmented Principle Component Transformation (SPCT). In SPCT, the highly correlated adjacent bands in hyperspectral data are grouped by edge detection. The principle components in each group are computed and each grouped is represented by a few components (eigenvectors) with high eigenvalues. The eigenvectors are then pruned according to their Bhattacharya distance between classes. However, the projection based dimension reduction methods are sensitive to noise. Therefore projection should be made using a large number of samples in order to overcome this problem. Although the simplicity and popularity of PCA, the features derived from PCA projection may not have better discrimination than the features in original data space [12]. Unlike to tranformation based algorithms, relevant features can be extracted by first order and second order statistics[3] or Best Bases Extraction algorithms [11] and etc..

Unlike feature extraction, feature selection is the process of moving to the K dimensional feature space from M ($K < M$) dimensional feature space by eliminating the irrelevant features as much as possible. The feature selection algorithms in the literature are usually performed [16] on four steps;

Step 1: Select an initial point in feature space

Start with an initial point in feature space. This selection is important because it may effect to find the solution in feature space.

Step 2: Search feature space

The search strategy of feature space is important because it determines the search direction. There exist 2^N possible feature subsets in N dimensional feature space. Generally speaking, there are two different search strategies for subset selection. The first one is *exhaustive search strategy* where all the subsets are investigated individually. This is an NP complete problem and suffers from computation complexity. The second one are *heuristic strategies*, which search the

feature space by heuristics methods. Basically, there are three types of heuristic method that is primarily works on addition or deletion approach [17].

Forward Selection starts with empty subset and successively expands by adding most relevant feature that will provide local improvement to the problem.

Backward Elimination starts with complete subset and successively deletes the most irrelevant feature from the current subset.

Stepwise Bidirectional Selection starts with null, full or randomly selected feature subset and adds the most relevant feature or removes the most irrelevant feature from the current subset.

Step 3: Evaluation of selected feature subset

The merit of the selected feature subset is evaluated by wrapper or filter model [Figure 2.1]. Evaluation functions which measure and determine the classification capability of individual features are used in filter model. In contrast, induction algorithm (learning accuracy) is executed to measure discrimination capability of the feature set in wrapper model. In wrapper model, data of the investigated feature subset is randomly divided into test and train set. An initial machine learning algorithm is trained with the train set and tested with the test set. The accuracy of the testing gives the merit of the investigated feature sub set. The wrapper model is slower than the filter function because the induction function is executed at every feature increment. However, this approach may give better results compared to filter models [18].

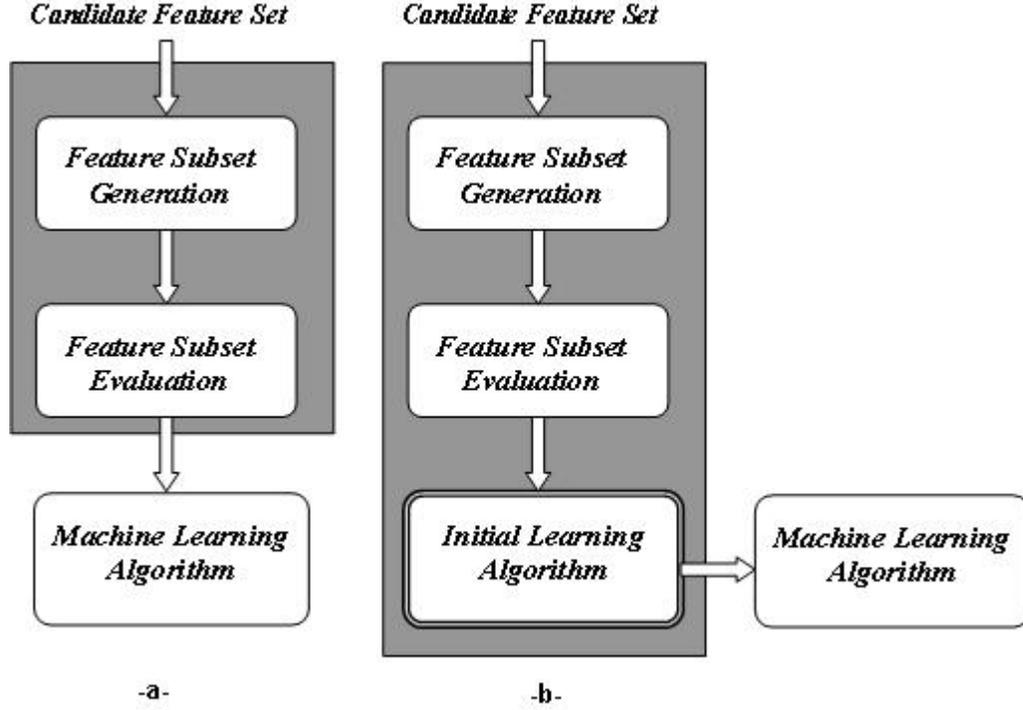


Figure 2.1: a) Filter and b) wrapper feature subset selection model

Step 4: Termination of feature selection

The feature subset selection process is terminated by predefined criteria. The process can be stopped when the addition of new feature does not improve the merit of feature subset (this is classification accuracy in wrapper model) or the process can be stopped at a given number feature subset size.

The critical step in feature selection algorithms is to evaluate the merit of the feature subset. Various evaluation functions may be used such as Mutual Information [19], Entropy, Information Gain, MDL, Gini, Relief, etc. [17, 16]. Hall [16] developed Correlation Based Feature Selection (CFS) based on the Pearson's Correlation. In CFS algorithm, the initial feature subset of size $k - 1$ is increased by the feature which gives the maximum merit M_s .

$$M_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k - 1)\bar{r}_{ff}}} \quad (2.1)$$

where $\overline{r_{cf}}$ is the mean class correlation and $\overline{r_{ff}}$ is the average feature-feature intercorrelation. The CFS algorithm aims to find the feature subset including the features that are highly correlated with the class but uncorrelated with each other. The algorithm is terminated when there is no improvement in the M_s value.

2.2 Best Bases Algorithm

Coifman and Wickerhauser developed a best basis algorithm for signal compression [6]. The best basis method firstly expands the signal into orthonormal wavelet or trigonometric basis in binary tree structure. Each basis vector has different location in time-frequency axis and some of them may be redundant for signal representation. The importance of each basis (node at binary tree) is evaluated at the second step by using a defined minimization criterion, which is the entropy for signal compression. At the third step, the binary tree is pruned from bottom to top by using the entropy values of nodes. The best-basis algorithm can be summarized as follows,

Step 1: Define a decomposition method (wavelet or trigonometric) and expand the signal into orthonormal basis vectors in binary tree structure.

Step 2: Evaluate the information cost (entropy) of each node by using the expansion coefficients. The entropy H of a sequence $\{p\}$ with $\sum p_i = 1$ can be calculated as

$$H(p) = - \sum_i p_i \log(p_i) \quad (2.2)$$

Step 3: Prune the binary tree by mother and child node comparison.

The nodes at the level, right before the deepest level in tree are firstly selected as parent nodes. The parent node is discarded if its children nodes have less cumulative information and then the cumulative value of children are set to parent's

value. Otherwise, the children are destroyed and the parent is kept. This parent is set as a child for the higher level. The survived nodes at the end of the whole pruning operation are the nodes that are the best basis for signal compression.

2.3 Local Discriminant Basis Algorithm

As stated before, the best-basis algorithm [6] is developed to get local information for signal representation and compression. Entropy values of the nodes are used as information cost at pruning the binary tree. Saito and Coifman [7] adapted this best-basis algorithm for classification of signals and images. The local discriminant basis of the signals are evaluated and used for classification purposes. Instead of using the entropy, they proposed to use a dissimilarity cost function which will maximize the difference in time-frequency energy distributions of classes. The LDB algorithm first decomposes the signal into orthonormal basis (nodes) by using wavelet trigonometric packets in binary tree structure. Each basis vector has different time-frequency locations. The discrimination power of each node is calculated by using an appropriate dissimilarity measure. The best basis are detected by pruning the binary tree using divide and conquer algorithm. When pruning the binary tree, a parent node is split only if the cumulative discriminative power of its children is greater than the discrimination of the parent node. The parent node in that situation is destroyed and the cumulative discrimination power of children is transferred to the parent node. Whenever the cumulative discrimination power of children is smaller than the parent node the child nodes are destroyed and the parent node with its discrimination value becomes a child of a node at higher level. The resulting best basis vectors are ordered by their discrimination power and used for constructing the classification machine. The LDB algorithm can be summarized as follows:

Step 1: Define a decomposition method (wavelet or trigonometric) and expand the signal into orthonormal basis in binary tree structure up to level J .

Step 2: Construct the time-frequency (TF) energy map of classes.

Step 3: Construct discrimination power tree D from the TF energy map by using an appropriate dissimilarity measure.

Step 4: Prune the J level tree by the following rule:

- start from the nodes in $j=J-1$ level
- for all k nodes in level j , do the following;
 - if $D(k) > [D(k)_{child1} + D(k)_{child2}]$
 - keep mother node, destroy children nodes
 - else
 - keep children nodes and set $D(k) = [D(k)_{child1} + D(k)_{child2}]$

Step 5: Order the selected best basis vectors (say M bases) by their discrimination value

Step 6: Use $K < M$ basis for constructing classification machine.

2.4 Shift Invariance in Signal Classification

The real time signals are not always aligned in time (or space). Figure 2.2 shows two impact acoustics signals sequence for demonstration. The phase difference between acoustic signals are obvious in Figure 2.2.

For a shift varying system, a shift at the input signal may results a big difference at the output. Therefore, the developed signal classification algorithms that use the features extracted from the output signals may not have the highest

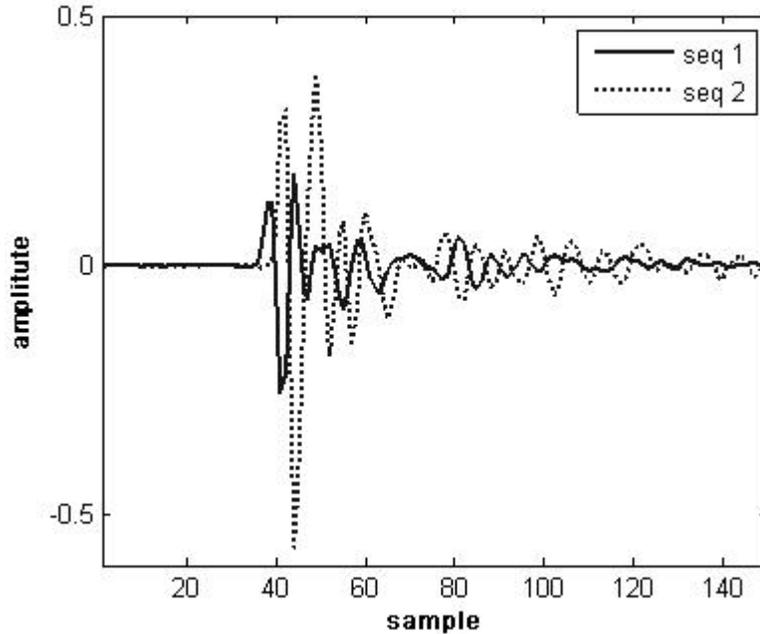


Figure 2.2: Two impact acoustic signal of hazelnut kernels from the same class.

accuracy. Shift invariance is usually introduced as requirement for robust classification. The importance of the shift invariance for classification is also emphasized in [7, 8, 20].

There are some shift invariant systems, but the wavelet and trigonometric packets that are used at LDB algorithm generates shift variant transformation. Some researchers used spin-cycle procedure to get shift invariance properties for signal classification purposes. [21, 22, 9]. In spin-cycle procedure, the input data set is expanded by shifting the original set in both directions in order to handle the possible shifts at the input signals. If the defined number of shift is τ then the data set is expanded by its $(-\tau, \tau + 1, \dots, 0, \dots, \tau - 1, \tau)$ versions and used as independent signals.

One drawback of spin-cycle procedure is increased computational complexity; because of the increased data size. This drawback must be taken account when the performance of the developed algorithm is considered.

Undecimated wavelet transform (UDWT) employed in this study; has shift invariance property. It is widely used in image denoising and enhancement ap-

plications [23, 24, 25]. It is also used for texture classification [26]. The UDWT can be used with the LDB algorithm to make it more robust against to the shifts for signal classification.

2.5 Dissimilarity Measure

The dissimilarity (distance) measure is used to compute the discrimination power of the basis vector(node). It has a direct effect on classification accuracy. Therefore, the best distance measure function should be investigated for a given problem. Let p and q are normalized energy distributions of signals belonging to *class 1* and *class 2*, respectively, where $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1$. The distance measure can be:

- The symmetric Kullback Leibler distance, which is also named as J-divergence

$$J(p, q) = I(p, q) + I(q, p) \quad (2.3)$$

$$I(p, q) = \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right)$$

- Euclidean distance

$$D(p, q) = \left\| p_i - q_i \right\|^2 = \sum_{i=1}^n (p_i - q_i)^2 \quad (2.4)$$

- Hellinger Distance

$$H(p, q) = \sum_{i=1}^n (\sqrt{p_i} - \sqrt{q_i})^2 \quad (2.5)$$

- Fisher distance

$$F = \frac{|\mu_1 - \mu_2|}{\sigma_1^2 + \sigma_2^2} \quad (2.6)$$

where μ_i and σ_i are the mean and standard deviation of the corresponding feature for class i .

2.6 Classification

In pattern recognition problems, classification can be defined as assigning the test pattern to a learned class. Various types of classifier are used in literature such as Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Multi Layer Perceptron (MLP), Linear Discriminant Analysis(LDA), Hidden Markov-Model(HMM), etc.. However the common point in these classifiers is the requirement to work with relevant and orthogonal features. The main objective of this study is the extraction of features for classification. Therefore, a single classifier, LDA, is selected and used for detailed analysis.

LDA is one of statistical classifiers commonly used in pattern recognition problems [27]. LDA generates the best hyperplane decision surface in M dimensional space. The orientation and location of the surface are determined by the vector w and bias w_0 , respectively.

$$g(x) = w^t x + w_0 \quad (2.7)$$

The discriminant function $g(x)$ gives the distance of the test pattern x to the decision surface where $g(x)$ takes positive value when x is on the positive side and takes negative value when $g(x)$ is on the negative side of the surface. The detailed description of the LDA classifiers can be found at [27]. The surface weight vector w used in LDA for two classes is

$$w = \left(\sum_1 + \sum_2 \right)^{-1} (\mu_1 - \mu_2) \quad (2.8)$$

where μ_i and \sum_i are the mean vector and covariance matrix of the class i , respectively.

CHAPTER 3

REDUCTION OF AFLATOXIN IN HAZELNUTS BY IMPACT ACOUSTICS AND HYPERSPECTRAL IMAGING

3.1 Aflatoxin in Food Stuff

Aflatoxin that is caused by *Aspergillus* type molds (*A. Flavus*, *A. Parasitius*) is one of the defects affecting the chemical composition of the food item. Aflatoxin may cause porcine, pulmonary edema, liver cancer and esophageal hyptesia. It is estimated that in USA 58 to 158 people per year are inflicted with liver cancer because of aflatoxin consumption [13]. Therefore, aflatoxin contamination levels of seeds (corn, pistachios, hazelnuts, pepper, etc..) is restricted by legislations. The level for aflatoxin contamination in seed is 20 ppb (ng/g) and 4 ppb in USA and European Union (EU), respectively. Aflatoxin in foods can be detected by high-performance liquid chromatography (HPLC), mass spectroscopy (MS) and enzyme-linked immunosorbent assay (ELISA).

3.2 Detection of Aflatoxin Contaminated Food Items by Non-invasive Methods

Several studies have been conducted to detect aflatoxin contaminated foods by non-invasive techniques. A spectrophotometer is commonly used to detect compound of food items under inspection and it is also used to estimate aflatoxin contamination in food kernels. A spectrophotometer consists of a *spectrometer*

to produce light and a *photometer* to collect the reflected (R) or transmitted (T) light in various spectral bands. Pearson [28] used spectral reflectance ratio (R735nm/R1005nm) for detecting highly contaminated corn kernels (>100 ppb) from low contaminated (<10 ppb) or uncontaminated ones under illumination and reached 95% correct classification rate. Pearson could also distinguish the high contaminated (>100 ppb) yellow corn kernels at a rate of 98% by using spectral absorbance at 750 nm and 1200 nm [28]. Hirano [29] used transmittance ratio (T700nm/T1100nm) of peanuts for identification of contaminated kernels from uncontaminated ones. The imaging with spectrophotometer is time consuming. Therefore the non-invasive food safety algorithms developed with spectrophotometer is not suitable for real-time operations.

Aflatoxin may also be present inside the kernel. In that case, it may not be possible to visualize the moldy effects on the surface of the kernel. Nuclear Magnetic Resonance (NMR) and X-Ray Imaging can be used to detect internal aflatoxin contamination. However, these are highly expensive techniques and take more than 10 minutes to evaluate each kernel.

The mold of *Aspergillus Flavus* produces kojic acid which is observed to bright greenish yellow fluorescence (BGYF) compound by peroxidase in the plant. These plants exhibited BGY particles under UV radiation and this radiation can be detected by machine vision. The number of exhibited Bright Greenish Yellow (BGY) particles is taken as an indication of aflatoxin contamination. The BGY fluorescence is also used by Tyson and Clark [30] for detecting aflatoxin contaminated pecans. They soaked the pecan halves in aflatoxin solution for two days. They measured the BGY fluorescence of the pecan halves under UV light at 320 nm and used fluorescence ratios at 440nm/490nm and 450nm/490nm for classification. Fersai [31] used the same method for detecting aflatoxin contaminated pistachio nuts by using the fluorescence ratio of 420nm/490nm. Fersai stated that the peak emission occurs at 490 nm for the BGY nuts. However, the human eye has only 30% percent sensitivity at 490nm compared to 550nm. This insensitivity causes fluorescence particles to appear as BGY although their peak sensitivities are at the region between blue and green not at yellow region. The BGY fluorescence is also used to detect aflatoxin contaminated corn kernel [32], figs [33, 34],

pistachio and Brazil nut [35, 36] and in some agricultural commodities [37].

However, Wilson detected that aflatoxin contaminated corn kernels do not exhibit BGY fluorescence all the time due to the insufficient amount of peroxidase in plant. Moreover, other types of fungi, besides the peroxidase, which don't produce aflatoxin, may yield kojic acid in foods and they may be regarded as aflatoxin contaminated foods when the BGY fluorescence is considered only. These findings are detailed in [38].

3.3 Aflatoxin Problem in Hazelnut

Tree nuts (almonds, pecans, pistachio, hazelnuts, etc) are used in food industry. However, environmental conditions and unsuitable processing procedures cause crack or damage to the nut shell. This damage decreases the nut quality and also increases the likelihood of mold infestation. For pistachio nuts, Pearson showed [39] that nearly all the aflatoxin contaminated pistachios are either damaged by birds or insects before harvesting or are early split ones. The damage at the shell of the kernels allows mold to diffuse into the kernels and cause aflatoxin. Pearson [40] used a machine vision system to classify pistachios nuts as stained (caused by early splitting), unstained or moderated stained with an average of 11% classification error rates. By removal of stained pistachio nuts aflatoxin contamination level of pistachio nut is reduced from 4.8- 8.6 interval to 0.04-2.5 ppb [41]. Unlike the pistachio nuts, hazelnut shell are damaged mostly at processing step before the drying phase which the contamination is mostly occurred in. We categorized the hazelnut into four groups to investigate the physical properties of shell for aflatoxin contamination [42]. A sample hazelnut image from these groups are shown in Figure 3.1

100 hazelnut kernels from each group are pruned to aflatoxin producing molds and stored in high humidity environment with $28^{\circ}C$ temperature for 20 days. At the end of the 20 days these kernels are analyzed for aflatoxin contamination. The contamination results (Table 3.1) showed that the intact shell form strong barriers against mold contamination during storage. However, any type of damage at any size on the shell eliminates this capability. This can be seen from the

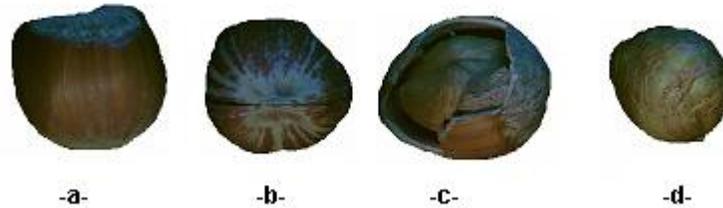


Figure 3.1: A sample image from a) Regular shell hazelnuts (ReH), b) Cracked shell hazelnuts (CrH), c) Broken shell hazelnuts (BrH) and d) In-shell hazelnuts (InH)

contamination level of the CrH, BrH and InH hazelnuts which are close the each other.

Table 3.1: The mean contamination levels of the hazelnuts from four groups

	ReH	CrH	BrH	InH
Cont.Level (ppb)	47	2730	2476	2494

Turkey is the major producer of hazelnut in the world. We produce 70% of the hazelnut in worldwide and we export 80% of this to European Union countries each year. The aflatoxin contamination in hazelnuts varies from year to year, and it is negligible in some years (2000-2003). In reality, 30 percent of the hazelnuts (Figure 3.2) are exported without any processing. The 35 percent is exported in shelled form and the 35 percent is exported after roasting and processing.

The hazelnuts in these three forms have different physical characteristics. Therefore, the hazelnut kernels in these categories should be investigated separately before storage or consuming. The risk for in-shell kernels can be eliminated by impact acoustic sorting systems. However, acoustic sorting can not be used for sorting the shelled kernels. A hyperspectral imaging system is used to address the risk of the shelled nuts (shelled, shelled&roasted) . The Shelled (SHD) and Shelled/Roasted (SRT) hazelnut kernels have different visual characteristics. Therefore these hazelnuts should be studied separately.

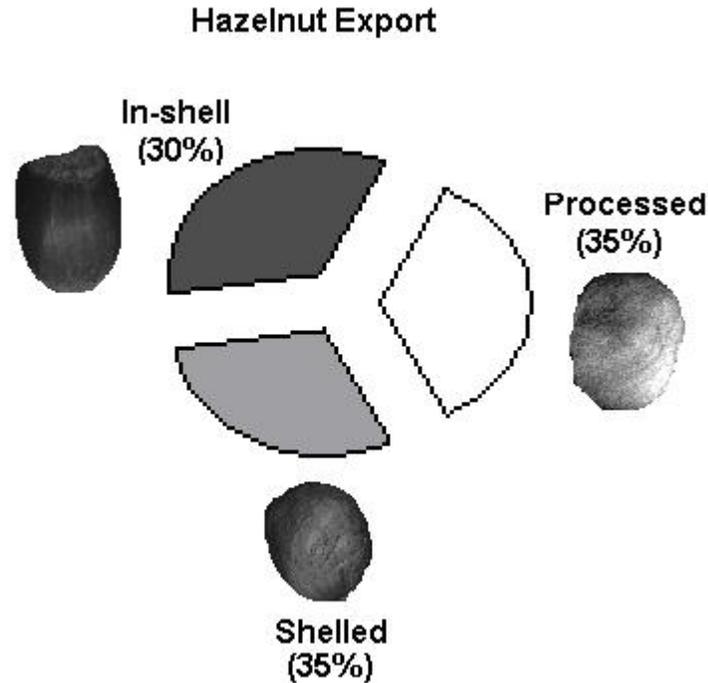


Figure 3.2: The forms of exported hazelnut in TURKEY

3.4 Impact Acoustic Sorter System

Pearson developed an impact acoustic sorter system to separate cracked and uncracked pistachio nuts. The system is currently working on pistachio processing assembly in California and also in Turkey. A similar system (Figure 3.3, modified from [43]) consisting of a pipe through which the nuts slide in, an impact plate that the nuts are dropped on, a microphone for impact sound acquisition and a PC for recording and processing the signals is used to get impact acoustic signal of hazelnut kernels.

A stainless steel plate with dimensions 7.5 x 15 x 2 cm is used as the impact plate. The impact plate is fixed to the ground with $120^{\circ}C$. This angle prevents the nuts from making multiple impacts. A microphone, sensitive to frequencies up to 20 kHz, is placed 5 cm from the impact plate. The impact acoustic signal is sampled at 44.1 kHz, processed and used for decision making.

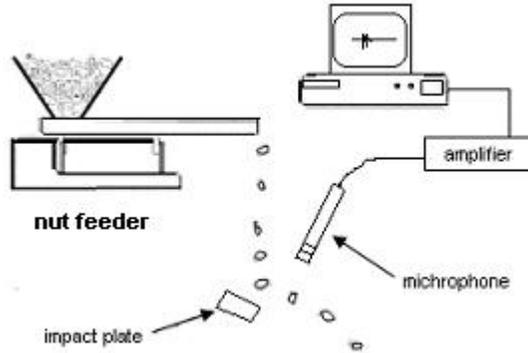


Figure 3.3: Schematic of experimental apparatus for collecting acoustic emissions from hazelnut kernels

This system is also used to separate empty hazelnuts from fully developed nuts in [44] by using 70 combined time and frequency features. Features that are widely used in speech processing algorithms are extracted from short time variances of signals, maximum signal amplitude, spectral peak locations and Weibull distribution parameters that fit to the envelope of the impact signal and used for classification. We called this feature set as non adaptive time-frequency (NATF) feature set.

Kalkan et.al [45] used impact acoustic signals for classifying cracked shell and regular shell hazelnuts by non-adaptive subband (NASB) features with 90% accuracy. In this work, impact acoustic signals were first decomposed into subbands by using undecimated wavelet transform. The subband signals were then divided into overlapping or non-overlapping segments of constant width. An energy feature was extracted for each segment in each subband. The relevance of features and the most relevant sub-bands for classification were investigated by wrapper or filter model. However, the non-adaptive feature extraction method may miss some of the relevant features for classification.

3.5 Multi/Hyperspectral Imaging System for Food Safety

A typical multi/hyperspectral imaging system consists of a CCD (charged coupled device) camera, a frame grabber device, a set of filters and an illumination system. The CCD cameras are solid state, silicon based devices. Some of the cameras have digital output. Frame grabber devices are used for the analog cameras in order to digitize the analog signal. Multiple images at different frequency bands are acquired by using a liquid crystal tunable filter, an acousto-optic tunable filter or by sequentially positioning band-pass filters in front of the lens of the camera.

In multi/hyperspectral imaging system, food items are usually visualized under UV (200-400nm), VIS (400-700nm) or NIR (700-2500nm) illumination. The light incident on the object is absorbed, reflected or transmitted through. These are accepted as the optical properties of the inspected material. The molecules of the objects are excited to higher energy states by absorbing the penetrating light. The excited molecules emit lights at higher wavelengths (fluorescence) when returning to their previous state. Some of the compounds in materials emit fluorescence in the visible region when exposed to UV illumination. The chemical properties of the investigated material are estimated by the properties of the reflected light (wavelength, intensity, etc...).

Tylor and McCure [46] used multispectral imaging system to detect leaf tissues. They stated the importance of three wavelengths 670, 800 and 990 for leaf tissue detection. Zeringue and Shih [47] used the reflectance at 435 nm for detecting the aflatoxin contaminated cotton lint. Park and Chen [48] stated the discriminative importance of the spectral image at 540 and 700 nm for separating the unwholesome carcasses from wholesome carcasses. The images at 566, 515 and 631 nm are used for fecal and ingesta detection at poultry carcasses [49]. Spectral image taken at 686 and 675nm are used for determining the main contamination and defect in apples [50].

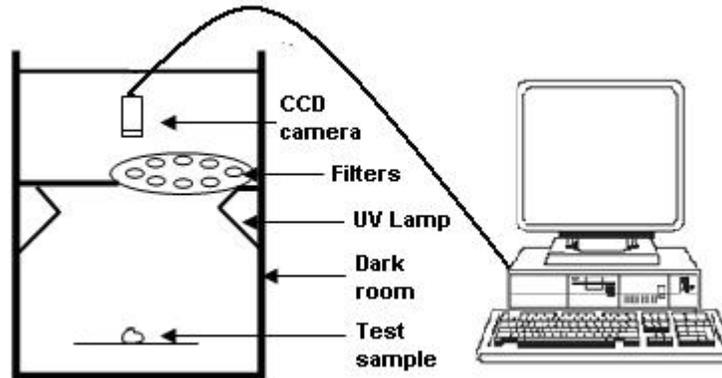


Figure 3.4: Hyperspectral imaging system with a rotating filter wheel

The hyperspectral imaging system (Figure 3.4) used in this study consists of

- An Imaging Source VIS DFK 41AF02 digital CCD camera
- 2 High Intensity Spectroline UV-A Lamps (peak intensity at 365 nm)
- A cabin to function as dark room
- A band pass filter set including the filters from 400 nm to 510 nm with 10 nm FWHM and 550nm and 600nm filters with 70 nm and 40 nm FWHM
- A computer for data acquisition and processing

CHAPTER 4

FEATURE EXTRACTION ALGORITHM BASED ON 2D LDB SEARCH

In signal classification applications, depending on problem and the dataset, time (spatial in 2D) domain and and/or frequency domain features are extracted and used for learning. The original LDB algorithm decomposes the time axis by Local Cosine Packets or frequency axis by wavelet packets to locate the discriminative features of the data. However, it is observed that, especially for non-stationary signals (speech, acoustics, EEG, vibration, etc.), both time and frequency domain features are important for classification [43, 10, 51].

The original LDB algorithm is modified to 2D structure and adapted to impact acoustic and hyperspectral data to get the exact location of the discriminative features. For impact acoustic data, the time-frequency plane features are extracted by combining the Local Cosine Packets and Wavelet Packet analysis to obtain time and frequency adaptation in an off-line step. The introduced technique requires no prior information on the relevant time and frequency locations. Pruning in both axis extracts the most discriminative features by combining the ones which do not provide extra information on their own. The extracted features are then selected by feature selection algorithm are used in classification. Similar approach is used for hyperspectral imaging where the spectral axis is used instead of time axis. LDB is implemented to get the most relevant spectral and spatial-frequency features of hyperspectral data. The 2D LDB algorithm for one dimensional impact acoustics signals and three dimensional hyperspectral images are explained separately.

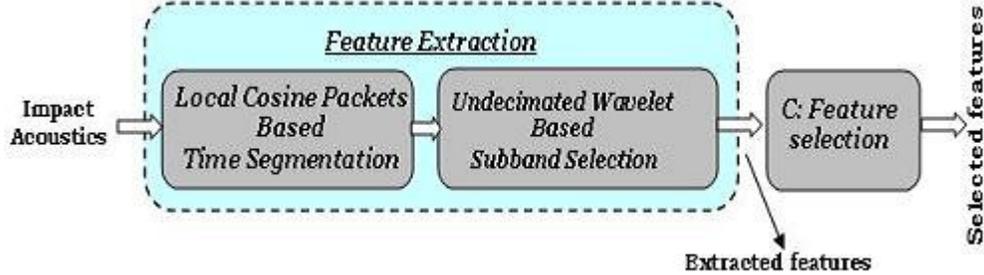


Figure 4.1: The block diagram of the 2D LDB based feature extraction for one dimensional signal

4.1 Feature Extraction from Acoustic Data by 2D Local Discriminant Bases Search

In 2D LDB algorithm, the time-frequency plane features are extracted by combining the Local Cosine Packets and Wavelet Packet analysis to obtain time and frequency adaptation in off-line step (Figure 4.1). The introduced technique requires no a priori information on the relevant time and frequency locations. It finds these locations automatically by pruning the time segments considering their discrimination potential. Then, the resulted time segment signals are decomposed into frequency subbands by undecimated wavelet transform. The extracted features are, then, selected by any feature selection algorithm and used for classification.

4.1.1 Adaptive segmentation and pruning in time axis

The non-stationary signals may have different characteristics in time. Therefore, these signals should be analyzed locally. In general, local information of the signal is extracted by Short Time Fourier Transform (STFT). This type of block transform generates side-lobe artifacts due to disjoint rectangular windows. On the other hand, usage of smooth windows removes the orthogonality. It is possible to construct orthogonal transforms with smooth and overlapping windows by trigonometric bases. Some researchers used Local Cosine Packets (LCP) because

of its advantages over STFT [52, 8]. Therefore, Local Cosine Packets (LCP) is used to partition the time axis in a dyadic binary tree structure (Figure 4.2).

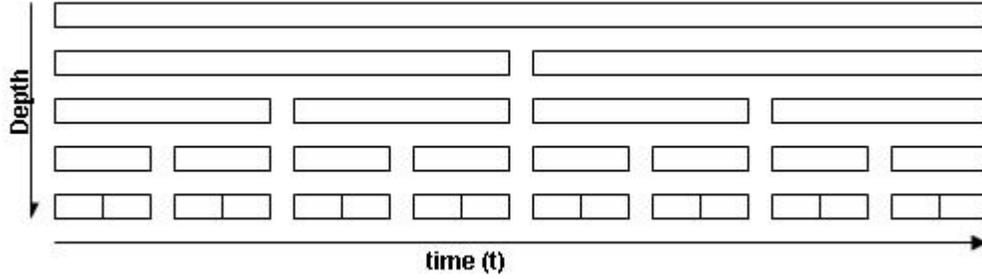


Figure 4.2: Time segmentation with LCP at 4 levels in dyadic tree structure

The LCP partition the time axis by using smooth bells [53] which are constructed using cut-off functions $r(t)$ that satisfy

$$|r(t^2)| + |r(-t)^2| = 1 \quad \text{for all } t \in R \quad (4.1)$$

$$r(t) = \begin{cases} 0 & \text{if } t \leq -1 \\ 1 & \text{if } t \geq 1 \end{cases} \quad (4.2)$$

An example of such a function $r(t)$ is

$$r(t) = \begin{cases} 0 & \text{if } t \leq -1 \\ \sin \left[\frac{\pi}{4} \left(1 + \sin\left(\frac{\pi t}{4}\right) \right) \right] & \text{if } -1 < t < 1 \\ 1 & \text{if } t > 1 \end{cases} \quad (4.3)$$

Each signal is represented with Local Cosine Packets within smooth windows (as in Eq. 4.3) in the tree structure (Figure 4.3). The resulting expansion coefficients are squared and then averaged over the signals in the given class. This provides an averaged energy spectrum of each class in a given time segment within the pyramidal tree.

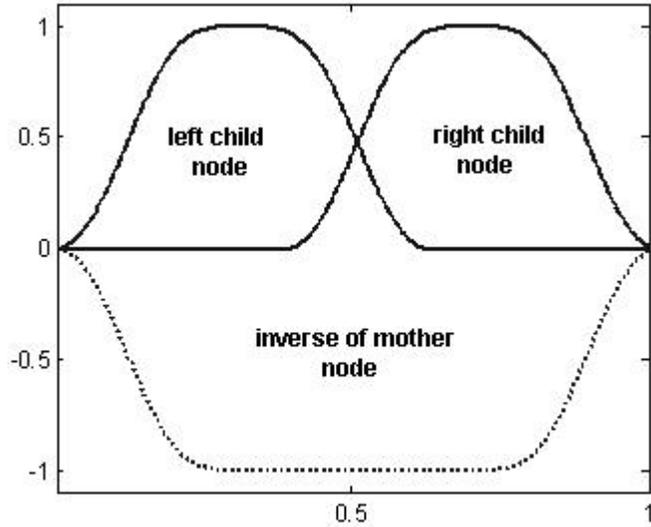


Figure 4.3: The smooth windows obtained by LCP

Let p_i and q_i be the mean energy spectra of cracked and regular classes, in a given time segment respectively. The distance between the average spectra is calculated with the Distance criterion J (Section 2.4) where 'n' in Equation 2.2 corresponds to the total number of time samples in a given node. This way, the distance is accumulated along the spectrum within all subspaces to get a single value representing each node of the tree. The resulting binary tree is, then, pruned from bottom to top according the following rule to find the nodes with maximum discrimination power:

Pruning algorithm #1
 if $J_{mother} \geq (J_{child1} + J_{child2})$
 keep mother
 else
 keep children

where J_{mother} and J_{child} are the discrimination power of the mother and children node and are computed by the Kullback-Leibler distance criteria. The algorithm keeps the mother node if it captures the discriminative power of the children nodes; otherwise it keeps the children nodes. The algorithm may construct different tilings for different dataset. This may be regarded as a robustness measure of the algorithm.

4.1.2 Adaptive segmentation and pruning in frequency axis

The signal portion in time segments is divided into frequency bands to get the specific frequency content of the signals. This segmentation in frequency axis helps to find and analyze the local patterns in signals. Fourier Transform and Discrete Wavelet Transform are two of the methods to decompose the frequency axis. The methods for decomposition should be selected according to nature of the data set. In that aspect, shift invariant decomposition system is highly required when the obtained signals are not time aligned. The Undecimated Wavelet Transform has shift invariant property and for classification purposes it is firstly used for texture classification [26]. In this study, similar approach with a filter bank is used to analyze the impact signals for classification. A filter f that satisfies the quadrature mirror filter condition

$$F(z)F(z^{-1}) + F(-z)F(-z)^{-1} = 1 \quad (4.4)$$

used to construct the pyramidal filter tree (Fig. 4.4), where $F(z)$ is the z -transform of f . The high-pass filter g is obtained by shift and modulation of f .

$$G(z) = zF(-z)^{-1} \quad (4.5)$$

The subsequent filters in the pyramidal tree are, then, generated by increasing the width of f and g at every step.

$$\begin{aligned} F_{i+1}(z) &= F(z^{2^i}) \\ G_{i+1}(z) &= G(z^{2^i}) \quad , i = 0, 1, 2, \dots, N \end{aligned} \quad (4.6)$$

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

$$\begin{aligned} f_{i+1}(k) &= [f]_{\uparrow 2^i} \\ g_{i+1}(k) &= [g]_{\uparrow 2^i} \end{aligned} \quad (4.7)$$

where the notation \uparrow_m denotes the up-sampling operation by a factor of m .

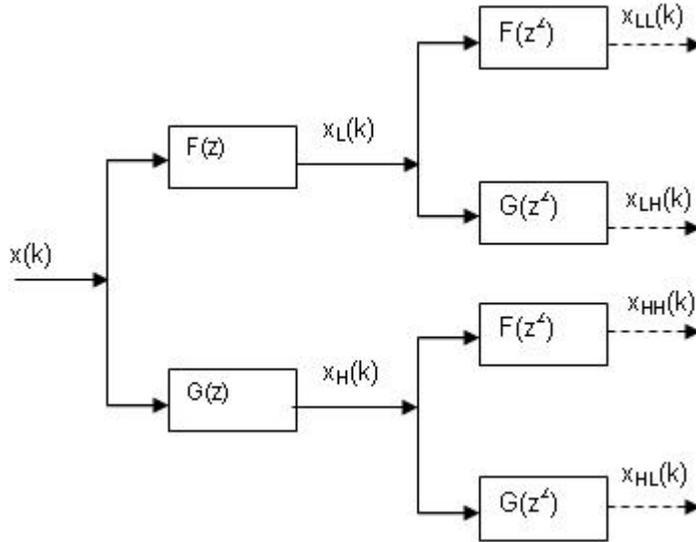


Figure 4.4: Pyramidal filter tree up to second level. L and H stands for Low and High band, respectively

The pyramidal filter tree is transformed into a filter bank (Figure 4.5) for which the filters are obtained by convolving the filters on the branch of the pyramidal tree.

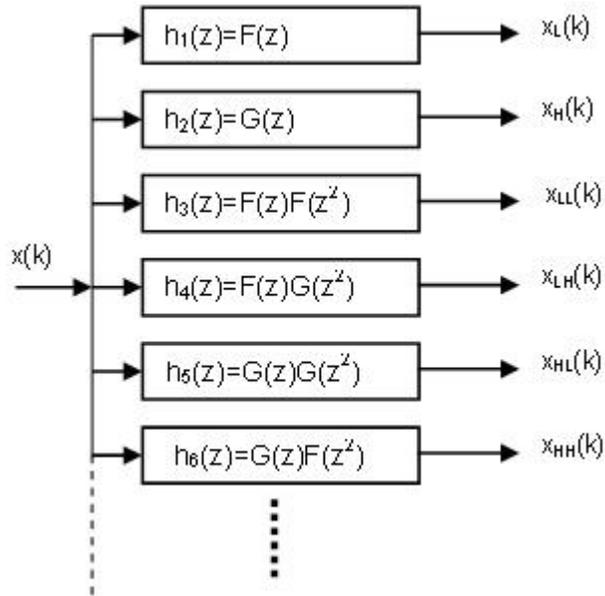


Figure 4.5: Filter bank

A third level filter band with Coiflet 5 tap wavelet and scale functions is shown in Figure 4.6.

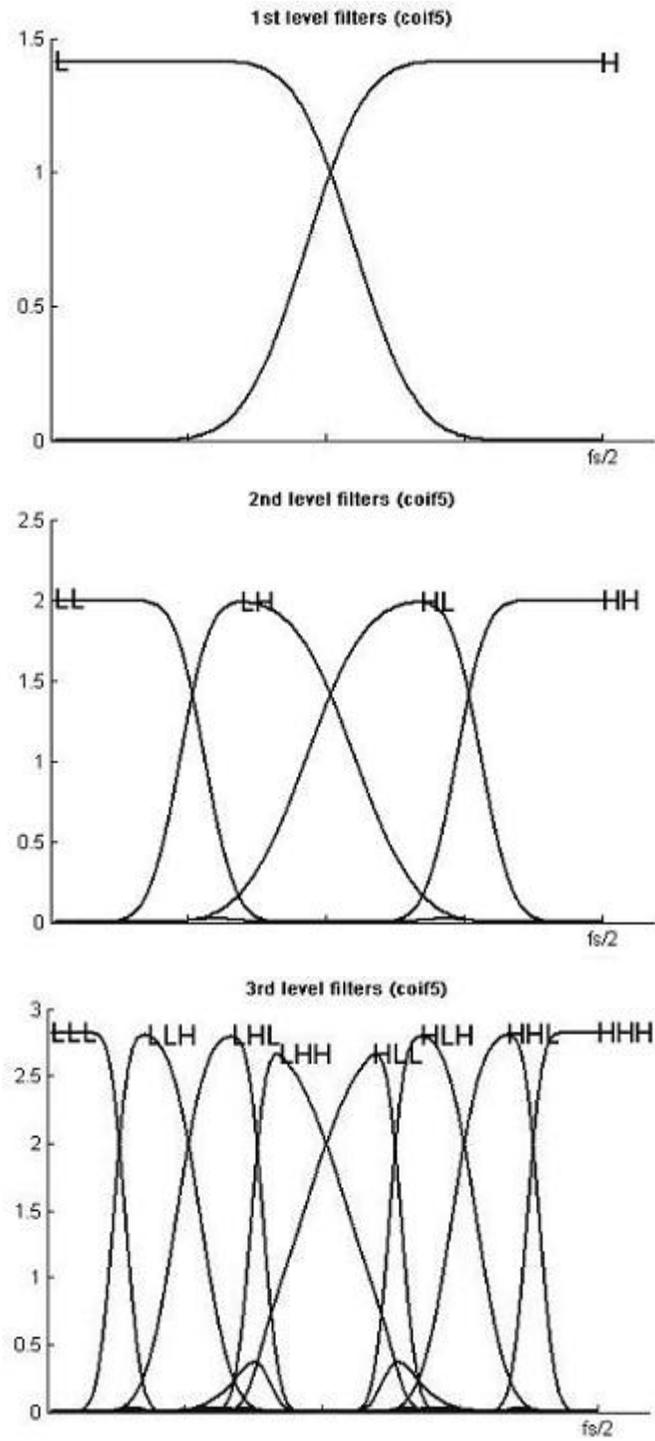


Figure 4.6: Undecimated filter banks up to third level decomposition

The obtained filter-bank is used to decompose the signals in time segments into sub-bands in frequency axis (Figure 4.7).

0kHz				fs/2 kHz			
L				H			
LL		LH		HL		HH	
LLL	LLH	LHL	LHH	HLL	HLH	HHL	HHH

Figure 4.7: The 3 level full wavelet sub-band tree

The signals usually have different energy distribution in each sub-band. The Euclidean distance between cumulative probability distribution of subband energies in Equation 2.3 is chosen as the discriminative measure and the constructed pyramidal sub-band tree is pruned from bottom to top by the following rule:

Pruning algorithm #2
if $d_{mother} < \max\{d_{child1}, d_{child2}\}$
set $\max\{d_{child1}, d_{child2}\}$ as d_{mother}
else
remove children

The characteristic of the filter may indirectly effect the discriminant band distribution. This may, also, change the classification accuracy of the system.

As a result, the discriminant time-frequency (TF) map is constructed by adaptive segmentation in time and frequency in order to localize the most discriminative patterns in signal.

4.1.3 Selection of extracted time-frequency features

The adaptive segmentation and pruning operation in both time and frequency axis by distance cost function revealed the location of the most discriminant energy features of the signals. This feature extraction step provides the best segmentation in two dimensions (time and frequency) but does not eliminate the irrelevant feature location. Therefore, the extracted features are then sorted by

feature selection algorithms and incrementally included to feature vector. These algorithms are Fisher Distance Based Feature Selection (FFS) and Correlation Based Feature Selection (CFS) algorithm [16]. The desired number of features is obtained when the minimum classification error is reached.

4.2 Feature Extraction from Hyperspectral Data by 2D Local Discriminant Bases Search

The LDB algorithm is also modified to 2D structure and applied to extract most discriminative features in hyperspectral data. A four step algorithm similar to the one in Section 4.4 is developed for hyperspectral data (Figure 4.8).

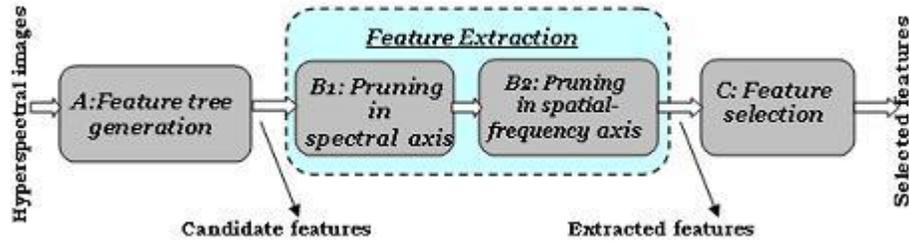


Figure 4.8: The block diagram of the 2D LDB based feature extraction for hyperspectral data

4.2.1 Feature Tree Generation

The first step in the algorithm is to obtain candidate feature set by generating two feature trees on spectral and spatial-frequency axis in order. In the first tree, the reflectance energies of spectral images are placed to the k th level of the tree from left to right. Figure 4.8 shows an illustration to $k=4$ levels binary spectral band tree with 16 spectral bands (SB). For the case of , the remaining nodes at the k th level can be set to null in order to complete the binary tree. The energy value of the mother nodes at the high levels were computed by summing the feature values of their branch nodes.

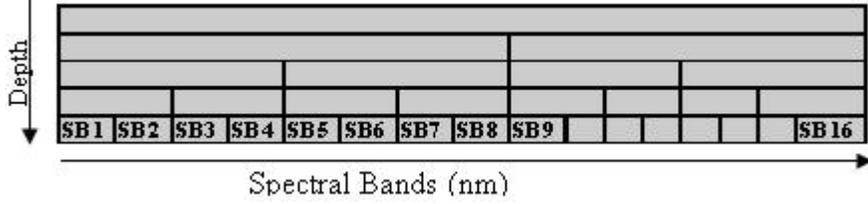


Figure 4.9: k=4 levels binary spectral band tree

The second feature tree is generated in spatial- frequency axis in quad tree structure by decomposing the raw spectral images into h level full wavelet subbands as in Figure 4.10. Wavelet transform retains the original image information and completely represents the image in subbands of (LL, LH, HL and HH) where first character shows the filtering (Low or High) through the x and second shows the filtering through the y direction of the image.

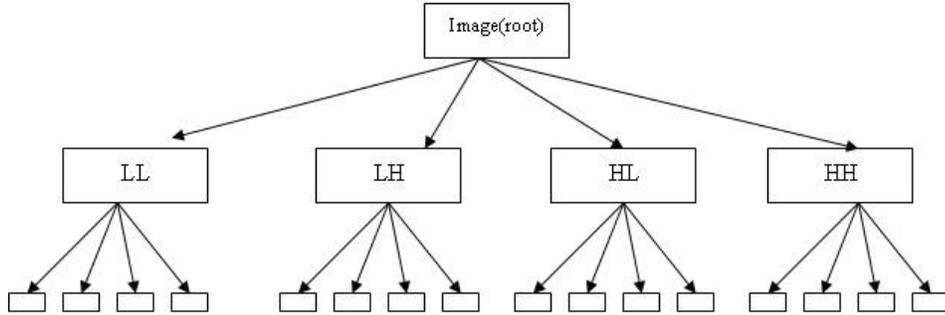


Figure 4.10: Full wavelet decomposition quad tree up to h=2 levels

The energy in each subband is computed and used as features for further analysis.

4.2.2 Adaptive Pruning in Spectral Axis

The feature extraction step (Figure 4.8) starts with pruning in spectral axis. The obtained binary spectral band feature tree (Figure 4.9) is pruned from bottom to top by the following rule:

Binary tree pruning algorithm
if $d_{mother} < \max\{d_{child1}, d_{child2}\}$
set $\max\{d_{child1}, d_{child2}\}$ as d_{mother}
else
remove children

where d_i is the distance of the i_{th} node feature between classes. We used Euclidean distance between cumulative probability distribution of the nodes. However, other distance metrics can be used as well.

The algorithm prunes the spectral bands (branch node) if their discrimination potential is lower than their mother node. This process fuses some of the spectral bands for better classification accuracy. The frequency subband energies of the pruned spectral bands are averaged in parallel before pruning in the spatial-frequency axis.

4.2.3 Adaptive Pruning in Spatial-Frequency Axis

The pruning algorithm in Section 4.5.2 is modified to four children mode because of the quad tree structure (Figure 4.10) in spatial-frequency axis and applied to prune the quad tree in a bottom-up manner.

Quad tree pruning algorithm
if $d_{mother} < \max\{d_{child1}, d_{child2}, d_{child3}, d_{child4}\}$
set $\max\{d_{child1}, d_{child2}, d_{child3}, d_{child4}\}$ as d_{mother}
else
remove children

where the $child1$, $child2$, $child3$ and $child4$ are the LL, LH, HL and HH wavelet subbands of the mother node, respectively. The algorithm keeps the mother node if its discrimination potential is higher than any of its four children nodes. Otherwise the children nodes survive as the nodes with high discrimination potential.

4.2.4 Selection of extracted spectral spatial-frequency features

The adaptive segmentation and pruning in both spectral and spatial-frequency axis based on the distance between the feature elements revealed the location of the discriminative features of the hyperspectral data. This feature extraction processes provides the best segmentation in both spectral and spatial-frequency axis but does not eliminate the irrelevant ones. Therefore, these features are also selected by feature selection algorithms (FFS, CFS and Wrapper) and fed into the classifier one by one to figure out the feature subset giving minimum classification error. In addition to these feature selection algorithm, a new feature set is generated by PCA algorithm and the transformed features are used in the classifier.

CHAPTER 5

DATASETS USED IN EXPERIMENT

5.1 Acquisition of Impact Acoustics Data

'Levant' type in-shell hazelnuts, which were collected from an orchard in Akcakoca, Turkey in August 2003 are used for impact acoustic sorter. The weights of the hazelnuts are measured and the ones less than 0.9 gram are taken as empty or undeveloped (EmH) nuts. Hazelnuts with weights over 0.9 grams are accepted as fully developed (FuH). The shells of the fully developed nuts are visually inspected are further classified as nuts with regular shell (ReH) and nuts with cracked shell (CrH). The weight histogram of the randomly selected 180 nuts from each group is depicted in Figure 5.1. It is seen that the fully developed nuts with cracked shell have similar weights with the fully developed nuts with regular shell.

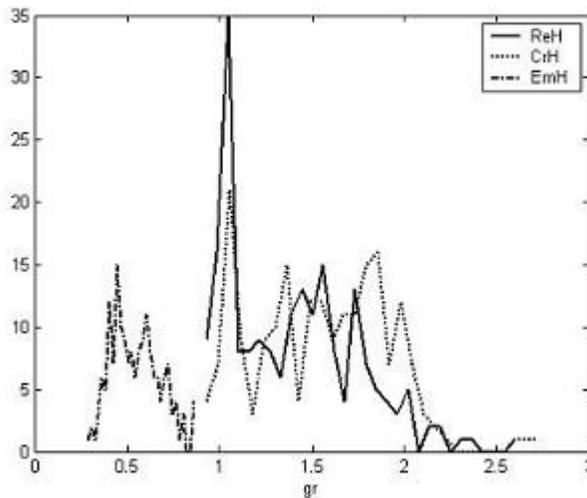


Figure 5.1: Weights of 'EmH', 'CrH' and 'ReH' hazelnuts

The impact acoustic data acquisition system is used to get impact acoustic signals of those classified hazelnuts. Typical signals of an empty or undeveloped hazelnut (EmH), a full hazelnut with regular (ReH) and cracked shell (CrH) are shown in Figure 5.2.

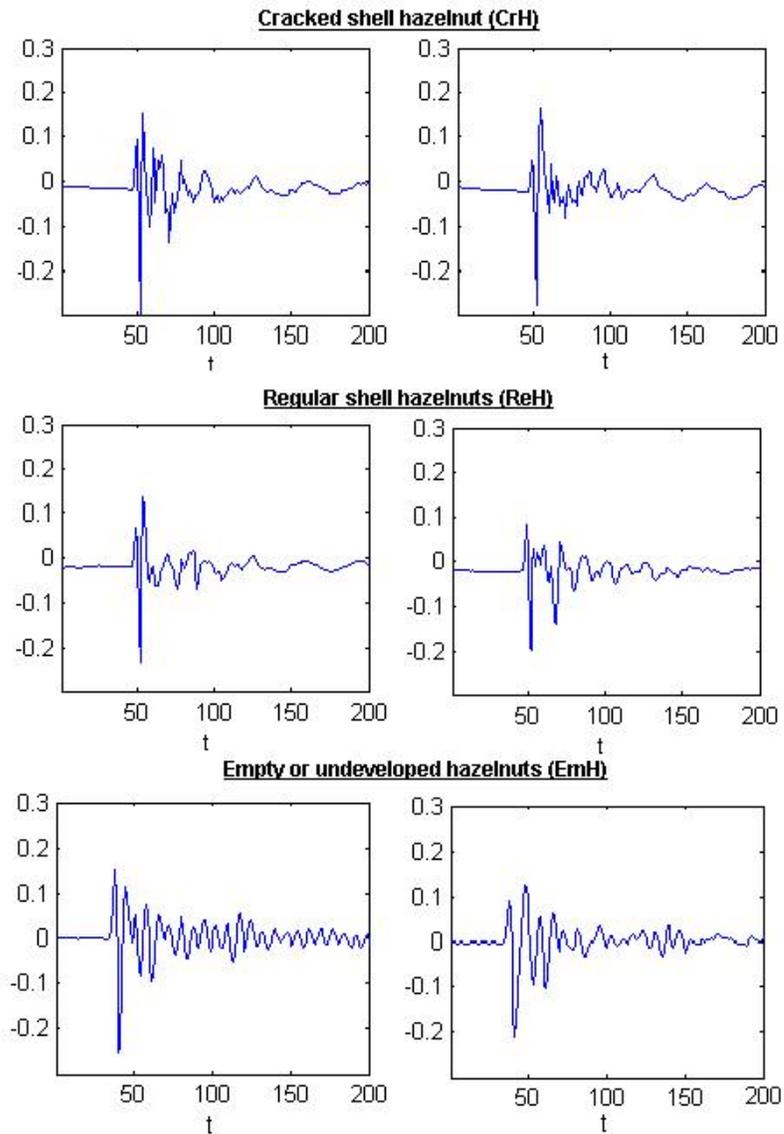


Figure 5.2: Typical impact acoustic signals of ReH, CrH and EmH hazelnuts

The impact instant of the acoustic signal is defined as the time of the first sample from left up to the sample whose amplitude is over the 15% of the absolute maximum amplitude. This impact instant is experimentally defined and the redundant samples before the impact instant are removed adaptively from the signals. The consecutive 768 samples from the impact instant are used as the impact acoustic signal for each hazelnut.

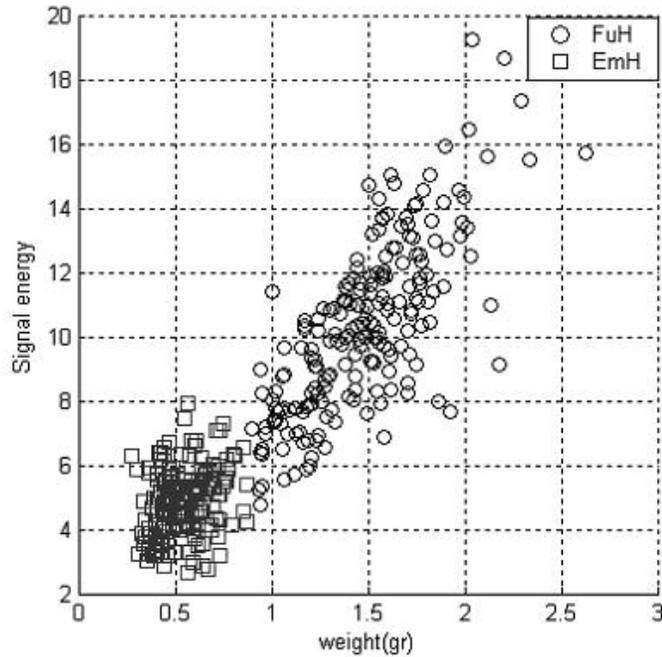


Figure 5.3: Signal energy of hazelnuts at various weights

It is observed that empty hazelnuts generate weaker signals compared to fully developed ones (FuH) and their impact signals tend to (computed by Equation 5.1) have lower energy values. The cracked hazelnut impact sounds have similar energy values to those with regular shell indicating similar weights but the cracked impact signals usually have a longer decay time due to the oscillations caused by the crack. Figure 5.3 shows the significant correlation between the nuts' weights and the total impact acoustic signal $x(j)$ energy levels of the nuts computed by

$$e = \sum_j |x(j)| \quad j = 1, 2, 3, \dots, M \quad (5.1)$$

where M is the total number of samples which is defined as 758.

It is possible to separate the EmH kernels from FuH ones at 96% classification accuracy by just using the raw signal energies as a feature. However, this is not valid for separation of ReH and CrH because they have similar weights and energy values. Therefore more advanced signal processing and feature selection techniques should be used to extract the relevant features for classification of ReH and CrH hazelnuts. The LDB algorithm is aimed to obtain such relevant features from the time-frequency domain of the signals (Figure 5.4). The impact acoustic signal of 1000 ReH and 1000 CrH hazelnuts were used for algorithm development.

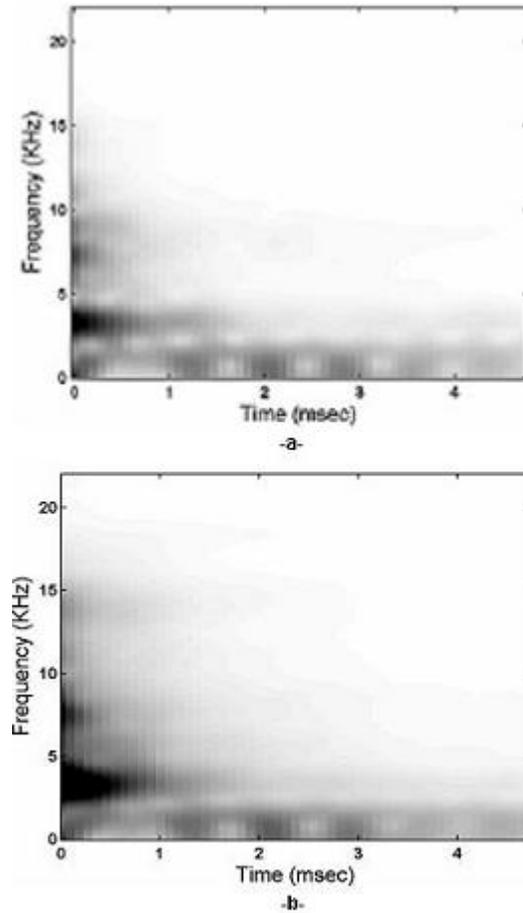


Figure 5.4: The averaged spectrogram of a) CrH and b) ReH hazelnuts

5.2 Acquisition of Hyperspectral Data

Shelled hazelnuts for hyperspectral imaging were collected from Ordu, Turkey in 2007. A few aflatoxin contaminated hazelnut kernels is encountered in the collected samples. Therefore an artificial contamination procedure is conducted to obtain aflatoxin contaminated hazelnut kernels. In stated in Section 3.2 that the in-shelled hazelnuts processed in two different category as Shelled (SHD) and Shelled/Roasted (SRT) nuts. The two types of hazelnuts have different visual characteristics because the roasting operation removes the inner skin of the InH hazelnuts. Therefore these hazelnuts were studied separately. The SHD and SRT hazelnuts were divided into three groups as shown in Figure 5.5. The first group was reserved for uncontaminated class (*UnCont*) where the second and third groups were artificially contaminated. The hazelnut kernels in the second group were incubated to aflatoxin producing molds (*FlavusCont*) whereas the kernels in the third group incubated to pure water (*WaterCont*).

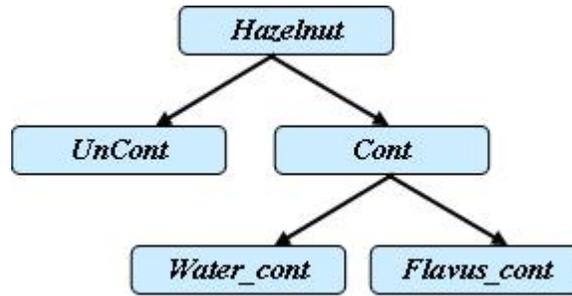


Figure 5.5: Categories of hazelnut used in the experiment

All the incubated kernels (*FlavusCont* and *WaterCont*) are kept in $28^{\circ}C$ with 90% humidity condition for 9 days. The mold infestation is observed at all the incubated hazelnuts. At the end of the 9th day, the hazelnuts in SRT category were roasted at $140^{\circ}C$ for 15 minutes. The roasting process removed the skin over the kernels. The hazelnuts in both SRT and SHD categories were sent to chemical analysis for aflatoxin contamination after hyperspectral image acquisition. The chemical analysis (Table 5.1 and Table 5.2) results showed high level of aflatoxin contamination at *FlavusCont* hazelnuts and respectively low level of aflatoxin at

some of the *WaterCont* hazelnut of both categories. Aflatoxin over 4 ppb is detected in two UnCont hazelnuts from SRT and one UnCont hazelnuts from SHD category. All the kernels in *WaterCont* group are naturally mold infected and these hazelnuts should not be consumed although a few of them contain aflatoxin.

Table 5.1: Number of aflatoxin contaminated kernels and the mean aflatoxin level (ppb) of the three group SRT hazelnuts

	SRT_UnCont	SRT_FlavusCont	SRT_WaterCont
# of SRT_Afla+ kernel	2	79	15
# of SRT_Afla- kernel	102	0	87
Mean Aflatoxin Level	0.7	2227	7,47

Table 5.2: Number of aflatoxin contaminated kernels and the mean aflatoxin level (ppb) of the three group SDT hazelnuts

	SDT_UnCont	SDT_FlavusCont	SDT_WaterCont
# of SDT_Afla+ kernel	1	49	4
# of SDT_Afla- kernel	81	0	100
Mean Aflatoxin Level	0.36	3418	26,25

The hazelnut samples under UV-A light are screened by capturing the reflected light from the sample by using the hyperspectral imaging system. The spectral images are taken by IC Capture (Imaging Source Inc) image acquisition toll. The exposure time of the camera is set to 0.33 sec for SRT and 2 sec for SHD hazelnuts in order to capture sufficient reflectance light. The SHD category hazelnuts do not reflect sufficient light because of the inner skin and moisture. Therefore, we could not get high quality images although we increase the exposure time from 0.33 sec to 2 sec with the same hyperspectral imaging system. Figure 5.6 shows some of the band features of hazelnut kernels from SRT and SHD category, respectively. The first three columns show the *FlavusCont* hazelnut images, the second third columns show the *WaterCont* hazelnuts, whereas the last 3 columns show the (*UnCont*) hazelnut images. The kernels were also visualized without

any filter, at a filter 550 nm (70 nm FWHM) and 600 nm (40 nm FWHM). These images are used at image preprocessing but not used for algorithm development.

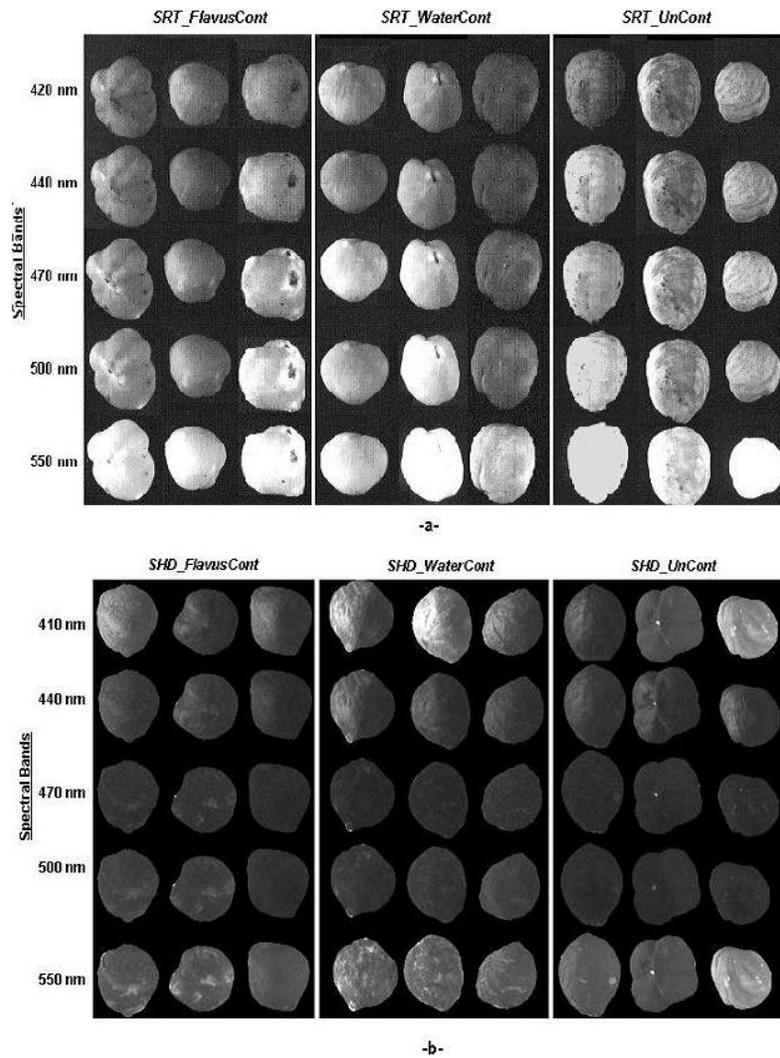


Figure 5.6: A few spectral band images for *FlavusCont*, *WaterCont* and (*UnCont*) group of a) SRT and b) SHD hazelnuts

5.3 Preprocessing the Hyperspectral Images

A median filter was applied to remove the noise at the spectral images (Fig 5.7-a). Secondly, a binary mask was generated to extract the hazelnut from background and the pixels of the regions which the inner skin was not removed during roasting. The images taken at 550 nm is appropriate for mask generation. This band clearly displays the distinction of the nut from background and unskinned region. The mask was further improved by erosion and dilation operation. These morphological operation removed the undesired defects due to thresholding. The generated mask was applied for all spectral images of the hazelnut (Fig 5.7-b).

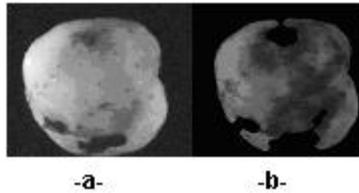


Figure 5.7: A raw spectral band image with unskinned region at the surface, b) masked image of 430 nm spectral band

Instead of a whole hazelnut image, the masked spectral images were divided into square regions (91x91 pixels) and each region was regarded as an independent sample and they were later on used for voting on the class membership of a given hazelnut kernel.

CHAPTER 6

RESULTS AND DISCUSSIONS ON IMPACT ACOUSTIC DATA CLASSIFICATION

It is possible to separate the empty *EmH* kernels from the fully developed *FuH* ones with 96% classification accuracy by just using the raw signal energies as a feature for the linear classifier. However, a classification by raw signal energies does not separate the regular shell hazelnuts *ReH* and cracked shell hazelnuts *CrH* because they have similar weights and energy values. Therefore more advanced signal processing and feature selection techniques should be used to extract the relevant features for classification of *ReH* and *CrH*. The developed LDB based algorithm is aimed to obtain such relevant features from the time-frequency domain of the signals and the classification results of the extracted features are compared.

A total of one thousand *ReH* and one thousand *CrH* hazelnut kernels are used in this study. Each hazelnut is dropped on the metal plate and the resulting acoustic signal consisting of 768 time samples is recorded. The one thousand acoustic signals for each class are randomly divided into five non-overlapping sets, each consisting of 200 records. Five pairs of *ReH* and *CrH* sets are then randomly formed. Each pair is used to construct the adaptive time-frequency (T-F) segmentation and select features. The features identified are then used with the remaining 1600 acoustic signals to determine the performance of the classifier. This procedure is repeated five times with the five different pairs of *ReH* and *CrH* sets.

The acoustic signals were analyzed up to a tree depth of four resulting in a smallest segment size of 48 time samples in the time domain. It is empirically found that this level provides a healthy balance between transient waveforms and the required spectral resolution to distinguish between subbands with different behavior. The signals were first represented by using LCP (Figure 4.1) over the pyramidal tree structure (Figure 4.2). The pyramidal tree was pruned by using the pruning algorithm of Section 4.1.1 and the adaptive time segmentation for classification purpose was obtained for different sets of signals as indicated in Figure 6.1. It was observed that different sets of signals may cause different segmentation in time. The segmentation of Figure 6.1-a were used in further analysis. In this case, the time axis is divided into seven segments.

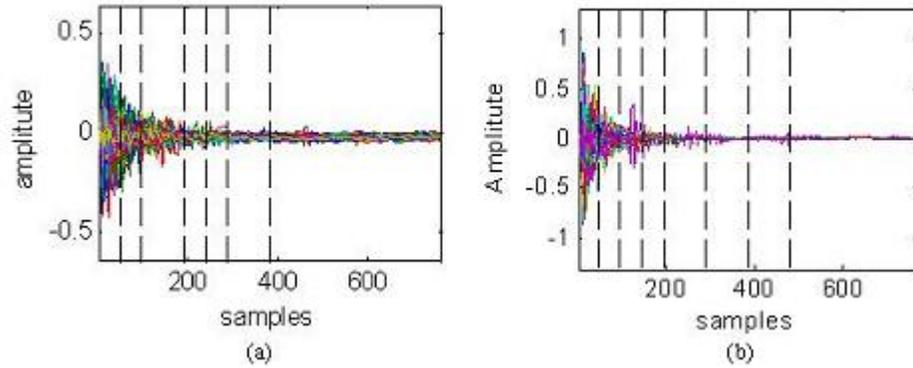


Figure 6.1: The adaptive time segmentation grids (dotted lines) of a-) of set1 and b-) of set2

In each time segment the signal was decomposed into sub-bands up to the fourth Wavelet decomposition level and the most relevant sub-bands were detected by using the procedures of Section 4.1.2.

A discriminative time-frequency map was generated in Figure 6.2 by combining the adaptively pruned trees both in time and frequency to visualize the most crucial T-F patterns.

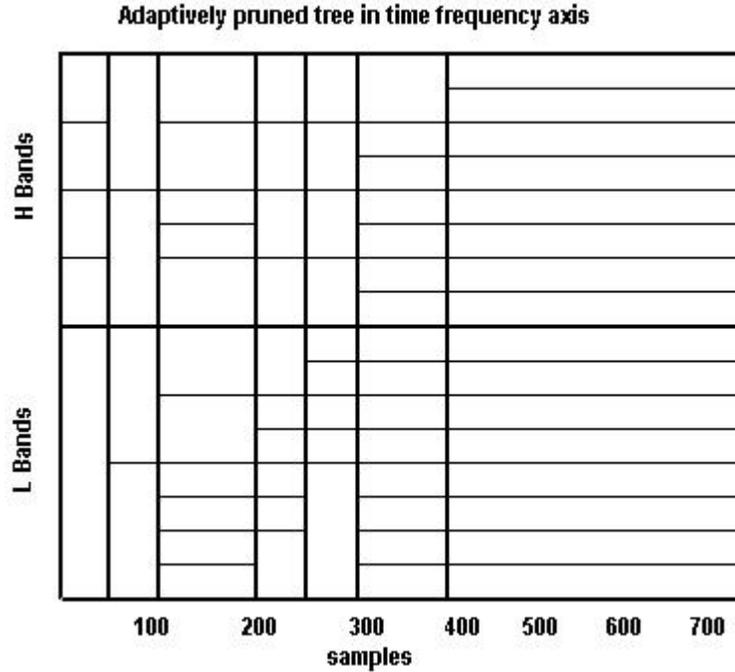


Figure 6.2: The location of most discriminative features in time-frequency axis

In our application, the algorithm usually generates a T-F map with around 70 sub-bands for various training data sets. For every signal in each training set, the energy value for each sub-band was computed resulting in two sets of feature vectors corresponding to cracked and healthy shell classes. The 70 features obtained were sorted by Fisher distance based (FFS) and Correlation Based Feature Selection (CFS) algorithms and then used for classification. We observed with all training data sets that the most discriminative feature locations (defined by FFS) were concentrated in the high frequency bands corresponding to the early and post impact regions as indicated in Figure 6.3. Among the 70 sub-bands, the 25 most discriminative ones are indicated by different shades of gray, with darker shades corresponding to higher discrimination levels.

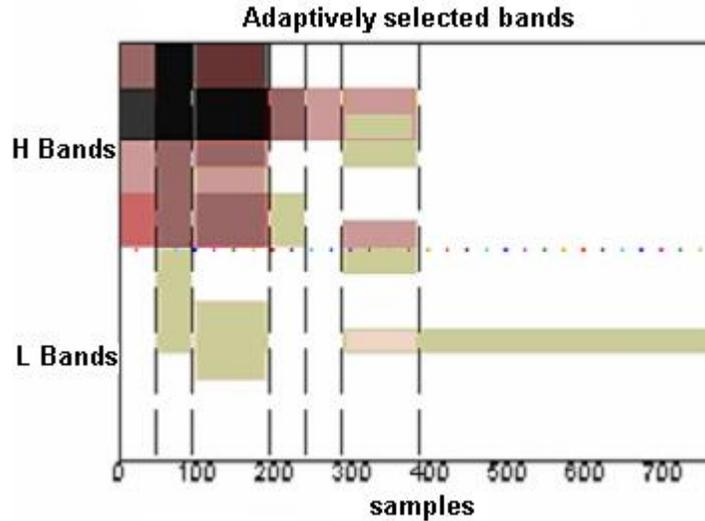


Figure 6.3: The time-frequency discrimination map of impact acoustic data. Darker regions indicate higher discrimination power

6.1 Classification

In order to assess the efficiency of the proposed algorithm, a comparison is made with the Non-Adaptive Time and Frequency features (NATF) [44], Non-Adaptive Sub-Bands features (NASB) and different order statistical features. Recall that in NATF method, 70 features were extracted from the short time variances of signal; maximum signal amplitude, spectral peak locations and Weibull distribution fit to the envelope of the impact signal and all are used for classification. In the NASB method, features were extracted from subband signals and the 20 most relevant features and the sub-bands including these features were manually selected. The time segmentation of Figure 6.1-a is employed to obtain a total of 28 statistical features including mean absolute energy, variance, skewness and kurtosis on each of the seven time segments.

The minimum number of features for classification was investigated by adding features one by one which were ranked by FFS and CFS feature selection algorithms. The feature selection step is repeated for all four different type of extracted feature sets. Related classification error curves are presented in Figure 6.4.

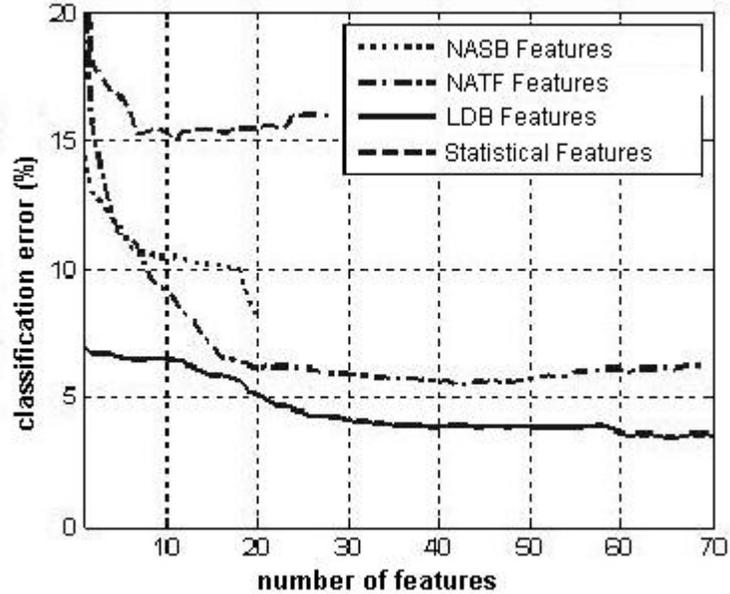


Figure 6.4: The classification error rates (%) with various numbers of features

The lowest classification error is achieved with the proposed LBD based approach. The minimal classification error rates achieved by each method are given in Table 6.1. It is observed that the lowest error is achieved by the first 64 features with an error level of 3.5% by our proposed approach. For the method of NATF, 43 out of 70 time and frequency domain features provided the minimum error level. Similarly, 20 features are used for the method of NASB. The statistical features gave poor classification error rates compared to other methods. The lowest error rate occurred when the first 7 features are used. Our proposed approach reaches an error rate around 4% after the first 30 features. Increasing the number of features provided marginal improvement of the error rate.

Table 6.1: Classification rate comparison of proposed LDB based method against the other methods

Method	Accuracy(%)
43 NATF features	94.47
7 statistical features	85.00
20 NASB features	91.80
64 LDB features	96.51

The ROC curves for the three methods are presented in Figure 6.5. It is observed that 64 and 30 dimensional LDB features provide higher detection of cracked hazelnuts for a given false alarm rate.

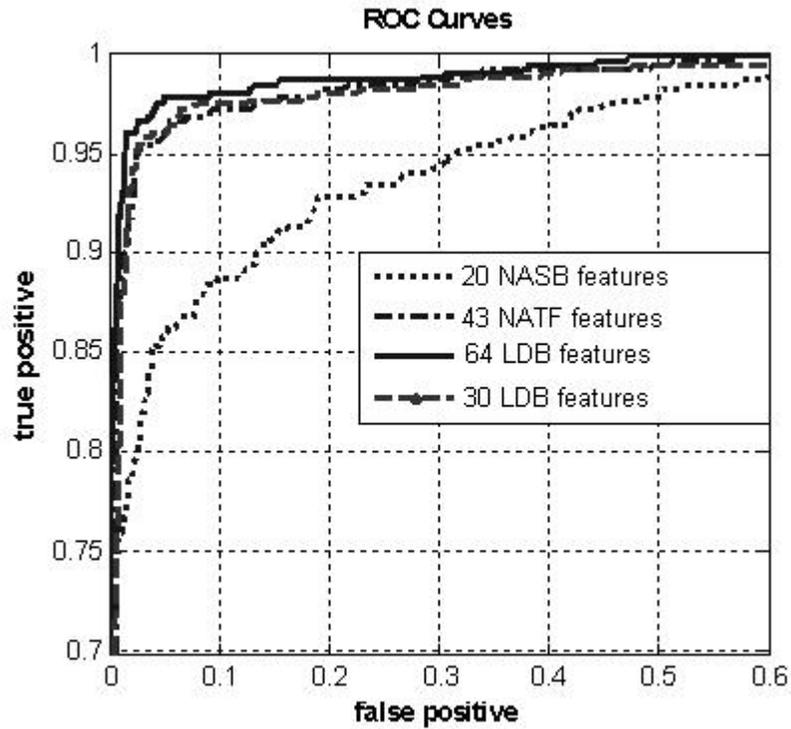


Figure 6.5: Receiver Operating Characteristics (ROC) curves

6.2 Filter Selection

Various types of wavelet filters (Daubechies, Coiflet, and Sym) are used for decomposition of the frequency axis and their effects on classification accuracy are observed. In Figures 6.6 (a) and (b) the classification error curves with various number of features are depicted in contour graphics format for the usage of various Daubechies and Coiflet Wavelets for frequency axis decomposition.

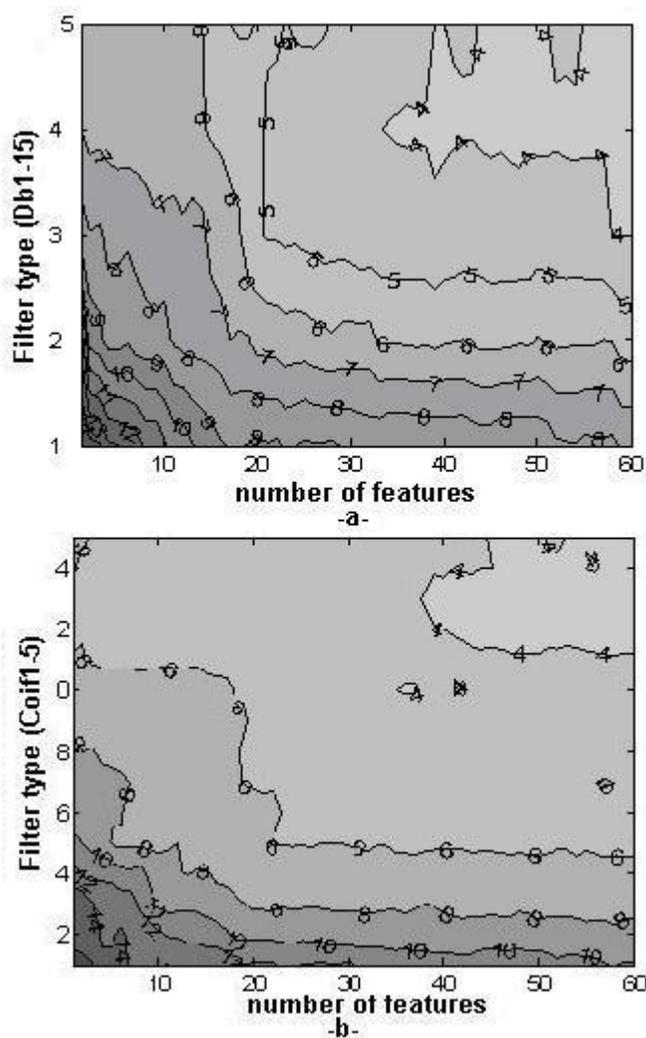


Figure 6.6: The effect of selected wavelets and feature dimension on classification accuracy. a) Daubechies, b) Coiflet

The x axis indicates the total number of features retained after sorting. The

y axis indicates the filter type used in subband decomposition. The numbers on the graph show the obtained classification accuracies (%). The higher filter types correspond to higher order filters. The darker regions in the contour graph give lower classification accuracy. It is observed that better classification error rates ($< 4\%$) are obtained when approximately 40 or more features are retained after decomposition with high order wavelet filters such as Daubechie's 12 to 15 tap and Coiflet's three to five tap filters. We selected one of the high order wavelet filters, Coiflet four tap, for further analysis. The discriminant band distribution of Figure 6.3 may slightly change depending on the wavelet filter.

6.3 Effect of noise on classification

In order to asses the robustness of our methods against disturbing effects, a zero mean Gaussian noise at various SNR level is added to the signal and classification performances are compared as shown in Figure 6.7.

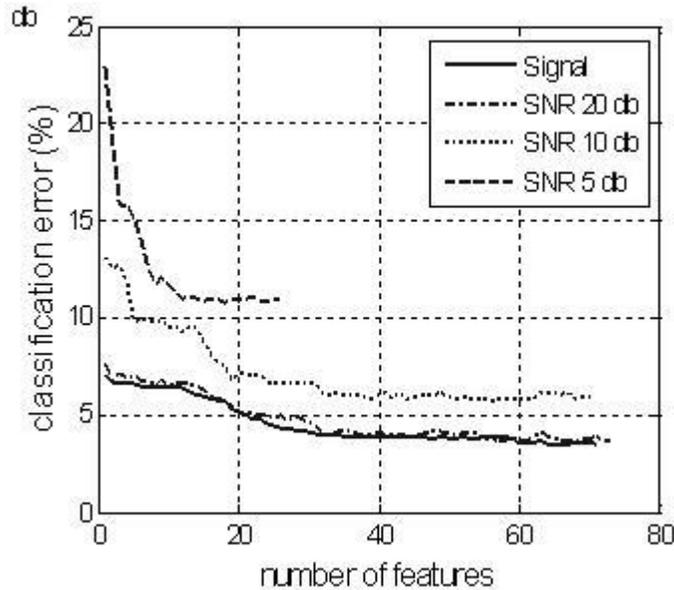


Figure 6.7: The classification error curves for noise disturbed impact acoustic signals

It is observed that the algorithm performs well for reasonable noise levels.

The algorithm usually selects low level subbands nodes when the signals are disturbed by high level noise. This can be justified by the fact that the energy of the impact acoustics is concentrated in the mid and lower bands of the spectrum as indicated in Figure 6.3. In order to keep the efficiency in classification, the algorithm selects features from lower bands with increasing noise level. This also results in a decrease in classification accuracy.

6.4 Effect of shift-invariance on classification

As indicated in the previous sections the main motivation for using UDWT against DWT is the shift invariance property of the UDWT. In order to justify our selection, the UDWT results with those obtained from the DWT and spin-cycle procedure of [Saito02] are compared. The spin-cycle procedure is introduced by [Saito02] to overcome the lack of shift invariance of the DWT and LCP. In particular, a signal is shifted to the left and right for a selected number of spins. For each shift, the signal is expanded into its DWT coefficients. These coefficients are either averaged or processed individually. It has been shown that the spin-cycle procedure provides significant improvements over the direct use of the DWT or LCP [Ince06, Saito02]. In The classification curves obtained from the DWT, the DWT with spin-cycle and the UDWT methods are shown in Figure 6.8.

As expected, the results obtained from DWT were poor. However, the DWT with spin-cycle and UDWT give better classification results. We note that the minimum error of spin-cycle method was slightly lower than UDWT but used more features. However, one should note that the data size of spin-cycle method is higher than that of UDWT. In real time applications it is difficult to obtain fast processing by this method.

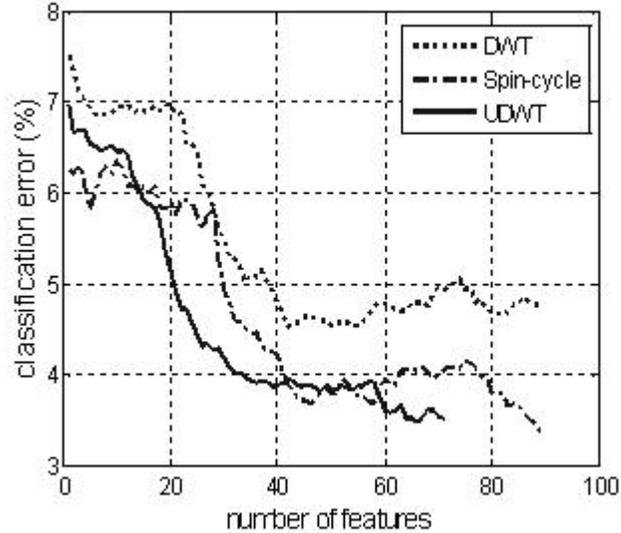


Figure 6.8: The classification error curves for evaluating the efficiency of shift invariance property. The spin-cycle curve stands for the results obtained from DWT supported 1-Spin-Cycle procedure

6.5 Effect of feature extraction algorithm on classification

A total of 210 features corresponding to 210 time-frequency bands are obtained before pruning both in time and frequency axis. The pruning operations not only extracted the most discriminative features; but also decreased the feature dimension from 210 to 70. The extracted features are also selected by FFS and CFS algorithms. Lower classification error curves are obtained with extracted feature set compared to candidate feature set of size 210 (Figure 6.9).

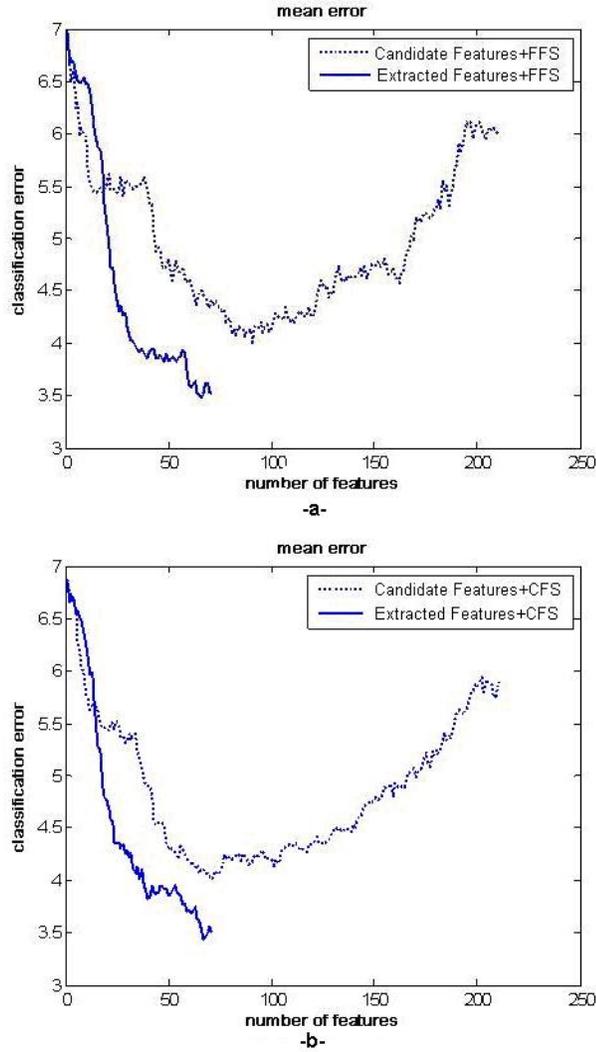


Figure 6.9: The classification error curves of a) FFS and b) CFS method with extracted and candidate feature set

The minimal error of 4% is achieved with around 70 candidate features and the classification error increased after addition of more features. This number coincides with the number of extracted features. It is observed from the Figure 6.9 that the proposed algorithm is successful at detecting relevant features in acoustic signals. Recall that the pruning algorithms in feature extraction merge the features in t-f axis for better classification. Therefore, lower classification error curves are achieved with 70 extracted features compared first 70 candidate features.

The feature selection algorithms of CFS and FFS on extracted feature set did not make a significant difference on classification error (Figure 6.10). However, the CFS has superior performance with the first 24 features.

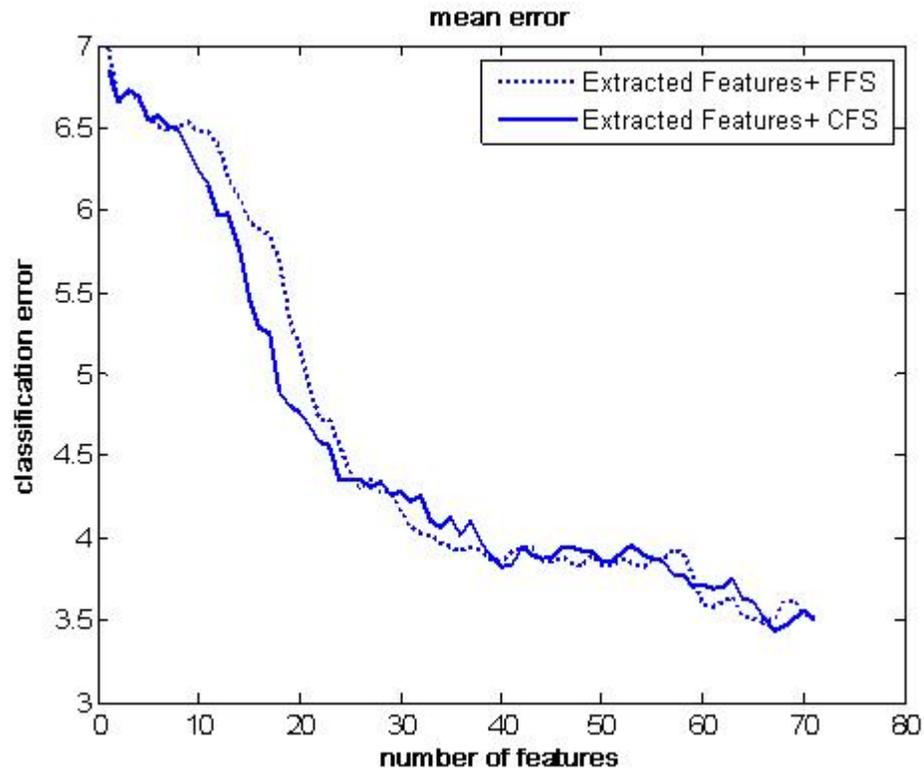


Figure 6.10: The classification error curves of CFS and FFS method with extracted feature set

6.6 Computational Complexity

Determining the best time-frequency segmentation of the signals and the bands to be retained for classification is relatively computationally demanding but this step has to be carried out only once, off-line. For online processing, the throughput of the algorithm in terms of nuts processed per second depends on the number of features used in classification. When the first 64 features providing the best classification rate is employed all 768 samples need to be processed. In this case 17.4 msec is required for signal acquisition of a single nut at a sampling rate of 44.1 kHz. The computations for feature extraction and classification require 13.1 msec on a dedicated P4 3GHz processor. In this case, up to 32 nuts can be processed in a second with classification error of 3.5%. In case an extra 0.5% classification error is tolerable, up to 45 nuts can be processed in a second with 30 features. We observed that only the first half of the signal is required to compute the first 19 features. The classification error achievable at this case is 5.3% and the throughput can be as high as 119 nuts/sec provided that the mechanical sorter system is able to keep up with signal processing.

CHAPTER 7

RESULTS AND DISCUSSIONS ON HYPERSPECTRAL DATA CLASSIFICATION

The developed algorithm is tested on Shelled (SHD) and Shelled/Roasted (SRT) hazelnut hyperspectral data, separately. The hazelnuts groups in both category (SHD and SRT) are considered for two different classification problems. In the first problem, the data sets (Table 5.1 and 5.2) are divided into mold contaminated *Cont* ($= \textit{FlavusCont} + \textit{WaterCont}$) and uncontaminated (*UnCont*) classes without considering the aflatoxin contamination (Figure 5.5). All of the kernels in *Cont* class are mold contaminated (but some of them do not contain measurable aflatoxin). In the second classification problem, we categorized the hazelnuts by just considering the aflatoxin contamination levels and the hazelnuts with over 4 ppb aflatoxin are accepted as aflatoxin contaminated (*Afla+*) and the remaining ones are accepted as aflatoxin free (*Afla-*). All the kernels in *FlavusCont*; some of the kernels from (*WaterCont*) and *UnCont* groups are assigned to *Afla+* class and the remaining kernels from from (*WaterCont*) and *UnCont* were assigned to *Afla-* class as in Figure 7.1

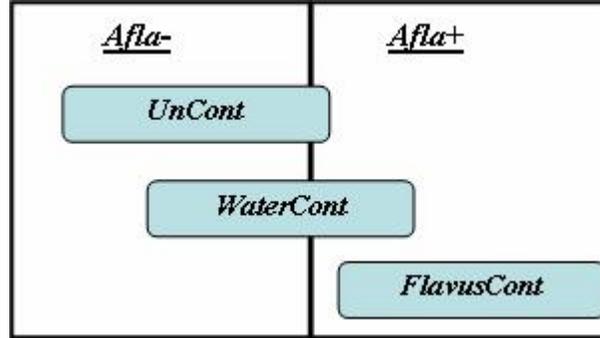


Figure 7.1: The schematic assignments of hazelnuts *Afla+* and *Afla-* groups

7.1 Problem 1: Classification of *SRT_Cont* and *SRT_UnCont* Hazelnuts

Initially, feature trees are generated first along the spectral then spatial-frequency axis. The reflectance energies of 12 spectral images (400-510) were placed on the 4th level of the binary tree (Figure 4.8) from left to right. The remaining four spectral band nodes (SB13-SB16) at the 4th level were set to null in order to complete the binary tree. Consequently, the spatial-frequency quad feature tree is generated by decomposing the spectral bands by full wavelet transform (Figure 4.9). We used Daubechies 8 tap filter for decomposition. Other wavelets can be used as well. For each spectral image, a total of 21 subband images are constructed by a two level decomposition (Figure 4.9). That gives a total of 252 spatial frequency patterns for 12 spectral bands. The nodes in trees are represented by their energies as features. The entropy distribution of the spatial-frequency features of spectral bands can be seen in Figure 7.2. It is observed that low spatial-frequency subbands have higher entropy than high frequency components.

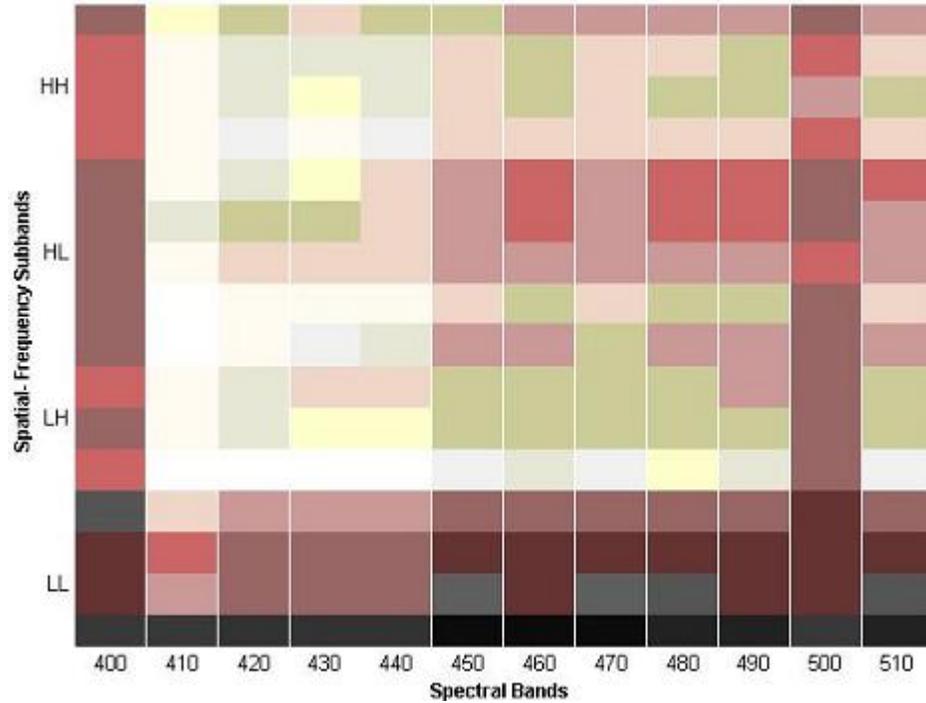


Figure 7.2: Entropy map of the spectral spatial-frequency features

After the generation of the feature trees, the feature extraction process is initiated by pruning the spectral bands (Section 4.2.2). Figure 7.3 shows the spectral band pruning. The spatial-frequency domain features (Figure 4.9) of the spectral bands are averaged according to the pruned tree in Figure 7.3 before pruning in spatial-frequency axis.

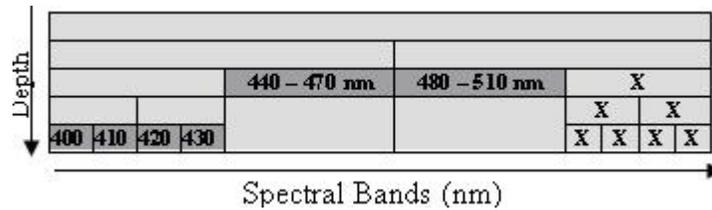


Figure 7.3: Spectral band pruning. The spectral bands (440-470) and (480 -510) are pruned. The null bands in the tree were ignored at pruning

The spatial-frequency subbands features are pruned after merging the frequency subband features of the pruned spectral bands (Section 4.2.3). The pruning in both axes revealed the location of the most discriminative features (Figure

7.4) of the hyperspectral data. A total of 12 spectral-frequency features were obtained after the pruning operations. These operations also decreased the feature dimension from 192 to 12. The location of 192 candidate features can be observed in Figure 7.2.

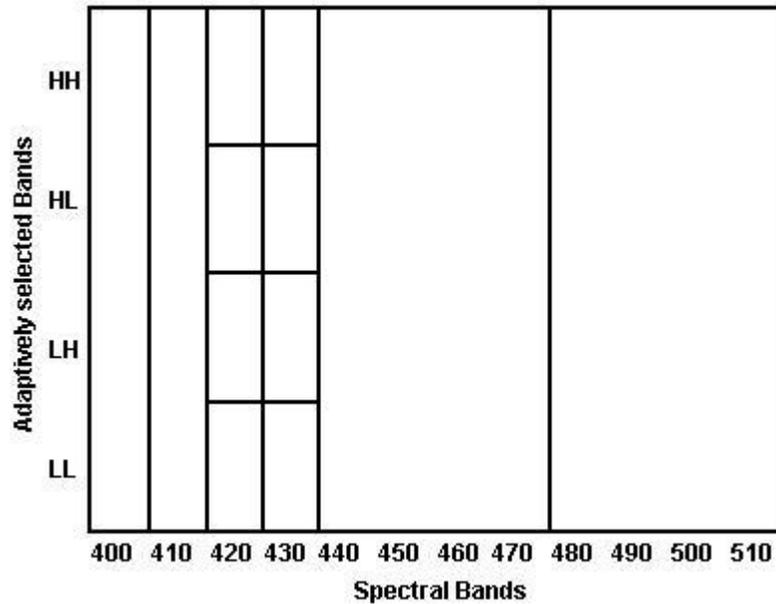


Figure 7.4: The location of the most discriminative features in spectral spatial-frequency axis

It is observed that the spectral bands (440-470) and (480-510) are pruned along the spectral axis and these spectral bands are not decomposed into spatial-frequency subbands. However, the spectral bands of 430 and 440 are decomposed into subbands in spatial-frequency axis.

The extracted 12 features are then ranked by four different feature selection algorithms and fed into the linear classifier incrementally.

- Fisher Distance Based Feature Selection(FFS)
- Correlation Based Feature Selection(CFS)
- PCA Based Feature Selection(PCA)
- Wrapper Based Feature Selection

In contrast to FFS and CFS, PCA does not work in incremental way and uses all the features to project to the new space. The wrapper model is just selects the feature combination giving the best classification. The FFS and CFS algorithms ordered the extracted features as in Figure 7.5. The darkness and the number on nodes indicate the relevance of the features in those nodes. The order of features may vary depending of the feature selection algorithm and this selection may effect the classification accuracy.

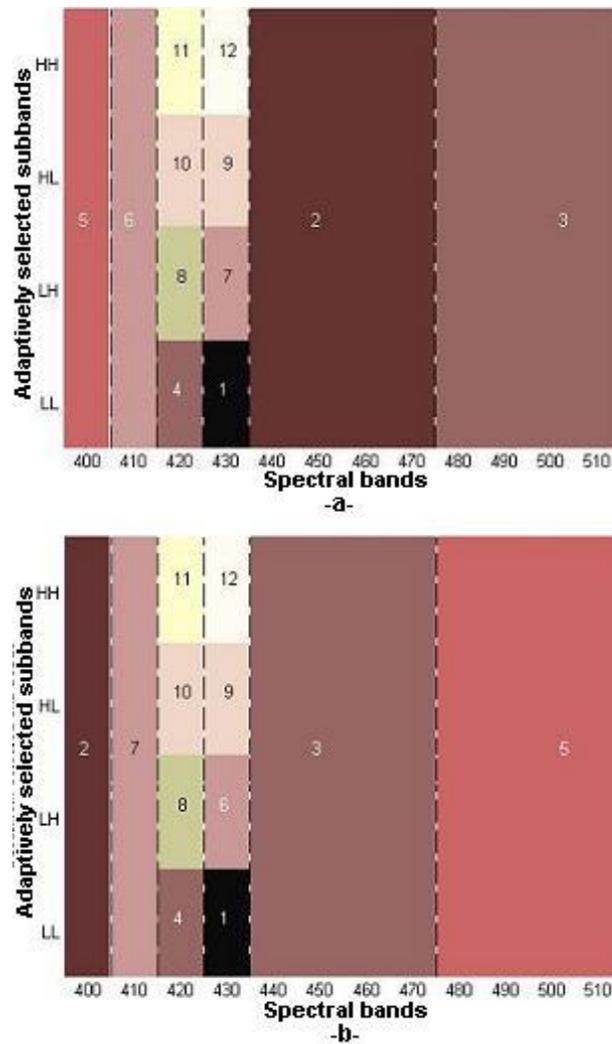


Figure 7.5: Spectral Spatial-Frequency feature map of SRT hazelnut data ranked by a) by FFS, b) by CFS algorithm

The selected features are added one by one to find out the optimal number of features for classification. The minimum classification error is obtained with the two features by PCA; four features by wrapper and five features by CFS and FFS (Table 7.1, Figure 7.6). In contrast to other feature selection algorithms, the feature number in PCA methods (Table 7.1) is the number of projected eigenvectors.

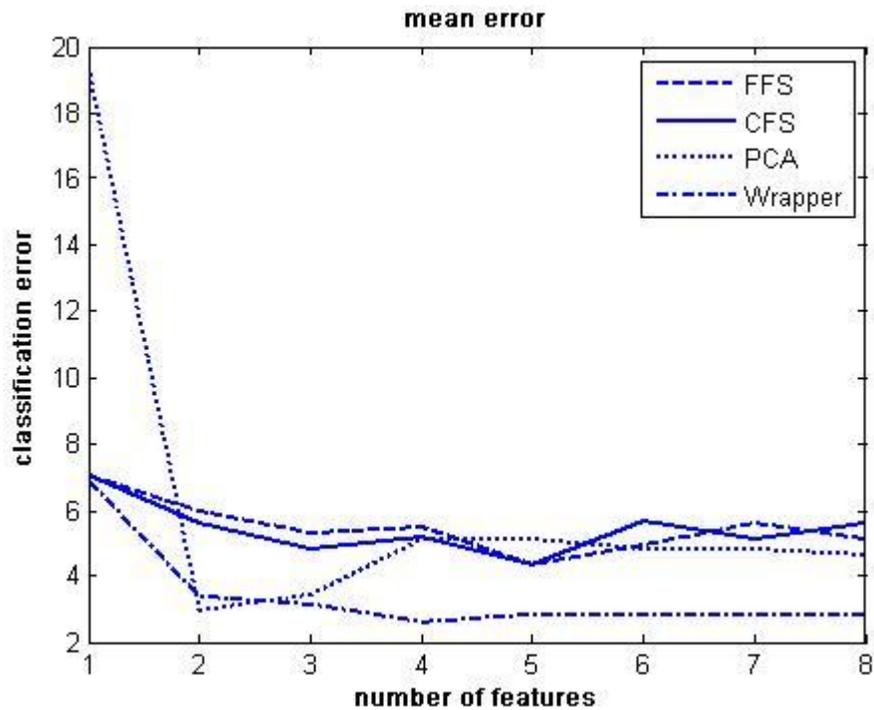


Figure 7.6: The classification error curves on FFS, CFS, PCA and Wrapper based selected ranked features

Table 7.1: Minimum classification error obtained by four feature selection algorithms. The number of features of methods giving the error is shown in brackets

	FSS(5)	CFS(5)	PCA(2)	Wrapper(4)
Error(%)	4.35	4.35	3.00	2.60

For a practical sorter system, it is preferable to reach to lowest error with fewer number of features. The lowest error is achieved by four wrapper based selected features. It is possible to get good results with PCA ordered features. However, It is impractical to use the PCA-processed features because the PCA uses all the features at hand for all unique projection although it gives lower classification error.

When the classification results with five FFS ordered features are analyzed, it observed that 10 of 181 *SRT_Cont* hazelnut and 3 of 104 *SRT_UnCont* hazelnuts were misclassified. However none of the misclassified *SRT_Cont* hazelnuts contain aflatoxin. The mean aflatoxin level of the test set including hazelnut from *SRT_Cont* and *SRT_UnCont* group is 608 ppb and the algorithm classified the kernels in to two classes whose aflatoxin contamination levels are 1095 ppb and 0.7 ppb.

The feature extraction step (Figure 4.7), which is the focus of proposed study, positively affects the success of the feature selection algorithm by providing the most discriminative features in hyperspectral data. This statement is validated (Figure 7.7, 7.8 and 7.9) by comparing the classification results when feature selection algorithms (CFS, FFS and PCA) are applied after feature extraction or applied to candidate feature set in the 192 dimensional space. This comparison could not be performed with wrapper model because of the extensive computation due to the high number of feature subset combination.

Lower classification error curves are obtained when working with extracted feature set than those with candidate feature set by all these feature selection algorithms. It is observed from the Figures 7.7, 7.8 and 7.9 that the 2D feature extraction algorithm generates high discriminative features by pruning both in spectral and spatial-frequency domains. It also enables us to select the sufficient spectral band by eliminating the irrelevant ones. This will decrease the image acquisition and processing cost of the food inspection and sorter systems.

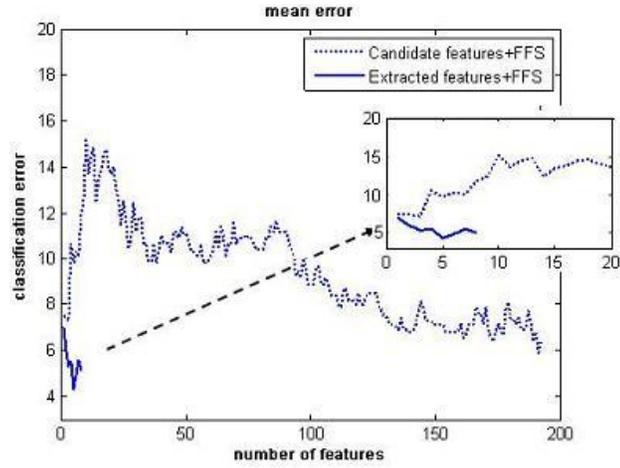


Figure 7.7: The classification error curves with the features selected from candidate and extracted features by FFS algorithm

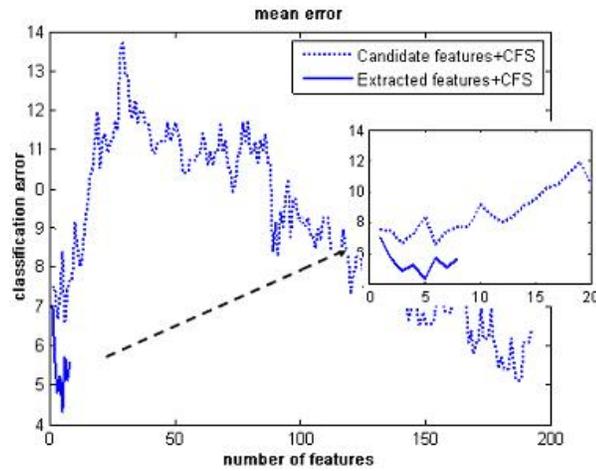


Figure 7.8: The classification error curves with the features selected from candidate and extracted features by CFS algorithm

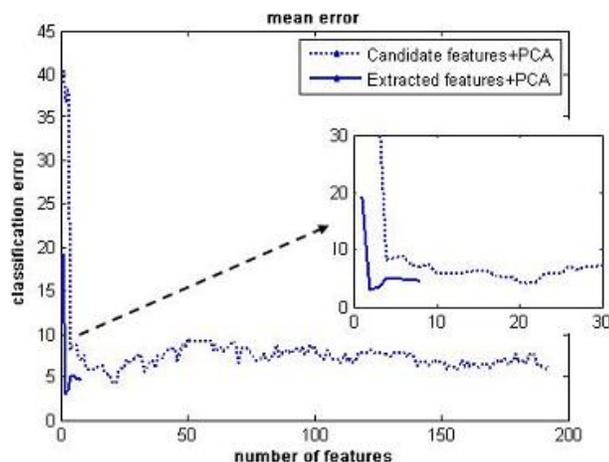


Figure 7.9: The classification error curves with the features selected from candidate and extracted features by PCA algorithm

7.2 Problem 2: Classification of SRT_Afla+ and SRT_Afla- Hazelnuts

In this problem, the SRT hazelnuts are categorized by assigning the ones over 4 ppb aflatoxin concentration to SRT_Afla+ and the remaining ones to SRT_Afla- groups. The average aflatoxin level of 96 SRT_Afla+ and 189 SRT_Afla- group hazelnuts are 1883 ppb and 0.06 ppb respectively. Same procedure in the first classification problem (Section 7.1) was applied for the new data set. The spectral pruning in feature extraction step pruned the spectral bands of 420-430, 440-450 and 480-510 but kept the spectral bands of 400, 410, 460, 470 nm (Figure 7.10) intact. The subbands in spatial-frequency axis of all spectral bands are completely pruned except the 420-430 nm spectral band.

The extracted features are then ranked by feature selection algorithms and used in linear classifier. The extracted feature map whose features are ranked by FFS and CFS algorithms are given in Figure 7.10. The darkness and the number on the nodes indicate the relevance of the features in those nodes.

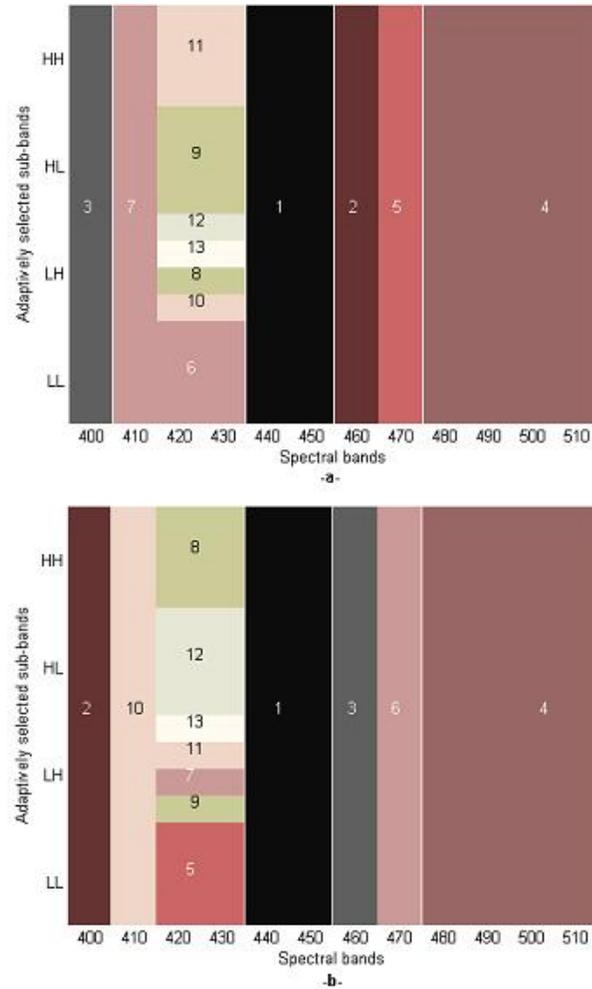


Figure 7.10: Spectral Spatial-Frequency feature map of *SRT_Afla+* and 189 *SRT_Afla-* data which were ranked by a) by FFS, b) by CFS algorithm.

The selected features are fed into the linear classifier in four fold validation as in Problem 1. However, lower classification accuracies are obtained compared to the classification of *SRT_Cont* and *SRT_UnCont* classes. The minimum classification error is obtained with the 6 features by PCA; 3 features by wrapper and four features by CFS and FFS (Table 7.2, Figure 7.11). The spectral pruning in feature extraction step pruned the spectral bands of 420-430, 440-450 and 480-510 but kept the spectral bands of 400, 410, 460, 470 nm (Figure 7.10) for the new data set.

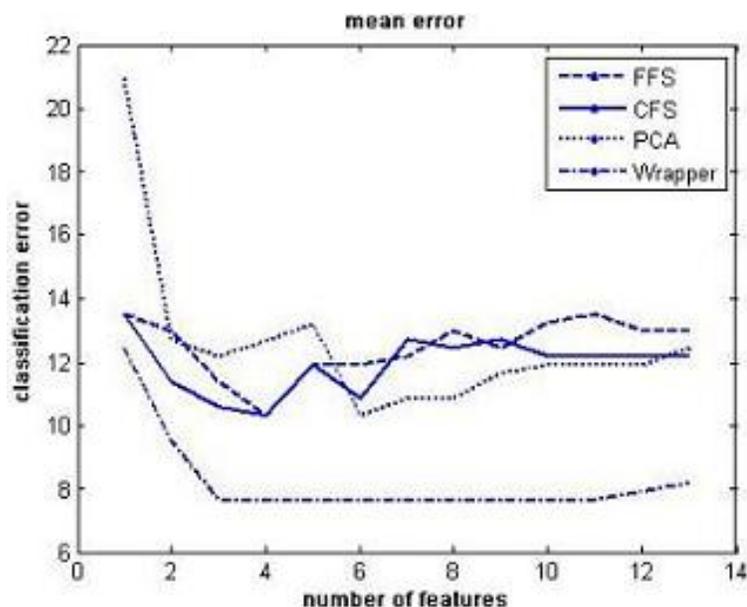


Figure 7.11: The classification error curves of *SRT_Afla+* and *SRT_Afla-* data by FFS, CFS, PCA and Wrapper based selected features

Table 7.2: Minimum classification error obtained by four feature selection algorithms.

	FSS(4)	CFS(4)	PCA(6)	Wrapper(3)
Error(%)	10.34	10.34	10.34	7.69

The lowest error of 7,69% is achieved with three features selected by Wrapper model. When the classification results with 4 FFS ordered features are analyzed, it is observed that the algorithm misclassified 3 of the 96 *SRT_Afla+* hazelnuts and 39 of the 189 *SRT_Afla-* hazelnuts. Two of the 3 misclassified *SRT_Afla+* hazelnuts are from *SRT_UnCont*; the remaining misclassified kernel is from *SRT_WaterCont* group, originally. Whereas, 38 of the 39 misclassified *SRT_Afla-* hazelnuts are from *SRT_WaterCont*, one of the 39 is from *SRT_UnCont* class, originally.

The effect of feature extraction step in the proposed algorithm can also be observed with *SRT_Afla+* and *SRT_Afla-* data set with three different feature selection algorithms (Figure 7.12, 7.13 and 7.14). It is observed that better classification error curves are obtained with extracted (pruned) features than candidate features.

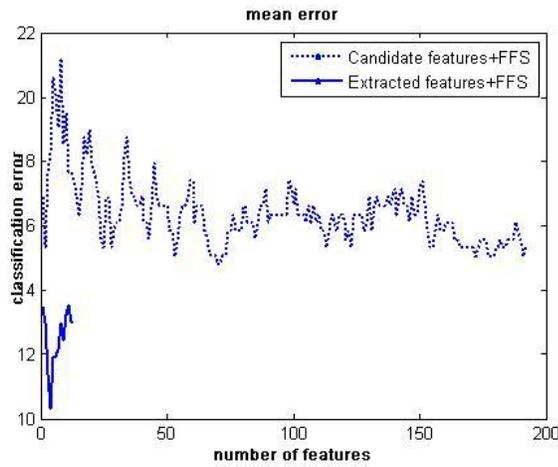


Figure 7.12: The classification error curves with the features selected from candidate and extracted features by FFS algorithm

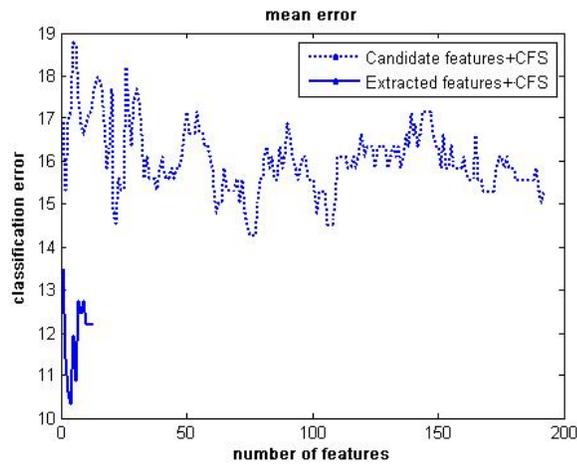


Figure 7.13: The classification error curves with the features selected from candidate and extracted features by CFS algorithm

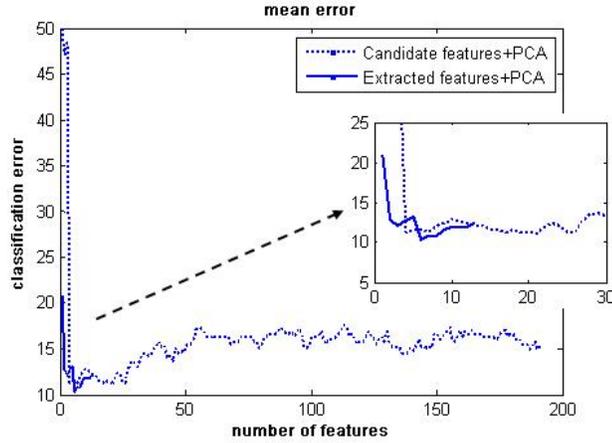


Figure 7.14: The classification error curves with the features selected from candidate and extracted features by PCA algorithm

The CFS based feature selection algorithm has similar performance compared to the FFS feature selection algorithm (Figure 7.6 and Figure 7.11). When the classification problems P1 and P2 of SRT hazelnuts are compared, better results are obtained in classifying *SRT_Cont* and *SRT_UnCont* kernels. The aflatoxin level of the SRT hazelnuts decreased to 0.7 ppb from 608 ppb by removal of the *SRT_Cont* hazelnuts and decreased to 0.84 ppb from 608 ppb by removal of the *SRT_Afla+* hazelnuts. It is recommended to separate *SRT_Cont* kernels from hazelnut lots to decrease aflatoxin level because the contaminated kernels *SRT_WaterCont* are likely contain aflatoxin. These nuts are also not preferred by consumers because of bad taste and appearance.

7.3 Problem 3: Classification of *SHD_Cont* and *SHD_UnCont* Hazelnuts

In this problem, the Shelled hazelnuts, SHD category, are divided into mold contaminated *SHD_Cont* and uncontaminated *SHD_UnCont* groups without considering the aflatoxin concentration. The average aflatoxin level of the 81 *SHD_UnCont* and 150 *SHD_Cont* group hazelnuts are 0.36 ppb and 938 ppb respectively. Same procedure in the first classification problem (Section 7.1) was

applied for the new data set. The hyperspectral images of the SHD category hazelnuts are lower than the SRT category hazelnuts. In contrast to SRT hazelnuts, we used the spectral band images of 550nm and 600 nm because of their positive effects on classification results of SHD hazelnuts. Recall that these spectral bands have higher FWHM range compared to the other spectral bands between 400 nm and 510 nm. Moreover, in contrast to the SRT hazelnuts, the algorithm did not prune the subbands both in spectral and spatial-frequency axis. The pruning algorithm only pruned the spectral bands of 460 and 470 nm. However, the algorithm pruned the subbands in LL and HH region of spatial-frequency axis of images (Figure 7.15). Figure 7.15 also shows the extracted feature location obtained 2D pruning algorithm and the features that is ranked by FFS based feature selection algorithm. The CFS algorithm ranked the features similar to FFS algorithm in that they selected the same features for the first 10.

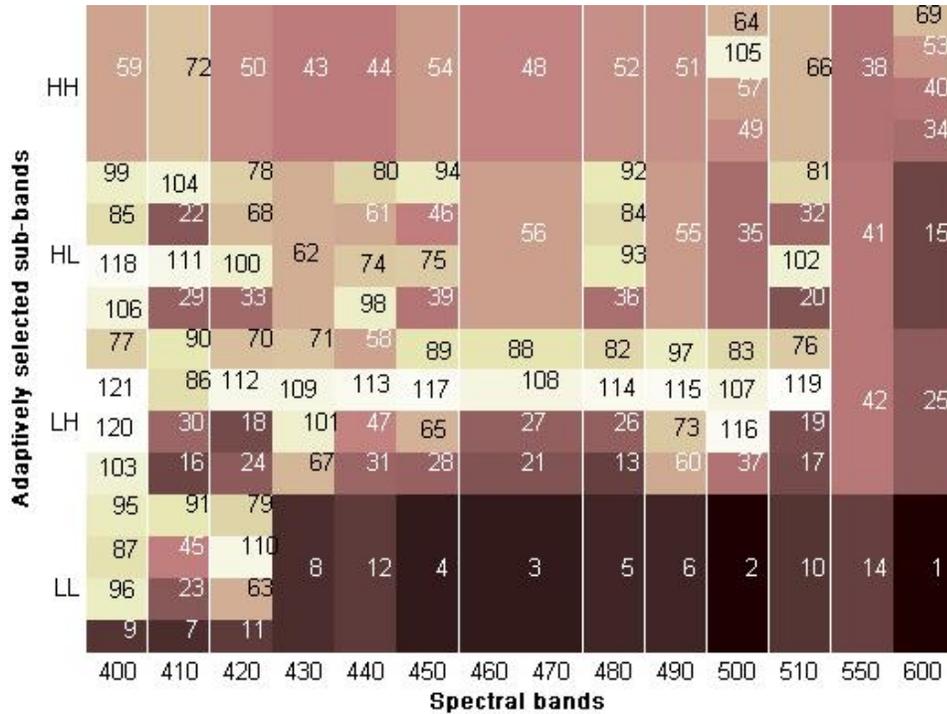


Figure 7.15: Spectral Spatial-Frequency feature map of *SHD_Cont* and *SHD_UnCont* data which were ranked by a) by FFS algorithm

Recall that the numbers and the color darkness on the map shows the order

of the selected features. It is observed that most discriminative feature is the LL subband of the spectral bands of 600 nm. When the classification results with the features ranked by FFS, CFS and PCA algorithm are compared, high classification errors are obtained with the first a few features and the addition of new features marginally effects the error (Figure 7.16).

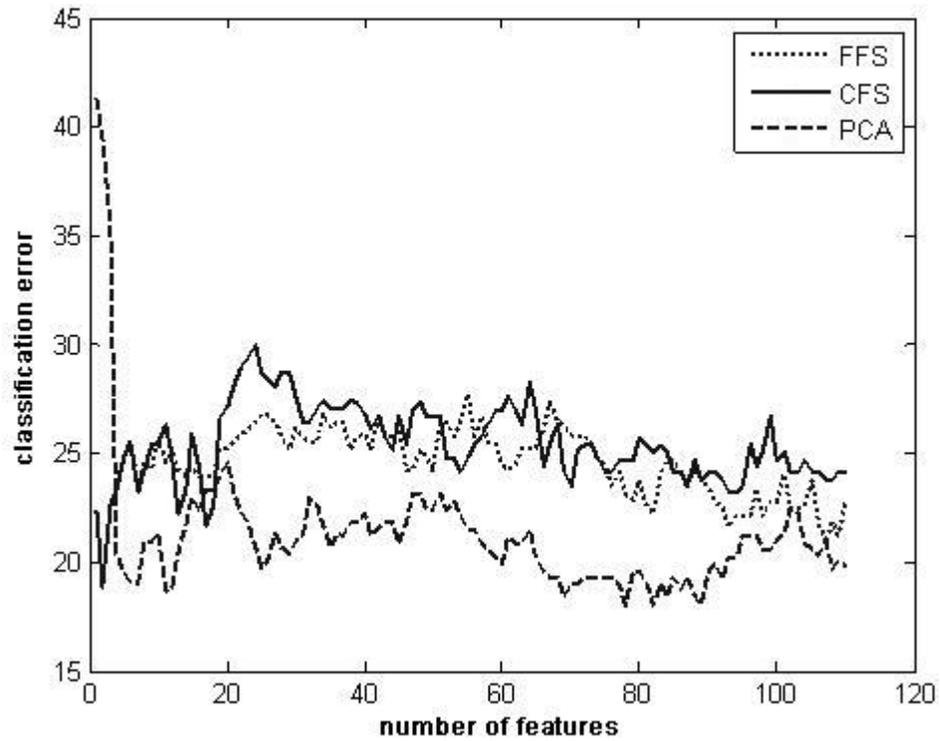


Figure 7.16: The classification error curves of *SHD_Cont* and *SHD_UnCont* data by FFS, CFS and PCA based selected features

The minimum classification error is obtained with the 78 PCA ranked; two wrapper, CFS and FFS ranked features (Table 7.3). The algorithm misclassified 10/81 *SHD_UnCont* and 37/150 *SHD_Cont* hazelnuts with FFS ranked features. However wrapper based selected features give better classification accuracy of 15.85% compared to other methods.

Table 7.3: Minimum classification error of *SHD_Cont* and *SHD_UnCont* data obtained by four feature selection algorithms. The number of features of methods giving the error is shown in brackets

	FSS(2)	CFS(2)	PCA(78)	Wrapper(2)
Error(%)	18.88	18.88	18.02	15.85

The effect of feature extraction step for classifying the *SHD_Cont* and *SHD_UnCont* data set with three different feature selection algorithms can be seen in Figures 7.17, 7.18 and 7.19. It is observed that better classification error curves are obtained with extracted (pruned) features than candidate features.

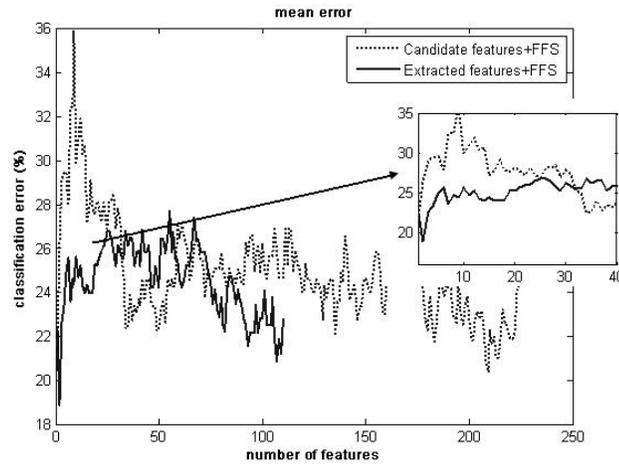


Figure 7.17: The classification error curves with the features selected from candidate and extracted features of SHD category hazelnuts by FFS algorithm

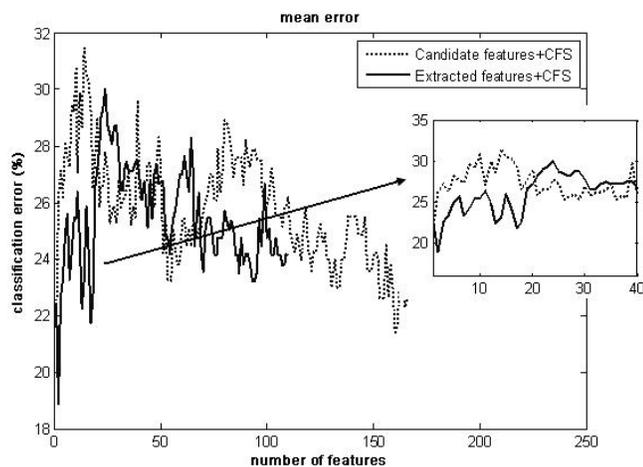


Figure 7.18: The classification error curves with the features selected from candidate and extracted features by of SHD category hazelnuts CFS algorithm

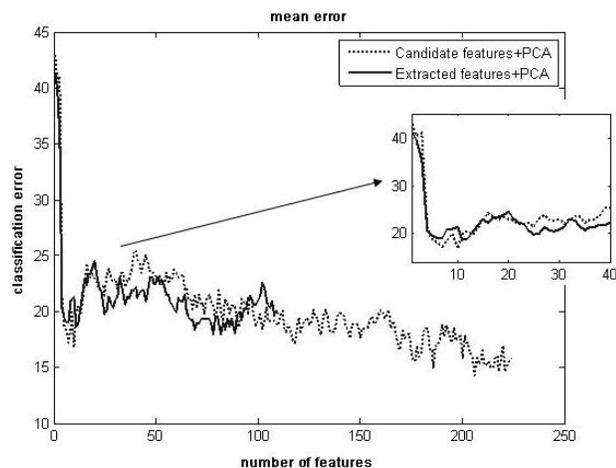


Figure 7.19: The classification error curves with the features selected from candidate and extracted features by of SHD category hazelnuts PCA algorithm

7.4 Problem 4: Classification of *SHD_Afla+* and *SHD_Afla-* Hazelnuts

In this problem, the SHD hazelnuts are categorized by assigning the ones with 4 ppb aflatoxin concentration to *SHD_Afla+* and the remaining ones to

SHD_Afla- groups. The average aflatoxin level of 53 *SHD_Afla+* and 179 *SHD_Afla-* group hazelnuts are 3146 ppb and 0.05 ppb respectively. The algorithm pruned the spectral bands of 480 and 490 nm in spectral axis. Similarly, all the spatial-frequency subbands of the 450, 510 and 550 nm images are completely pruned(Figure 7.20). Figure 7.20 also shows the order of the features that are selected by FFS algorithm. As in stated before the algorithm does not make a comprehensive pruning to the low quality images (such as SHD hazelnut images). The most discriminative feature in the map is located in the spectral band of 510 nm.

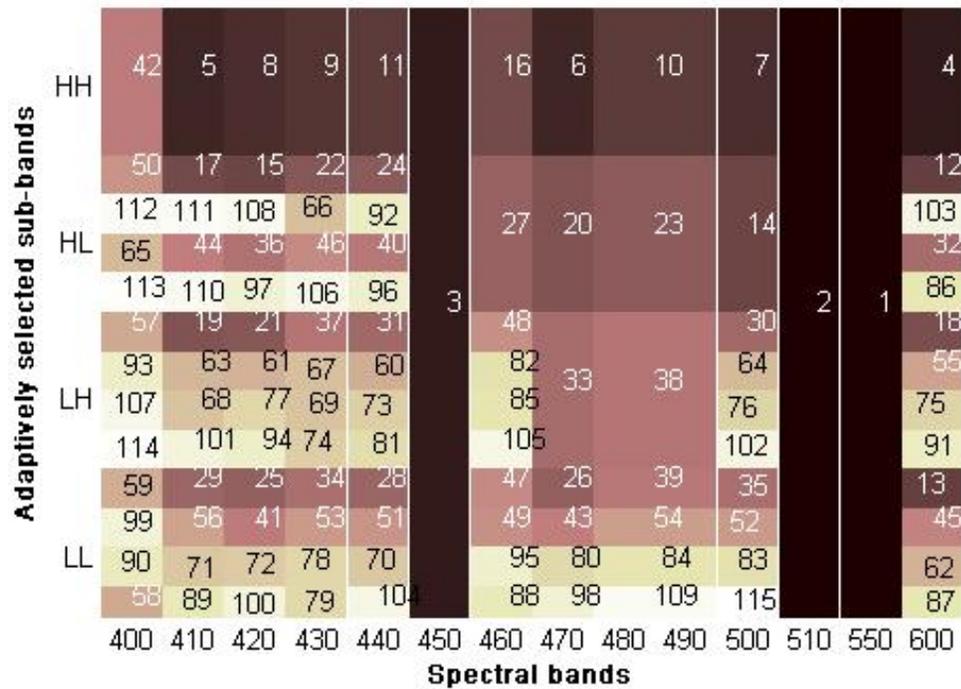


Figure 7.20: Spectral Spatial-Frequency feature map of *SHD_Afla+* and *SHD_Afla-* data which were ranked by a) by FFS algorithm

The lowest classification error of 17% is obtained with the features that are selected by wrapper approach. The CFS and FFS algorithm results similar error curve which are worst then the curve obtained by PCA selected features(7.21). Minimum classification error for *SHD_Afla+* and *SHD_Afla-* data set is obtained with 115 PCA ranked, 95 FFS ranked, 18 CFS ranked and two wrapper

ranked features (Table 7.4).

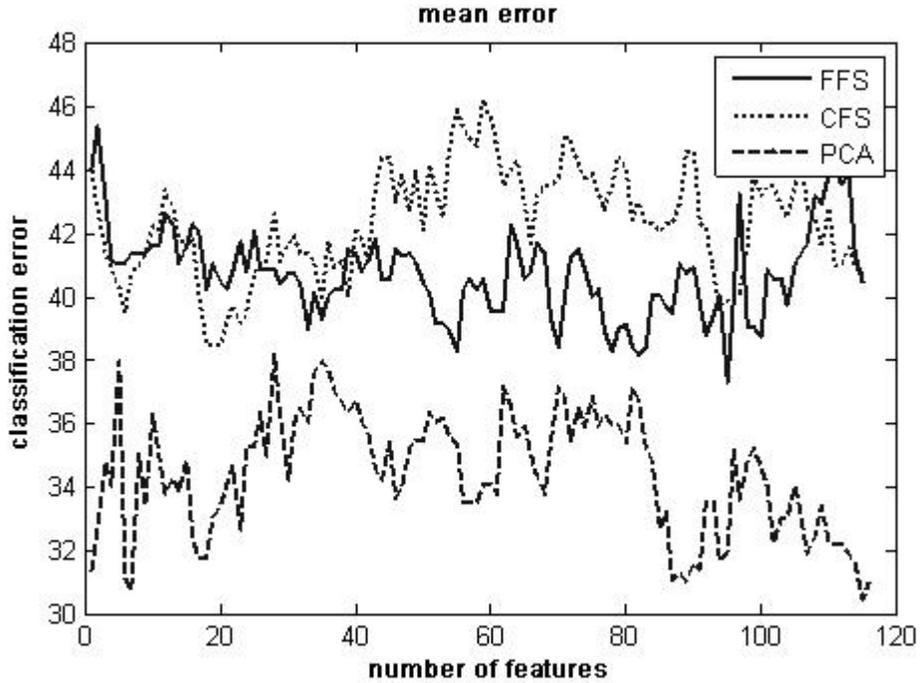


Figure 7.21: The classification error curves of *SHD_Afla+* and *SHD_Afla-* data by FFS, CFS and PCA based selected features

Table 7.4: Minimum classification error of *SHD_Afla+* and *SHD_Afla-* data obtained by four feature selection algorithms. The number of features of methods giving the error is shown in brackets

	FSS(95)	CFS(18)	PCA(115)	Wrapper(2)
Error(%)	37	38	30	17

The feature extraction step in the algorithm does not positively effects the performance of feature selection algorithm at classification of the *SHD_Afla+* and *SHD_Afla-* hazelnuts (Figure 7.22, 7.23 and 7.24). This is thought to be from the low quality of images. The decrease in quality may suppress the possible discriminative features in the images.

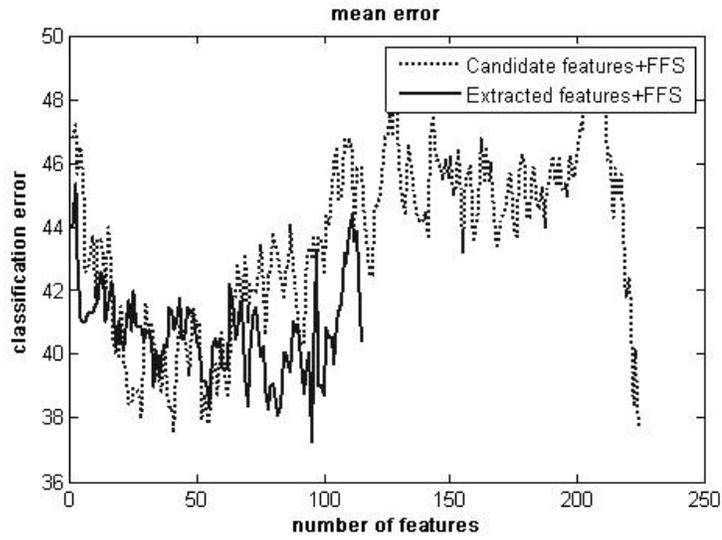


Figure 7.22: The classification error curves with the features selected from candidate and extracted features of SHD category hazelnuts by FFS algorithm

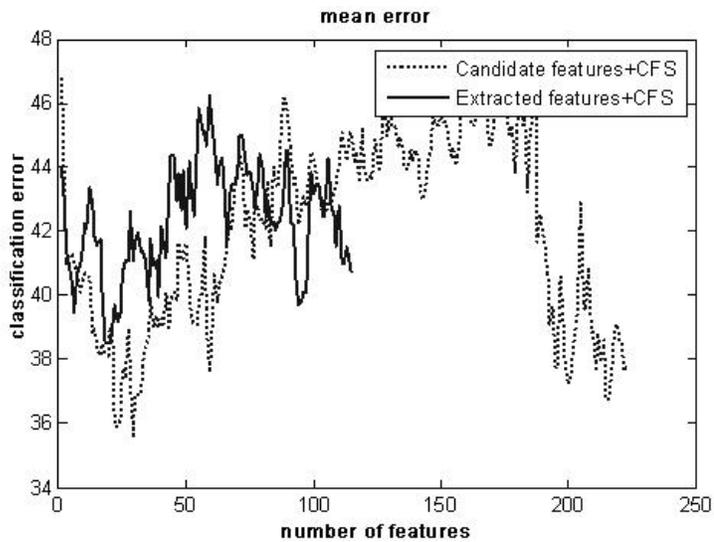


Figure 7.23: The classification error curves with the features selected from candidate and extracted features of SHD category hazelnuts by CFS algorithm

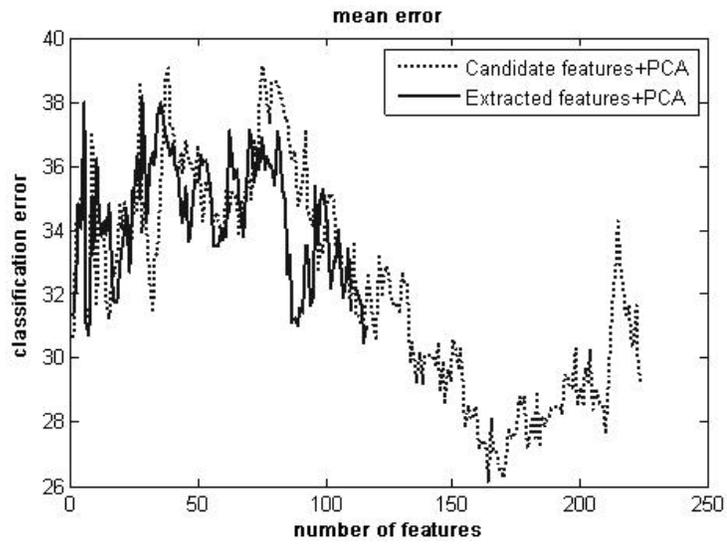


Figure 7.24: The classification error curves with the features selected from candidate and extracted features of SHD category hazelnuts by PCA algorithm

CHAPTER 8

CONCLUSIONS

In this thesis, an algorithm is developed to extract discriminative features from acoustic and hyperspectral data by using Local Discriminant Bases search. The original LDB algorithm is adapted to two dimensional searches to extract the most discriminant features in data. The feature extraction process decreases the feature dimension by both eliminating the irrelevant ones and/or by merging the ones that do not provide extra information on their own. Another dimension reduction is performed by selecting the extracted features by various feature selection algorithms. Our aim is to achieve high classification accuracy with fewer number of features. The number of extracted feature for minimum error error is obtained by classification accuracy and a basic LDA classifier is used to get this accuracy. One can try other non-linear classifiers, like neural networks or Support Vector Machine, to achieve better classification accuracies with the extracted features.

The developed algorithm is implemented on both impact acoustic and hyperspectral data of hazelnut kernels for classification. This application is stemmed from the need of separating contaminated or potentially contaminated hazelnut kernels by non-invasive and fast methods that can perform real time in practical systems. The usage of fewer number of features with less amount of data enables the production of low cost and high throughput sorting systems.

We conducted an experimental study on detecting the effect of hazelnut shell on aflatoxin contamination and it observed that shell of hazelnut form a strong barrier against mold contamination. However, any damage on shell eliminates this capability. Therefore the hazelnuts with cracked shell should be separated from the regular shell hazelnut by using impact acoustic signal. The relevant features in acoustic signal are extracted by searching the local discriminant bases

first in time, then in frequency axis. The extracted features are then selected by feature selection algorithms and used in linear classifier. A classification accuracy of 96.5% is obtained with 64 LBD features. We observed that the most discriminative feature locations were concentrated in the high frequency bands corresponding to the early and post impact regions as indicated in Figure 6.3. The LDB based features give better classification accuracy compared to other types of features. In addition, the features obtained by feature extraction step positively effect the classification compared to candidate feature set. The developed algorithm can be implemented to real-time sorting devices due to its high throughput up to 25 nuts/sec. This limitation in throughput stemmed from the time required for each acoustic signal, not from the computation time.

Unlike to InH hazelnuts, the risk in SHD and SRT category hazelnuts can not be eliminated by impact acoustic. We used hyperspectral imaging for classification in shelled (SHD) and shelled&roasted (SRT) categories. Two different classification problems are defined for each category hazelnuts. The first problem is based on classifying contaminated (*Cont*) or uncontaminated (*UnCont*); second one is on classifying aflatoxin contaminated (*Afla+*) or aflatoxin free (*Afla-*) hazelnuts. The developed algorithm for hyperspectral images starts by obtaining a candidate feature set by generating two feature trees on spectral and spatial-frequency axis in order. The features trees are then pruned to get the most discriminative ones in spectral and spatial-frequency axis.

For the separation problem of *Cont* and *UnCont* group hazelnuts in SRT category, the candidate feature set of 192 dimensions is decreased to 12 by adaptive pruning in both axes. Four different feature selection algorithms are performed on extracted feature set to decrease the feature dimension if possible. A classification accuracy of 95.6% is achieved by the first five CFS or FFS ranked features. A better accuracy of 97.4% is achieved by the first four wrapper based ranked features. The most discriminant feature is defined as the LL subband of the spectral bands of 430 nm. For aflatoxin *Afla+* versus *Afla-* separation problem in the same category, the algorithm decreased to feature dimension from 192 to 14 in which the most discriminative one is the pruned form of the spectral bands of 440 and 450 nm. Relatively poor classification accuracies are obtained compared to

the first problem of this category. A classification accuracy of 89.66% is obtained by FFS, CFS and PCA ordered features. However, better accuracy of 92.31% is obtained by the first wrapper based ranked features. The extracted features positively affect the classification accuracies compared to candidate feature set. These can be observed from the Figures 7.7- 7.9 and Figures 7.12-7.14. The aflatoxin level of the SRT hazelnuts decreased to 0.7 ppb from 608 ppb by removal of the SRT Cont hazelnuts and decreased to 0.84 ppb from 608 ppb by removal of the *SRT_Afla+* hazelnuts. It is recommended to separate *SRT_Cont* kernels from hazelnut lots to decrease aflatoxin level because the contaminated kernels are high risk ones and these nuts are also not preferred by consumers because of bad taste and appearance.

The system does not produce good results for the SHD category hazelnuts as with the SRT category hazelnuts. As stated in Chapter 5, the hazelnuts in SHD category do not reflect sufficient light with existing hyperspectral imaging system under UV illumination even at high exposure times. Therefore, the images of SHD category hazelnuts are very low quality compared to SRT category hazelnuts. We included the images of the spectral bands at 550 and 600 nm to increase the information we have. For the *UnCont* and *Cont* group classification problem of SHD category hazelnuts, the feature extraction procedure did not decrease the feature dimension to reasonable levels. The most discriminative features are concentrated on LL region of spectral bands. Minimum classification accuracy of 18% is achieved with two FFS and CFS ordered features, whereas the same accuracy is obtained with 78 PCA ordered features. The lowest accuracy of 15.85 % is achieved with two wrapper based ranked features. The extracted features marginally decreased the error curves compared to candidate feature set (Figure 7.17-7.19). For the *SHD_Afla+* and *SHD_Afla-* classification problem, worse classification accuracies around 30% are achieved with CFS, FFS and PCA ranked features. The lowest classification accuracy 17% for *SHD_Afla+* and *SHD_Afla-* classification is achieved with two wrapper based selected features.

8.1 Future work

The limitations in this study originated from the hyperspectral imaging system with UV illumination. The existing illumination conditions are not suitable for SHD category hazelnuts. These nuts can be illuminated by light sources performing in IR, NIR or VIS region.

We used a rotating filter wheel to skip to other filters of same FWHM. The reflectance of a combined filter is estimated by combining the reflectance of individual filters. An electronically tunable filter with tunable FWHM may help taking more spectral images at any pass band region. This may help to identify the best spectral bands for any specific classification problem.

Final classification accuracy of the system is obtained with the selected features among the extracted feature set. The feature extraction procedure did not take the correlation between features in the data into account. The feature-feature correlation is considered in feature selection step. The feature selection and extraction algorithm may be combined to get most discriminant as well as independent features.

The developed algorithm is fast, robust and can be applied to real-time food grading and sorting systems by just using the identified filters for imaging.

The feature extraction algorithm can be applied to other hyperspectral data and we are also working on satellite images as well as other food items.

REFERENCES

- [1] **Hsu, P.**, 2004. Feature Extraction of HyperSpectral Images Using Matching Pursuit, *Proceedings of the XXth ISPRS Congress, Istanbul*.
- [2] **Landgrebe, D.A.**, 2003. Signal Theory Methods in Multispectral Remote Sensing, *John Wiley & Sons*.
- [3] **Lee, C. and Landgrebe, D.**, 1993. Analyzing High-dimensional Multispectral Data, *IEEE Trans. Geoscience and Remote*, **31**, 792–800.
- [4] **Jolliffe, I.**, 2002. Principal Component Analysis, second ed, *Springer*.
- [5] **Jia, X. and Richards, J.**, 1999. Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification, *IEEE Trans Geoscience and Remote Sensing*, **37(1)**, 538–542.
- [6] **Coifman, R. and Wickerhauser, M.V.**, 1992. Entropy-based Algorithms for Best Basis Selection, *IEEE Trans. Inform. Theory*, **38**, 713–719.
- [7] **Saito, N. and Coifman, R.R.**, 1994. Local Discriminant Bases, in Mathematical Imaging: Wavelet Applications in Signal and Image Processing II, *Proceedings of the SPIE*, **2003**, 2–14.
- [8] **Ince, N., Tewfik, A. and Yagcioglu, S.**, 2005. Classification of Movement EEG With Local Discriminant Bases, *Proceedings of the IEEE ICASSP*.
- [9] **Ince, N., Arica, S. and Tewfik, A.**, 2006. Classification of Single Trial Motor Imagery EEG Recordings With Subject Adapted Non-dyadic Arbitrary Time-frequency Tilings, *J. Neural Eng*, **3**, 235–244.

- [10] **Yang, H., S.V.Vuuren and Hermansky, H.**, 1999. Relevance of Time-Frequency Features for Phonetic Classification Measured by Mutual Information, *Proceeding of IEEE ICASSP*.
- [11] **Kumar, S., Ghosh, J. and Crawford, M.**, 2001. Best-Bases Feature Extraction Algorithms for Classification of Hyperspectral Data, *IEEE Trans Geoscience and Remote Sensing*, **39**, 1368–1379.
- [12] **Cheriyadat, A. and L.M.Bruce**, 2003. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data, *Proceedings of the IEEE IGARSS*, **6**, 3420–3422.
- [13] **C.R.Dichter**, 1984. Risk Estimates of Liver Cancer Due to Aflatoxin Exposure from Peanuts and Peanut Products, *Food Chemistry and Toxicology*, **22(6)**, 431–437.
- [14] **Jimenez, L. and Landgrebe, D.**, 1999. Hyperspectral Data Analysis and Feature Reduction Via Projection Pursuit, *IEEE Transactions on Geoscience and Remote Sensing*, **37(6)**, 2653–2667.
- [15] **Jain, A., Duin, R. and Mao, J.**, 2000. Statistical Pattern Recognition: A Review, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **22(1)**, 4–37.
- [16] **Hall, M.A.**, 1998. Correlation-based Feature Selection for Machine Learning, *PhD Thesis, Hamilton, NZ: Waikato University, Department of Computer Science*.
- [17] **Zhao, J., Wang, G., Wu, Z., Tang, H. and Li, H.**, 2002. The Study on Technologies for Feature Selection, *Proceedings of the International Machine Learning and Cybernetics*.
- [18] **L.Talavera**, 2005. An Evaluation of Filter and Wrapper Methods for Feature Selection in Categorical Clustering, *Lecture Notes in Computer Science, Springer Berlin / Heidelberg*, **3646**.

- [19] **Peng, H., Long, F. and Ding, C.**, 2005. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy, *IEEE Transactions on pattern analysis and machine intelligence.*, **27 (8)**, 1226–1238.
- [20] **Ince, N., Tewfik, A. and Arica, S.**, 2007. Extraction Subject-specific Motor Imagery Time-frequency Patterns for Single Trial EEG Classification, *Computers in Biology and Medicine*, **37(4)**, 499–508.
- [21] **Saito, N.**, 2002. Discriminant Feature Extraction Using Empirical Probability Density and A Local Basis Library, *Pattern Recognition*, **35**, 1842–1852.
- [22] **Coifman, R.R. and Donoho, D.**, 2002. Translation Invariant De-noising, Wavelets and Statistics, *Lecture Notes in Statistics, Springer-Verlag*, 125–150.
- [23] **Gyaourova, A., Kamath, C. and Fodor, I.**, 2002. Undecimated Wavelet Transforms for Image De-Noising, *LLNL technical report, UCRL-ID-150931*.
- [24] **Stefanou, H., Kakouros, S., Cavouras, D. and M.Wallace**, 2005. Wavelet-based Mammographic Enhancement, *Proceedings of the 5th International Network Conference (INC)*.
- [25] **Figueiredo, M. and Nowak, R.**, 2007. An EM Algorithm for Wavelet-based Image Restoration, *IEEE Transaction on Image Processing*, **12(8)**.
- [26] **Unser, M.**, 1995. Texture Classification and Segmentation Using Wavelet Frames, *IEEE Transactions on Image Processing*, **4(11)**.
- [27] **Duda, R. and Hart, P.**, 1973. Pattern Classification and Scene Analysis, *John Wiley & Sons, Inc.*
- [28] **Pearson, T., Wicklow, D., Maghirang, E., Xie, F. and Dowell, F.**, 2001. Detecting Aflatoxin in Single Corn Kernels by Using Transmittance and Reflectance Spectroscopy, *Transactions of the ASAE*, **44(5)**, 1247–1254.

- [29] **Hirano, S., Okawara, N. and Narazaki, S.**, 1998. Near Infra Red Detection of Internally Moldy Nuts, *Bioscience, Biotechnology, and Biochemistry*, **62(1)**, 102–107.
- [30] **T.W.Tayson and R.L.Clark**, 1974. An Investigation of Ilorescence Properties of The Aflatoxin -Infected Pecans, *Transactions of the ASAE*, **17**, 942–944.
- [31] **Fersaie, A., McClure, W.F. and Monroe, R.J.**, 1978. Development of Indices for Sorting Iranian Pistachio Nuts According to Fluorescence, *Journal of Food Science*, **43(5)**, 1550–1552.
- [32] **Wicklow, D.**, 1999. Influence of Aspergillus Flavus Strains on Aflatoxin and Bright Greenish Yellow Fluorescence of Corn Kernels, *The American Phytopathological Society, Plant Disease*, **83(2)**.
- [33] **Steiner, W.E., Rieker, R.H. and Battaglia, R.**, 1988. aflatoxin Contamination in Dried Figs: Distribution and Association with Fluorescence, *J. Agric. Food Chem.*, **36(1)**.
- [34] **Doster, M.A. and Michailides, T.J.**, 1998. Production of Bright Greenish Yellow Fluorescence in Figs Infected by Aspergillus Species in California Orchards, *The American Phytopathological Society, Plant disease*, **82(6)**.
- [35] **Hadavi, E.**, 2005. Several Physical Properties of Aflatoxin-Contaminated Pistachio Nuts: Application of BGY Fluorescence for Separation of Aflatoxin-Contaminated Nuts, *Food Additives and Contaminants*, **22(11)**, 1144–1153.
- [36] **Steiner, W.E., Brunschweiler, K., Leimbacher, E. and Schneider, R.**, 1992. Aflatoxins and Fluorescence in Brazil Nuts and Pistachio Nuts, *J. Agrlc. Food Chem.*, **40**, 2453–2457.
- [37] **Bothast, R.J. and Hesseltine, C.W.**, 1975. Bright Greenish-Yellow Fluorescence and Aflatoxin in Agricultural Commodities, *Applied MICROBIOLOGY*, 337–338.

- [38] **Jacks, T.**, 2005. Evaluation of Kojic Acid for Determining Heme and Non-heme Haloperoxidase Activities Spectrofluorometrically, *Analytical Letters*, **38**, 921–927.
- [39] **Pearson, T.C.**, 1987. Separating Early Split From Normal Pistachio Nuts For Removal of Nuts Contamination On The Tree With Aflatoxin, *Ms Thesis, University of California*.
- [40] **Pearson, T.C.**, 1996. Machine Vision System for Automated Detection of Stained Pistachio Nuts, *Lebensm.-Wiss. u.-Technol.*, **29**, 203–209.
- [41] **Pearson, T.C. and Schatzki, T.F.**, 1998. Machine Vision system for Automated Detection of Aflatoxin-Contaminated Pistachios, *Journal of Agric. Food Chem.*, **46**, 2248–2252.
- [42] **Kalkan, H., Yardimci, Y., Ince, F., Tewfik, A., Senyuva, H., Arici, M., Pearson, T., Cetin, E. and Onaran, I.**, 2007. Separation of Damaged Shell Hazelnut by Impact Acoustics, *XII International IUPAC Symposium on Mycotoxins and Phycotoxins, Istanbul*.
- [43] **Cetin, A., Pearson, T. and Tewfik, A.**, 2004. Classification of Closed- And Open- Shell Pistachio Nuts Using Voice-Recognition Technology, *Transaction of the ASAE*, **47(2)**, 659–664.
- [44] **Onaran, I., Dulek, B., Pearson, T., Yardimci, Y. and Cetin, E.**, 2005. Detection of Empty Hazelnuts From Fully Developed Nuts by Impact Acoustics, *13th EUSIPCO, Antalya*.
- [45] **Kalkan, H. and Yardimci, Y.**, 2006. Classification Of Hazelnut Kernels By Impact Acoustics, *IEEE MLSP, Dublin*.
- [46] **Taylor, S.K. and McCure, W.F.**, 1989. Measuring the Distribution of Chemical Components, *Proceedings of the Second International NIR Conference, Tsukuba, Japan*, 393–404.

- [47] **Zeringue, H. and Shih, B.**, 1998. Extraction and Separation of the Bright-Greenish-Yellow Fluorescent Material from Aflatoxigenic *Aspergillus* spp. Infected Cotton Lint by HPLC-UV/FL, *Agric. Food Chem.*, **46(3)**, 1071–1075.
- [48] **Park, B., Chen, Y.R. and Nguyen, M.**, 1998. Multi-spectral Image Analysis Using Neural Network Algorithm for Inspection of Poultry Carcasses, *Journal of Agricultural Engineering*, **69**, 351–363.
- [49] **Park, B., Lawrance, K.C., Windham, W.R. and Smith, D.P.**, 2002. Multi-Spectral Imaging System for Fecal and Ingesta Detection on Poultry Carcasses, *J. Food Proc. Engr.*, **27(5)**, 311–327.
- [50] **Mehl, P.M., Chen, Y.R., Kim, M.S. and Chan, D.E.**, 2004. Development of Hyperspectral Imaging Technique for the Detection of Apple Surface Defects and Contaminations, *Journal of Food Engineering*, **61**, 67–81.
- [51] **Dharwarkar, G.**, 2005. Using Temporal Evidence and Fusion of Time-Frequency Features for Brain- Computer Interfacing, *Ms Thesis, Waterloo, Ontario, Canada*.
- [52] **Saito, N. and Coifman, R.R.**, 1994. Local discriminant bases, *Mathematical Imaging: Wavelet Applications in Signal and Image Processing II Proc. of SPIE*, **2303**, 2–14.
- [53] **Wickerhauser, M.**, 1994. Adapted Wavelet Analysis from Theory to Software, *Massachusetts: A.K. Peters*.

CURRICULUM VITAE

Habil Kalkan was born in Erzincan in 1979. He graduated from Ege University, Department of Electrical and Electronics Engineering in 2002. He received his M.S. degree from METU, Informatics Institute in 2005. He has started to his PhD education in Informatics Institute in 2005 and he has worked as a visiting researcher at University of Minnesota, Department of Electrical Engineering and Kansas, USDA-ARS-GMPRC in 2006. Since 2002, he has been working as research assistant in METU. His current research interests are in the field of signal and image processing, pattern recognition and machine learning.