TIME-FREQUENCY AND TIME-SCALE ANALYSIS, DECOMPOSITION AND CLASSIFICATION OF ADVENTITIOUS PULMONARY SOUNDS

by

Sezer Ulukaya

B.S., Electrical and Electronics Engineering, Ankara University, 2008

M.S., Electrical and Electronics Engineering, Bahçeşehir University, 2011

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Electrical and Electronics Engineering

Boğaziçi University

2017

TIME-FREQUENCY AND TIME-SCALE ANALYSIS, DECOMPOSITION AND CLASSIFICATION OF ADVENTITIOUS PULMONARY SOUNDS

APPROVED BY:

Prof. Yasemin Palanduz Kahya ..... (Thesis Supervisor)

Prof. Emin Anarım .....

Prof. Nizamettin Aydın .....

Prof. Çiğdem Eroğlu Erdem .....

Prof. Murat Saraçlar .....

DATE OF APPROVAL: 27.09.2017

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere and deepest gratitude to my thesis supervisor Prof. Yasemin P. Kahya for the sustained support during my Ph.D study and related research, for her patience, motivation, and immense knowledge. Her encouragement and advices helped me to enlighten this long journey. Her guidance helped me in all the time of research, discussions and writing of this thesis. It has been an honor for me to be her Ph.D. student.

I would like to thank my thesis committee members: Prof. Emin Anarım, Prof. Nizamettin Aydın, Prof. Çiğdem Eroğlu Erdem and Prof. Murat Saraçlar for their valuable time, interest, and constructive comments.

I am very happy and lucky for being a member of BULAL and BUSIM research laboratories. I would like to thank all the members of these laboratories for their supportive and motivational friendship. I would like to thank İpek Şen, Görkem Serbes, Oya Çeliktutan Dikici, Mehmet Yamaç, Erinç Dikici, Sinem Aslan, Can Altay, Alican Gök and Güney Kayım for their friendship, supports, fruitful discussions and times.

I would like to thank my colleagues and administrative staff at Trakya University for their support during my Ph.D. studies with a special thank to the members of Electrical and Electronics Engineering Department.

Finally, my deepest thanks go to my family for their endless love and support. I am sure without their motivation and encouragement, this thesis would have never been written.

This thesis was supported by the Ph.D. scholarship (2211) from the Scientific and Technological Research Council of Turkey (TÜBİTAK) and Boğaziçi University Research Fund under grant number 16A02D2.

ABSTRACT

TIME-FREQUENCY AND TIME-SCALE ANALYSIS, DECOMPOSITION AND CLASSIFICATION OF ADVENTITIOUS PULMONARY SOUNDS

Pulmonary diseases affect the quality of life and disturb the patients throughout their life. Due to some disadvantages of auscultation with a traditional stethoscope, computerized lung sound analysis has become a necessity. In this thesis, novel non-dyadic overcomplete wavelet based methods are proposed to decompose, detect and classify primary indicators (crackle and wheeze) of pulmonary diseases using various machine learning algorithms. Crackle (explosive and discontinuous), wheeze (musical and continuous) and normal lung sounds are classified using Rational Dilation Wavelet Transform based extracted features and compared with related works. It is shown that the proposed method is more successful and faster than its competitors. Moreover, in an ensemble learning scheme it is shown that the optimal representations of signal of interest can be achieved employing the proposed method. Resonance based decomposition using Tunable Q-factor Wavelet Transform and Morphological Component Analysis techniques are proposed to decompose adventitious lung sounds and to localize crackles successfully. The proposed method is compared with related works on adventitious lung sound decomposition and is shown to perform better than other methods in terms of root mean square error, crackle localization accuracy and visual validation. Within class problem in wheeze type classification is explored using non-dyadic wavelet based features and adaptive peak energy ratio metric. It is shown that either using fixed parameter settings in wavelet transform or fixed time-frequency (TF) based features, the optimum representation and high performance can not be achieved. After repetitive experiments, it is shown that by using the proposed novel wavelet based methods, optimum and better TF and time-scale representation can be achieved.

ÖZET

SOLUNUM EKSESLERİNİN ZAMAN-SIKLIK VE ZAMAN-ÖLÇEK ANALİZİ, AYRIŞTIRILMASI VE SINIFLANDIRILMASI

Akciğer hastalıkları, yaşam kalitesini etkiler ve hastaları yaşamları boyunca rahatsız eder. Geleneksel stetoskobun bazı dezavantajlarından dolayı bilgisayarlı solunum sesi analizi ihtiyaç haline gelmiştir. Bu tezde, akciğer hastalıklarının birincil göstergelerini (çıtırtı ve üfürüm) çeşitli makina öğrenme algoritmaları ile ayrıştırmak, saptamak ve sınıflandırmak için, yeni diyadik olmayan tamamlanmış dalgacık dönüşümü tabanlı yöntemler önerilmiştir. Çıtırtı (patlayıcı ve kesikli), üfürüm (müzikal ve sürekli) ve normal akciğer sesleri, Kesirli Genişleyen Dalgacık Dönüşümü tabanlı öznitelikler kullanılarak sınıflandırılmış ve literatürde ilgili çalışmalarla karşılaştırılmıştır. Önerilen yöntemin rakiplerinden daha başarılı ve hızlı olduğu gösterilmiştir. Dahası, bir topluluk öğrenme şemasında, ilgili sinyalin eniyilenmiş gösterimlerinin önerilen yöntemi kullanarak gerçekleştirilebileceği gösterilmiştir. Uyarlanabilir Q-faktörlü Dalgacık Dönüşümü ve Morfolojik Bileşen Analizi teknikleri kullanılarak rezonansa dayalı ayrıştırma, akciğer ekseslerinin ayrıştırılması ve çıtırtıların başarıyla lokalize edilmesi için önerilmiştir. Önerilen yöntem, akciğer ekses ayrıştırma üzerine ilgili çalışmalarla karşılaştırılmış ve kök ortalama karesi hatası, çıtırtı lokalizasyon doğruluğu ve görsel doğrulama açısından diğer yöntemlerden daha iyi performans göstermiştir. Üfürüm tipi sınıflandırmadaki sınıfiçi problem, diyadik olmayan dalgacık tabanlı özellikler ve uyarlamalı tepe enerji oranı metriği kullanılarak araştırılmıştır. Dalgacık dönüşümünde sabit parametreli kurulum veya sabit zaman-frekans (ZF) tabanlı özniteliklerin kullanılması durumunda, eniyilenmiş gösterim ve yüksek performans elde edilemediği gösterilmiştir. Yapılan yoğun deneyler sonucunda, önerilen yeni dalgacık tabanlı yöntemler kullanılarak, eniyilenmiş ve daha iyi ZF ve zaman ölçekli gösterim elde edilebileceği gösterildi.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a	Scale parameter
$\mathbf{A}$	Unknown random mixing matrix
a_{ij}	Random mixing coefficients
b	Shifting parameter
e	Energy
f_c	Center frequency
$g_0(n)$	High pass filter
$G_0(\omega)$	Frequency response of high pass filter
$g(x)$	Activation function
$h_0(n)$	Low pass filter
$H_0(\omega)$	Frequency response of low pass filter
$I(N)$	Impurity
i_x	Class label of the test sample
J	Number of levels
K	Number of samples in a time segment
$lcm(q, s)$	Least common multiple of q and s
L_p	Lagrangian term
n	Background vesicular (residual) sound
p	Positive integer which sets the Q-factor and redundancy
$p(c\|x)$	Posterior probability of class (c) given the test sample x
$P_{psd}(f)$	Periodogram, power spectral density
$P(w)$	Power spectrum of data segment
q	Positive integer which sets the Q-factor and redundancy
r	Over-sampling rate
$Red(p, q, s)$	Redundancy of the wavelet transform
s	Positive integer which sets the Q-factor and redundancy
Se	Shannon entropy
$S(\tau, f)$	Continuous version of Stockwell transform

$U(w_1, w_2)$	Objective function U
$X(f)$	Fourier transform of $x(k)$
$x(k)$	Data samples in a time segment
x_1	Oscillatory waveform
x_2	Transient waveform
$\mathbf{W}$	Inverse of matrix $\mathbf{A}$
w_J	Class J
$w[n]$	Hamming window
λ_i	Parameter which sets the energy of resonance terms
μ	Mean
$\phi(t)$	Scaling function
$\mathbf{\Psi_i}$	Transformation matrix
$\psi(t)$	Wavelet function
σ	Standard deviation
σ_{res}	Residual standard deviation
$\theta(\omega)$	Transition function

LIST OF ACRONYMS/ABBREVIATIONS

ALS	Adventitious Lung Sounds
AWGN	Additive White Gaussian Noise
CALS	Continuous Adventitious Lung Sounds
COPD	Chronic Obstructive Pulmonary Disease
CWT	Continuous Wavelet Transform
DALS	Discontinuous Adventitious Lung Sounds
DT	Decision Tree
DWT	Discrete Wavelet Transform
ELM	Extreme Learning Machines
EMD	Empirical Mode Decomposition
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
FT	Fourier Transform
IC	Independent Component
ICA	Independent Component Analysis
IMF	Intrinsic Mode Function
KNN	K Nearest Neighbor
LOOCV	Leave-One-Out Cross Validation
LS	Lung Sounds
MAD	Median Absolute Deviation
MCA	Morphological Component Analysis
MFCC	Mel Frequency Cepstral Coefficients
MP	Monophonic
NB	Naive Bayes
PER	Peak Energy Ratio
PLP	Perceptual Linear Prediction
PP	Polyphonic

PSD	Power Spectral Density
RADWT	Rational Dilation Wavelet Transform
RMSE	Root Mean Square Error
S Transform	Stockwell Transform
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
TF	Time Frequency
TP	True Positive
TQWT	Tunable Q-factor Wavelet Transform
WT	Wavelet Transform

1. INTRODUCTION

1.1. Aim, Background and Motivation

Asthma, bronchiectasis, fibrosing alveolitis, pneumonia, chronic obstructive pulmonary disease (COPD) and interstitial fibrosis are some of the pulmonary diseases that degrade the quality of life. COPD, which is one of the most critical pulmonary diseases, was ranked as the sixth mortality cause in 1990 and will become the third in 2020 [1]. Moreover, according to [2], asthma or COPD affect 1 in 12 people around the world and these two lung diseases may overlap on 15 % of the obstructive lung disease population. In literature normal respiratory sounds are named as vesicular sounds. On the other hand, adventitious sounds, which are the primary indicators of lung dysfunctions, are superimposed on vesicular sounds. The adventitious sounds are divided into two types, namely as continuous (wheeze) and discontinuous (crackle) pulmonary sounds. The number of crackles per breath is related to the severity of the disease, and the timing, duration and types of crackles (fine or coarse) in a breath cycle may be different in various lung diseases. For example, coarse crackles exist in bronchiectasis whereas fine crackles are common symptoms of interstitial fibrosis and pneumonia. The presence of wheezes usually indicates a pulmonary disorder such as asthma and COPD. Wheeze characteristics such as pitch frequency and duration are related to the degree of airway obstruction. In low and middle income countries access to modern health services is limited and expensive. Moreover, lack of medical experts is another issue in these regions and because of these issues cost effective computerized lung acoustics systems are needed. Therefore, in this thesis the first aim is to propose novel and faster wavelet based methods which may be employed in real-time cost effective computerized lung acoustics systems. As discussed above, in order to automatically analyse pulmonary diseases using computerized systems, proper detection of crackles and wheezes is very important. Therefore, as a second aim, in this thesis novel wavelet based methods are used to extract discriminative features and classify them with high accuracy as compared with related studies. Proper detection of crackles and wheezes can be seen as focusing on the lung sound data at a micro level. The features

(hidden atomic information in lung sound) extracted from micro level may be added to model based features extracted at macro level for improving the accuracy of the diagnostic classification systems. Stemmed from time-varying characteristics of lung tissue and chest wall, most lung sound signals are non-stationary in character, independent from time scale and time frequency analysis domain. This is valid for adventitious lung sound types, especially crackles. Most of the time, adventitious lung sounds have high-frequency components very tight in time domain, low-frequency components very tight in frequency domain which are superimposed on low frequency vesicular sounds. Hence a suitable analysis method for detecting them should supply information about good frequency resolution along with good time resolution, the first to localize the low frequency entities, and the second to resolve the high frequency entities. Therefore classical Fourier transform (FT), which assumes that the analysed signal is stationary and does not contain any time information, is not appropriate to analyse most of the biomedical signals. Short Time FT (or windowed FT) partially overcomes the drawback of FT by considering an analysis window that has fixed time-frequency resolution. However, presenting a time-scale description of signals, wavelet transform (WT) has finer frequency resolution at low frequencies, but also has finer time resolution at high frequencies. Moreover, in all previous discrete wavelet transform (DWT) based methods, constant low Q-factor wavelets, which have limited frequency resolution, have been used. These types of low Q-factor wavelets are adequate in the analysis of piecewise smooth signals but for more oscillatory signals like wheeze signals, a DWT with better frequency resolution is needed. Therefore in this thesis, unlike the previous studies, novel non-dyadic overcomplete wavelet transform in which the Q-factor of the analysis and synthesis filters can be adjusted according to the properties of signal of interest, is proposed as the feature extractor.

Previous works focus on stationary-nonstationary separation of lung sounds. However, wheezes and vesicular sounds are both stationary compared to crackles. DWT with low Q-factor has limited frequency resolution and is incapable of modeling wheezes. Wheezes are oscillatory (high Q-factor), crackles are transient (low Q-factor) waveforms and both crackles and wheezes may co-exist in the same lung sound data. The importance of detection and separation of adventitious lung sounds

in the discrimination of healthy and pathological subjects, is the motivation behind the study presented in this thesis since there is also an overlap of frequency ranges of vesicular (normal) and adventitious lung sound types. Therefore, unlike linear or frequency based filtering, resonance based nonlinear decomposition is proposed to separate adventitious lung sounds. Moreover, previous works suffer from being unable to decompose low and high frequency components of the same lung sound type into the same channel and deforming the waveform of the crackles whose time domain parameters are vital in computerized diagnostic classification systems.

In Chapter 2, unlike the previous studies which are based on FT or dyadic DWT and have limited TF resolution, a non-dyadic overcomplete WT, in which the Q-factor of the analysis and synthesis filters can be adjusted according to the properties of signal of interest, is proposed as the feature extractor.

In Chapter 3, resonance based decomposition of lung sounds that aims to separate wheeze, crackle and vesicular sounds into three individual channels while automatically localizing crackles for both synthetic and real data is proposed.

In Chapter 4, unlike previous studies which use fixed TF resolution based on Fourier transform, we propose an optimal (better TF resolution) and adaptive (automatic and tunable) wavelet based technique to discriminate monophonic (MP) and polyphonic (PP) wheezes (primary indicators of lung diseases) in a more robust and objective manner.

Moreover, the proposed non-dyadic overcomplete WT, being a rational (based on non-dyadic dilations), fully discrete, near shift-invariant and invertible transform with acceptable redundancy and higher accuracy may be a robust candidate to be employed in real-time diagnostic classification systems.

1.2. Outcomes and Organization of the Thesis

Outcomes of the studies have been published in [3–10].

In Chapter 1, motivation, aim and structure of the thesis are presented. Outcomes and objectives of the research are also introduced.

In Chapter 2, an overcomplete non-dyadic wavelet based method is proposed to discriminate pulmonary sounds using various machine learning techniques. Crackle, wheeze and normal lung sound discrimination is vital in diagnosing pulmonary diseases. Previous works suffer from limited frequency resolution and lack of ability to deal with oscillatory signals (wheezes). The main objective of this chapter is to propose a novel wavelet based lung sound classification system that is capable of adaptively representing crackle, wheeze and normal lung sound signal time-frequency properties. A method which is based on Rational Dilation Wavelet Transform (RADWT) is proposed to classify lung sounds into three main categories, namely, normal, wheeze and crackle. Six different feature extraction methods are used with five different classifiers all of which are compared with the proposed method on 600 lung sound episodes in a cross validation scheme. Six statistical subset features are extracted from raw features and fed into classifiers. After comparative evaluation of the proposed method, an ensemble learning scheme is built to increase the performance of the proposed method. It is shown that performance of the proposed method is superior to previous methods in terms of accuracy. Moreover, its computational time is far less than its nearest competitor (S transform). It is shown that the proposed method is able to cope with oscillatory type signals as well as transient sounds performing 95.17 % average accuracy for energy subset and 97.38 % ensemble average accuracy showing a promising time-frequency tool for biological signals. The proposed method performs better even using only one subset of extracted features. It provides better time-frequency resolution for all types of signals of interest and is less redundant than continuous wavelet transform (CWT) and significantly faster than its nearest competitor.

In Chapter 3, resonance based decomposition of lung sounds that aims to separate wheeze, crackle and vesicular sounds into three individual channels while automatically localizing crackles for both synthetic and real data is presented. Previous works focus on stationary-non stationary discrimination to separate crackles and vesicular sounds disregarding wheezes which are stationary compared to crackles. However, wheeze

sounds include important cues about the underlying pathology. Using two different threshold methods and synthetic sound generation scenarios in the presence of wheezes, resonance based decomposition performs 89.5 % crackle localization recall rate for white Gaussian noise and 98.6 % crackle localization recall rate for healthy vesicular sound treated as noise at low signal-to-noise ratios. Besides, an adaptive threshold determination, which is independent from the channel at which it is applied, is used and is found to be robust to noise. The proposed method is compared with Independent Component Analysis (ICA) and Empirical Mode Decomposition (EMD) methods in terms of normalized Root Mean Square Error (RMSE), crackle localization accuracy and visual validation performance and found to be successful on all of the metrics using synthetic and real patients' data including overlapping wheezes and crackles. The proposed method is also experimented on merely wheeze, merely crackle and merely vesicular sounds containing cases to represent the decomposition ability of the proposed method. Moreover, an energy based classification metric is proposed using candidate decomposed channel energies to estimate the label of the experimented case.

In Chapter 4, adaptive non-dyadic wavelet and peak energy ratio based automatic method is proposed to discriminate wheeze types. The previous works mainly focused on using fixed time-frequency resolution based on Fourier Transform in classifying wheeze types. The main objective of this chapter is to discriminate wheeze types in an adaptive and optimal way by showing effectiveness of the proposed system over fixed parameter valued settings. An adaptive and automatic Rational Dilation Wavelet Transform (RADWT) based peak energy ratio (PER) parameter selection method is proposed as feature extractor. Distribution of PER values and accuracies of different classifier kernels are provided for validation. It is shown that wheeze types can not be represented in time-frequency domain optimally by using fixed scale and shifting wavelet parameters. Extracted best representing PER values obtained from decomposed energy sub-bands fed into support vector machine classifier and proposed adaptive wavelet based method outperformed fixed parameter sets achieving 86 % accuracy. It is concluded that, without using proposed adaptive wavelet based method, best time-frequency representation of wheezes can not be achieved using a specific (fixed) parameter set.

In Chapter 5, a summary of the proposed contributions and conclusions of the thesis is represented. Future directions of the proposed contributions are also introduced in this chapter of the thesis.

2. RESPIRATORY SOUND CLASSIFICATION

2.1. Introduction

Stethoscope is a traditional tool in diagnosing respiratory dysfunctions and disorders. However, it is regarded to have low diagnostic value due to its limited frequency response which attenuates frequencies greater than 120 Hz and due to the subjectivity [11] involved in the evaluation of the auscultated sounds. Moreover, the traditional stethoscope offers no option to record pulmonary data for further analysis. Interdisciplinary efforts in medicine and engineering as summarized in [12], which aim to make auscultation a more valuable diagnosis tool, use advanced machine learning and signal processing algorithms to be utilized in treatment follow-ups and remote diagnosis.

Lung sounds (LS) are believed to be produced by the turbulent flow in the lung airways, and are essentially classified as adventitious (abnormal) sounds and vesicular (normal) sounds [11, 13]. The normal breath sounds heard over the chest wall, which are synchronous with air flow in the airways, are defined as vesicular sounds. The frequency spectrum of vesicular sounds in healthy people has 200-600 Hz dominant frequency range. Adventitious lung sounds (ALS), which are specific markers of various respiratory diseases, are superimposed on vesicular sounds. Discontinuous adventitious lung sounds (DALS) and continuous adventitious lung sounds (CALS) are the two main categories of ALS. In literature, crackles and wheezes can be exemplified as the most studied components of the CALS and DALS, respectively [14, 15].

Crackles are non-musical instantaneous bursts, explosive in nature and divided as coarse (lower pitch) or fine (higher pitch). Crackles are believed to be generated by abnormally closed airway openings [16]. The frequency spectrum of crackles varies between 200 and 2000 Hz range while their duration is usually less than 20 ms. In most of the pulmonary diseases, the severity of the disease has a strong correlation with the number of crackles per breath. Moreover, different lung diseases can be diagnosed by using the occurrence times, durations and types of crackles within a

breath cycle [16]. For example, in bronchopneumonia and bronchiectasis coarse crackles exist typically whereas the common symptoms of pneumonia and interstitial fibrosis are fine crackles [17]. Crackles in chronic obstructive pulmonary disease are coarse and occur in early to mid inspiration whereas crackles in fibrosing alveolitis are fine and occur in late inspiration [16]. Due to their transient waveform and scattered energy content over frequencies (200-2000 Hz), crackles have barely noticeable impact on the total power spectrum [17]. The frequency and time domain characteristics of vesicular and crackle sounds overlap in both domains.

Wheezes are oscillatory waveforms, which last more than 80-250 ms, and represent narrow-horizontal lines in the time-frequency domain ($>$ 100 Hz). Time-frequency representation of two samples of crackles and wheezes is illustrated in Figure 2.1. Various pulmonary diseases such as asthma and chronic obstructive pulmonary disease, can be diagnosed by using the presence of wheezes. As given in [18] and [19], the degree of airway obstruction may be related to wheeze properties such as duration and pitch frequency.

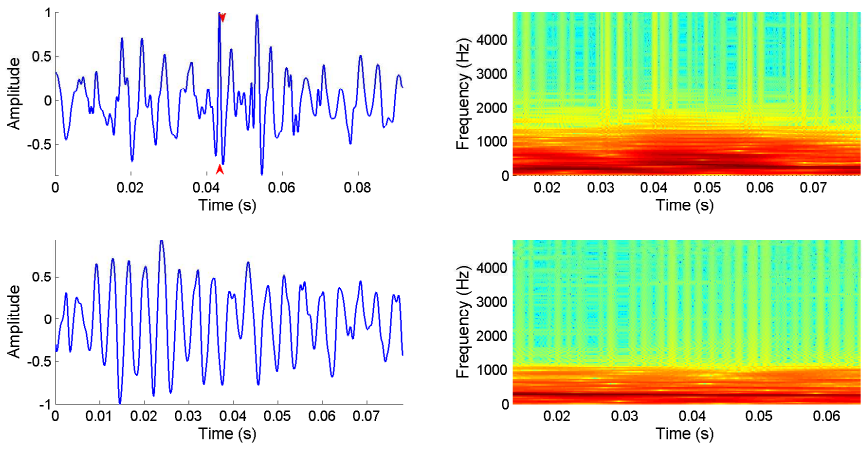

Figure 2.1: Time domain (left) and time-frequency domain (right) representations of crackle (upper) and wheeze (lower) sounds. Red arrow indicates crackle location.

The motivation of the proposed LS, CALS and DALS discrimination system is to overcome the subjectivity and low-performance drawbacks of traditional non-automatic systems while increasing the performance of automated systems by enhancing the frequency selectivity of wavelet filters resulting in a better separation of overlapped com-

ponents of these three lung sounds in time-frequency domain.

In previous studies, crackle/non-crackle and wheeze/non-wheeze episode discrimination has been extensively examined and a summary can be found in [14, 20]. In addition to binary discrimination approaches, there were also a few three class (crackle, wheeze and normal classes) discrimination studies such as [21–26]. In [21], Discrete Wavelet Transform (DWT) coefficients and Artificial Neural Networks (ANN) based classification system were used to solve a six-class (squawk, stridor and rhonchus in addition to crackle, normal and wheeze classes) problem on 265 episodes using mean and standard deviation based statistical features. 100 % and 94.02 % accuracies were achieved for the training and validation sets, respectively. However, when the episode number was increased from 422 to 5786, the general accuracy decreased to 59.15 % on the validation set. In [22], the power spectral density (PSD) features, obtained from three lung sound types, were fed into Genetic Algorithm for feature selection. The selected features were forwarded to Multilayer Perceptron neural network and 91.7 % average accuracy was achieved when 96 subjects were used. In [23], the data was modelled with maximum-likelihood approach and, Hidden Markov Model (HMM) based classification was employed resulting in an average classification rate of 83 % on 1544 episodes. In [24], using Multilayer Perceptron neural network on 20 test epochs, confidence levels of 90 %, 87 % and 89 % were obtained for normal, wheeze and crackle classes, respectively. In [25], a combination of Mel Frequency Cepstral Coefficients (MFCC) and Gaussian Mixture Model (GMM) were used for classification resulting in 98.75 % and 52.5 % accuracies for the reference (50 epochs) and cross-validation (24 epochs) set, respectively. Using 225 samples and S-transform (which is a phase corrected version of continuous wavelet transform (CWT) with a scalable Gaussian window) based statistical features (mean and standard deviation), the study in [26] reached overall classification accuracy of 94.99 % with k Nearest Neighbor (k-NN), 96.85 % with support vector machine (SVM) and 98.52 % with extreme learning machines (ELM) classifiers. A summary of the related works in literature is shown in Table 2.1.

Table 2.1: An overview of the methods related with three groups of lung sound classification problem

Approach	**Features extracted**	**Dataset**	**Validation**	**Model**	**Classifier**	**Accuracy (%)**
Kandaswamy *et. al.*, 2004 [21]	statistical	265 segments	5-fold CV	DWT coefficients	ANN	94.02
Güler *et. al.*, 2005 [22]	GA based PSD	48 training, 48 testing	2-fold CV	Welch spectrum	MLP-NN	91.70
Matsunaga *et. al.*, 2009 [23]	MFCC	1544 segments	leave-one-out	HMM	ML	83
Abbas *et. al.*, 2010 [24]	PSD	279 training, 60 testing	Unseen	Fourier Transform	MLP-NN	88.67
Mayorga *et. al.*, 2010 [25]	MFCC	50 training, 24 testing	4-fold CV	GMM	Bayes rule	52.50
Palaniappan *et. al.*, 2015 [26]	statistical	225 segments, 48 subjects	60 % train, 40 % test	S-transform	knn, svm, elm	98.52

Feature extraction in a classifier is one of the most critical steps since the features' discrimination ability is more important than their number. The works [22] and [24] used Fourier Transform (FT) based power spectral density features for classification. Moreover, the study in [22] used an extra feature selection step based on Genetic Algorithm. These methods, however, lacked localizing waveform features in time domain. In [23, 25] MFCC features were extracted where these features were designed to represent human auditory perception and used frequently in speech recognition. However, as depicted in Table 2.1 the lowest accuracies were obtained using MFCC features, the reason being that lung sound characteristics need not be person specific as human voice [27]. The works of [21] and [26] used wavelet transform based features where wavelet based features had better time-frequency resolution than the others. However, the wavelets employed in [21] had low Q-factors and this resulted in poor frequency selectivity, which was a dramatic drawback in the modelling of wheeze signals. In [21] a shrinkage denoising technique was additionally used with the performance decreasing severely with increasing number of episodes. In [26], best results were reported in literature; however, database was relatively small and all the episodes had the same, fixed length while crackle and wheeze sounds had different durations in real cases.

In literature, wavelet transform based methods have been successfully employed for feature-extraction and/or de-noising in the pulmonary sounds [17, 21, 28]. The meaning of the Q-factor in wavelet terminology is the ratio of bandpass filter's center frequency to its bandwidth. In the previous DWT based systems, mostly, constant low Q-factor wavelets having poor frequency-resolution were employed and satisfactory results were obtained in the analysis of piecewise smooth signals. On the other hand, a DWT having better and controllable frequency resolution must be utilized in the analysis of oscillatory signals like wheezes to achieve optimum time-frequency representation of signal of interest. Hence in this thesis, unlike previous pulmonary signal processing systems, Rational Dilation Wavelet Transform (RADWT) [29], whose analysis and synthesis filters' Q-factors can be tuned with respect to the signal of interest, was proposed for feature extraction. Shannon entropy, standard deviation, energy, maximum/minimum, mean and skewness/kurtosis values of each decomposed sub-band

were calculated as statistical feature-subsets. Later, these statistical feature-subsets were given into Decision Tree (DT), Naive Bayes (NB), k-NN, SVM, and ELM models with the final aim of classifying wheeze, crackle and normal lung sounds. Experimental results showed that with the high Q-factor analysis (proposed method hereafter) that was implemented as leaving one lung sound sample out for each training and testing phase (also called as leave-one-out cross validation - LOOCV), higher average accuracy, normal signal, wheeze and crackle classification accuracies were obtained when compared with the low Q-factor wavelet analysis and the other previously suggested methods in literature.

In this chapter, all indicated previous approaches were experimented on our data set and compared extensively with our proposed method. The flowchart for all the experimented feature extraction and classification methods is given in Figure 2.2.

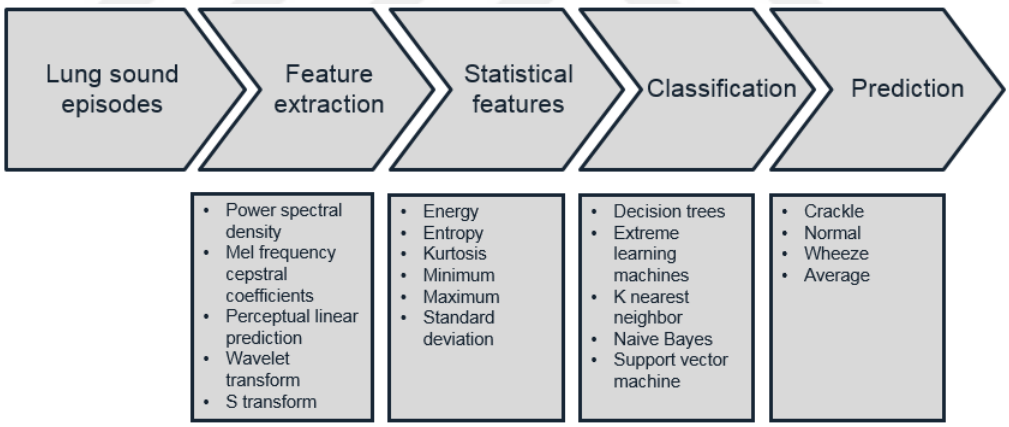

Figure 2.2: The flowchart of the comparative evaluation of the proposed method with literature.

This chapter is organized as follows; Section 2.2 introduces data acquisition system. Feature extraction methods, methodology and classification methods are given in Sections 2.3-2.6. Section 2.7 and 2.8 introduce ensemble learning and detailed analysis of proposed system. Section 2.9 represents the results and finally, Section 2.10 presents discussion and summary.

2.2. Data Acquisition System and Dataset

The lung sounds used in this thesis were recorded with a 14-channel data acquisition device which was designed in Boğaziçi University Lung Acoustics Laboratory (BU-LAL) and a comprehensive description of the device may be found in [30]. The data acquisition device was comprised of 14 air-coupled electret microphones (SONY ECM-44 BPT) that were located on the posterior chest wall. Microphone locations [30] were determined by a pulmonary physician by mirroring around spine line, with seven microphones covering each lung area. A pre-amplifier filter unit with a passband of 80-4000 Hz was employed for attenuating heart sounds and friction noise. A laptop computer was used to store and visualize the data with a 12-bit data acquisition card (NI DAQCard-6024E). A pneumotachograph (Validyne CD379) was used to measure airflow to synchronize with respiratory cycle. The sampling rate was 9600 Hz and each recording period lasted for 15 seconds. All subjects gave an informed consent before recording. The Istanbul Yedikule Teaching and Research Hospital, Chest Diseases and Thoracic Surgery Department was the collaborating partner and the CALS and DALS were taken from patients who were under treatment in the hospital. The data acquisition procedure had the approval of the second Ethical Committee on Clinical Research of Istanbul (in compliance with the Declaration of Helsinki).

An expert labelled the crackle and wheeze segments both by visual inspection of the time expanded waveforms and by auditory verification. A total of 200 crackle segments, consisting of a wide spectrum of crackle frequencies from fine to coarse, were used in the dataset. A total of 200 wheeze segments, of which 110 being monophonic and the rest being polyphonic, were used in the dataset. The remaining 200 normal segments were recorded from healthy subjects forming a total dataset of 600 segments. The dataset used in this chapter of thesis was obtained from 19 male and 11 female subjects. Healthy subjects had no pulmonary dysfunction history, were at the age of 27 ± 7 and were non-smokers. Wheeze segments were taken from four asthma and three chronic obstructive pulmonary disease patients (two of them have Chronic Heart Failure) who were at the age of 50 ± 17 complaining about shortness of breath and/or cough. Crackle segments were taken from eight Bronchiectasis and five Interstitial

Lung Disease patients who were at the age of 65 ± 9 complaining about pain in the lungs and cough or shortness of breath.

2.3. Feature Extraction Methods

2.3.1. Proposed Rational Dilation Wavelet Transform (RADWT)

Wavelet transform (WT) is a signal analysis/synthesis method, which converts a time-domain signal into a time-scale representation consisting of wavelet coefficients. In the continuous wavelet transform (CWT), due to the effect of scaling and shifting operations, a continuous two-dimensional (time-scale) representation of the signal of interest is achieved instead of the classical one-dimensional frequency domain representations, such as the Fourier Transform. Non-stationary biomedical signals (for example wheeze lung sounds [31]) may be successfully processed with continuous WT due to its two dimensional time-scale representation advantage which provides time information. However, performing the CWT on a signal leads to redundant information because of the continuous change of scale and shifting parameters. This redundancy results in high computational complexity, which makes the real-time applications very difficult, and decreases the performance of feature extraction and machine learning phases of biomedical systems. Discrete scale and shifting parameters may be employed in order to achieve the computational efficiency and easy invertibility. Mostly dyadic discrete wavelet transform (DWT), which utilizes dyadic (powers of two) scaling and shifting parameters, is used to avoid sacrificing information contained in the signal. DWT which is used in [21] on lung sound classification problem, is a low-Q factor method and therefore has some drawbacks. At each DWT decomposition level, the output of filters are down-sampled in order to avoid redundancy. However, down-sampling operation causes aliasing and shift-variance which causes undesirable sensitivity to the phase-shifts occurring in the input signal. Additionally, dyadic nature of the DWT limits the frequency resolution at higher frequencies resulting in limited performance for processing oscillatory signals like wheezes. In order to highlight the non-oscillatory (crackles) and oscillatory (wheezes) properties of lung signals, a discrete overcomplete wavelet transform, in which the frequency selectivity of the sub-bands can be set with

respect to the characteristics of signal of interest, is needed. Therefore in this thesis, the RADWT ([29, 32]), which has finer and adjustable frequency resolution with acceptable redundancy, was proposed as a suitable feature extractor for processing lung sounds.

In literature, many overcomplete WTs such as the double-density WT and the dual-tree complex WT (CWT), are used in signal processing applications [33–35]. These WTs reach overcompleteness by increasing only the number of samples taken in time for some or all frequency bands, and the sampling rate stays the same resulting in insufficient frequency resolution. Additionally, in these WTs, the redundancy factor is fixed and in some applications this rate of over-completeness would be unnecessary. However, the RADWT can attain over-completeness by increasing sampling in both time and frequency which enables obtaining the optimum time-scale representation with controllable redundancy factors. Unlike most discrete WTs that use FIR-based orthonormal wavelet bases [36], the RADWT is based on a frequency-domain (FFT based) design which does not employ rational transfer functions and offers greater design flexibility. Moreover, the RADWT is a rational (based on non-dyadic dilations), fully discrete, near shift-invariant and invertible transform. The non-dyadic (rational) behaviour of the RADWT yields a range of Q-factors and redundancy factors.

By using the RADWT, which is also performed through an iterated two-channel filter-bank structure like the DWT, desired dilation rates can be achieved with controllable redundancies. In the RADWT, the Q-factor of wavelets is built upon three positive integers p, q and s satisfying $1 \leq p < q$ and $p/q + 1/s \geq 1$, where p and q are co-prime. In RADWT, the a (scale) and b (shifting) parameter set $\{a, b\}$ are allowed to take values from $\{q^j/p^j, spn/q\}_{j,n\in\mathbb{Z}}$, which are controlled by p, q and s [32].

In RADWT, the relation between the scaling($\phi(t)$)/wavelet($\psi(t)$) functions and the low($h_0(n)$)/high($g_0(n)$) pass filters are given as,

$$\phi(t) = (q/p)^{1/2} \sum_{n\in\mathbb{Z}} h_0(n)\phi\left(\frac{q}{p}t - n\right) \tag{2.1}$$

and

$$\psi(t) = (q/p)^{1/2} \sum_{n\in\mathbb{Z}} g_0(n)\phi\left(\frac{q}{p}t - n\right) \tag{2.2}$$

where $h_0(n)$ and $g_0(n)$ denote the low-pass and high-pass filters, respectively.

Mathematically, the frequency responses of $h_0(n)$ ($H_0(\omega)$) and $g_0(n)$ ($G_0(\omega)$) are given as,

$$H_0(\omega) = \begin{cases} \sqrt{pq} & \omega \in [0, (1-\frac{1}{s})\frac{\pi}{q}] \\ \sqrt{pq}\theta(\frac{\omega-a}{b}) & \omega \in [(1-\frac{1}{s})\frac{\pi}{q}, \frac{\pi}{q}] \\ 0 & \omega \in [\frac{\pi}{q}, \pi] \end{cases} \tag{2.3}$$

and

$$G_0(\omega) = \begin{cases} 0 & \omega \in [0, (1-\frac{1}{s})\pi] \\ \sqrt{s}\theta_c(\frac{\omega-pa}{pb}) & \omega \in [(1-\frac{1}{s})\frac{\pi}{q}, \frac{p}{q}\pi] \\ \sqrt{s} & \omega \in [\frac{p}{q}\pi, \pi] \end{cases} \tag{2.4}$$

where

$$a = \left(1-\frac{1}{s}\right)\frac{\pi}{p}, b = \frac{1}{q} - \left(1-\frac{1}{s}\right)\frac{1}{p} \tag{2.5}$$

the transition function $\theta(\omega)$ is,

$$\theta(\omega) = \frac{1}{2}(1+cos(\omega))\sqrt{2-cos(\omega)} \quad \text{for} \quad \omega \in [0,\pi] \tag{2.6}$$

and $\theta_c(\omega)$ is

$$\theta_c(\omega) := \sqrt{1 - \theta^2(\omega)} \tag{2.7}$$

The transition function, $\theta(\omega)$, which is used to construct the transition bands of $G_0(\omega)$ and $H_0(\omega)$, originates from Daubechies' orthonormal wavelet filters with two vanishing moments. The reason behind using wavelet filters having small number of vanishing moments is to obtain a transform with good time-frequency localization properties.

FFT based circular convolution for low-pass and high-pass filtering are used in the implementation of the RADWT. The RADWT provides perfect reconstruction for discrete signals of any length when the length of the signal at each level is a multiple of the least common multiple of q and s (denoted as lcm(q, s)). Otherwise, circular convolution filtering operation does not support the perfect reconstruction property. In case the signal length does not satisfy this property, zero-padding operation is applied to the input signal in order to obtain the next multiple of lcm(q, s).

The redundancy of the RADWT is found as follows when the iterated filter-bank (number of levels go to infinity) is considered,

$$\mathrm{Red}(p, q, s) = \lim_{j \to \infty} \mathrm{Red}_j(p, q, s) = \frac{1}{s} \frac{1}{1 - p/q} \tag{2.8}$$

2.4. Related Feature Extraction Methods in Literature

In this section, relevant approaches in literature which were experimented on our dataset and compared extensively with the proposed method using the reported best parameters in related works, will be summarized.

2.4.1. Non-parametric Power Spectral Density (PSD)

The work of [22] and [24] used PSD based features for lung sound classification problem and in this thesis the reported best parameters (which are detailed in the last paragraph) were used to make a comparison with the proposed method. $x(k)$ being data samples in a time segment, PSD, namely periodogram ($P_{psd}(f)$), is computed as follows;

$$P_{psd}(f) = \frac{1}{K}\left|\sum_{k=0}^{K-1} x(k)e^{-j2\pi fk}\right|^2 = \frac{1}{K}\left|X(f)\right|^2 \tag{2.9}$$

where Fourier transform (FT) of $x(k)$ is denoted by $X(f)$ and K is the number of samples in a time segment. Welch method is used as a non-parametric power spectrum estimation method which allows data segment overlapping and windowing prior to computation of periodogram and is based on the idea of averaging modified periodograms [37].

Respiratory sounds were windowed with length 256 of Hanning window and 50% overlapping ratio was applied as stated in [22]. 256 point FFT was employed in Welch method and then logarithm of the 129 point-spectrum was fed into classifiers either directly or by calculating Shannon entropy, standard deviation, energy, minimum/maximum and skewness/kurtosis values of each log-spectrum.

2.4.2. Perceptual Linear Prediction (PLP)

The PLP method was first introduced to be used in speaker-independent automatic speech recognition [38]. We proposed to use this method in lung sound acoustics to exploit person independent discriminatory spectral characteristics of the diseases [10]. The method warps the spectra to minimize personal differences while holding the key features [39]. The feature set was extracted as follows: Input signal ($x[n]$)

was transformed into frequency domain using 256 point FFT;

$$X(w) = \sum_{n=0}^{N-1} x[n]w[n]e^{-j2\pi\frac{k}{N}n} \tag{2.10}$$

where $k = 0, 1, \cdots, N-1$ and $x[n]$ is the lung sound segment. Lung sounds were processed with 256 point Hamming window ($w[n]$) with 50% overlap. Since the sampling frequency was 9600 Hz, 50% overlap guaranteed 10 ms shift to encode sudden adventitious sound changes (i.e crackles). Frequency spectrum in Hz was transformed into Bark scale by combining FFT bins into Bark bins, thus critical frequency grouping was performed. After inverse FFT was computed, autoregressive parameters were estimated by n^{th} order all pole model solving Yule-Walker equations. In a recent work [40], PLP method was used with the order of 13 while in our experiments the orders of 3 to 13 were tested.

2.4.3. Mel Frequency Cepstral Coefficients (MFCC)

MFCC was proposed in [41] and used for automatic speech and speaker recognition in literature. The idea behind the MFCC was to mimic human auditory system using non-linear frequency scale called mel scale. Input lung sound ($x[n]$) was transformed into frequency domain using the same formula given in Equation 2.10. The power spectrum of each segment ($P(w)$) was calculated using $P(w) = |X(w)|^2$. Mel filter-bank is composed of M triangular filters (mel windows) which are converted from Hz to mel scale using $f_{mel}(f) = 1125ln\,(1 + f/700)$. Mel filters are placed denser in low frequency regions than high frequency regions. The logarithm of the energy in each mel filter is computed by applying mel filter-bank to the power spectrum $P(w)$. Finally, discrete cosine transform is applied to M filter outputs to compute MFCC [41]. MFCC does not provide temporal information since it is based on FT. The related works in literature with MFCC method are [23, 25, 40] which used the orders of 5, 13 and 20 for MFCC coefficients, respectively. In this thesis, the orders of 3 to 13 and 20 were experimented for MFCC to compare this method with the proposed method.

2.4.4. Stockwell Transform (S Transform)

S transform was proposed in [42] which aimed to exploit the advantages of the continuous wavelet transform (CWT) by combining FT features. In S transform, by dilating and translating a scalable Gaussian window, finer time and frequency resolution is achieved. S transform is a phase corrected version of CWT by multiplying mother wavelet with a phase factor and gives Fourier spectrum by averaging over time. S transform was proposed in [26] to classify lung sounds achieving favourable results. Given $x(t)$ lung sound episode, continuous version of S transform is written as

$$S(\tau, f) = \int_{-\infty}^{+\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau-t)^2 f^2}{2}} e^{-2\pi f i t} dt \tag{2.11}$$

where the terms other than $x(t)$ in the integral form the mother wavelet. The output of the S transform in discrete form is a complex matrix and needs to be handled by taking the absolute value. The drawback of this algorithm in our experiments was that since the lung sounds may have variable duration, the extracted statistical features (min., max., etc.) from output matrix also resulted in variable length features. This condition violated the classifier input formation and forced us to do experiments for fixed length (1024 points for each crackle, wheeze and normal samples) lung sounds to compare with other methods. For wheeze and normal classes 1024-point-segments were extracted randomly, for crackle class 1024-point-segments were extracted by centering the crackle location without removing successive crackles as in the original data.

2.5. Calculation of Statistical Features

In order to feed into classifiers, Shannon entropy, standard deviation, energy, maximum/minimum, mean and skewness/kurtosis values of each vector or matrix obtained from feature extraction methods were calculated as follows;

$$Mean(\mu) = \frac{1}{N}\sum_{1}^{N} x_i \tag{2.12}$$

$$Standard\ deviation(\sigma) = \left(\frac{1}{N-1}\sum_{1}^{N}(x_i - \mu)^2\right)^{\frac{1}{2}} \tag{2.13}$$

$$Skewness(s) = \frac{E(x_i - \mu)^3}{\sigma^3} \tag{2.14}$$

$$Kurtosis(k) = \frac{E(x_i - \mu)^4}{\sigma^4} \tag{2.15}$$

$$Shannon\ entropy(Se) = -\sum_{i} x_i^2 log(x_i^2) \tag{2.16}$$

$$Energy(e) = \frac{1}{N}\sum_{i}^{N} |x_i|^2 \tag{2.17}$$

Max. (maximum) and min. (minimum) of outputs of each feature extraction method were also used as additional features. Although eight features were calculated, best six subset features are given in 2.9 Results section.

2.6. Classification Methods

2.6.1. K Nearest Neighbour (k-NN)

k-NN is a non-parametric approach when the distribution of the data is unknown [43]. It is based on the idea that given the training samples ($y_{1\cdots N}$) and their class labels ($i_{1\cdots M}$) find the class label of the test sample (i_x) among the k nearest data points where M is the number of classes, N is the number of samples with $N \geq M$. Majority voting is employed to determine the most probable class label (i_x);

$$i_x = \underset{c}{argmax} \sum_{(y_i,i)} I(c == i) \tag{2.18}$$

where c is one of the class labels, i is the label of the i^{th} nearest samples in the neighbourhood of test sample [44]. In practical cases, k is chosen an odd number to overcome a tie condition. If k is chosen a small value it is much affected by noise, however if k is chosen a big value the accuracy decreases since the samples far away from the local region are taken into account. In the experiments, k values from 1 to 10 were employed with *Euclidean* and *city block* distance metrics.

2.6.2. Naive Bayes (NB)

Bayes' rule is designed to estimate the class of the test sample (c) given the test sample x using

$$p(c|x) = \frac{p(c)p(x|c)}{p(x)} \tag{2.19}$$

If it is assumed that the input features are independent of each other given the class label c_j will result in Equation 2.20.

$$p(x|c) = p(x_{1\cdots N}|c_j) = \prod_{i=1}^{N} p(x_i|c_j) \tag{2.20}$$

where N is the number of input features x_i. This simplification decreases the computational load and diminishes the curse of dimensionality problem as the dimensionality of features increase [45]. In conclusion, the Bayes' rule aims to estimate the class label (c_j) by maximizing the $p(c_j)\prod_{i=1}^{N} p(x_i|c_j)$ posterior probability.

2.6.3. Decision Trees (DT)

DT is a non-parametric approach which is composed of nodes, branches and leafs. Once the tree is trained using the training data, the same model and learned parameters are used in the testing. Starting from the root (base of the tree) at each node a test function is employed and after splitting, a branch is followed by test samples depending on the result of testing. This procedure is finalized at the leafs and label of the test sample is assigned by the trained leaf. The split for branching is based on the impurity score of the node. At the root node starting with N samples, for node i, N_i is the number of samples that reach node i. N_i^c is the number of samples that reach node i from class c and obtained using $\sum_c N_i^c = N_i$. Majority of the samples of the same class determines the label of the samples at the leafs by choosing maximum of $p_i^c = \frac{N_i^c}{N_i}$ which is called probability of belonging to class c [43]. If all the samples belong to same class at the leaf node p_i^c is 1 and it can be stated that the leaf node is 100 % pure. Impurity (I) can be measured either by entropy or Gini index

$$I(N) = -\sum_{c=1}^{K} p_i^c log_2 p_i^c \tag{2.21}$$

or

$$I(N) = \frac{1}{2}\left(1 - \sum_{c=1}^{K}(p_i^c)^2\right) \tag{2.22}$$

where K is the number of classes [46]. Evaluation of the DT algorithm in the experiments was done using various number of minimum leaf sample sizes and various type node splitting algorithms, without pruning and restricting the tree growth.

2.6.4. Support Vector Machine (SVM)

SVM is a maximum margin and a kernelized classifier that aims to find optimal separating hyperplane among training samples. If the problem is linearly separable, $f(x)$ being linear separating function passing through midpoint of the classes, testing is done by looking at the sign of the function such that $f(x_{test}) > 0$ belongs to the class i, or class j otherwise. When the problem is non-linear rather than fitting a non-linear model, the problem is mapped to another space where linear separation is possible using this kernel trick. The optimal separating hyperplane ensures the maximum accuracy with better generalization choosing the maximum margin between closest members of the classes to the hyperplane. SVM, not only provides best classification performance on the training samples, but also maximizes the margin aims to leave as much space as possible for better classification of testing data [44]. Once the maximum margin hyperplane is determined, the SVM classifier aims to minimize [47] Equation 2.23 with

$$L_p = \frac{1}{2}\left\|\vec{w}\right\| - \sum_{m=1}^{K} a_m y_m \left(\vec{w}\vec{x_m} + b\right) + \sum_{m=1}^{K} a_m \tag{2.23}$$

respect to $\vec{w}$ (normal vector) and b and maximize with respect to $a_m \geq 0$, where K is the number of training samples, and a_m, $m = 1, \ldots, K$, are Lagrange coefficients, y_m is the class label and $\vec{x_m}$ is the data vector. Taking the derivative of L_p with respect to a_m as zero, L_p is named as Lagrangian. The optimal hyperplane is constructed by

the vectors w and the fixed term b.

For training and testing the classifier, LIBSVM [48] implementation was used with radial basis and linear function kernels. Best kernel width (g) and cost (C) parameters were determined using a grid search with various values.

2.6.5. Extreme Learning Machine (ELM)

Feed-forward neural networks have been used in literature extensively because this algorithm is capable of non-linear mappings from direct inputs and provide robust models for the data which is difficult to analyze using parametric methods [49]. However, this method suffers from slow training/learning speed and is prone to get stuck in local minima due to gradient descent based learning algorithms. Back propagation (BP) learning algorithm (used previously in the works [21], [22] and [24]) has some disadvantages when employing gradient based learning: when learning rate is small convergence is too slow and convergence to a local minimum and over-learning introduce poor generalization performance. In order to alleviate these disadvantages a new classification method called extreme learning machines (ELM) was proposed [49]. ELM (used previously in [26]) is based on random assignment of input weights and hidden node biases without the need for differentiable activation function and setting up the output layer weights analytically through basic inverse matrix operations. ELM has advantages such as having extremely fast learning, better generalization capability than BP and straightforward convergence as compared to classical BP learning algorithms. ELM is modelled using single layer feed forward neural networks with N hidden nodes and an activation function ($g(x)$);

$$\sum_{i=1}^{N} \beta_i g_i(x_j) = \sum_{i=1}^{N} \beta_i g_i(w_i x_j + bi) = o_j \tag{2.24}$$

where $j = 1, \ldots, N$, w_i is the weight that links the i^{th} hidden node with input nodes, β_i is the weight that links the i^{th} hidden node with output nodes (o_j), and b_i is the i^{th}

hidden node threshold. In this thesis, sigmoid activation function ($f_{sig} = \frac{1}{1+e^{-x}}$) was used and number of hidden neurons were chosen by trying a range of values among 1 to 500.

2.7. Ensemble Learning

Ensemble learning is the process which combines the opinions of multiple simpler learning models for solving a particular problem with improved performance. It is seen that various ensemble methods were successfully employed in literature with different names such as ensemble of neural networks [50], mixture of experts [51], combination of multiple classifiers [52], classifier fusion [53], classifier ensembles [54], etc. The main reason behind using the ensemble learning is its ability to increase the generalization performance of machine learning systems. It is known that the test set performance of individual classifiers having similar training performances may change due to the deficiencies in their generalization ability. In such instances, combining the predictions of individual classifiers by majority voting method may minimize the risk of an unfortunate choice of a weak performing model. Weighted majority voting can be chosen to prevent a tie situation such as when classifier outputs result in 2-2-1 prediction (ie. two of the classifiers predict wheeze/normal and one of the classifiers predicts crackle given the test sample) for three classes in five individual classifier scenarios like in this thesis. Therefore, in this thesis to increase the classification accuracy of proposed method, weighted majority voting method was applied to a fused dataset consisting of the five feature subsets (energy, entropy, standard deviation, kurtosis and minimum) having highest individual average accuracy rates mentioned in 2.8 Results section. Therefore, prior to weighted majority voting stage, a new dataset was created by concatenating these individual feature subsets resulting in a fused dataset. In weighted majority voting, if an evidence exists that some individual classifiers are more eligible than others, weighting the decisions of those eligible classifiers and then applying majority voting may increase the overall accuracy of the ensemble system.

Assume that the decision of classifier h_t on class w_j is represented as $d_{t,j}$, such that $d_{t,j}$ is 1, if h_t chooses w_j and 0, otherwise. Additionally, suppose that the gen-

eral classification accuracy of each classifier is known and weight w_t to classifier h_t in proportion to its classification accuracy is set. By using this notation, the classifiers whose predictions are combined through weighted majority voting [55] will select w_J (class J), if

$$\sum_{t=1}^{T} \omega_t d_{t,J} = \max_{j=1}^{C} \sum_{t=1}^{T} \omega_t d_{t,j} \tag{2.25}$$

that is, if the total weighted vote collected by w_j is greater than the total vote collected by any other class. In this thesis the normalized weights (their sum is 1) were calculated by using the following equation,

$$\omega_t = \frac{a_t}{\sum_{t=1}^{T} a_t} \tag{2.26}$$

where a_t is the general classification accuracy of classifier h_t and T is the number of individual classifiers.

2.8. Detailed Analysis of Proposed System

600 lung sound signal episodes consisting of 200 wheezes, 200 crackles and 200 normal sounds with variable duration (the duration of sample episodes varied between 80 and 200 ms assuring that each included at least one crackle or wheeze) were analysed using high Q-factor RADWT to extract informative features. In doing so, we aimed to use the distinctive spectral properties of normal signals, wheezes and crackles for classification. As reported in literature, crackle signals having transient behaviour and short duration, may show themselves in a wide frequency range (200-2000 Hz) in spectrum [14,16]. Conversely, wheeze signals having oscillatory time domain behaviour, show themselves in a narrower frequency range (100-1000 Hz) [13,15]. Additionally, the wheeze, crackle and normal signals have opposite shape energy distributions; energy

atoms of crackle signals spread over a wide frequency spectrum while wheeze signal energy atoms exist in local and narrow frequency ranges. In the case of normal signals, energy atoms exist over a wide frequency range like the crackles but at lower frequency values (200-600 Hz). In our classification system, high Q-factor decomposition was performed on all 600 lung sound episodes. p, q, and s values were set to 6, 7 and 5 respectively. The dilation factor was obtained as 1.17 and the redundancy was 1.40. Various p, q, and s parameters were tested and the optimum combination was chosen empirically after exhaustive trials. 30 levels of wavelet decomposition resulted in one group of approximation and 30 groups of detail coefficients. Besides, the low Q-factor analysis was also employed with the aim of emphasizing the difference between high and low Q-factor analysis. p, q, and s values were set to 2, 3 and 2 respectively. The dilation factor was obtained as 1.5 with a redundancy value of 1.5. These parameters resulted in wavelet filters having low Q-factor and poor frequency selectivity. The decomposition was employed for 8 levels resulting in one group of approximation and 8 groups of detail coefficients. As seen in Figure 2.3, the typical frequency characteristics of pulmonary signals (normal, wheeze and crackle) could not be differentiated sufficiently (parts a, c and e) with the conventional low Q-factor analysis. By contrast, distinguishing spectral properties of lung sounds could be emphasized clearly (parts b, d and f) when the pulmonary sounds were decomposed with high Q-factor filters, as a result of the better frequency selectivity property of high Q-factor wavelet bases. After employing high Q-factor analysis, as discussed in the next section, energy, entropy, standard deviation, kurtosis and minimum feature subsets were extracted from each sub-band. These extracted features were fused at the feature level resulting in a row feature vector for each sample. Then, all the fused features were trained and tested using individual classifiers which could be listed as k-NN, NB, DT, SVM and ELM. In order to exploit the discrimination ability of the proposed model, ensemble learning method was applied to the outputs of individual classifiers. The individual classifier predictions were combined using weighted majority voting in order to assign the final labels of samples. Ensemble accuracies were computed and the testing was finalized in a leave-one-out cross validation (LOOCV) scheme. The pictorial representation of the complete proposed system can be viewed in Figure 2.4.

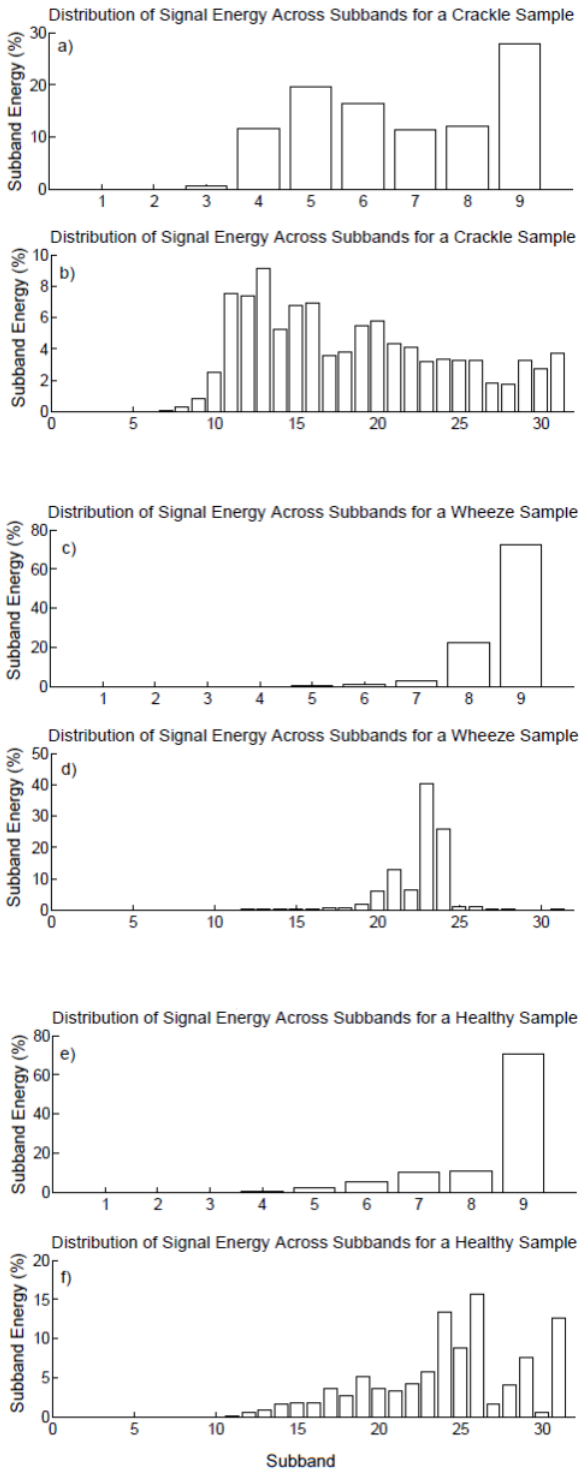

Figure 2.3: Distribution of signal energy over sub-bands for low (a, c and e) and high (b, d and f) Q-factor analysis for three classes.

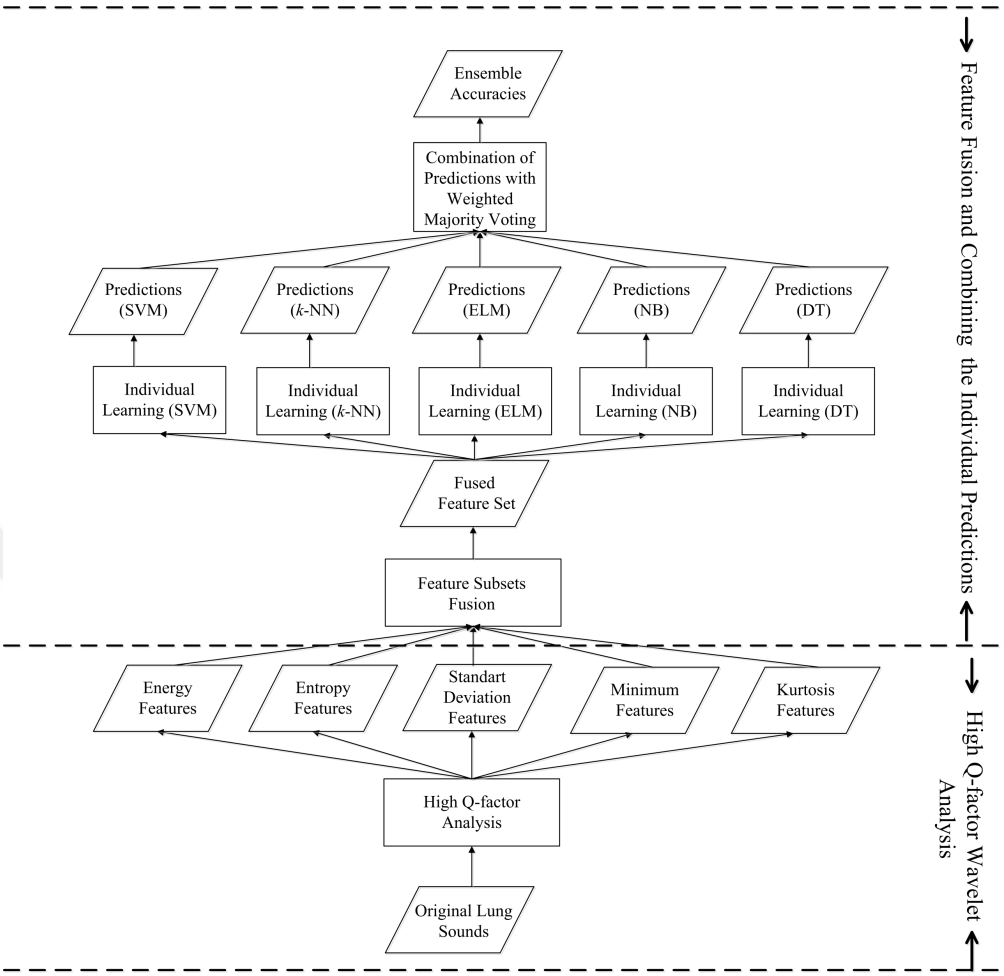

Figure 2.4: The design of the detailed proposed system with ensemble learning.

2.9. Results

The experiments were performed on a lung sound database with the aim of finding optimal classifier and feature subset extraction method configuration. Consequently, experiments using five different classifiers and six different feature extraction methods were performed. Additionally, six different statistical feature subsets were calculated from each feature extraction method such as energy, entropy, minimum, maximum, standard deviation and kurtosis of raw features. The classification methods were used in a LOOCV scheme and the results were shown in Tables 2.2 to 2.7. At each table the individual crackle, wheeze, normal lung sound classification accuracies and the average of overall accuracies were presented in percentage for different feature extraction methods. As illustrated in Table 2.7, the best average classification accuracy was obtained using the proposed feature extraction method (High Q-factor) with 95.17 % when the

SVM was used as classifier and energy of each subband as feature subset. As seen in Tables 2.5 and 2.7 the best individual classification performance for wheeze sound was achieved by the proposed feature extraction method resulting in 98.00 % accuracy. On the other hand, best crackle and normal sound classification accuracies were achieved by the S transform as 96.00 % and 96.50 %, respectively. When the performance of feature subsets, energy, entropy, minimum, maximum, standard deviation and kurtosis, were considered, it is seen in Tables 2.2 to 2.7 that the best average accuracy values were obtained with the energy feature subset in four of six feature extraction methods. When the energy of each subband obtained from high Q-factor analysis was employed as feature subset, the highest average accuracy was achieved. Conversely, the worst average accuracy was achieved when the entropy subset features were obtained from PSD. As seen in Tables 2.2 to 2.7, wavelet based methods performed better general accuracy performances (95.00 % for S transform, 95.17 % for the proposed method) than the Fourier Transform based algorithms highlighting their finer time-frequency resolution property. In Figure 2.5, best average accuracy of all feature extraction methods were depicted. As shown in the left part of Figure 2.5a S transform and the proposed feature extraction method present break even performances while PSD presenting the worst performance. In the right part of Figure 2.5a, classification errors of all feature subsets related with feature extraction methods are averaged and represented. It is clear that the proposed high Q-factor method achieved the lowest error rate among all feature extraction methods. In Figure 2.5b, average computation time for each feature extraction method is presented. The experiments were carried out 100 times and the average computation time of 100 trials was given.

As illustrated in Figure 2.5b, proposed high Q-factor feature extraction method was superior to S transform in terms of computation time by alleviating redundant information of S transform and in terms of slightly better classification performance. Additionally, high Q-factor approach showed better classification performance than low-Q method solving the poor frequency resolution problem of traditional dyadic DWT. Even though MFCC and PLP had lower computation times than the proposed method, these methods showed poor classification performances as compared to high Q-factor method. In each column-group of Figure 2.6, minimum (in this sense minimum means

Table 2.2: Power spectral density (PSD) features correct classification rates (in %) for five different classifiers using six different feature subsets.

Type \ Feature	Energy					Entropy					Std				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	86.00	90.50	90.00	83.50	90.50	48.50	51.50	43.50	46.50	48.00	60.00	72.50	68.00	60.00	73.00
Wheeze	46.50	76.00	74.00	53.00	71.50	39.00	48.50	52.00	46.50	47.50	49.00	67.00	68.50	53.00	57.00
Normal	45.00	19.50	19.50	49.50	27.50	54.00	51.50	60.00	52.00	56.50	39.50	21.00	17.50	39.00	18.00
Average	59.17	62.00	61.17	62.00	63.17	47.17	50.50	51.83	48.33	50.67	49.50	53.50	51.33	50.67	49.33

Type \ Feature	Min					Max					Kurtosis				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	68.50	76.00	75.50	65.50	77.50	81.00	82.00	84.00	82.00	84.00	67.50	71.00	76.50	61.00	71.00
Wheeze	42.00	39.50	56.50	41.50	49.00	61.00	73.00	77.00	56.50	67.50	63.50	47.00	50.50	67.00	48.50
Normal	43.00	56.00	40.50	46.00	44.50	48.00	62.00	48.00	53.50	64.50	45.50	70.50	61.00	44.00	70.00
Average	51.17	57.17	57.50	51.00	57.00	63.33	72.33	69.67	64.00	72.00	58.83	62.83	62.67	57.33	63.17

Table 2.3: Mel Frequency Cepstral Coefficient (MFCC) features correct classification rates (in %) for five different classifiers using six different feature subsets.

Type \ Feature	Energy					Entropy					Std				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	55.00	79.50	76.00	74.00	63.50	54.50	74.00	47.00	52.00	46.50	85.50	88.50	94.00	92.00	63.00
Wheeze	86.50	95.00	85.50	92.00	93.50	82.50	84.00	80.50	84.00	85.50	91.00	44.50	52.00	67.00	77.00
Normal	81.50	92.50	89.00	80.00	85.50	54.00	47.00	71.00	52.50	29.50	40.50	87.50	79.50	71.00	70.00
Average	74.33	89.00	83.50	82.00	80.83	63.67	68.33	66.17	62.83	53.83	72.33	73.50	75.17	76.67	70.00

Type \ Feature	Min					Max					Kurtosis				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	88.00	89.50	92.00	81.50	91.00	75.00	75.00	80.50	78.00	77.50	32.50	42.50	32.00	48.00	36.00
Wheeze	93.50	87.00	78.50	85.00	83.00	87.00	77.50	78.50	78.00	78.50	76.00	78.00	71.50	62.50	74.00
Normal	92.50	92.50	93.00	83.50	93.50	87.00	87.50	89.00	84.50	91.50	57.50	68.50	69.50	47.00	41.50
Average	91.33	89.67	87.83	83.33	89.17	83.00	80.00	82.67	80.17	82.50	55.33	63.00	57.67	52.50	50.50

Table 2.4: Perceptual linear prediction (PLP) features correct classification rates (in %) for five different classifiers using six different feature subsets.

Type \ Feature	Energy					Entropy					Std				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	90.50	94.00	90.00	89.50	81.00	50.50	53.50	49.00	48.50	50.00	84.50	89.50	89.50	86.50	85.50
Wheeze	90.00	79.00	87.50	91.00	69.50	46.50	41.00	57.00	52.00	41.00	95.50	72.00	78.00	74.50	83.50
Normal	92.00	95.50	90.50	90.00	68.50	86.50	94.50	81.50	90.50	78.50	40.00	76.00	90.50	76.00	80.50
Average	90.83	89.50	89.33	90.17	73.00	61.17	63.00	62.50	63.67	56.50	73.33	79.17	86.00	79.00	83.17

Type \ Feature	Min					Max					Kurtosis				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	75.50	92.50	90.00	87.00	89.00	71.50	87.00	89.50	75.00	82.00	53.00	59.50	54.00	42.00	47.00
Wheeze	92.00	86.00	86.00	83.50	90.50	91.00	87.50	85.50	78.00	84.50	79.00	81.00	57.00	79.50	68.50
Normal	74.00	89.50	94.00	91.00	90.00	77.00	89.00	94.00	84.50	84.00	36.50	47.50	83.00	40.50	41.00
Average	80.50	89.33	90.00	87.17	89.83	79.83	87.83	89.67	79.17	83.50	56.17	62.67	64.67	54.00	52.17

Table 2.5: S-transform features correct classification rates (in %) for five different classifiers using six different feature subsets.

Feature / Type	Energy					Entropy					Std				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	94.00	94.00	92.50	88.50	94.50	80.50	81.00	83.50	78.00	79.50	77.00	82.50	78.50	84.00	73.50
Wheeze	95.00	96.50	94.50	91.50	96.50	76.50	73.00	76.50	73.00	74.00	83.50	88.00	84.50	82.00	86.50
Normal	91.50	94.50	89.00	89.50	92.50	71.50	72.00	72.00	67.00	80.50	79.50	83.50	81.00	77.50	64.00
Average	93.50	95.00	92.00	89.83	94.50	76.17	75.33	77.33	72.67	78.00	80.00	84.67	81.33	81.17	74.67

Feature / Type	Min					Max					Kurtosis				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	96.00	89.50	75.00	81.00	72.50	87.00	81.00	58.00	80.00	72.50	76.50	79.00	78.50	74.50	78.50
Wheeze	91.00	95.00	89.50	82.00	95.00	83.00	88.50	75.00	86.50	89.50	77.00	78.50	77.00	76.50	79.50
Normal	92.00	96.50	73.50	80.50	65.00	86.50	90.00	74.00	88.00	74.00	73.50	72.50	71.50	76.50	74.00
Average	93.00	93.67	79.33	81.17	77.50	85.50	86.50	69.00	84.83	78.67	75.67	76.67	75.67	75.83	77.33

Table 2.6: Low Q-factor features correct classification rates (in %) for five different classifiers using six different feature subsets.

Feature / Type	Energy					Entropy					Std				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	92.00	89.50	91.50	92.00	91.50	88.00	90.00	89.00	87.50	88.50	80.50	83.00	64.00	87.00	79.50
Wheeze	86.00	86.50	74.50	87.50	84.50	83.00	85.00	75.00	84.00	83.00	87.50	85.00	96.50	80.50	89.00
Normal	85.50	90.50	94.50	86.50	90.00	87.00	88.00	93.00	89.50	90.50	90.50	93.50	83.00	89.50	92.50
Average	87.83	88.83	86.83	88.67	88.67	86.00	87.67	85.67	87.00	87.33	86.17	87.17	81.17	85.67	87.00

Feature / Type	Min					Max					Kurtosis				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	85.00	84.50	70.00	79.50	83.00	72.00	75.50	62.50	73.00	76.00	73.50	86.00	80.00	82.50	81.00
Wheeze	73.50	81.00	79.00	74.50	82.50	82.50	87.00	86.00	77.00	83.00	60.50	61.00	43.50	64.00	58.50
Normal	76.00	83.50	75.50	75.00	81.00	72.50	79.50	71.00	65.50	78.50	76.00	74.00	91.00	72.00	79.50
Average	78.17	83.00	74.83	76.33	82.17	75.67	80.67	73.17	71.83	79.17	70.00	73.67	71.50	72.83	73.00

Table 2.7: High Q-factor features correct classification rates (in %) for five different classifiers using six different feature subsets.

Feature / Type	Energy					Entropy					Std				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	92.50	95.00	82.50	84.50	89.50	91.00	92.00	90.00	82.50	90.00	92.00	91.50	82.50	85.50	93.00
Wheeze	96.50	97.00	97.00	93.50	94.00	95.00	98.00	97.00	88.00	95.50	96.00	96.50	98.00	91.00	66.00
Normal	91.50	93.50	96.00	89.50	82.50	89.00	89.50	91.50	82.50	85.50	93.00	95.00	89.00	83.00	85.50
Average	93.50	95.17	91.83	89.17	88.67	91.67	93.17	92.83	84.33	90.33	93.67	94.33	89.83	86.50	81.50

Feature / Type	Min					Max					Kurtosis				
	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**	***k*-NN**	**SVM**	***NB***	**DT**	**ELM**
Crackle	90.00	92.00	75.00	87.00	93.00	80.00	78.50	73.50	82.50	80.50	72.00	86.50	85.00	86.50	79.50
Wheeze	95.50	96.50	96.00	90.00	94.00	93.00	92.50	94.50	90.50	84.50	78.00	82.50	67.00	80.50	77.50
Normal	93.00	91.50	87.50	91.50	89.50	84.00	88.50	86.50	79.50	76.50	93.00	91.00	92.50	83.00	89.00
Average	92.83	93.33	86.17	89.50	92.17	85.67	86.50	84.83	84.17	80.50	81.00	86.67	81.50	83.33	82.00

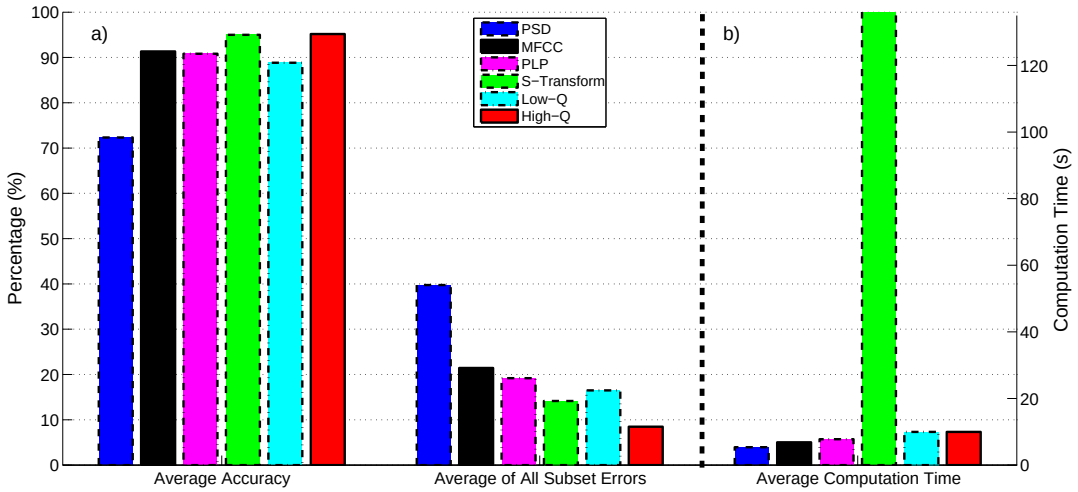

Figure 2.5: Best average accuracy (a-left), subset error (a-right) and computation time (b) for the six different feature extraction methods.

the best individual general accuracies that are obtained for various classifiers, which are used in error rate calculation) error rates achieved with the feature subsets are depicted. In each column-group, the performances of all experimented feature extraction methods are also given. The rightmost yellow-bar in each column-group refers to the mean of the average error rates obtained with the tested feature extraction methods and indicates the average performance of relevant statistical feature subset. Additionally, in the rightmost column-group of the figure (labelled as *average of methods* in the x-axis), the mean average error rates of statistical feature subsets are shown and this presents the average performance of feature extraction methods. As illustrated in the yellow-bars of each feature subset column-group, the minimum average error rates were obtained with minimum, maximum and energy feature-subsets while the highest average error rates were obtained with kurtosis and entropy, respectively. As summarized in the rightmost column-group of Figure 2.6, mean of average error rates for each feature extraction method was minimum when the proposed high Q-factor method was applied, and the S transform and the low Q-factor (dyadic wavelet) methods followed the proposed method supporting the idea that wavelet based methods had finer time-frequency resolution than Fourier transform based methods. PSD had highest error rate for almost all feature subsets showing the poorest performance as seen in rightmost column-group of the Figure 2.6. In Figures 2.7 to 2.9, classifier performances related

to feature extraction methods for crackle, wheeze and normal lung sounds are given. For all lung sound types, the minimum (in this sense minimum means the individual crackle, wheeze or normal lung sound classification accuracies that are obtained for various feature sub-sets and the highest of these sub-set related accuracies are used in error rate calculation) error rates are given in the first five column-groups. For all lung sound types, each one of these first five column-groups represents a classifier performance. In the first five column-groups, the rightmost yellow bar presents the mean value of error rates obtained with various feature extraction methods and this displays the performance of the related classifier. Additionally, in the rightmost (sixth) column-group of Figures 2.7 to 2.9 (labelled as *Average of Methods* in the x-axis), the mean minimum error rates of feature extraction methods obtained with various classifiers are given, and this illustrates the performance of feature extraction method for each lung sound type. As depicted in Figures 2.7 to 2.9, when the average performance of each classifier is considered, the best performances were achieved with SVM classifier while the worst performances were achieved with DT classifier for the crackle, wheeze and normal sound cases, respectively. As illustrated in rightmost column-group of Figure 2.7, when the mean of minimum error rates are considered, best performances were obtained with the S transform, proposed high Q-factor method and low-Q method while the worst performances were obtained with the PSD for crackles, respectively. As

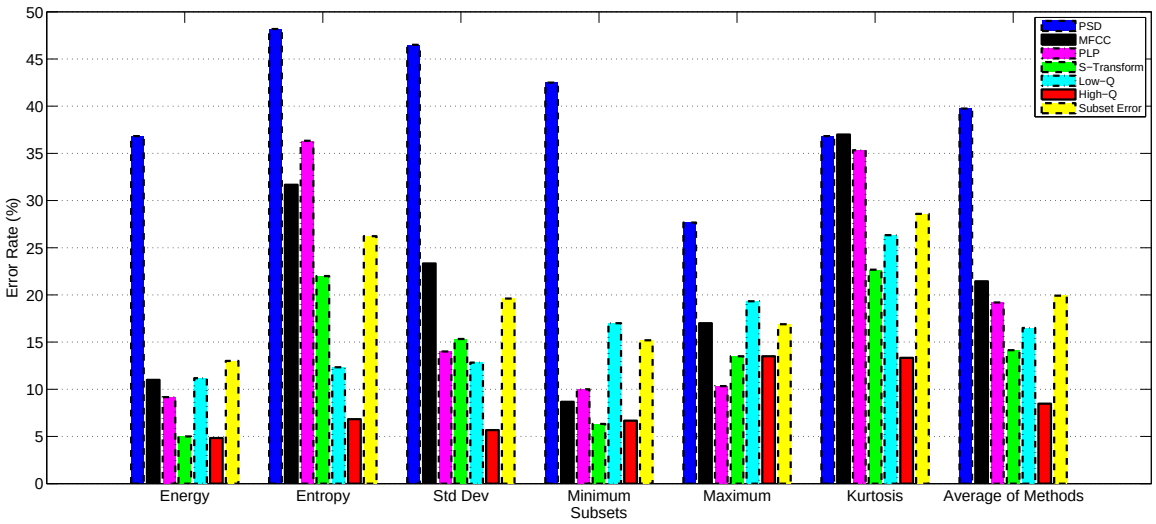

Figure 2.6: Minimum average error rate for six individual feature subsets of the six different feature extraction methods.

summarized in the rightmost column-group of Figure 2.8, when the mean of minimum error rates are considered for wheezes, the best performances were obtained with the proposed high Q-factor method, S transform and MFCC method while the worst performances were obtained with the PSD, respectively. This shows the weakness of low-Q method in localizing frequency characteristics of oscillatory waveforms (i.e wheezes). As marked in the rightmost column-group of Figure 2.9, mean of error rates for normal sounds were minimum for the proposed method and PLP method while the maximum for PSD method, respectively. As depicted in the *average of methods* part of the *classifiers* axis in Figure 2.9, all the feature extraction methods except PSD were able to model normal lung sounds with similar error rates. As shown in Figures 2.7 to 2.9, SVM was the best classifier in terms of average error rates for all feature extraction methods with ELM as the second best. To sum up, as depicted in Figures 2.5 and 2.7 to 2.9, proposed high Q-factor method had lower error rates and was able to track the time-frequency behaviour of transient (crackle) and oscillatory (wheeze) waveforms. Moreover, its computational load was less than the S transform (nearest competitor of the proposed method) and approximately same with others.

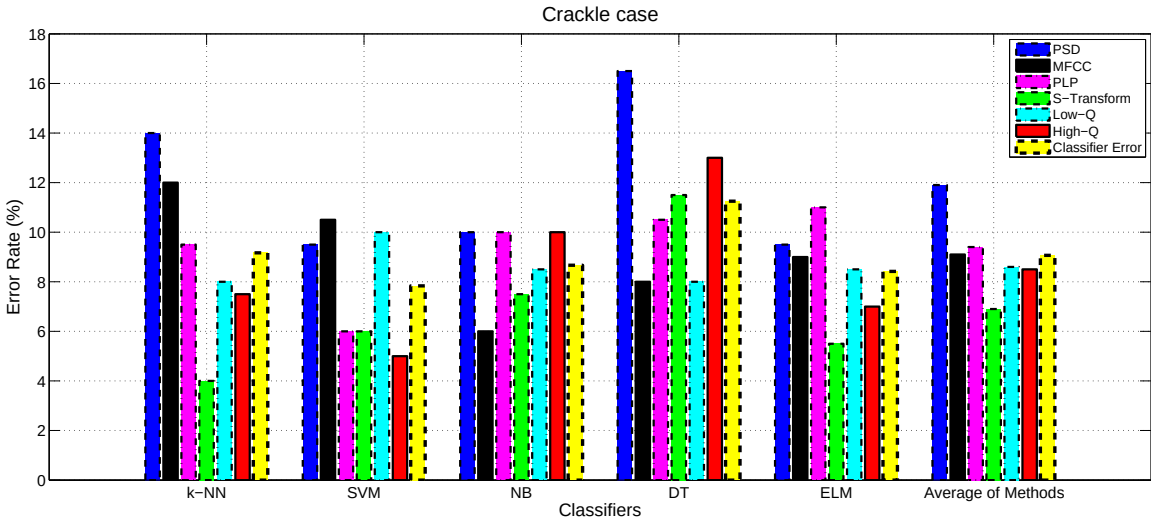

Figure 2.7: Minimum average crackle error rate for five individual classifiers of the six feature extraction methods.

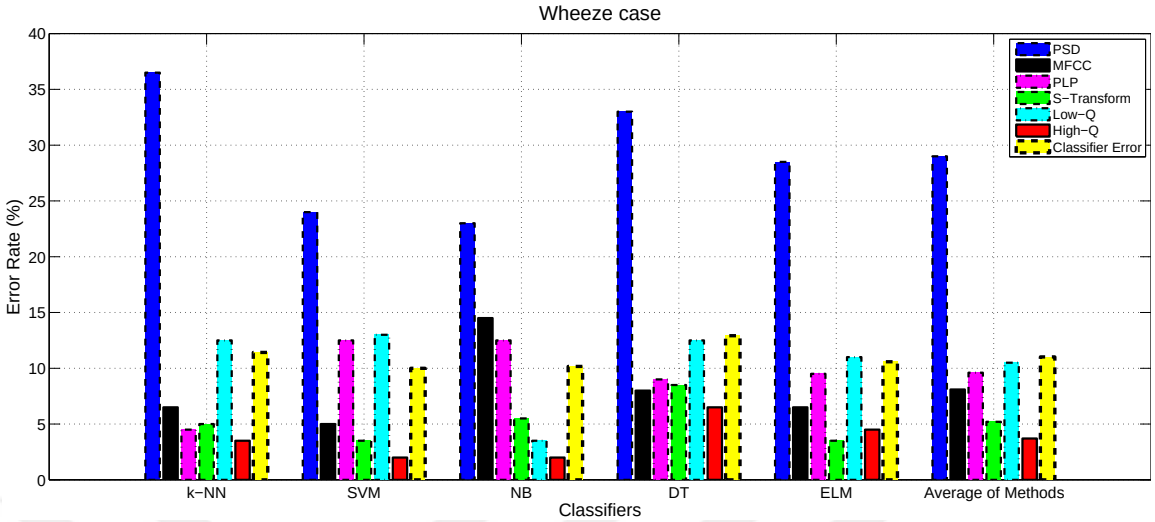

Figure 2.8: Minimum average wheeze error rate for five individual classifiers of the six feature extraction methods.

As depicted in Table 2.8, the complete algorithm of Figure 2.4 was run ten times to produce generalizable results and 2.21 % improvement in accuracy showed the success of the proposed system with ensemble learning. Moreover, best results were achieved with least standard deviation for wheezes achieving an average accuracy of 99.05 %. This shows the ability of the proposed system for localizing wheezes and highlighting the difference between the spectral characteristics of normal lung sounds and crackles.

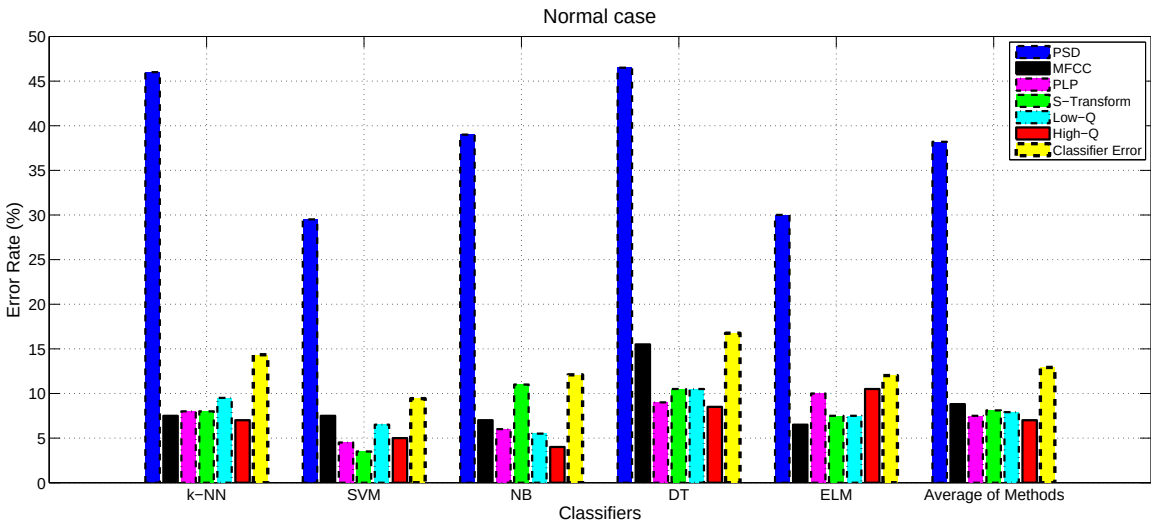

Figure 2.9: Minimum average normal error rate for five individual classifiers of the six different feature extraction methods.

Table 2.8: Ensemble accuracy results (in %) for the proposed method

Type \ Iteration	1	2	3	4	5	6	7	8	9	10	Average	Std
Crackle	95.50	95.50	95.50	96.00	94.50	96.00	95.50	95.00	96.00	95.50	95.50	0.45
Wheeze	99.50	99.00	99.00	99.00	99.00	99.00	99.00	99.00	99.00	99.00	99.05	0.15
Normal	97.00	97.50	98.00	98.00	97.50	97.50	97.50	97.50	97.50	98.00	97.60	0.30
Average	97.33	97.33	97.50	97.67	97.00	97.50	97.33	97.17	97.50	97.50	97.38	0.18

2.10. Discussion and Summary

A comprehensive experimental study was conducted using six feature extraction methods and five classifiers. Six subset features were extracted from the raw features and fed into the classifiers in order to compare the performances of feature extraction methods and classifiers. The proposed high Q-factor wavelet transform based method performed well both in terms of classification accuracy and computational time as compared to other methods. The proposed method was able localize the time-frequency characteristics of both oscillatory and transient signals of interest however low-Q method had limited frequency resolution especially in modelling oscillatory signals. S transform was the second powerful method in terms of accuracy, however, it had some structural disadvantages such as having the highest computational load. Additionally, S transform must be applied to fixed length signals (1024 point in our case) to obtain the same number of features that will be used in classifiers while the proposed method can be applied to variable length signals. PSD had the lowest accuracies for almost all of the cases as seen in Figures 2.5 and 2.7 to 2.9. PSD is a Fourier based method and therefore, its output is not a matrix but a vector having no time information. Therefore, using the subset features (minimum, maximum, etc.) in classification decreases its performance. As a solution raw features were fed into classifiers to observe the performance. The raw feature classification results (classification accuracy for crackle was 88.5 %, for normal was 92 %, for wheeze was 82.5 % and average accuracy was 87.67 %) were better than its subset form but still poorer as compared to other methods. One can argue that, classical Fourier transform (FT), since it does not contain time information, is not adequate to cope with this problem. MFCC and

PLP are windowed Fourier based methods that are able to alleviate the drawback of FT. However, these methods have fixed time-frequency resolution that results in an improved accuracy up to a certain extent as compared to wavelet based methods. On the other hand, the proposed method has the ability to tune its wavelet bases aiming at localizing TF (time frequency) characteristics of wheeze, crackle and normal lung sounds and as a result shows finer time-frequency resolution and higher accuracy performance than the Fourier based methods. Moreover, the proposed method is 18 times faster than its nearest competitor (S transform) in terms of computational load.

As a robust classifier, SVM has also served as a powerful classifier in this thesis with ELM. However, ELM has an important disadvantage: it needs to be randomly initialized at each new trial and this randomness results in non-robust predictions when compared with robust predictions of SVM. NB and k-NN classifiers, which have shown even better performances in some cases, are less robust than SVM and ELM in the overall performance.

Detection of crackle and wheeze sounds within a breath cycle significantly plays a vital role in differential diagnosis [16, 18, 56]. When a localized wheeze is missed, this results in misdiagnosis of asthma and accompanied by mistreatment and dense hospital visits [15]. Moreover, asbestosis and idiopathic pulmonary fibrosis can be detected earlier than when radiological findings become apparent using computerized analysis of pulmonary sounds. The beginning of pneumonia, crackles are observed in mid inspiration phase; however, in the recovery period crackles appear at end inspiration phase of the breath cycle [15]. The proposed system is a robust candidate to discriminate abnormalities within a breath cycle, in order to be employed in differential pulmonary disease diagnosis systems, more successfully and faster than related methods in literature.

3. RESPIRATORY SOUND DECOMPOSITION AND DETECTION

3.1. Introduction

Although stethoscope is a vital instrument in the medical diagnosis, it does not transmit the whole frequency content to the expert since it attenuates frequencies above 120 Hz [11]. To minimize disadvantages of the subjective auscultation via stethoscope, computerized analysis of lung sounds (LSs) has become an emerging area in the interdisciplinary cooperation between engineering and medicine. LS are believed to be produced by the turbulent flow in the lung airways, despite the fact that the explicit setup of lung sound production is unidentified [57]. Lung sounds (LS) can be categorized into two classes, vesicular (normal) sounds and adventitious (abnormal) sounds. Vesicular sounds can be defined as the normal breath noise heard over the chest wall and synchronous with air flow in the airways. In healthy people, the frequency range of normal lung sounds is 200-600 Hz. Adventitious lung sounds (ALS) are superimposed on vesicular sounds and are typical indicators of various pulmonary diseases. ALS are divided into two main categories such as continuous adventitious lung sounds (CALS) and discontinuous adventitious lung sounds (DALS). The most studied components of the CALS are the wheezes and of the DALS are the crackles in literature. Crackles are non-musical sudden bursts and are explosive in nature. Crackles are classified as coarse (lower pitch) or fine (higher pitch). Crackles are believed to be generated by abnormally closed airway openings [16]. The time span of crackles is usually less than 20 ms and their spectrum has 200 to 2000 Hz frequency range. The number of crackles per breath is related to the severity of the disorder, and the timing, duration and types of crackles in a breath cycle may be different in various lung disorders [16]. For example, coarse crackles exist in bronchopneumonia and bronchiectasis whereas fine crackles are common symptoms of interstitial fibrosis and pneumonia [17]. On the other hand, wheezes are musical waveforms with duration of more than 80-250 ms and they exhibit distinct peaks in the frequency domain ($>$ 100 Hz). The presence of

wheezes usually indicates a pulmonary disorder such as asthma (AS) and chronic obstructive pulmonary disease (COPD). According to [18] and [19] wheeze characteristics such as pitch frequency and duration are related to the degree of airway obstruction. Stemmed from time-varying characteristics of lung tissue and chest wall, most lung sound signals are non-stationary in character, independent from time scale and time frequency analysis domain. This is valid for adventitious lung sound types, especially crackles. Most of the time, adventitious lung sounds have high-frequency components very tight in time domain, low-frequency components very tight in frequency domain which are superimposed on low frequency vesicular sounds. Hence a suitable analysis method for detecting them should supply information about good frequency resolution along with good time resolution, the first to localize the low frequency entities, and the second to resolve the high frequency entities. Therefore classical Fourier transform (FT), which assumes that the analyzed signal is stationary and does not contain any time information, is not appropriate to analyze most of the biomedical signals. Short Time FT (or windowed FT) partially overcomes the drawback of FT by considering an analysis window that has fixed time-frequency resolution. However, presenting a time-scale description of signals wavelet transform has finer frequency resolution at low frequencies, but also has finer time resolution at high frequencies [58].

In order to automatically analyze pulmonary diseases using computerized systems, proper detection of crackles and wheezes is very important. Wheezes and crackles are superimposed on normal lung sounds, either during inspiration or expiration, and they represent varying duration, intensity and frequency content as stated above. To isolate fine crackles from vesicular sounds, stationary-nonstationary separating filter (ST-NST) is proposed in [59]. In [60], so as to isolate crackles from vesicular sounds, a modified version of ST-NST filter is used. In [17], using DWT, the crackle detection performance outperforms ST-NST filter based detectors. For crackle detection case, an Empirical Mode Decomposition (EMD) based method is used to highlight crackle information within vesicular sounds in [61] but it is limited due to mere visual validation. In another study in [62], EMD is employed to denoise explosive lung sounds (i.e. crackles). In the study of [63], various Independent Component Analysis (ICA) algorithms are compared with the aim of isolating discontinuous adventitious

lung sounds from vesicular lung sounds. Due to their non-stationary characteristics for both crackle and wheeze signals, wavelet transform based signal processing methods were also used in [17, 31, 64–67]. For crackle detection the wavelet transform based stationary-nonstationary filter (WTST-NST) is proposed in [64] and gives better performance than ST-NST based filters with a computational burden. In [65], a wavelet based crackle detection system is realized in digital signal processor to quantify crackles in pulmonary sounds. In [66], wavelet packet transform based filter (WPST-NST) is proposed to separate crackles in a more accurate and faster way than previous WTST-NST based filters. In [67], continuous wavelet transform (CWT) is proposed to separate wheezes from background lung sound using scalogram. A more robust and improved wheeze type classification system is proposed in [31] and proposed wavelet based bicoherence features showed statistically significant results in different lung disorders such as asthma and COPD. As noticed, ST-NST ([59, 60]), WTST-NST [64] and WPST-NST [66] based crackle separation systems are incapable of separating wheezes from background sound since both of the sounds are stationary as compared to crackles. On the other hand, CWT [67], and bicoherence based wavelet system [31] just deal with wheeze sounds disregarding crackles whereas the two phenomenon may exist successively [68]. To construct automatic systems for crackle and wheeze signals detection which are significant indicators of pulmonary diseases, limited extent discrete wavelet based methods were proposed separately in literature. In all of these methods, the wavelet transforms with constant low Q-factor filters, which have limited frequency resolution, have been used. However, when we look at the morphological properties of wheezes and crackles, we see that they show opposite time behaviors. Crackles are sudden bursts and explosive in nature and can be classified as low Q-factor signals; on the other hand wheezes are sinusoidal like waveforms which can be classified as high Q-factor signals. Therefore, by using Morphological Component Analysis (MCA) based resonance signal decomposition and tunable Q-factor wavelet transform (TQWT) [69], we planned to decompose lung sound into low Q-factor signals (crackles), high Q-factor signals (wheezes) and residual (vesicular) signals simultaneously. In our method, by employing wavelets with properly tuned Q-factors, optimum representations for signal of interest are obtained resulting in denoised versions of crackle and wheeze candidate

channel signals.

3.2. Data Acquisition System and Dataset

The dataset used in this thesis is recorded by the 14-channel data acquisition system designed in Boğaziçi University Lung Acoustics Laboratory (BU-LAL). For detailed information about the system, please refer to [30]. The data acquisition system is made up of 14 air-coupled electret microphones (SONY ECM-44 BPT) attached on the posterior chest wall, an analog amplifier filter unit with a pass band of 80 to 4000 Hz and a gain of 100, a Fleisch type pneumotachograph (Validyne CD379) to measure the flow rate simultaneously for synchronization. The sampling rate is 9600 samples per second and each data acquisition session lasts 15 seconds. An informed consent is taken from all the subjects before recording. The data acquisition procedure of this thesis has been approved by the Second Ethical Committee on Clinical Research of Istanbul (which is in compliance with the Declaration of Helsinki).

To evaluate the proposed method the CALSs and DALSs which are found in COPD and Asthma patients are used. By visually inspecting the time expanded waveforms together with auditory verification, an expert labelled the wheeze and crackle locations.

3.3. Generation of Simulated Adventitious Sounds

In order to test the performance of the proposed method several synthetic wheezes and crackles are generated. Synthetic crackles are generated using the mathematical equation proposed in [70] as follows

$$y(t) = sin(4\pi t^{\alpha}) \text{ where } \alpha = \frac{log(0.25)}{log(t_0)} \tag{3.1}$$

and $y(t)$ possessing two cycles and its first positive t-intercept at t_0. A modulating function, $m(t)$, is applied to shift the power of $y(t)$ to the beginning of the waveform

as follows

$$m(t) = 0.5\left\{1 + cos\left[2\pi(t^{0.5} - 0.5)\right]\right\} \tag{3.2}$$

multiplying $m(t)$ and $y(t)$. Synthetic wheezes are generated using the mathematical equation proposed in [71] as follows

$$c_{1,k}(t) = sin\left[2\pi f_c(k)t + 0.6sin(2\pi 15t)\right] \tag{3.3}$$

$$c_{2,k}(t) = sin\left[2\pi f_c(k)t + 2\pi\mu(k)t^2\right] \tag{3.4}$$

where $c_{1,k}(t)$ is a monophonic wheeze which slightly varies in a frequency range and $c_{2,k}(t)$ is a monophonic wheeze with frequency sweeping linear frequency modulated signal, f_c is the center frequency, $\mu(k)$ is the slope and t is the duration of the signal.

3.4. Decomposition Methods

3.4.1. Proposed Resonance Based Decomposition

Tunable Q-factor Wavelet Transform (TQWT) method depends on two parameters: Q-factor and r (over-sampling rate) [69]. The parameter r is a measure of overlapping degree between successive band-pass filters [72]. Increasing r enough, will yield approximately fully-discrete version of CWT. After the determination of the optimal Q-factor, resonance based decomposition of given lung sound data into candidate crackle waveform, candidate wheeze waveform and candidate residual waveform (background sound) is explored using MCA based TQWT method.

Oscillatory components (wheezes) and transient components (crackles) may contain both low and high frequencies such that linear time invariant (LTI) filters are not able to handle this problem since the decomposition is not frequency based. Addi-

tionally unlike linear filtering (frequency based decomposition), neither low nor high resonance components obey the superposition rule turning the decomposition problem into non-linear case which is exactly supported by MCA. Assume $y = x_1 + x_2 + n$, where y is the observed lung sound component, x_1 is the oscillatory waveform, x_2 is the transient waveform and n is the background vesicular (residual) sound. By using both low and high Q-factor TQWT and MCA it is expected to decompose y signal into three sub-bands by exploiting the tunable wavelets and resonance based non-linear decomposition, respectively. Sparse representation of x_1 and x_2 can be found in Ψ_1 and Ψ_2 transformation matrices. Given y, at first sparse w_1 and w_2 coefficient vectors must be found which meet the requirement $y = \Psi_1^* w_1 + \Psi_2^* w_2$. Using basis pursuit approach, the sparse w_1 and w_2 coefficient vectors can be determined by l_1 norm minimization. The problem can be defined as minimization of an objective function U;

$$U(w_1, w_2) = \|y - \Psi_1^* w_1 - \Psi_2^* w_2\|_2 + \lambda_1 \|w_1\|_1 + \lambda_2 \|w_2\|_1 \tag{3.5}$$

with respect to coefficient vectors w_1 and w_2. $\|y - \Psi_1^* w_1 - \Psi_2^* w_2\|_2$ term is the energy of the residual (vesicular) sound. Estimation of the x_1 and x_2 is as follows $\hat{x_1} = \Psi_1^* w_1$ and $\hat{x_2} = \Psi_2^* w_2$, respectively. λ_i is the weighting parameter which sets the energy of the resonance terms. The optimization problem can be solved using the Split Augmented Lagrangian Shrinkage Algorithm (SALSA) [73]. Once the decomposition is employed, the critical step is the selection of the suitable threshold level to localize or at the worst case detect candidate abnormality waveforms. It is important to determine a threshold independent from the band on which it will be applied. We propose to use adaptive threshold determination by using decomposed vesicular channel $\hat{n} = y - \hat{x}_1 - \hat{x}_2$. Median absolute deviation (MAD) [74] measures the residual standard deviation (σ_{res}) in wavelet coefficients and/or residual channel samples by ensuring the independency of the threshold from the crackle (transient) channel using $\sigma_{res} = median(|\hat{n}| / 0.6745)$. As a second choice, Teager energy operator T;

$$T(\hat{x}_2) = \hat{x}_2(k)^2 - \hat{x}_2(k+1) \times \hat{x}_2(k-1) \tag{3.6}$$

is used to determine crackle locations [17] from transient channel ($\hat{x}_2$) by amplifying the signal components having sharp changes while suppressing background activity.

3.4.2. Empirical Mode Decomposition (EMD)

EMD is an adaptive, complete and orthogonal technique that is able to decompose nonlinear and nonstationary components directly using the data itself [75]. Intrinsic mode functions (IMFs), which are oscillating waveforms obtained after a sifting process, constitute the data using local time scale characteristics [75].

Two requirements must be satisfied when IMFs are employed [75]: Firstly, in the entire dataset, the number of zero crossings must be equal (at most by one difference) to the number of extrema. Secondly, mean of the envelopes derived from local maxima and minima is zero at any part of the data.

After the sifting process, original data $x(t)$ is presented using linear combinations of IMFs as below [75]:

$$x(t) = \sum_{i=1}^{N} IMF_i + n(t) \tag{3.7}$$

where N is the number of overall IMFs and $n(t)$ is the residual constant data.

EMD is an iterative method that original data can be recovered/denoised by summing the related IMF components. The first IMF is the highest oscillatory component extracted from data, namely, IMFs are computed by decreasing frequencies [75].

EMD in lung sound literature is employed to decompose crackle sounds from the background vesicular sounds [61, 62]. This method has two drawbacks, namely, end effect and mode mixing effect. End effect generates distortions at the boundaries of the IMFs leading to false oscillations. Mode mixing effect is stemmed from the fact

that each IMF has its own frequency, however crackles and wheezes may have both low and high frequency components. The mode mixing effect causes to decompose some of the components of the crackles and wheezes to another IMF channels which resulted in poor decomposition ability.

3.4.3. Independent Component Analysis (ICA)

Assume,

$$y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \tag{3.8}$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \tag{3.9}$$

$$y_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \tag{3.10}$$

where y is the observed lung sound component, x_1 is the oscillatory waveform, x_2 is the transient waveform, x_3 is the background vesicular (residual) sound and a_{ij} is the random mixing coefficients.

The above mentioned ICA model can be rewritten as follows;

$$y = Ax \tag{3.11}$$

where x is the random vector of statistically independent, unknown sources that have non-Gaussian distribution, A is the unknown random mixing matrix and y is the observed random vector.

The aim of ICA model is to estimate unknown sources x using their linearly mixed measurements y. After the estimation of unknown A, its inverse, W is estimated by $x = Wy$ resulting in estimates of independent components (ICs), unknown sources x [76].

Centering and whitening are the two necessary preprocessing steps in ICA algorithms. Centering means subtracting the mean of the observed data and whitening aims to decorrelate the components of the observed data and to make their variances equal unity.

In this thesis, employed ICA model is linear. According to related work [63], source numbers are assumed to be equal to sensor numbers and mixing is assumed to be linear. Fast ICA estimates independent sources by maximizing negentropy contrast function employing fixed point iterative algorithm [76, 77]. Extended Infomax ICA estimates super- and sub-Gaussian sources by maximizing the entropy and using natural gradient optimization algorithm [78]. ICA has two drawbacks, sign or polarity of the waveform and order of the estimated components may change.

3.5. Experimental Setup

Experiments are conducted in three scenarios. At first, using Equations $3.1 - 3.4$, synthetic adventitious sounds are generated, and then additive white Gaussian noise (AWGN) is injected as vesicular sound for various $(8 - 14)$ dB levels. This scenario is called Method 1 in Tables 3.1, 3.2, 3.3, 3.4. In the second scenario, the generated synthetic sounds are added onto inspiratory vesicular sounds which are obtained from healthy subjects for various $(0 - 6)$ dB levels which are compatible with [61]. This scenario is called Method 2 in Tables 3.1, 3.2, 3.3, 3.4. The variety of the synthetic adventitious sounds is provided by using different frequencies and amplitudes for combination of wheezes and different types (coarse or fine) and amplitudes for combination of crackles. It is ensured that mono or polyphonic and coarse and/or fine types of synthetic signals are included in the generated database. More than one synthetic monophonic wheezes with different amplitudes are added together to generate polyphonic wheezes. In order to prevent suppression of crackles by generated wheezes, only crackles are considered as signal and inspiratory vesicular sound as noise to set signal-to-noise ratio (SNR). In total 88 wheezes and 352 crackles are generated and used in the experiments. After that, by using the generated crackle locations as ground truth, two threshold based algorithms which are MAD and Teager energy operator are

compared with each other automatically, respectively. In the third scenario, proposed system is experimented on the real data from asthma and COPD patients to see its viability. Crackles are sharp bursts and explosive in nature and can be seen as low Q-factor signals; on the other hand wheezes are sinusoidal like waveforms which can be seen as high Q-factor signals. In decomposition, for the candidate crackle containing channel, Q-factor, r and decomposition level are chosen as 1, 3 and 10 while for candidate wheeze containing channel they are chosen as 4, 3, and 30, respectively after various trials.

The threshold is determined in an adaptive manner using the residual channel information which is independent from the transient (crackle) channel. In an automated way, the candidate crackle locations in transient channel are determined by comparing peaks which exceed threshold level using the generated signal locations as ground truth.

In the ICA part, the same above mentioned workflow is followed. The only modification is, to decompose a given lung sound segment into three channels, three measurements are needed. More clearly, unlike TQWT and MCA based decomposition, three linearly mixed measurements are given to the ICA algorithms. Fast ICA ([76, 77]) and extended Infomax ICA ([78, 79]) algorithms are employed to decompose measurements into estimated sources. The measurements are preprocessed using centering and whitening approaches before estimating independent components (ICs). Fast ICA algorithm needs the number of ICs to be estimated as an input parameter and uses symmetric decorrelation approach and non-quadratic function $G(u) = u^3$ [63]. Extended Infomax ICA assumes sub- and super-Gaussian distributions to model mixed measurements [78] and converges using gradient based algorithms. As compared to Fast ICA, extended Infomax ICA has input parameters to tune for decomposition problem such as learning rate, block size, number of iterations and step size. As a result, extended Infomax ICA is more parametric and complicated than fast ICA needing optimization of parameters.

In the EMD part, original EMD algorithm ([75]) is employed. Each decomposition is validated visually. In order to alleviate drawbacks of the EMD, end-effect-seen

parts are discarded from the IMFs and successive IMFs are averaged to reduce mode mixing effect. Individual and averaged successive IMFs are both experimented to represent the decomposition ability of the EMD with and without end-effect-seen parts. Best experimental results are reported in 3.6 Results section.

3.6. Results

The experiments using MAD and Teager threshold methods are represented in Tables 3.1, 3.2, 3.3, 3.4 for proposed TQWT-MCA, fast ICA, extended Infomax ICA and EMD based decompositions, respectively. FN is the term for false negative, namely miss rate, FP is the term for false positive namely false alarm. As shown in Tables 3.1, 3.2, 3.3, 3.4, in Method 1 (Adding AWGN as noise), MAD significantly performs better than Teager operator for both rates. Moreover, 8 dB true positive (TP) rate for MAD is 89.5 % while for Teager 75.6 % besides resulting in more false alarms as can be seen in Method 1 in Table 3.1. MAD in Method 2 (Healthy vesicular as noise) also shows similar performance resulting in lower false alarm and miss rates than Teager as represented in Tables 3.1, 3.2, 3.3, 3.4.

Table 3.1: Localization results of synthetic crackles using MAD and Teager methods for various SNR dB values of AWGN (Method 1) and healthy vesicular (Method 2) sounds as noise using proposed TQWT-MCA method.

		Method 1				**Method 2**			
Threshold	**dB** / **Rate**	**8**	**10**	**12**	**14**	**0**	**2**	**4**	**6**
MAD	**FN**	37	28	8	3	5	2	1	0
	FP	8	3	1	0	11	7	4	3
Teager	**FN**	86	61	29	32	26	14	5	1
	FP	114	45	27	13	48	25	14	16

As represented in Tables 3.1, 3.2, 3.3, 3.4, proposed system has superior crackle (transient adventitious sounds) localization performance than fast ICA, extended Infomax ICA and EMD methods in the presence of wheezes (continuous adventitious sounds). Besides, 8 dB true positive (TP) rate for MAD is 89.5 % using Method 1 and 98.6 % using Method 2 in Table 3.1 while its (proposed method) nearest competitor

EMD method has 87.8 % TP rate for MAD using Method 1 and 94.0 % TP rate for MAD using Method 2 resulting in more false positives as represented in Table 3.4. Extended Infomax ICA and Fast ICA performed 86.1 and 83.8 % TP rate for MAD using Method 1 and 92.1 and 90.1 % TP rate for MAD using Method 2 resulting in more false positives as represented in Tables 3.2, 3.3, respectively.

Table 3.2: Localization results of synthetic crackles using MAD and Teager methods for various dB values of AWGN (Method 1) and healthy vesicular (Method 2) sounds as noise using fast ICA method.

		Method 1				**Method 2**			
Threshold	**dB** / **Rate**	**8**	**10**	**12**	**14**	**0**	**2**	**4**	**6**
MAD	**FN**	57	37	26	20	35	23	14	8
	FP	69	45	33	28	63	35	22	12
Teager	**FN**	93	67	41	37	57	42	27	14
	FP	127	78	51	42	120	75	43	38

Table 3.3: Localization results of synthetic crackles using MAD and Teager methods for various SNR dB values of AWGN (Method 1) and healthy vesicular (Method 2) sounds as noise using Infomax ICA method.

		Method 1				**Method 2**			
Threshold	**dB** / **Rate**	**8**	**10**	**12**	**14**	**0**	**2**	**4**	**6**
MAD	**FN**	49	31	16	11	28	17	9	5
	FP	62	35	19	10	51	29	16	11
Teager	**FN**	84	59	36	34	49	39	23	13
	FP	119	56	43	27	111	71	38	27

Sensitivity and precision rates are calculated to compare the performance of the proposed method with related methods in literature. Sensitivity, namely recall rate is defined as follows; $Sensitivity = TP/(TP + FN)$. Additionally, precision rate is defined as follows; $Precision = TP/(TP + FP)$.

Since MAD performs better than Teager method in all cases, in order to represent precision and sensitivity rates clearly, only MAD results are depicted in Figures 3.1,

Table 3.4: Localization results of synthetic crackles using MAD and Teager methods for various SNR dB values of AWGN (Method 1) and healthy vesicular (Method 2) sounds as noise using EMD method.

		Method 1				**Method 2**			
Threshold	**dB / Rate**	**8**	**10**	**12**	**14**	**0**	**2**	**4**	**6**
MAD	**FN**	43	29	12	7	21	11	7	3
	FP	51	30	17	9	37	23	13	8
Teager	**FN**	87	62	27	28	35	24	14	9
	FP	118	49	36	29	67	43	31	23

3.2, 3.3 and 3.4. In Figures 3.1 and 3.2, the sensitivity rates of four decomposition methods vs. various SNR dB levels are depicted. As can be seen in Figures 3.1 and 3.2, proposed system performs better crackle localization performance then the ICA and EMD methods with increasing noise (at low SNRs) where EMD performs the second best sensitivity rate in the presence of wheezes. As depicted in Figure 3.2, sensitivity rate of the proposed system is more stable than other methods showing the robustness to high noise. The same stability can not be seen in Figure 3.1, because AWGN distorts the waveform of the lung sounds randomly at every frequency value where the power of the healthy vesicular generally concentrate on 200-600 Hz. low frequency range.

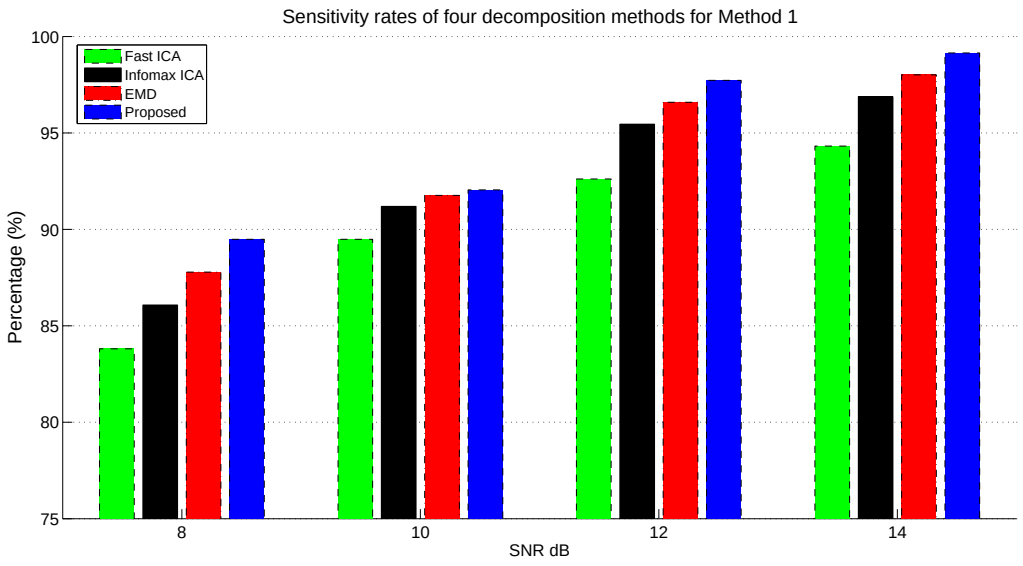

Figure 3.1: Sensitivity rates of four decomposition methods for Method 1

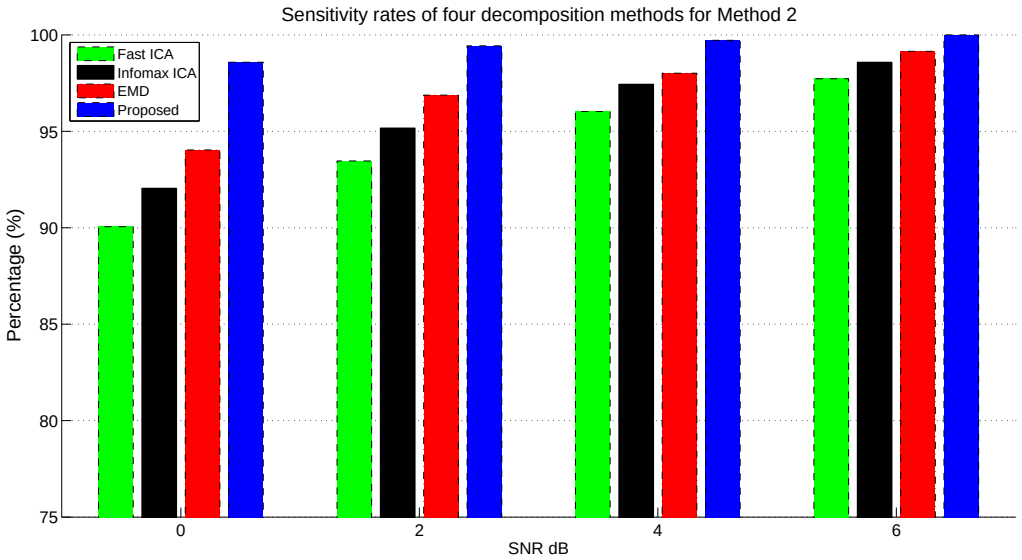

Figure 3.2: Sensitivity rates of four decomposition methods for Method 2

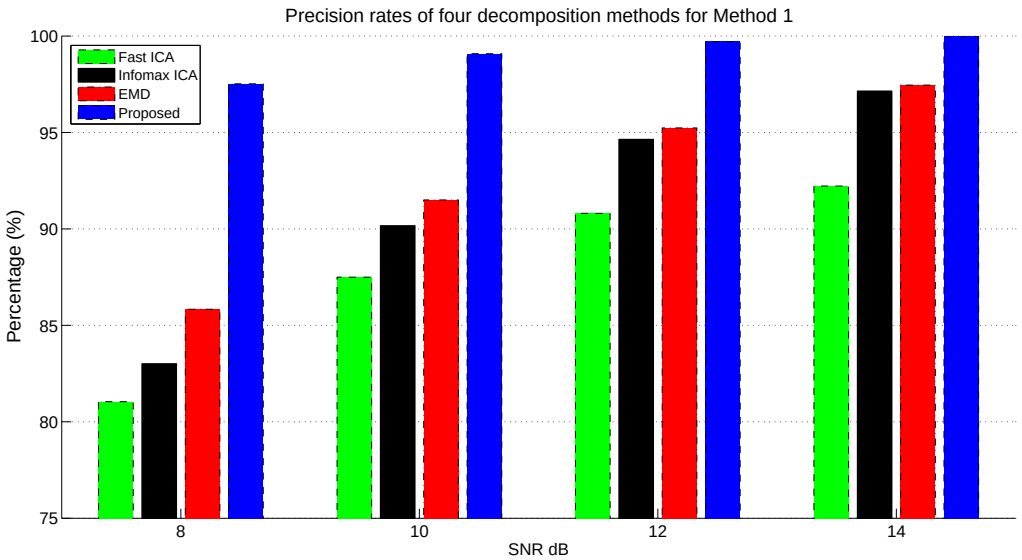

Figure 3.3: Precision rates of four decomposition methods for Method 1

In Figures 3.3 and 3.4, the precision rates of four decomposition methods vs. various SNR dB levels are represented. As can be seen in Figures 3.3 and 3.4, proposed system performs better precision rate performance then the ICA and EMD methods with increasing noise (at low SNRs). Moreover, in Figures 3.3 and 3.4, it is represented that the range of precision rate difference between the proposed method and the other methods is increasing with decreasing SNR dB level (high noise) showing less false

positive (FP) generation ability of the proposed method.

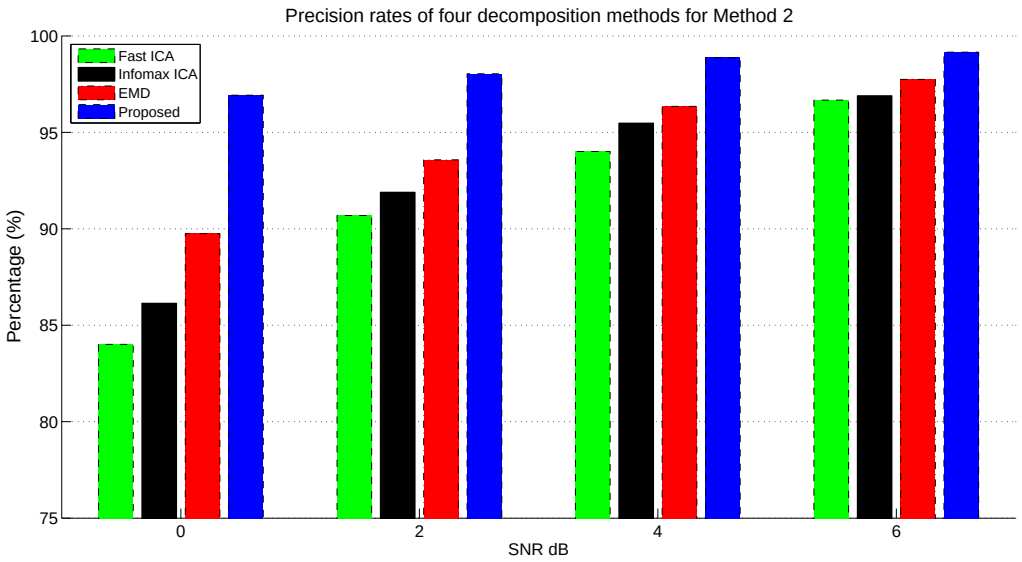

Figure 3.4: Precision rates of four decomposition methods for Method 2

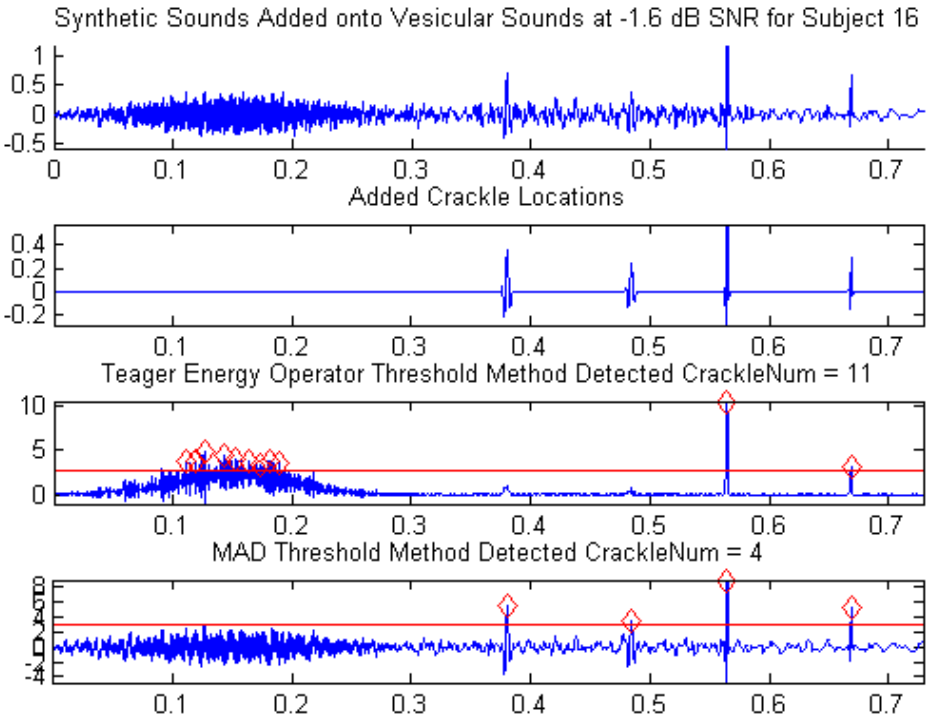

Figure 3.5: Generated sounds added onto vesicular sound of healthy subject at -1.6 dB SNR, lower two sub-figures represent the localization of crackles for Teager and MAD methods using extended Infomax ICA method, respectively.

These results are probably stemmed from the fact that proposed resonance based system introduces neither linear nor frequency based filtering. Unlike ICA which assumes linear mixtures, chest wall and lungs generate convolved nonlinear mixtures whose solution is provided by proposed nonlinear resonance based decomposition technique. When the pictorial representation of extended Infomax ICA decomposition results are inspected in Figure 3.5, it is seen that since either crackles or wheezes have both low and high frequency components, estimated Infomax ICA components could not be able to decompose low frequency components of crackles and wheezes into different channels which is achieved by proposed technique. In Figure 3.6, EMD decomposition result on synthetic adventitious sounds added onto vesicular sounds is depicted. When one compare the same results of Infomax ICA in Figure 3.5, it can be concluded that EMD distorts the crackle waveforms more than Infomax ICA and this produces higher RMS error.

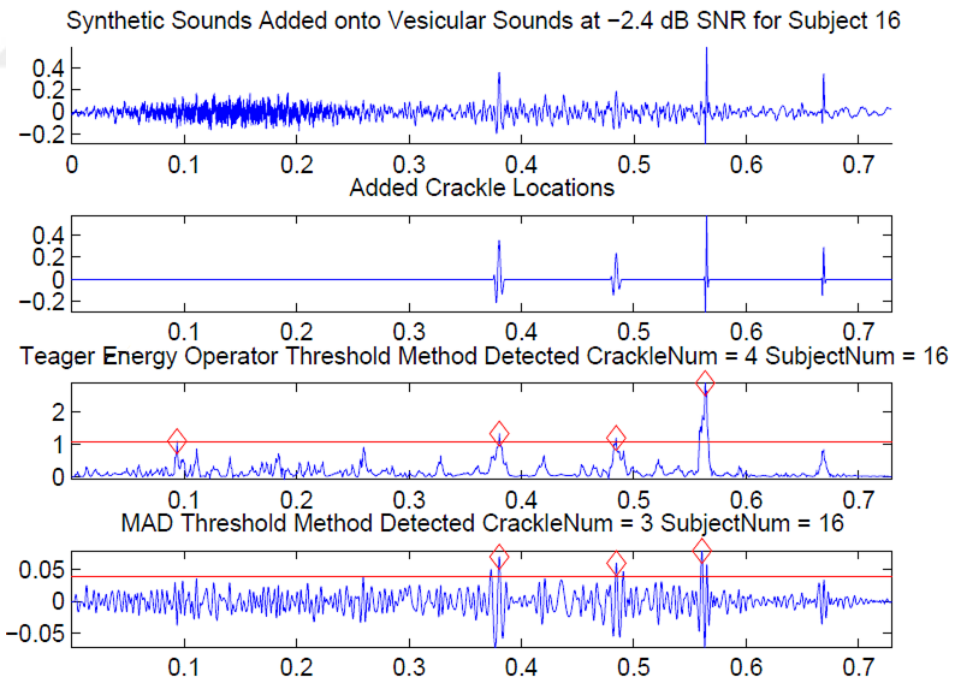

Figure 3.6: Generated sounds added onto vesicular sound of healthy subject at -2.4 dB SNR, lower two sub-figures represent the localization of crackles for Teager and MAD methods using EMD method, respectively.

In Figure 3.7, the performances of MAD and Teager for a low SNR (-2.4 dB) are depicted using proposed method. One can conclude that the lower amplitude middle crackle in Figure 3.7, vanishes and using only time expanded waveform analysis it would be difficult to analyse crackles under low SNR.

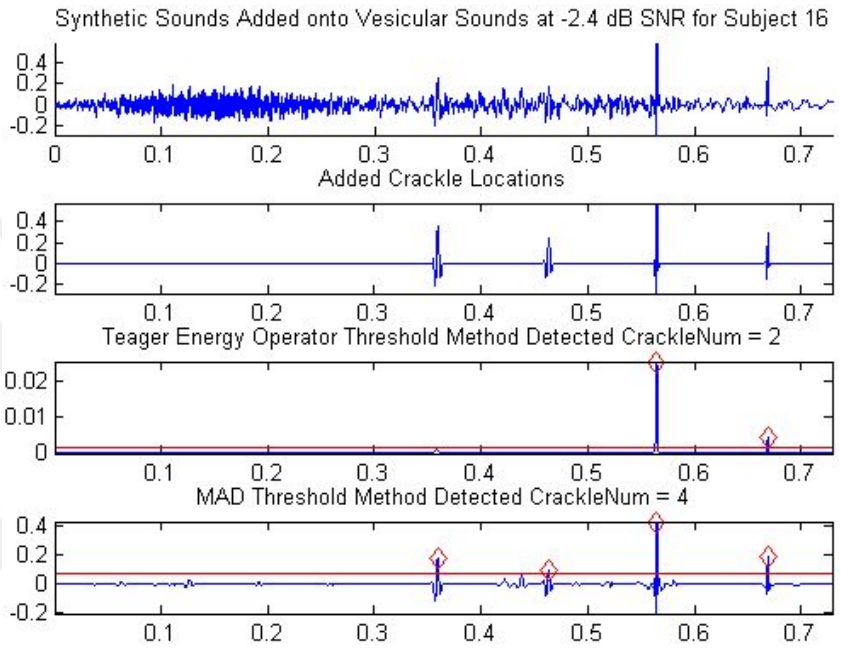

Figure 3.7: Generated sounds added onto vesicular sound of healthy subject at -2.4 dB SNR, lower two sub-figures represent the localization of crackles for Teager and MAD methods using proposed method, respectively.

It can be concluded that the proposed method is able to localize crackles without generating false positives as extended Infomax ICA and without distorting the crackle waveform as EMD methods.

In Figure 3.8, the problem is harder since wheeze and crackle data overlap and the data is collected from a real COPD patient. By looking at Figure 3.8, it can be concluded that resonance based signal decomposition method performs successfully in the real scenario case. However, in Figure 3.9 the decomposition result of ICA method performed poorly, as noticed there are false alarms for crackles and the algorithm misses the crackles at the end of the segment (between 0.3 and 0.35 s) and just before 0.25 s.

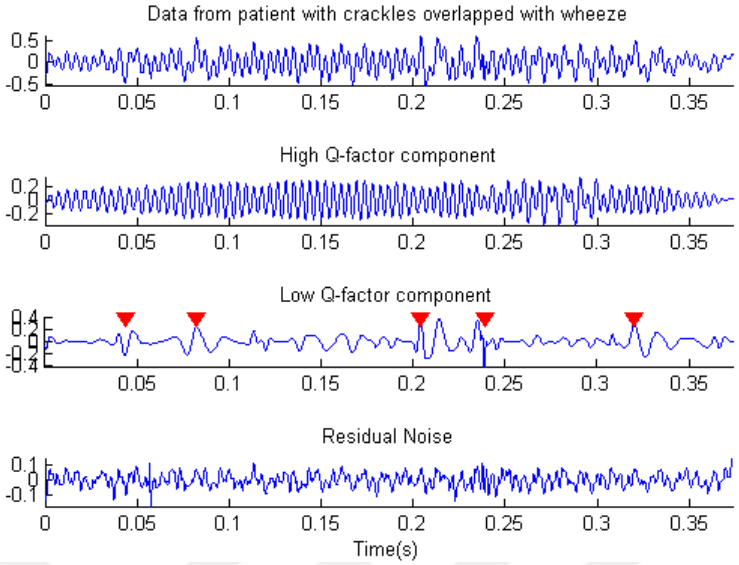

Figure 3.8: Proposed resonance based decomposition result for real patient data with crackles overlapped with wheeze. Crackle locations are determined using MAD method and validated visually.

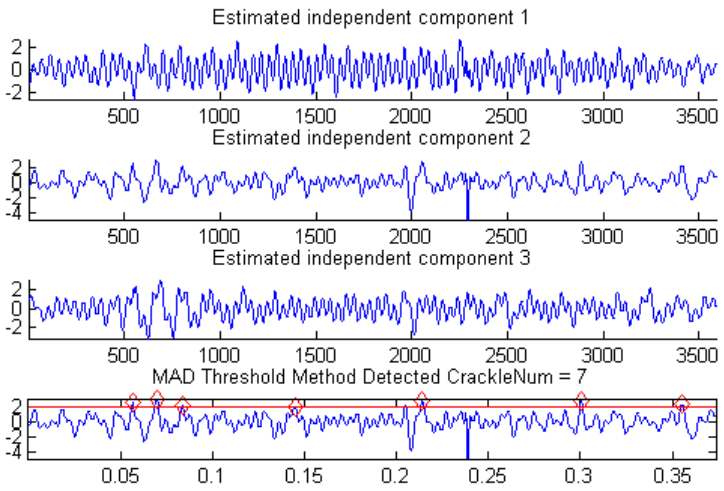

Figure 3.9: Infomax ICA based decomposition result for real patient data with crackles overlapped with wheeze. Crackle locations are determined using MAD method and validated visually.

In Figure 3.10, EMD decomposition result for real COPD patient is depicted. It can be concluded by looking at Figure 3.8 that EMD missed the crackles around 0.05 and 0.25 s and produced one false alarm around 0.3 s which shows the drawback of

the algorithm stemmed from frequency based filtering. It can be seen in Figure 3.11 that since EMD is a frequency based filtering method, its IMFs are varied from high to low frequency oscillating waveforms. In order to recover candidate crackle and wheeze channels, several IMFs are averaged together (as seen in Figure 3.10) or employed separately (as seen in Figure 3.11) and best representing waveform is selected after various experimental crackle localization trials.

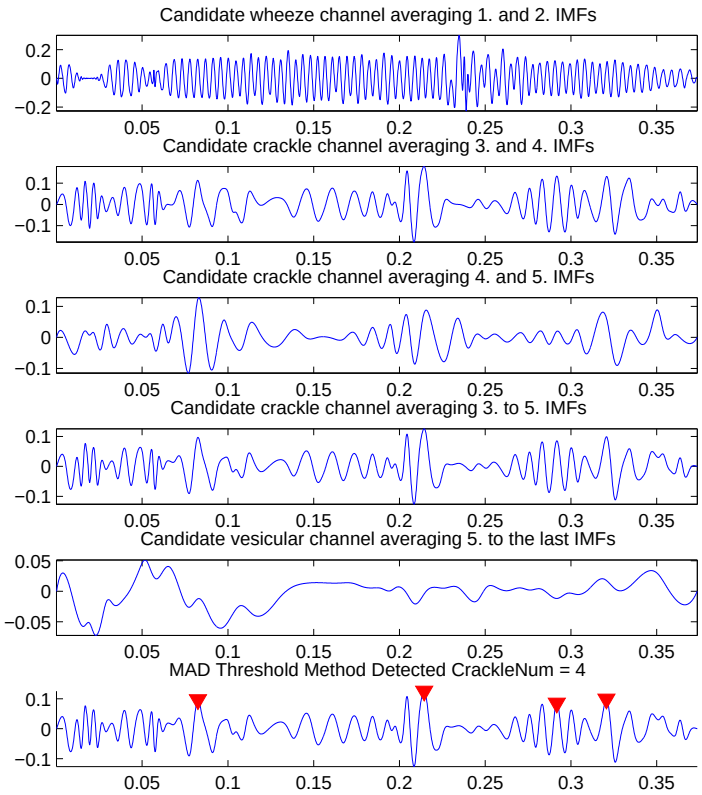

Figure 3.10: EMD based decomposition result for real patient data with crackles overlapped with wheeze. Crackle locations are determined using MAD method and validated visually.

It is seen that by averaging IMFs, candidate crackles and wheezes are better recovered than separate IMFs. Moreover, even when several IMFs are averaged together, the waveform of the crackle is distorted in case EMD is employed. Several crackle parameters (2 CD, LDW) are extracted from time domain crackle waveform which is a very critical step in diagnostic classification may severely be affected by this distortion.

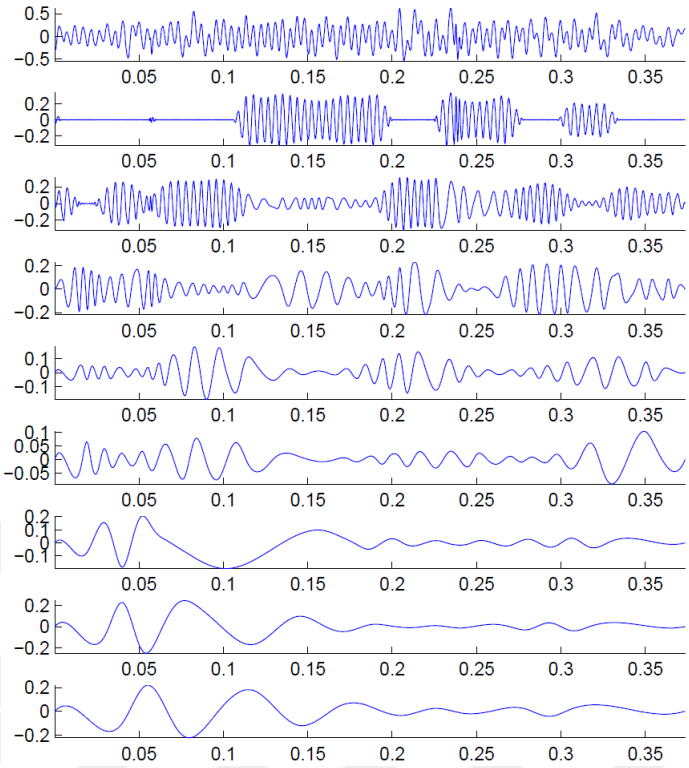

Figure 3.11: Full decomposition result of data from patient using EMD

Root Mean Square Error (RMSE) for proposed TQWT-MCA, ICA and EMD methods is computed to represent the decomposition ability of the methods and to represent the amount of distortion as an objective metric. Ground truth crackle channels are subtracted from estimated crackle channels to represent decomposition ability of the methods when computing RMSE. In Figure 3.12, total RMS errors for decomposition methods are depicted for various noise levels (Method 1 - AWGN). As seen, proposed method achieved the minimum total RMS error under AWGN. In Figure 3.13, total RMS errors for decomposition methods is depicted for various noise levels (Method 2 - Healthy vesicular as noise). As seen, proposed method achieved the minimum total RMS error for Method 2. In Figures 3.12, 3.13 it can be pointed out that RMS error of proposed method is lower when Method 2 is used, because when Method 1 is used, noise is injected randomly at every frequency range unlike low frequency vesicular sound addition in Method 1. Thus, proposed algorithm has difficulty in decomposing AWGN as compared to vesicular sound. As represented in Figure 3.12

while both EMD and Infomax ICA methods showed break even performance under AWGN, in Figure 3.13 Infomax ICA showed better performance than EMD in terms of total RMS error.

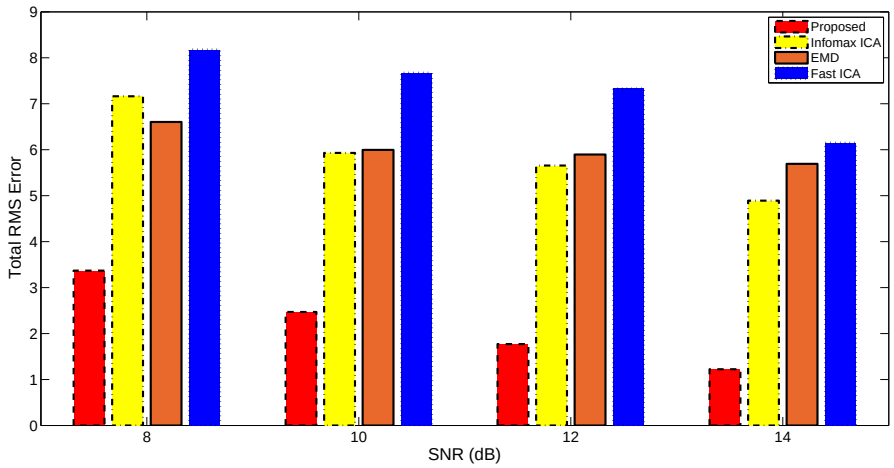

Figure 3.12: Total RMS errors of four decomposition methods for Method 1

As represented in Figures 3.1, 3.2, 3.3 and 3.4, EMD has better sensitivity and precision rates than ICA based methods, however, as shown in Figures 3.12 and 3.13, EMD has poorer RMSE performance as compared to Infomax ICA. One reason for this is since EMD is a frequency based decomposition method some of the frequency components of the crackles are decomposed into another channels and the waveform of the crackle is severely distorted resulting in higher RMS error as compared to Infomax ICA method. Since wheeze and crackle sounds have both high and low frequency components and the mixture is non-linear, neither Infomax ICA nor Fast ICA may be able to decompose low and high frequency components of the lung sound into the targeted candidate wheeze or crackle channel. Moreover, as represented in Figures 3.12, 3.13 Fast ICA algorithm showed the worst performance.

In Table 3.5, crackle localization performances of the EMD, Fast ICA, Infomax ICA and proposed method are represented in terms of sensitivity and specificity rates on 135 lung sound segments obtained from real patients. It is seen that proposed method is more successful than related works in both metrics. During the experiments

it is seen that when the crackles are superimposed onto wheezes it is hard to decompose crackles because of the waveform deformation even when the proposed method is used. Moreover, on real life recordings, there may be crackle like noises which degrades the performance of the system by generating false alarm.

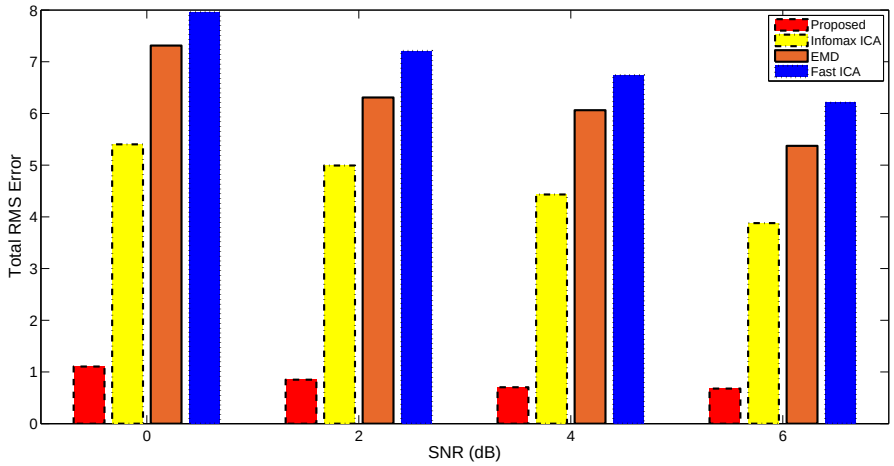

Figure 3.13: Total RMS errors of four decomposition methods for Method 2

Table 3.5: Performance comparison of proposed method with related methods on crackle and wheeze containing real patient data using MAD threshold. Sensitivity and precision rates are given as %.

Rate / Method	Sensitivity	Precision
EMD	81.5	75.3
Fast ICA	77.8	70
Infomax ICA	79.3	72.3
Proposed	85.9	80.6

Crackles and wheezes in a breath cycle may exist successively or overlapped [11, 68], or only one of them may appear which is more likely. The robustness of the proposed algorithm on overlapped or successive cases is represented by visual validation, RMSE calculation and in terms of crackle localization accuracy. The proposed algorithm is also tested on separate cases (only crackle or only wheeze containing cases). In Figures 3.14, 3.15 the decomposition ability of the algorithm on only crackle cases is depicted.

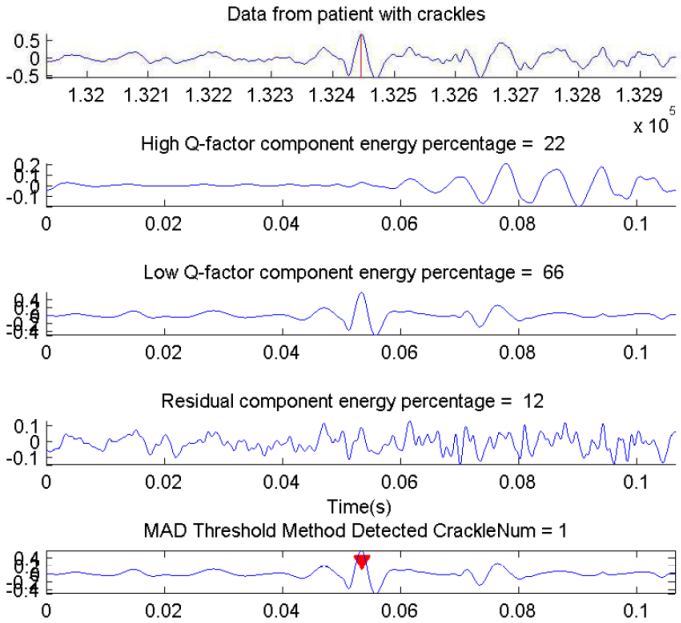

Figure 3.14: Decomposition result of only crackle data from patient using proposed method

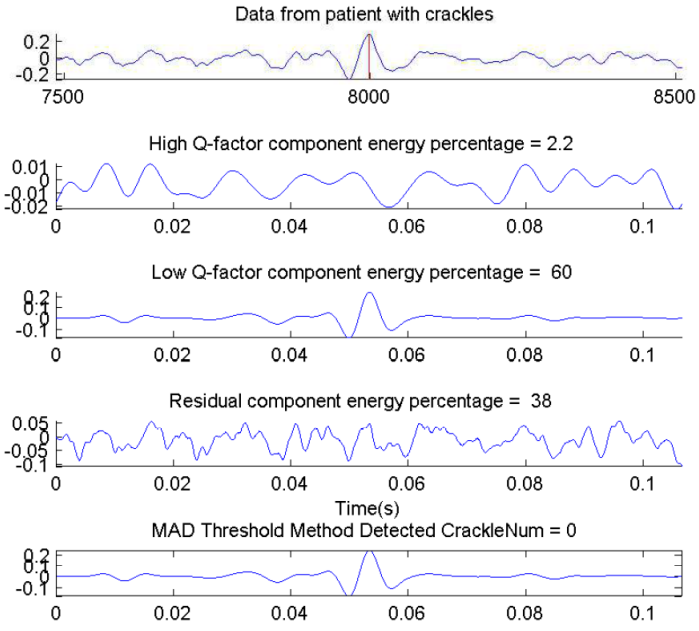

Figure 3.15: Decomposition result of only crackle data from patient using proposed method

As seen at the top of each figures ground truth location of the crackle is marked. The energy of each candidate channel is also computed and shown that when the lung

sound only contains crackle component, the energy of the candidate crackle channel is higher (66 % and 60 % of the energy in Figures 3.14, 3.15, respectively).

In Figures 3.16, 3.17 the decomposition ability of the proposed algorithm on only wheeze cases is depicted. It is represented that since wheezes have high energy and oscillating waveform most of the energy (83 % of the energy in both figures) is concentrated on candidate wheeze channel.

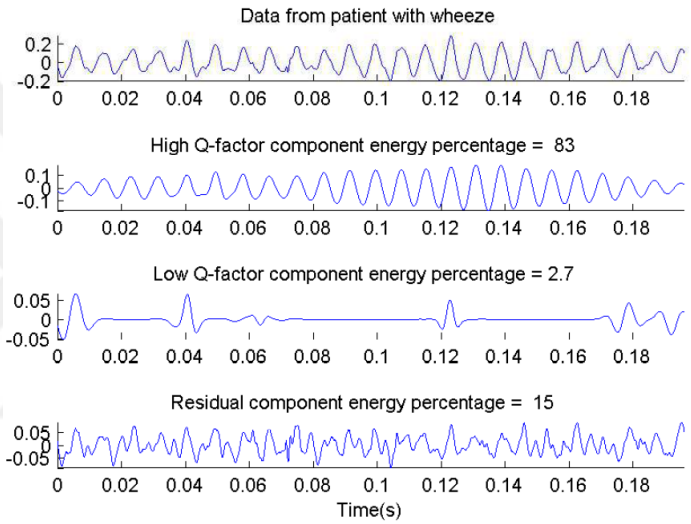

Figure 3.16: Decomposition result of only wheeze data from patient using proposed method

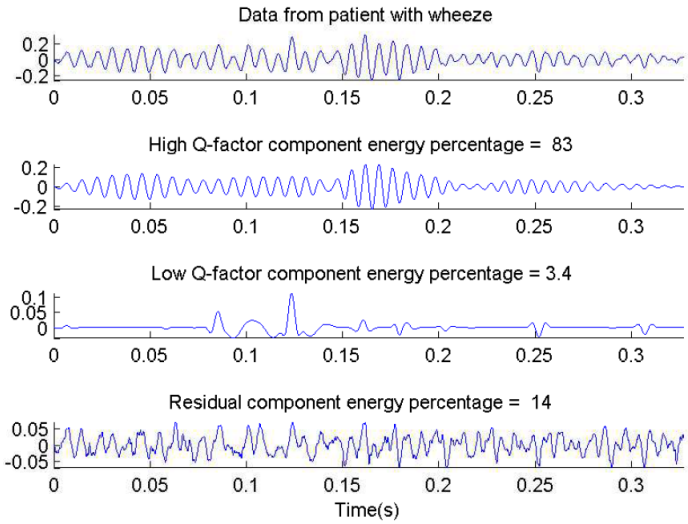

Figure 3.17: Decomposition result of only wheeze data from patient using proposed method

The performance of the proposed algorithm is also tested on the healthy vesicular (background) sound and the decomposition result is depicted on Figure 3.18. Although most of the energy is concentrated on vesicular channel, it is hard to say that the proposed algorithm is also successful in decomposing vesicular only data.

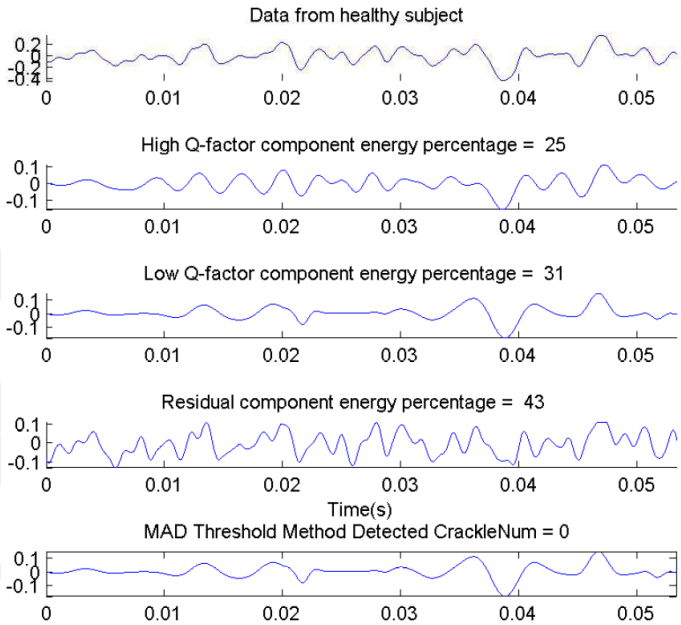

Figure 3.18: Decomposition result of only vesicular data from healthy subject using proposed method

In Tables 3.6, 3.7, the decomposition ability of the proposed method on patient lung sound segments containing only crackle and only wheeze adventitious sounds is explored in terms of energy of the decomposed candidate channels, respectively.

In Table 3.6, mean and standard deviation (std) of energy of the candidate channels and energy distribution based classification accuracy of proposed method on 330 lung sound segments is represented. As seen mean of the energy distribution of the candidate crackle channels is 65.9 % of the total energy when the segment contains only crackle. When maximum of the energies of decomposed channels is used to estimate the label of the given segment, 88.2 % accuracy is reached on only crackle containing segments.

Table 3.6: Energy distribution (in %) of each candidate channel and energy based classification accuracy (in %) of proposed method on only crackle containing segments.

	Crackle Only		
Channel / Rate	**Crackle**	**Vesicular**	**Wheeze**
Mean	65.9	20.7	13.4
Std	17.4	17.9	9.9
Accuracy	88.2	11.2	0.6

In Table 3.7, mean and standard deviation values of the energies of each decomposed channel are represented. As represented, 76.7 % of the mean energy is concentrated on decomposed wheeze channel showing the effectiveness of the proposed method on 231 wheeze segments. Moreover, when maximum of the energy of the decomposed channels is selected to estimate the label of the given only wheeze containing segment, 93.1 % accuracy is reached.

Table 3.7: Energy distribution (in %) of each candidate channel and energy based classification accuracy (in %) of proposed method on only wheeze containing segments.

	Wheeze Only		
Channel / Rate	**Crackle**	**Vesicular**	**Wheeze**
Mean	11.8	11.5	76.7
Std	12.9	15.6	18.5
Accuracy	2.1	4.8	93.1

3.7. Discussion and Summary

In this chapter, respiratory sounds are decomposed into candidate crackle, vesicular (background) and wheeze channels simultaneously without any prior information about respiratory sound types. The proposed resonance based TQWT-MCA algorithm represents better performance than related methods in respiratory sound decomposition (EMD ([61,62]), Fast ICA ([63]) and Infomax ICA ([63])) on both synthetic and

real patient data. Performance of the methods is compared using RMS error, crackle localization accuracy (specificity and sensitivity) and visual validation metrics. Two thresholding methods are employed and MAD outperforms Teager operator method.

Morphological properties of the crackles and wheezes represent adverse time domain behaviors. The proposed method is able to decompose continuous (wheeze) and discontinuous (crackle) respiratory sounds by tuning the wavelet bases to highlight signal of interest. Then MCA is able to decompose low and high frequency components of the signal of interest into the same candidate channel by exploiting the morphological differences obtained using TQWT method. Since the adventitious lung sound generation mechanism is affected by lung tissue and chest wall, linear or frequency based filtering methods are less successful in modeling the underlying pulmonary sound types with linearity assumption. The experimental results showed that extended Infomax ICA performed better than Fast ICA method coherent with the study in [63]. ICA based methods have some drawbacks that decrease the performance of the model. When the sign (polarity) and the order of ICs are changed this will result in higher RMS errors and possible false detection generation. Crackle polarity and time domain parameters are helpful and important in diagnosing pulmonary diseases [80, 81].

EMD has also some drawbacks that degrade the overall system performance. Mode mixing and end-effect generate false detections and result in higher RMS errors. Although mode mixing effect is tried to be removed using ensemble IMFs and end-effect parts removed in the experiments, EMD generates higher RMS error by deforming the waveform of the crackle than extended Infomax ICA and proposed method.

In order to be employed in a diagnostic pulmonary disease classification system, a method should provide automatic and simultaneous decomposition ability, less RMS error and higher accuracy. As concluded in [61], EMD based method was not automatic. According to [62], the proposed EMD based filter dealt with explosive (transient) lung sounds disregarding the cases that wheezes may be present in the lung sound data. When the lung sound recording includes wheeze and crackle types successively or overlapped on the lung sound data, this will change the order and number of the

IMFs that are used to recover the adventitious lung sound only. Moreover, when the included wheeze is mono or polyphonic, this condition will also change the EMD model formation to recover signal of interest. The same problem is valid for the ICA based algorithms because the order of the estimated ICs may be changed depending on the lung sound types. Blind and automatic decomposition is exactly supported by the proposed method being independent from whether the lung sound component includes wheeze only, crackle only and both wheeze and crackle containing cases without using any prior information. Therefore, the proposed method is a robust and successful candidate to be employed in a diagnostic pulmonary disease classification system as a preprocessing and detection unit when compared with related methods in literature.

4. ADAPTIVE WHEEZE TYPE CLASSIFICATION

4.1. Introduction

Auscultation with a traditional stethoscope is a convenient tool with low diagnostic value since it has limited frequency response, attenuating frequencies above 120 Hz [11], is subjective in nature, depending on the physician's expertise, and is unable to record sounds for further analysis. The need for a patient specific clinical decision support system has become vital during recent years in the diagnosis of lung disorders and for diminishing healthcare expenses [82].

Lung sounds may be categorized into two basic groups: vesicular and adventitious sounds. Adventitious sounds which are usually indicators of various lung diseases are either discontinuous , i.e. crackles, or continuous, i.e. wheezes. Unlike crackles, wheezes are musical and continuous in nature and have narrow representations in frequency domain.

A lung sound segment is accepted as wheeze according to American Thoracic Society (ATS) and Computerized Respiratory Sound Analysis (CORSA) if its main frequency is higher than 400 and 100 Hz and its duration is longer than 250 and 100 ms, respectively [83,84]. On the other hand, in the studies of [12,85], reported minimum duration is 80 ms.

Wheezes are closely related with diseases such as asthma and chronic obstructive pulmonary disease (COPD) [19]. The severity of the disease may be related to the duration, number and main frequency of wheezes within a respiration cycle [18, 19]. Monophonic (MP) wheezes comprised of either single pitch frequency or multiple pitch frequencies starting and ending at different times stem from single bronchial narrowing and may related with asthma [31, 86]. Polyphonic (PP) wheezes composed of harmonically unrelated multiple pitch frequencies starting and/or ending simultaneously originate from multiple central bronchial compression and are commonly related with

COPD [31, 86]. A MP and a PP wheeze sample time-frequency (TF) domain may be depicted in Figure 4.1. Despite advances made in the analysis of lung sounds, discrimination of multiple MP and PP wheezes is still an open problem [87] since both of them are sinusoidal.

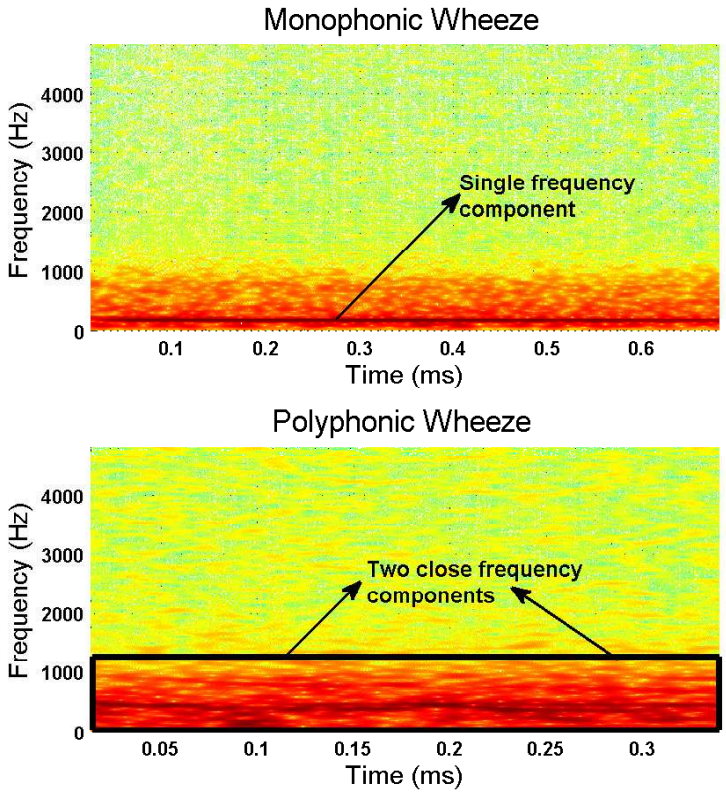

Figure 4.1: Time-frequency domain representation of MP (top) and PP (lower) wheezes (Best viewed in color).

In [31], it is reported that there are statistically significant differences between MP and PP wheezes of the same pathology (asthma or COPD) using wavelet based features, paving the way for classification studies. In literature very few studies can be found in MP-PP wheeze classification. In [88], nine monophonic-polyphonic wheezes are detected using spectrogram based peak continuity resulting in 89 % accuracy. In [89], 92 % F_1 score is reached using dominance spectrogram based on instantaneous frequency on normal, monophonic, polyphonic and stridor classes whereas 72 % using the classical spectrogram on 155-70 wheezes. A recent work [90] using time domain based higher order statistics reached 91 % classification accuracy using 102 wheezing sounds.

Our previous study [8] on this classification problem has led us to explore robust and discriminative features for PP wheezes. In this thesis, unlike previous studies which use fixed TF resolution based on Fourier transform, we propose to determine an optimal (better TF resolution) and adaptive (automatic and tunable) wavelet based technique to discriminate MP and PP wheezes in a more robust and objective manner. Properties of the database are described in Section 4.2, while Section 4.3.1 gives details of the wavelet based method and Section 4.3.2 introduces the proposed method, respectively. Section 4.4 and Section 4.5 contain experimental results and discussion, respectively.

4.2. Data Acquisition and Database

The 14-channel data acquisition system [30] designed at Boğaziçi University Lung Acoustics Laboratory (BU-LAL) was utilized to record wheeze sounds. Sampling rate was 9600 Hz and each data recording session lasted 15 seconds. Each subject had a nose clip and a flow-meter was employed to measure airflow. An informed consent was taken from all subjects before data acquisition. The data acquisition procedure had the consent of the second Ethical Committee on Clinical Research of Istanbul (in accordance with the Declaration of Helsinki). Wheeze sound was collected from asthma and COPD patients who were under treatment at the Istanbul Yedikule Teaching Hospital for Chest Diseases and Thoracic Surgery. Database was comprised of seven male and four female subjects at the age of 52 ∓ 19. Wheeze sounds were labeled by visual verification of time expanded waveforms and auditory inspection by an expert. The database consisted of 147 MP and 153 PP wheezes, where the duration of each segment is at least 80 ms being consistent with literature.

4.3. Proposed Adaptive Techniques

4.3.1. Rational Dilation Wavelet Transform (RADWT)

According to the definition, wheezes that occur with a single peak or with the harmonics of a single basal peak are called MP wheezes, while those with variable peaks that differ in harmonics are called PP wheezes [87]. Due to the similarities between MP

and PP wheezes, discriminating multiple MP wheezes from PP wheezes is still an open and important task. In MP wheezes when the severity of pathology is very strong, a fundamental (basal peak with high energy) signal can occur with accompanying harmonics (peaks with lower energy). This MP pattern may be confused with PP wheezes in which various peaks with relatively close energies show up. In order to discriminate the MP and PP wheezes in time-scale domain, a wavelet transform, in which the frequency selectivity of the sub-bands can be adjusted, is needed. Therefore in this thesis, the RADWT [29], which has finer and adjustable frequency resolution with acceptable redundancy, is proposed as a suitable feature extractor for processing wheeze sounds.

The RADWT [29] is a frequency-domain (FFT based) design transform which does not employ rational transfer functions and offers greater design flexibility. Moreover, the RADWT is a rational (based on non-dyadic dilations), fully discrete, approximately shift-invariant and easily invertible transform. The non-dyadic (rational) behaviour of the RADWT yields to attain a range of Q-factors and redundancy factors. In the RADWT, the Q-factor of wavelets, which controls the frequency resolution of transform, is built upon three positive integers p, q and s satisfying $1 \leq p < q$ and $p/q + 1/s \geq 1$, where p and q are co-prime.

In RADWT, the relation between the scaling ($\phi(t)$)/wavelet ($\psi(t)$) functions and the low ($h_0(n)$)/high ($g_0(n)$) pass filters can be given as,

$$\phi(t) = (q/p)^{1/2} \sum_{n \in \mathbb{Z}} h_0(n) \phi\left(\frac{q}{p}t - n\right) \tag{4.1}$$

$$\psi(t) = (q/p)^{1/2} \sum_{n \in \mathbb{Z}} g_0(n) \phi\left(\frac{q}{p}t - n\right) \tag{4.2}$$

Mathematically, the frequency responses of $h_0(n)$ ($H_0(\omega)$) and $g_0(n)$ ($G_0(\omega)$) are given as,

$$H_0(\omega) = \begin{cases} \sqrt{pq} & \omega \in [0, (1-\frac{1}{s})\frac{\pi}{q}] \\ \sqrt{pq}\theta(\frac{\omega-a}{b}) & \omega \in [(1-\frac{1}{s})\frac{\pi}{q}, \frac{\pi}{q}] \\ 0 & \omega \in [\frac{\pi}{q}, \pi] \end{cases} \tag{4.3}$$

and

$$G_0(\omega) = \begin{cases} 0 & \omega \in [0, (1-\frac{1}{s})\pi] \\ \sqrt{s}\theta_c(\frac{\omega-pa}{pb}) & \omega \in [(1-\frac{1}{s})\frac{\pi}{q}, \frac{p}{q}\pi] \\ \sqrt{s} & \omega \in [\frac{p}{q}\pi, \pi] \end{cases} \tag{4.4}$$

where

$$a = \left(1-\frac{1}{s}\right)\frac{\pi}{p}, b = \frac{1}{q} - \left(1-\frac{1}{s}\right)\frac{1}{p} \tag{4.5}$$

the transition function $\theta(\omega)$ is,

$$\theta(\omega) = \frac{1}{2}(1+cos(\omega))\sqrt{2-cos(\omega)} \quad \text{for} \quad \omega \in [0, \pi] \tag{4.6}$$

and $\theta_c(\omega)$ is

$$\theta_c(\omega) := \sqrt{1-\theta^2(\omega)} \tag{4.7}$$

The transition function, $\theta(\omega)$, which is used to construct the transition bands of $G_0(\omega)$ and $H_0(\omega)$, originates from Daubechies' orthonormal wavelet filters with two vanishing moments. As it can be seen from above equations, the bandwidth, center frequency and transition bands of high-pass and low-pass filters are determined by using p, q and s values. As the q/p ratio approaches one, higher number of decomposition levels are needed. Therefore, the number of subbands (J) must also be considered as an important parameter in analysis.

4.3.2. Adaptive Peak-Energy-Ratio Parameter Selection Method

A single peak and at least two peaks must be obtained when the time-scale representations of MP and PP wheezes are investigated respectively. However, the location, amplitude and bandwidth of these peaks differ for each sample due to the physiological properties of the lung and the mechanism of the pathology. This results in a need for an adaptive and automatic algorithm that can locate peaks in TF domain for processing MP and PP wheezes. In the proposed algorithm, the RADWT is applied to MP and PP wheezes by using a set of various p, q, s and J values, which are given in Table 4.1 with an aim to achieve an optimum representation. In this sense optimum representation is defined as two distinct and non-consecutive peaks, where one belongs to basal peak in MP wheezes or first peak in PP wheezes and the other belongs to weak harmonics in MP wheezes or second peak in PP wheezes. Then a metric named as the peak-energy-ratio (PER) is defined as,

$$\text{PER} \quad = \quad \frac{\text{Energy of first peak}}{\text{Energy of second peak}} \tag{4.8}$$

In Figure 4.2, energy distribution of wheezes given in Figure 4.1 across sub-bands and peaks are represented. The RADWT is applied to a wheeze with one of 22 various p, q, s and J combinations as given in Table 4.1. For each set, two distinct and non-consecutive peaks are found and the PER is calculated. As a result, for one signal, 22 different PER values are obtained. The minimum PER value is selected as the indicator

of best representation because it means that two peaks are correctly located while preserving maximum amount of their energies. In order to quantify the performance of the proposed method, the chosen minimum PER metrics are employed as features for discriminating MP and PP wheezes. Support Vector Machines (SVM) [48] is used as the classification method with leave one out cross validation scheme in a grid search for parameter (C, γ) optimization.

4.4. Results

In Figure 4.2, for the same MP and PP wheezes given in Figure 4.1, the two distinct peaks and corresponding p, q, s and J values are presented. It is seen that for MP wheezes high PER values are obtained. In contrast, for PP wheezes, relatively small PER values are obtained. In Figure 4.3, whisker plot of PER values with respect to wheeze types is depicted. As represented, the median values of PER metrics for monophonic and polyphonic wheezes are 22.01 and 2.41, respectively.

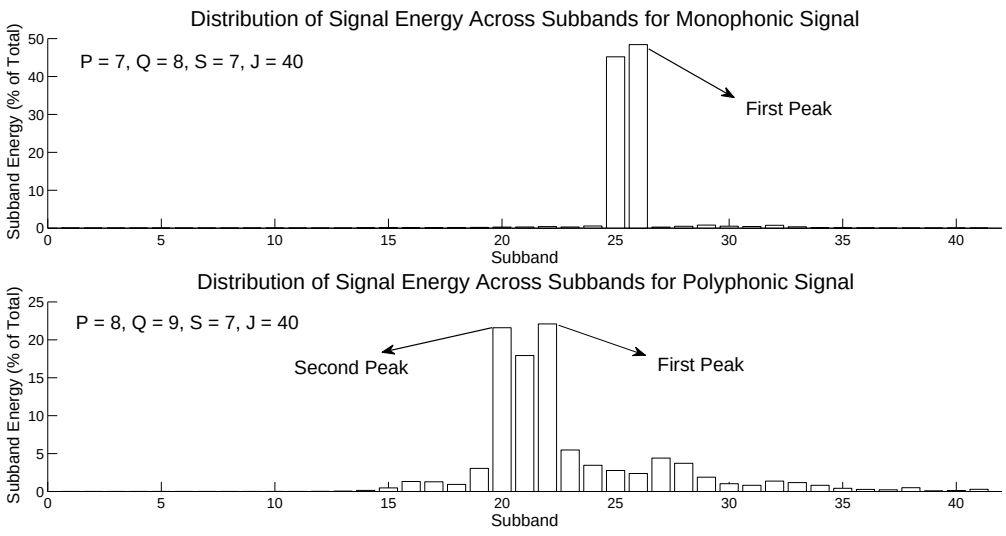

Figure 4.2: Energy distribution of MP and PP wheezes

During the experiments with the proposed algorithm the distribution of p, q, s and J values (parameter set) related with selected minimum PER metric is investigated. It is seen that, in order to achieve the best TF representation, which is obtained with the algorithm given in Section 4.3.2, a specific parameter set can not be obtained.

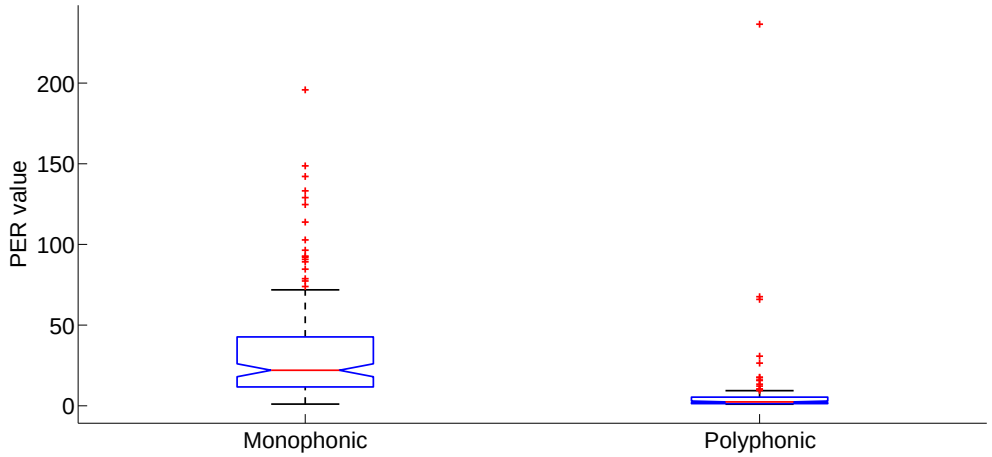

Figure 4.3: Comparison of PER values with respect to wheeze types when the optimum parameters employed.

In Figure 4.4, the percentages (number of samples for a specific parameter set/ number of total samples) for the distribution of total, MP and PP wheezes changing with various p, q, s and J values are given. The order of p, q, s and J values used in Figure 4.4 is the same with the order of Table 4.1. For example the first set number in Figure 4.4 corresponds to the first column in Table 4.1 ($p = 2, q = 3, s = 2, J = 8$). It is seen that, for almost all parameter sets, except the first set (low Q-factor, low frequency resolution), optimum TF representation for a wheeze sample is achieved. Namely, the percentages (number of samples for a specific parameter set/number of total samples) are spread over all 22 parameter sets given in Table 4.1 when the proposed method is used as depicted in Figure 4.4.

In Table 4.2, classification accuracy of fixed p, q, s and J parameter set (the same order as in Table 4.1) on the whole dataset is given without optimal choices. For example, fixed set of p, q, s and J values ($p = 2, q = 3, s = 2, J = 8$) are experimented on the whole dataset. It is shown that best classification accuracy (81.7 %) is obtained with 19. parameter set in Table 4.1 using SVM radial basis function (RBF) kernel. This is because in Table 4.1 on the left-side the relatively low Q-factor combinations and on the right side relatively higher Q-factor combinations are represented which means frequency resolution increases from left to right.

Table 4.1: Various p, q, s and J values used in analysis.

Set #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
p	2	3	4	5	5	5	6	6	7	7	7	8	8	8	8	8	9	10	10	10	11	11
q	3	4	5	6	6	6	7	7	8	8	8	9	9	9	9	9	10	11	11	11	12	12
s	2	2	3	4	5	5	5	6	5	6	7	3	4	5	6	7	5	6	7	8	7	8
J	8	10	15	20	25	30	30	35	35	35	40	35	35	35	35	40	35	40	45	50	45	50

Table 4.2: Classification results (in %) of fixed p, q, s and J parameters using support vector machines with different kernels.

Kernel Type	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
Linear	56	50.7	57	54.3	50.7	56.3	54.3	50.3	52.7	51.3	74.7	60.7	63	61.7	54	54.3	56	76.7	76.3	76	57	56
RBF	56	54.3	62	72.3	76.3	77.3	75	79	77.3	76.3	79.3	75.3	79.3	76.3	79.3	77.7	75.7	80.3	81.7	80	76.3	75

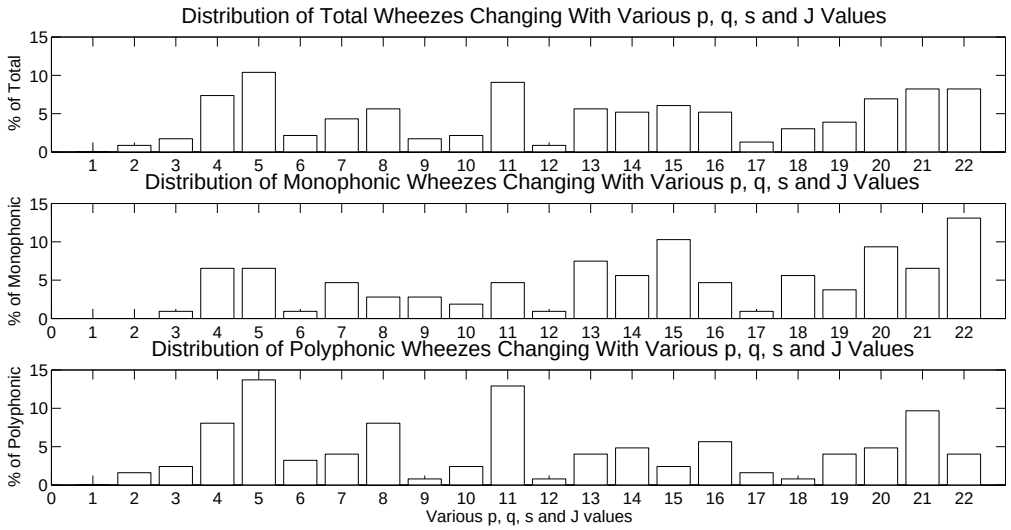

Figure 4.4: Distribution of total, MP and PP wheezes with respect to various p, q, s and J values

In Figure 4.5, frequency responses and wavelet time domain representations at several scales for low-Q (limited frequency resolution) and high-Q (higher frequency resolution) is depicted. It is represented that in the right lower part of the Figure 4.5, the waveform of the wavelet is more oscillatory than in the left lower part of the figure. As noticed, when high Q-factor filters are used, oscillatory wavelets similar to wheeze signals in the time domain and better frequency resolution for low and middle frequency bands in frequency domain can be achieved. Therefore, it is concluded that an adaptive and automatic system is needed for optimum localization of different peaks due to subject specific TF properties of wheezes. Additionally, the accuracy of SVM classifier is obtained with the proposed method as 82.6 % and 86 % when the linear and RBF kernels are employed, respectively. This shows that PER metric can be used as an indicator for discriminating MP and PP wheezes and better TF representation can be achieved with the proposed method when compared to fixed parameter setting. Our previous study [8] on MP-PP classification problem had reached 75.8 % overall accuracy when time domain and Fourier transform based features were combined. When compared with these findings it is clear that proposed adaptive wavelet based method outperforms Fourier transform based method which has fixed TF resolution.

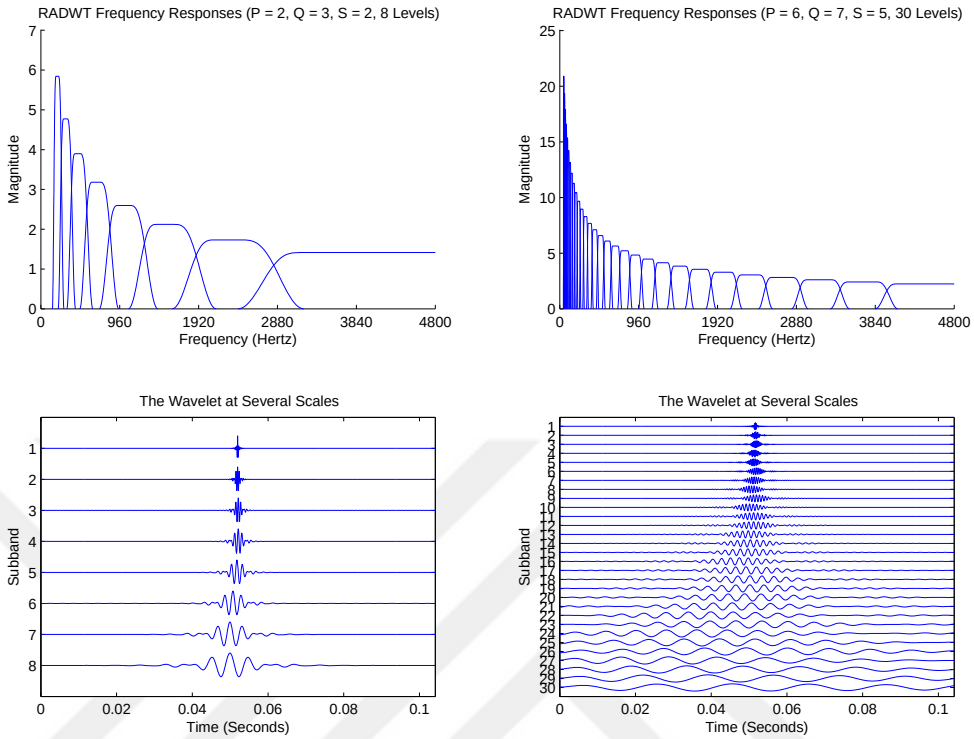

Figure 4.5: The wavelets at several scales and corresponding frequency responses used in low Q-factor (left) and high Q-factor (right) analysis.

4.5. Discussion and Summary

Wheeze type classification is an important problem in diagnostic lung disease classification since it may be related with Asthma or COPD. It is reported that Asthma or COPD affects 1 in 12 people around the world and these two lung diseases may overlap on 15 % of the obstructive lung disease population [2].

Wheeze/non-wheeze classification problem is extensively studied in literature as summarized in [8]. Unlike wheeze/non-wheeze classification, MP-PP wheeze classification is a within class problem since both of the classes are oscillatory and periodic. From this point of view, it is clear that time-frequency resolution must be high enough to exploit the discriminative characteristics of the wheeze types. For example, zero

crossing based time domain techniques could not be able to capture the discriminative features because of the periodicity of the wheeze types.

As detailed in the introduction part, the database used in this thesis is at least twice extensive as in related works in literature. Although it is not possible to compare related works with the proposed method due to issues on data privacy and publicly unavailable implementations, it can be argued that proposed method achieved better TF resolution than Fourier and time-domain based methods compared to our previous results [8]. Moreover, our method performed better than approaches where fixed parameters are employed in wavelet analysis. For example, dyadic discrete wavelet transform which is a low-Q factor transform is not suitable for this type of problem because of the limited frequency resolution. When the fixed parameter setting is employed the best obtained accuracy is 4 % lower than the proposed adaptive method. Classification accuracy of confusion matrix on MP and PP wheezes is 81.6 % and 90.9 % respectively demonstrating the effectiveness of the proposed method in optimal TF representation of PP wheezes. Our previous study based on time domain and Fourier transform suggested the exploration of stronger features to better represent distinctive characteristics of polyphonic type of wheezes.

RBF kernel gives better results than linear kernel, it was suggested to use non-linear kernel when the number of features (one PER value) is small [48]. It is seen that, when the linear kernel is employed computation load is at least ten times higher than RBF kernel.

In this thesis, non-dyadic wavelet based automatic and adaptive method is proposed to deal with MP-PP wheeze classification problem providing better TF resolution as compared to fixed parameter setting. Moreover, it is shown that proposed adaptive non-dyadic wavelet based method is successful than dyadic wavelet transform.

5. CONCLUSION AND FUTURE PERSPECTIVES

Lung sounds provide important clues about the underlying lung dysfunction. The classical stethoscope has some limitation such as attenuating frequencies above 120 Hz, being unable to record lung sounds for remote analysis and depending on the medical physician's expertise and experience. The need for an automatic and cheap analysis tool has become vital during recent years in the diagnosis of lung diseases especially in low and middle income countries. The cooperation between engineering and medicine produced electronic stethoscope products which benefit medical doctors by providing objective measurements however, there is a lot of research to be done. Electronic stethoscopes are not cheap for personal use and need to be smart including advanced algorithmic solutions. Faster and smart solutions can be obtained by implementing novel proposed wavelet based algorithms on field programmable gate arrays or chips. The proposed algorithms can also be implemented in differential diagnosis systems that are employed in medical decision support systems. Detection of crackle and wheeze sounds or abnormalities within a breath cycle significantly plays a vital role in differential diagnosis. For example, when a localized wheeze is missed, this results in misdiagnosis of asthma and accompanied by mistreatment and dense hospital visits wasting resources. The proposed novel algorithms are robust candidates to discriminate abnormalities within a breath cycle, in order to be employed in differential pulmonary disease diagnosis systems, more successfully and faster than related methods in literature. Lung abnormalities such as crackle and wheeze sounds may overlap and be successively within a breath cycle. Localization of these types of stationary and non-stationary waveforms is hard to achieve using Fourier based methods because of fixed time-frequency resolution property. On the other hand, the proposed algorithms have the ability to tune their wavelet bases aiming at localizing time-frequency characteristics of normal and abnormal lung sounds and as a result show finer time-frequency resolution and higher accuracy performance than the Fourier based methods. Moreover, the proposed algorithms are faster than their competitors being the suitable candidate for real-time applications. Traditional approach on differential diagnosis is to fit a model on the lung sound recording and then to use the extracted features for

predicting the lung diseases. However, the performance of the system may be limited when the low level information is not incorporated into the system. At macro level, fitting a model and extracting general features may limit the performance of the system. However, in micro level the atomic parts of the abnormalities (location in a breath cycle, type, duration and frequency of the abnormality) may help to exploit the discriminative characteristics of the lung diseases. Moreover, these micro level features can be combined with general model features to improve the diagnostic classification accuracy of the decision support systems.

In Chapter 2, a comprehensive experimental study was conducted using six feature extraction methods (including the proposed method) and five classifiers. Six subset features were extracted from the raw features and fed into the classifiers in order to compare the performances of feature extraction methods and classifiers. Contrary to previous studies which use constant low Q-factor wavelets and Fourier based methods that have limited frequency resolution and are unable to handle oscillatory behaviour of wheezes, adjustable Q-factor wavelets are proposed that can accommodate the signal of interest. Results show that by using high Q-factor wavelets, higher average accuracy, crackle, wheeze and normal signal classification rates than the related works in literature are achieved in a LOOCV scheme. The proposed method has shown better performance even using only one subset of extracted features. It provides better time-frequency resolution for all types of signals of interest and is less redundant than continuous wavelet transform and significantly faster than its nearest competitor (S Transform). DT, NB and k-NN classifiers, which have shown even better performances in some cases, are less robust than SVM and ELM in the overall performance.

In Chapter 3, both synthetic and real crackles are localized in the presence of both low SNR noise and wheezes. Previous methods focused on crackle separation from the point of stationary-non stationary discrimination. However, both crackles and wheezes have low and high frequency components and may exist successively. By using two threshold methods it is shown that both synthetic and real crackles are decomposed and localized with high accuracy using the proposed method. An adaptive threshold is determined independently from the channel on which it will be applied

providing robustness to noise variations. MAD threshold method has performed better than Teager energy operator method. Proposed method is compared with ICA and EMD methods in lung sound literature and shows higher sensitivity and precision rates for various SNR dB levels than ICA and EMD based methods. Moreover, decomposition ability of the methods is determined using normalized root mean square error (RMSE) metric for various dB levels and it is shown that minimum total RMS error is achieved with the proposed method. It can be concluded that the proposed method is able to localize crackles without generating false positives as extended Infomax ICA and without distorting the crackle waveform as EMD methods. The robustness of the proposed algorithm on overlapped or successive cases is represented by visual validation, RMSE calculation and in terms of crackle localization accuracy. The proposed algorithm is also tested on real data containing merely wheezes or crackles and is shown to be successful without any prior information.

In Chapter 4, non-dyadic wavelet based automatic and adaptive method is proposed to deal with MP-PP wheeze classification problem providing better TF resolution as compared to fixed parameter methods. Fixed parameter setting is defined as same p, q, s and J values are applied to the whole dataset. It is shown that by selecting proposed minimum peak-energy-ratio metric, higher classification accuracy than fixed parameter setting is achieved. It is also shown that by using time domain and/or Fourier based methods the classification accuracy is lower than the proposed method. MP-PP wheeze classification is more complicated than wheeze/non-wheeze classification (highly studied in literature) because the former is a within class problem (both types of wheezes are periodic). Moreover, it is shown that proposed adaptive non-dyadic wavelet based method is more successful than dyadic wavelet transform whose time-frequency resolution is limited.

5.1. Future Perspectives

As a future work in Chapter 2, the superimposed crackle and wheeze samples, which decrease the performance of the proposed system, need to be explored in more detail. Algorithms that can separate the crackle and wheeze containing signal com-

ponents will increase the accuracy of the classification system. Particle swarm optimization, independent vector analysis and non-negative matrix factorization techniques can be used to solve optimization problems, decompose multivariate data and separate spectral components of lung sounds. Additionally, it will be a challenge to implement the proposed system in real-time on a field programmable gate array as a lung sound classification system which can be employed in the diagnosis of lung diseases such as interstitial fibrosis, pneumonia, and chronic obstructive pulmonary disease.

As a future work in Chapter 3, sparse representations of decomposed lung sound channels can be obtained and sparse representations may be used to encode the sensor data. As an another future work, novel de-noising methods can be used as a preprocessing step in the exploration of diseases related with lung dysfunctions. Deep learning can also be employed to better understand and model the lung sound data in a supervised scheme. However, since the training process needs a very high number of patient data, this process may take a longer time to collect data. Lung sound spectrograms can be thought as images and auto-encoders may be employed to estimate the class labels or predict the abnormalities related with diseases using small set of training data in an unsupervised concept.

As a future work in Chapter 4, once the wheeze types are accurately determined at micro level, these findings can be associated with lung diseases such as asthma and COPD at macro level [91]. Again, deep learning concept may be employed to highlight wheeze regions on the spectrogram images of lung sounds. Moreover, the effect of preprocessing steps, such as de-noising, and ensemble learning methods at the classifier and feature extraction level (with additional TF features) may be explored.

REFERENCES

1. Vestbo, J., S. S. Hurd, A. G. Agustí, P. W. Jones, C. Vogelmeier, A. Anzueto, P. J. Barnes, L. M. Fabbri, F. J. Martinez, M. Nishimura *et al.*, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary", *American Journal of Respiratory and Critical Care Medicine*, Vol. 187, No. 4, pp. 347–365, 2013.

2. Postma, D. S. and K. F. Rabe, "The asthma–COPD overlap syndrome", *New England Journal of Medicine*, Vol. 373, No. 13, pp. 1241–1249, 2015.

3. Ulukaya, S., G. Serbes and Y. P. Kahya, "Overcomplete discrete wavelet transform based respiratory sound discrimination with feature and decision level fusion", *Biomedical Signal Processing and Control*, Vol. 38, pp. 322–336, 2017.

4. Ulukaya, S., G. Serbes and Y. P. Kahya, "An Adaptive and Automatic Parameter Selection Method based on Rational Dilation Wavelet Transform for Wheeze Type Classification", *Proceedings of the 8th International Workshop on Biosignal Interpretation (BSI 2016)*, pp. 163–166, IEEE, 2016.

5. Ulukaya, S., G. Serbes and Y. P. Kahya, "Resonance based Respiratory Sound Decomposition Aiming at Localization of Crackles in Noisy Measurements", *Engineering in Medicine and Biology Society (EMBC), 2016 38th Annual International Conference of the IEEE*, pp. 3688–3691, IEEE, 2016.

6. Ulukaya, S., G. Serbes, I. Sen and Y. P. Kahya, "A Lung Sound Classification System based on the Rational Dilation Wavelet Transform", *Engineering in Medicine and Biology Society (EMBC), 2016 38th Annual International Conference of the IEEE*, pp. 3745–3748, IEEE, 2016.

7. Ulukaya, S., G. Serbes, I. Sen and Y. P. Kahya, "Classification of Respiratory Sounds Using Spectral Features", *Biomedical Engineering Meeting (BIYOMUT),*

2016 20th National, IEEE, 2016.

8. Ulukaya, S., I. Sen and Y. P. Kahya, "Feature extraction using time-frequency analysis for monophonic-polyphonic wheeze discrimination", *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 5412–5415, 2015.

9. Ulukaya, S., I. Sen and Y. P. Kahya, "A novel method for determination of wheeze type", *Signal Processing and Communications Applications Conference (SIU), 2015 23th*, pp. 2001–2004, IEEE, 2015.

10. Ulukaya, S. and Y. P. Kahya, "Respiratory sound classification using perceptual linear prediction features for healthy-pathological diagnosis", *Biomedical Engineering Meeting (BIYOMUT), 2014 18th National*, pp. 1–4, IEEE, 2014.

11. Gavriely, N. and D. W. Cugell, *Breath Sounds Methodology*, CRC Press, 1995.

12. Pasterkamp, H., S. S. Kraman and G. R. Wodicka, "Respiratory sounds: advances beyond the stethoscope", *American Journal of Respiratory and Critical Care Medicine*, Vol. 156, No. 3, pp. 974–987, 1997.

13. Sovijarvi, A. R. A., L. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi and J. Earis, "Characteristics of breath sounds and adventitious respiratory sounds", *European Respiratory Review*, Vol. 10, No. 77, pp. 591–596, 2000.

14. Reichert, S., R. Gass, C. Brandt and E. Andrès, "Analysis of respiratory sounds: state of the art", *Clinical medicine. Circulatory, Respiratory and Pulmonary Medicine*, Vol. 2, pp. 45–58, 2008.

15. Bohadana, A., G. Izbicki and S. S. Kraman, "Fundamentals of lung auscultation", *New England Journal of Medicine*, Vol. 370, No. 8, pp. 744–751, 2014.

16. Piirila, P. and A. R. A. Sovijarvi, "Crackles: recording, analysis and clinical significance", *European Respiratory Journal*, Vol. 8, No. 12, pp. 2139–2148, 1995.

17. Sankur, B., E. C. Güler and Y. P. Kahya, "Multiresolution biological transient extraction applied to respiratory crackles", *Computers in Biology and Medicine*, Vol. 26, No. 1, pp. 25–39, 1996.

18. Baughman, R. P. and R. G. Loudon, "Lung sound analysis for continuous evaluation of airflow obstruction in asthma", *CHEST Journal*, Vol. 88, No. 3, pp. 364–368, 1985.

19. Meslier, N., G. Charbonneau and J. L. Racineux, "Wheezes", *European Respiratory Journal*, Vol. 8, No. 11, pp. 1942–1948, 1995.

20. Palaniappan, R., K. Sundaraj and N. U. Ahamed, "Machine learning in lung sound analysis: a systematic review", *Biocybernetics and Biomedical Engineering*, Vol. 33, No. 3, pp. 129–135, 2013.

21. Kandaswamy, A., C. S. Kumar, R. P. Ramanathan, S. Jayaraman and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients", *Computers in Biology and Medicine*, Vol. 34, No. 6, pp. 523–537, 2004.

22. Güler, I., H. Polat and U. Ergün, "Combining neural network and genetic algorithm for prediction of lung sounds", *Journal of Medical Systems*, Vol. 29, No. 3, pp. 217–231, 2005.

23. Matsunaga, S., K. Yamauchi, M. Yamashita and S. Miyahara, "Classification between normal and abnormal respiratory sounds based on maximum likelihood approach", *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, pp. 517–520, IEEE, 2009.

24. Abbas, A. and A. Fahim, "An automated computerized auscultation and diagnostic system for pulmonary diseases", *Journal of Medical Systems*, Vol. 34, No. 6, pp.

1149–1155, 2010.

25. Mayorga, P., C. Druzgalski, R. L. Morelos, O. H. Gonzalez and J. Vidales, "Acoustics based assessment of respiratory diseases using GMM classification", *Engineering in Medicine and Biology Society (EMBC), 32nd Annual International Conference of the IEEE*, pp. 6312–6316, IEEE, 2010.

26. Palaniappan, R., K. Sundaraj, N. Huliraj and S. Revadi, "A telemedicine tool to detect pulmonary pathology using computerized pulmonary acoustic signal analysis", *Applied Soft Computing*, Vol. 37, pp. 952–959, 2015.

27. Gross, V., A. Dittmar, T. Penzel, F. Schuttler and P. Von Wichert, "The relationship between normal lung sounds, age, and gender", *American Journal of Respiratory and Critical Care Medicine*, Vol. 162, No. 3, pp. 905–909, 2000.

28. Serbes, G., C. O. Sakar, Y. P. Kahya and N. Aydin, "Pulmonary crackle detection using time–frequency and time–scale analysis", *Digital Signal Processing*, Vol. 23, No. 3, pp. 1012–1021, 2013.

29. Bayram, I. and I. W. Selesnick, "Frequency-domain design of overcomplete rational-dilation wavelet transforms", *Signal Processing, IEEE Transactions on*, Vol. 57, No. 8, pp. 2957–2972, 2009.

30. Sen, I. and Y. P. Kahya, "A multi-channel device for respiratory sound data acquisition and transient detection", *Engineering in Medicine and Biology Society (EMBC), 27th Annual International Conference of the IEEE*, pp. 6658–6661, IEEE, 2005.

31. Taplidou, S. and L. J. Hadjileontiadis, "Analysis of wheezes using wavelet higher order spectral features", *Biomedical Engineering, IEEE Transactions on*, Vol. 57, No. 7, pp. 1596–1610, 2010.

32. Selesnick, I. W. and I. Bayram, "Oscillatory plus transient signal decomposition us-

ing overcomplete rational-dilation wavelet transforms", *SPIE Optical Engineering+ Applications*, pp. 74460V–74460V, International Society for Optics and Photonics, 2009.

33. Selesnick, I. W., "The double density DWT", *Wavelets in Signal and Image Analysis*, pp. 39–66, Springer, 2001.

34. Selesnick, I. W., R. G. Baraniuk and N. C. Kingsbury, "The dual-tree complex wavelet transform", *IEEE Signal Processing Magazine*, Vol. 22, No. 6, pp. 123–151, 2005.

35. Chui, C. K. and W. He, "Compactly supported tight frames associated with refinable functions", *Applied and Computational Harmonic Analysis*, Vol. 8, No. 3, pp. 293–319, 2000.

36. Daubechies, I., *Ten Lectures on Wavelets*, Vol. 61, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.

37. Proakis, J. G. and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice-Hall, Inc., NJ, USA, 1996.

38. Hermansky, H., "Perceptual linear predictive PLP analysis of speech", *The Journal of the Acoustical Society of America*, Vol. 87, No. 4, pp. 1738–1752, 1990.

39. Ellis, D. P. W., "PLP and RASTA (and MFCC, and inversion) in Matlab", 2005, `http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/`, accessed at September 2017.

40. Sengupta, N., M. Sahidullah and G. Saha, "Lung sound classification using cepstral-based statistical features", *Computers in Biology and Medicine*, Vol. 75, pp. 118–129, 2016.

41. Davis, S. and P. Mermelstein, "Comparison of parametric representations for mono-

syllabic word recognition in continuously spoken sentences", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 28, No. 4, pp. 357–366, 1980.

42. Stockwell, R. G., L. Mansinha and R. Lowe, "Localization of the complex spectrum: the S transform", *Signal Processing, IEEE Transactions on*, Vol. 44, No. 4, pp. 998–1001, 1996.

43. Alpaydin, E., *Introduction to Machine Learning*, The MIT Press, 2014.

44. Wu, X., V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip *et al.*, "Top 10 algorithms in data mining", *Knowledge and Information Systems*, Vol. 14, No. 1, pp. 1–37, 2008.

45. Marsland, S., *Machine Learning: An Algorithmic Perspective*, CRC press, 2015.

46. Duda, R. O., P. E. Hart and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2012.

47. Vapnik, V. N., *The Nature of Statistical Learning Theory*, Springer-Verlag New York Inc., 1995.

48. Chang, C. C. and C. J. Lin, "LIBSVM: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, pp. 1–27, 2011.

49. Huang, G.-B., Q.-Y. Zhu and C.-K. Siew, "Extreme learning machine: theory and applications", *Neurocomputing*, Vol. 70, No. 1, pp. 489–501, 2006.

50. Hansen, L. K. and P. Salamon, "Neural network ensembles", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 12, No. 10, pp. 993–1001, 1990.

51. Jacobs, R. A., M. I. Jordan, S. J. Nowlan and G. E. Hinton, "Adaptive mixtures of local experts", *Neural Computation*, Vol. 3, No. 1, pp. 79–87, 1991.

52. Woods, K., W. P. Kegelmeyer and K. W. Bowyer, "Combination of multiple classifiers using local accuracy estimates", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 19, No. 4, pp. 405–410, 1997.

53. Kuncheva, L. I., J. C. Bezdek and R. P. Duin, "Decision templates for multiple classifier fusion: an experimental comparison", *Pattern Recognition*, Vol. 34, No. 2, pp. 299–314, 2001.

54. Drucker, H., C. Cortes, L. D. Jackel, Y. LeCun and V. Vapnik, "Boosting and other ensemble methods", *Neural Computation*, Vol. 6, No. 6, pp. 1289–1301, 1994.

55. Polikar, R., "Ensemble based systems in decision making", *IEEE Circuits and Systems Magazine*, Vol. 6, No. 3, pp. 21–45, 2006.

56. Rodgers, G. W. *et al.*, "A proof of concept study of acoustic sensing of lung recruitment during mechanical ventilation", *Biomedical Signal Processing and Control*, Vol. 32, pp. 130–142, 2017.

57. Wilkins, R. L., J. E. Hodgkin and B. Lopez, *Fundamentals of Lung and Heart Sounds*, Mosby St. Louis, MO, 2004.

58. Akay, M., *Time Frequency and Wavelets in Biomedical Signal Processing*, IEEE Press Series in Biomedical Engineering, 1998.

59. Ono, M., K. Arakawa, M. Mori, T. Sugimoto and H. Harashima, "Separation of fine crackles from vesicular sounds by a nonlinear digital filter", *Biomedical Engineering, IEEE Transactions on*, Vol. 36, No. 2, pp. 286–291, 1989.

60. Arakawa, K., H. Harashima, M. Ono and M. Mori, "Non-linear digital filters for extracting crackles from lung sounds", *Frontiers of Medical and Biological Engineering: The International Journal of the Japan Society of Medical Electronics and Biological Engineering*, Vol. 3, No. 4, pp. 245–257, 1990.

61. Charleston-Villalobos, S., R. González-Camarena, G. Chi-Lem and T. Aljama-Corrales, "Crackle sounds analysis by empirical mode decomposition", *Engineering in Medicine and Biology Magazine, IEEE*, Vol. 26, No. 1, pp. 40–47, 2007.

62. Hadjileontiadis, L. J., "Empirical mode decomposition and fractal dimension filter", *IEEE Engineering in Medicine and Biology Magazine*, Vol. 26, No. 1, pp. 30–39, 2007.

63. Castañeda-Villa, N., S. Charleston-Villalobos, R. González-Camarena and T. Aljama-Corrales, "Assessment of ICA algorithms for the analysis of crackles sounds", *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 605–608, IEEE, 2012.

64. Hadjileontiadis, L. J. and S. M. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter", *Biomedical Engineering, IEEE Transactions on*, Vol. 44, No. 12, pp. 1269–1281, 1997.

65. Kahya, Y. P., S. Yerer and O. Cerid, "A wavelet-based instrument for detection of crackles in pulmonary sounds", *Engineering in Medicine and Biology Society (EMBC), Proceedings of the 23rd Annual International Conference of the IEEE*, Vol. 4, pp. 3175–3178, IEEE, 2001.

66. Bahoura, M. and X. Lu, "Separation of crackles from vesicular sounds using wavelet packet transform", *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, Vol. 2, pp. 1076–1079, IEEE, 2006.

67. Taplidou, S., L. J. Hadjileontiadis, I. K. Kitsas, K. Panoulas, T. Penzel, V. Gross and S. M. Panas, "On applying continuous wavelet transform in wheeze analysis", *Engineering in Medicine and Biology Society (EMBC), 26th Annual International Conference of the IEEE*, Vol. 2, pp. 3832–3835, IEEE, 2004.

68. Lehrer, S., "Understanding Lung Sounds (audio CD)", *Saunders, Philadelphia*, 2002.

69. Selesnick, I. W., "Wavelet transform with tunable Q-factor", *Signal Processing, IEEE Transactions on*, Vol. 59, No. 8, pp. 3560–3575, 2011.

70. Kiyokawa, H., M. Greenberg, K. Shirota and H. Pasterkamp, "Auditory detection of simulated crackles in breath sounds", *CHEST Journal*, Vol. 119, No. 6, pp. 1886–1892, 2001.

71. Lozano, M., J. A. Fiz and R. Jané, "Performance evaluation of the Hilbert–Huang transform for respiratory sound analysis and its application to continuous adventitious sound characterization", *Signal Processing*, Vol. 120, pp. 99–116, 2016.

72. Selesnick, I. W., "Sparse signal representations using the tunable Q-factor wavelet transform", *SPIE Optical Engineering+ Applications*, pp. 81381U1–81381U15, International Society for Optics and Photonics, 2011.

73. Afonso, M. V., J. M. Bioucas-Dias and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization", *Image Processing, IEEE Transactions on*, Vol. 19, No. 9, pp. 2345–2356, 2010.

74. Quiroga, R. Q., Z. Nadasdy and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering", *Neural Computation*, Vol. 16, No. 8, pp. 1661–1687, 2004.

75. Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 454, pp. 903–995, The Royal Society, 1998.

76. Hyvärinen, A. and E. Oja, "Independent component analysis: algorithms and applications", *Neural Networks*, Vol. 13, No. 4, pp. 411–430, 2000.

77. Hyvärinen, A. and E. Oja, "A fast fixed-point algorithm for independent compo-

nent analysis", *Neural Computation*, Vol. 9, No. 7, pp. 1483–1492, 1997.

78. Lee, T.-W., M. Girolami and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources", *Neural Computation*, Vol. 11, No. 2, pp. 417–441, 1999.

79. Bell, A. J. and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution", *Neural Computation*, Vol. 7, No. 6, pp. 1129–1159, 1995.

80. Vyshedskiy, A., R. M. Alhashem, R. Paciej, M. Ebril, I. Rudman, J. J. Fredberg and R. Murphy, "Mechanism of inspiratory and expiratory crackles", *CHEST Journal*, Vol. 135, No. 1, pp. 156–164, 2009.

81. Hoevers, J. and R. G. Loudon, "Measuring crackles", *CHEST Journal*, Vol. 98, No. 5, pp. 1240–1243, 1990.

82. Sen, I., M. Saraclar and Y. P. Kahya, "Computerized diagnosis of respiratory disorders", *Methods of Information in Medicine*, Vol. 53, No. 4, pp. 291–295, 2014.

83. Society, A. T., "Updated nomenclature for membership reaction", *American Thoracic Society News*, Vol. 3, pp. 5–6, 1977.

84. Sovijarvi, A., F. Dalmasso, J. Vanderschoot, L. Malmberg, G. Righini and S. Stoneman, "Definition of terms for applications of respiratory sounds", *European Respiratory Review*, Vol. 10, No. 77, pp. 597–610, 2000.

85. Pramono, R. X. A., S. Bowyer and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review", *PloS One*, Vol. 12, No. 5, p. e0177926, 2017.

86. Forgacs, P., "The functional basis of pulmonary sounds.", *CHEST Journal*, Vol. 73, No. 3, pp. 399–405, 1978.

87. Nagasaka, Y., "Lung sounds in bronchial asthma.", *Allergology International*, Vol. 61, No. 3, pp. 353–363, 2012.

88. Jain, A. and J. Vepa, "Lung sound analysis for wheeze episode detection", *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 2582–2585, IEEE, 2008.

89. Jin, F., S. Krishnan and F. Sattar, "Adventitious sounds identification and extraction using temporal–spectral dominance-based features", *Biomedical Engineering, IEEE Transactions on*, Vol. 58, No. 11, pp. 3078–3087, 2011.

90. Naves, R., B. H. G. Barbosa and D. D. Ferreira, "Classification of lung sounds using higher-order statistics: A divide-and-conquer approach", *Computer Methods and Programs in Biomedicine*, Vol. 129, pp. 12–20, 2016.

91. Sen, I., M. Saraclar and Y. P. Kahya, "A Comparison of SVM and GMM-based classifier configurations for diagnostic classification of pulmonary sounds", *Biomedical Engineering, IEEE Transactions on*, Vol. 62, No. 7, pp. 1768–1776, 2015.