CLASSIFICATION AND DETECTION OF WHEEZES IN RESPIRATORY SOUNDS

by

ÇAĞLAYAN ŞERBETÇİ

B.S., Electrical and Electronics Engineering, Bogazici University, 2017

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

Graduate Program in System and Control Engineering
Boğaziçi University
2020

ACKNOWLEDGEMENTS

First all, I would like to express my sincere appreciation to my supervisor Prof. Dr. Yasemin Kahya for her support and guidance throughout this study. Without her valuable advice, encouragement, and understanding, this study would never be completed successfully. I also would like to thank Asst. Prof. Ipek Şen for her kind support and valuable feedback.

Without Electrosalus Biomedical Inc. and its members, the data used in this thesis would not have existed. So, I would like to thank them for their support.

This thesis was written in Covid-19 days, and I surprisingly had a chance to finish this study at my home with my parents, Osman and Kadriye Şerbetçi. Once again, I realized their continuous and unrequited support deeply. You are my first and life-long teachers. Thank you.

Last, I would like to thank my fiance as the second reader of this thesis. I am grateful for her valuable comments and endless support. I also thank my little brother and friend Volkan for relief and joy he gave during this study.

ABSTRACT

CLASSIFICATION AND DETECTION OF WHEEZES IN RESPIRATORY SOUNDS

Analyzing respiratory sounds and detecting anomalies in them with intelligent computer algorithms has opened a new era for auscultation that has 250 years of history. These algorithms can overcome the drawbacks of conventional stethoscopes and support medics about auscultation.

In this thesis, a new intelligent algorithm to detect wheezes superimposed on vesicular sounds is developed and presented. Detection of wheezes with intelligent algorithms is one of the hot topics currently being researched by many researchers. They are continuous musical adventitious respiratory sounds. Their duration, intensity, and phase in respiratory sounds give essential information for the diagnosis and prognosis of respiratory diseases.

In this study, one of the aims is to determine the best discriminative features among nine features which are mostly used in other researches. The other aim is to find the best-performed machine learning classifier to classify wheezes and normal respiratory sounds.

Last, we created a novel detection algorithm is presented to detect correctly the wheeze interval in recorded respiratory sounds by employing selected machine learning model to respiratory sounds.

ÖZET

SOLUNUM SESLERİNDE BULUNAN HIRILTI SESLERİNİ SINIFLANDIRMA ve TESPİT ETME

Solunum seslerini akıllı bilgisayar algoritmaları ile analiz etmek ve içindeki normal olmayan solunum seslerini teşhis etmek 250 yıllık geçmişe sahip oskultasyon yönteminde yeni bir çağ başlatmıştır. Bu algoritmalar oskultasyon konusunda geleneksel steteskopların problemlerini çözebilir ve sağlık çalışanlarını destekleyebilirler.

Bu tezde, hırıltı seslerini normal solunum seslerinin arasından tespit etmeyi sağlayan yeni bir akıllı algoritma geliştirildi ve sunuldu. Hırıltı seslerini akıllı algortimalar ile tesbit etmek günümüzde bir çok araştırmacı tarafından çalışılmaktadır. Hırıltılar sürekli ve normal olmayan solunum sesleri olarak tanımlanabilirler. Müzikal bir yapıya sahiptirler. Süresi, yoğunluğu ve nefes alış verişinini hangi evresinde olduğu verisi, akciğer hastalıklarının teşhisi ve durumu hakkında önemli bilgiler verir. Bu çalışmadaki hedeflerden biri daha önceki araştırmalarda hırıltı belirlemede kullanılan 9 tane ses özelliği arasıdan, en ayırıcı olanlarını belirlemektir. Ayrıca, bu özellikleri kullanarak hırıltı seslerini ayırt edebilen en başarılı makine öğrenmesi algoritmasını da bulunması hedeflendi.

Son olarak, en başarılı sonuçları veren makine öğrenmesi modelini kullanarak hırıltı seslerini normal solunum seslerinden ayırt etmeyi sağlayan yeni bir algoritma geliştirildi.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

b	Bias term in SVM hyperplane equation
$c_m(k)$	Spectral unpredictability
$E(X)$	Expected value
$H_\alpha(X)$	Renyi entropy
$K(x, \acute{x})$	Kernel function
k	k parameter in k-NN
L_{TI}	The number of windows for average calculation
L_{ASE}	Averages of spectra fragments
m	Window index
n	Sample count
N	Window size
p_i	Sound samples
$P_m(k)$	Power of m'th window
$r_m(k)$	Magnitude of Discrete Fourier Transform of m'th window
s_x	Sample standard deviation
t	T-test t value
w	p-dimensional hyperplane vector
X	Real-valued random variable
x	Input vector
x_i	Data vector
α	Order of Renyi entropy
γ	Scale parameter in Kernel functions
ϵ_t	Error term in AR parameters
ζ	Regularization parameter in SVM
μ	Mean vector
σ	Standard deviation
Σ	Covariance matrix

ϕ_n	Auto-Regressive model coefficient
$\phi_m(k)$	Phase of Discrete Fourier Transform of m'th window
ω	Angular frequency

LIST OF ACRONYMS/ABBREVIATIONS

2CD	Two Cycle Duration
2D	Two Dimensional
ASE	Audio Spectral Envelope
AUC	Area Under Curve
AR	Auto-Regressive
ATS	American Thoracic Society
CORSA	Computerized Respiratory Sound Analysis
COPD	Chronic obstructive pulmonary disease
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
FN	False Negative
FPDB	FFT Peak Baseline Difference in dB
FP	False Positive
GMM	Gaussian Mixture Model
IDW	Initial Deflection Width
IoT	Internet of things
LAL	Lung Acoustic Laboratory
LDA	Linear Discriminative Analysis
LPC	Linear Predictive Coding
k-NN	k-Nearest Neighbour
MCI	Mean Crossing Irregularity
MFCC	Mel frequency cepstral coefficients
MLP	Multilayer Perceptron
MPEG	Motion Picture Experts Group
PCA	Principal Component Analysis
PPV	Positive Predictive Value
PSD	Power Spectral Density
RBF	Radial Basis Function

SVM	Support Vector Machine
TDW	Total duration width
TI	Tonality Index
TN	True Negative
TP	True Positive
VQ	Vectorial Quantification

1. INTRODUCTION

1.1. Background

Over the centuries, mankind has been trying to understand human physiology and trying to diagnose diseases by trying many kinds of methods. To understand the nature of the lungs and heart, it was apparent that body sounds play a crucial role in examining patient. Listening to the chest wall and understanding the sound of lungs and hearth opened a new way to cure the lung and heart diseases [1].

Before the invention of the stethoscope in 1816, physicians used to listen to body sounds with "immediate auscultation" in order to uncover possible respiratory and heart diseases. A French doctor, Rene' The'ophile Hyacinth Laënnec, was the first physician using a conic paper to amplify the chest sound. Laënnec rolled a piece of conic paper and put slim side on the chest wall of the patient. When he gave ear the other side of the conic paper, he noticed that sound was louder and more apparent than the earlier method "immediate auscultation." After the discovering the exceptional benefit of conic paper in auscultation, he developed the idea by inventing monaural wooden tubes in Figure 1.1 which were later named stethoscope [2]. The word "stethoscope" is combined with two Greek words which are stethos and scopos. Stethos means "chest or breast" and the suffix -scope used for the instruments for seeing. After the invention, this iconic device has changed and improved many times to become the present form [1].

For decades, conventional stethoscope auscultation has been the primary tool for monitoring the lungs in order to detect possible respiratory diseases. Though it is a widely used method, it has many disadvantages that lower the diagnostic quality because of its old architecture between the body surface and physician's ear. The most prominent drawback of the conventional stethoscope is the frequency response of the device. It only amplifies lower than 112 Hz and attenuates higher frequencies [3]. On

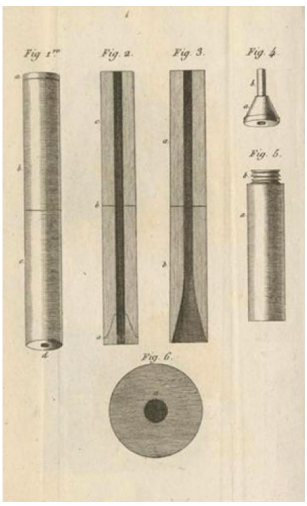

Figure 1.1. Laënnec's stethoscope: Monaural wooden tubes

the other hand, this study [4] shows frequency of normal respiratory sounds spreads up to 2000 Hz and this value approaches to 6000 Hz in pathological conditions. Therefore, traditional stethoscopes are not capable of detecting high pitched adventitious sounds superimposed on vesicular sounds.

The weakness of human hearing capability is the other drawback of a traditional stethoscope. Individuals mainly focus on what they desire to hear while neglecting other sounds. This inclusive ability in our auditory system is described as selective hearing or selective auditory attention [1]. Even though selective hearing is a facilitator of human life, it can be a difficulty for physicians in detecting adventitious respiratory sounds. Typically, adventitious sounds can easily coincide with other abnormal sounds in vesicular sounds so that physicians have a probability of dismissing some of the abnormal sounds that occur simultaneously or sequentially. One study states that family physicians in the United States are only able to detect about 80% of wheeze in recorded a series of recorded pulmonary sounds [5].

Advances in computer systems in today's world lead to rapid development in healthcare. Recording health data and graphical representations of health analysis open improved ways in diagnosing diseases and making assessments. Unfortunately,

conventional stethoscopes have not such technology for recording respiratory sounds or making graphical output representations [3].

Several electronic stethoscopes have been introduced to fulfill the need for a better auscultation tool for the last two decades. These modern stethoscopes have clear advantages in recording respiratory sounds and amplifying them. After the starting of usage of the electronic stethoscopes, digitized respiratory sound analysis techniques have been developed fast and created structured and objective ways to overcome drawbacks mentioned above.

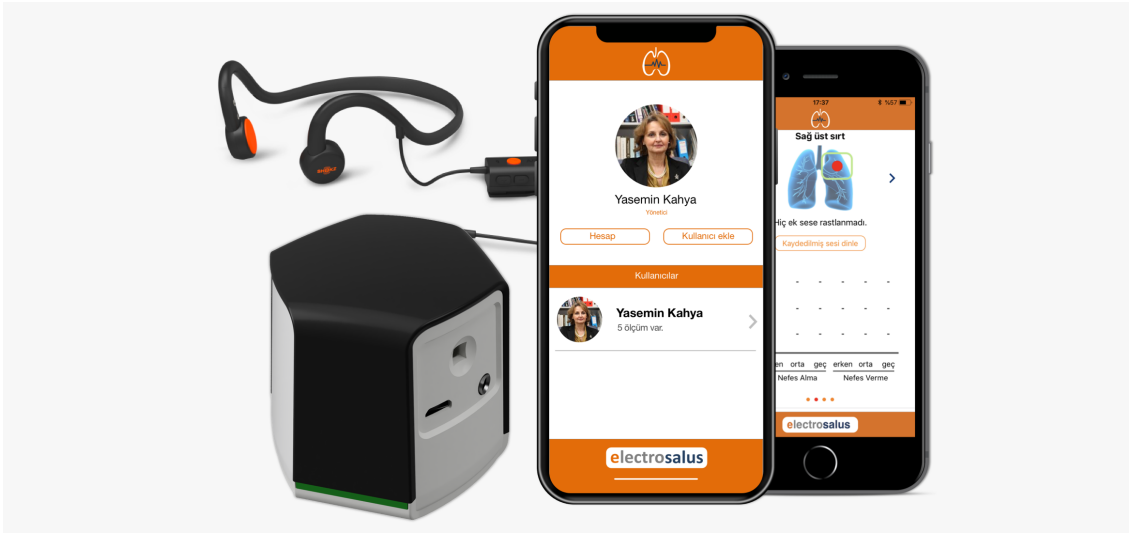

Figure 1.2. An electronic Stethoscope

Electronic stethoscopes and new pulmonary sound analysis techniques can solve the information loss problem, which is originated from the frequency response of traditional stethoscope. They also increase the diagnostic value of auscultation by using intelligent algorithms, which detect and classify anomalies in respiratory sounds. Also, remote medical screening may be possible even by non-medics who are unable to reach doctors and medical facilities. Respiratory sounds can be stored for further assessments of individuals and for teaching purposes in respiratory lectures at medical schools by the help of electronic stethoscopes.

These studies [3, 6, 7] show the improvement of the development of electronic stethoscopes and possible abilities they offer improving healthcare in pulmonary specialty.

1.2. Respiratory Sounds

Breathing or "external respiration" was developed around 428 million years ago on land [8] for making gas exchange with the internal environment of livings by transferring air into and out of the lungs. In human physiology, respiratory sounds are defined as all sounds corresponding to respiration, which includes adventitious sounds, breath sounds, snoring sounds, couch sounds, and respiratory muscle sounds [9]. Figure 1.3 shows the basis of the respiratory sound terminology and their classification.

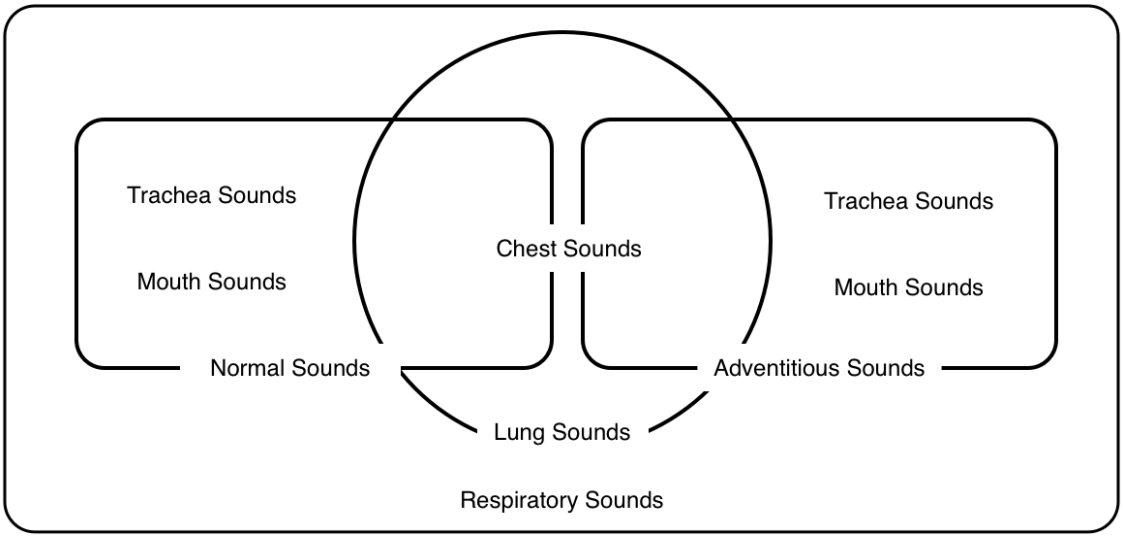

Figure 1.3. Basis of the respiratory sound terminology

1.2.1. Lung sounds

All respiratory sounds which are heard over the chest wall are defined as lung sounds, which include the adventitious sounds and normal sounds, as seen in Figure 1.3. Airflow into and out of airways in the lungs generates these sounds, which have two-

phase, inspiratory and expiratory phases.

Lung sounds differentiate for every individual because they are highly correlated with body anatomy and physiology. Moreover, many factors such as gender, age, body mass and height, the location that sounds are heard at, affect these sounds [10].

1.2.2. Normal Respiratory Sounds

Respiratory sounds consist of both normal and adventitious sounds, which are heard over the chest wall and trachea. Normal breath sounds or normal respiratory sounds are nonmusical sounds which are recorded over chest of healthy subjects. These sounds usually include no adventitious sound. They are generally heard in the inspiration phase rather than the expiration phase, and they peak on low-frequency components so that high-frequency components are not audible. However, the amplitude of normal breath sounds is affected by the body and chest locations where they are heard [3].

The normal respiratory sounds above 300 Hz originate from the turbulent airflow vertex; on the other hand, the source of sound below 300 Hz is uncertain [11]. On the other hand, nearly 500 Hz (131–552.5 Hz) is the median frequency (F50) for normal breath sounds and their peak is between 100 and 200 Hz [10]. The chest wall and the lungs attenuate the normal respiratory sounds heard, and they act as low pass filters so that most of the energy of normal respiratory sounds drops above 300 Hz [11]. Because of the internal structure of the human anatomy, it is hard to hear normal respiratory sounds especially when they mix with the cardiovascular system sounds and muscle sounds. Fortunately, the normal sounds above 800 Hz can be recorded with the proper microphones such as electret microphones.

Figure 1.4 shows graphical representations of the 15 seconds healthy individual's normal breath sounds according to time.

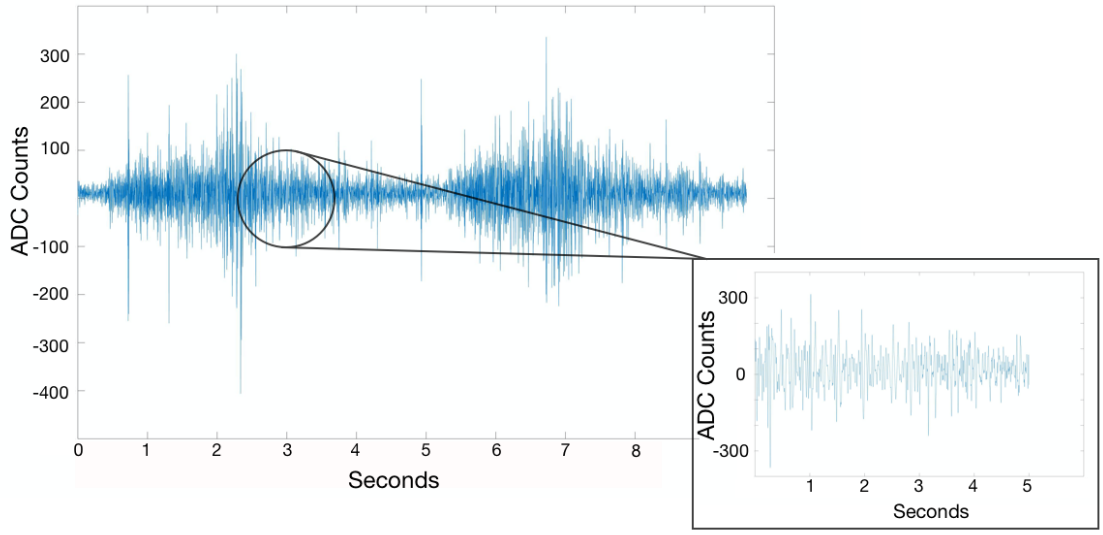

Figure 1.4. Plot of a Normal Respiratory Sound waves

1.2.3. Adventitious sounds

In certain pathological conditions of lungs and airways, respiratory sounds contain additional abnormal sounds when individuals inhale and exhale. These unusual sounds can also be called adventitious sounds. There are two main types of adventitious sounds, which are continuous, e.g., wheezes, and discontinuous adventitious sounds, e.g., crackles [9]. On the other hand, some adventitious sounds like squawks can possess both continuous and discontinuous characteristics. Adventitious sounds mostly are the indication of pulmonary diseases, and they are essential for earlier detection and prognosis of respiratory diseases.

1.2.3.1. Crackles. Crackles are nonmusical, brief, explosive, and intermittent adventitious sounds heard mostly during inspiration rather than during expiration [10]. They are examples of discontinuous adventitious sounds. According to Paul Forgacs, they are caused by sudden opening and closing of airways to equalize gas pressures inside and outside of airways. He thought that opening and closing of each airway induces

a single crackle and he was the first person who theorized the crackles as above in 1967 [12].

Later, stress-relaxation quadrapoles were submitted as a mathematical model of crackles by Fredberg and Holford. Their theory enhances Forgacs theory. Forgacs theory defend that crackles are triggered by the gas pressure change inside lung airways. Fredberg and Holford introduce the elastic stress change inside and near airway surfaces as a reason of the crackles. Stress change in airways produces the dynamic events which generates sound waves. They also showed the authentic waveform of crackle signals, as seen in figure 1.5 [13].

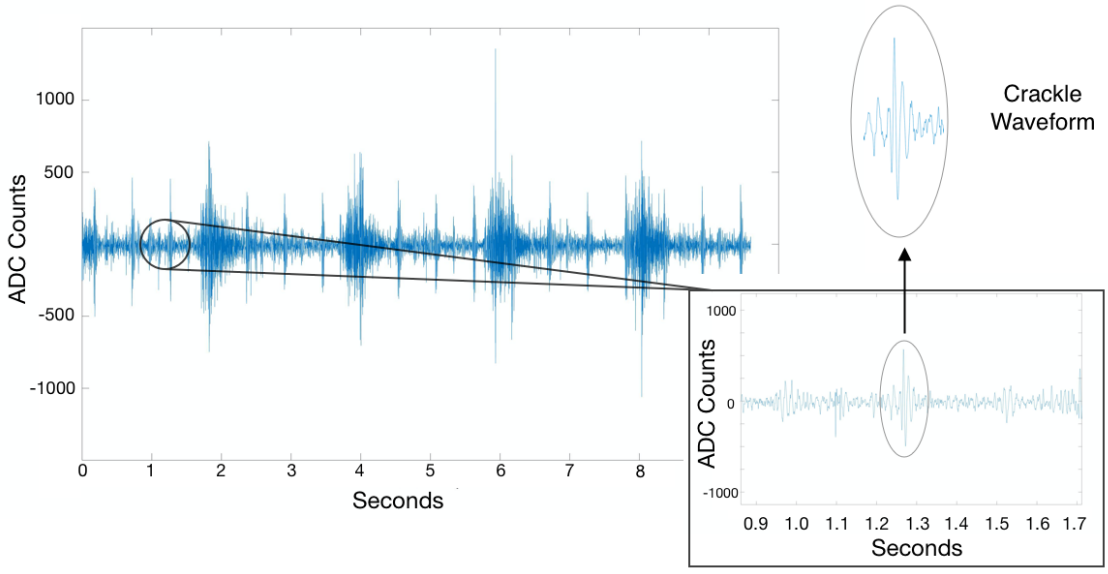

Figure 1.5. Plot of a pathological respiratory sound with crackles and a crackle wave

Another study [14] also showed that the properties of airways affect the type of crackles. Shorter airways generate crackles of shorter duration. Moreover, crackles' duration and pitch differentiate them as course crackles and fine crackles. Fine crackles have relatively short duration in contrast to coarse crackles. Moreover, low-frequency components are dominant in course crackles rather than fine crackles [15].

Crackles may also exist in breath sounds of healthy people [16, 17]. In one study [16], the lung sounds of 44 volunteers who are non-smokers and have no lung disease are recorded over the anterior chest. More than 50 per cent of healthy lung sounds contains crackles during the inspiration phase from a low-level lung volume. Additionally, it is observed that the occurrence of crackles may disappear in repeated auscultation after coaching due to the increased lung volume. These indications show that occurring of crackles are not always the indication of pathological lung diseases. The number of crackles decreases naturally because of deflated airways opening.

Number of detected crackles, their specific location and their phase during the respiratory cycle (inspiration and expiration) are essential symptoms for the degree of lung disease and airways disorder. Because of their short duration and low intensity, their detection and characterization is not simple by traditional normal auscultation. Furthermore, the frequency range of crackles is specified as 100 to 2000 Hz. The frequency of crackles are highly correlated with the diameter of airways. Airways with higher diameter produces low-frequency crackles and vice-versa [10].

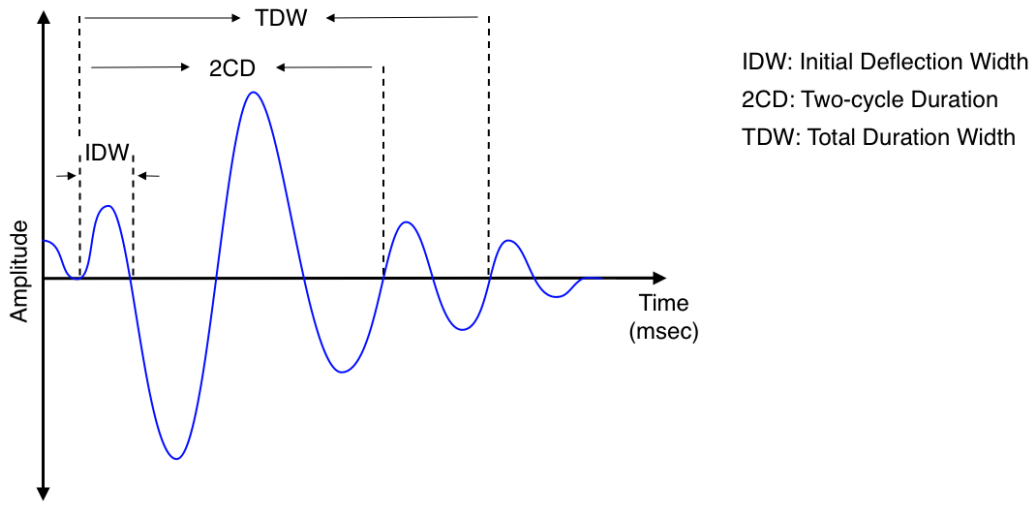

Figure 1.6. A crackle waveform and its measurements

Subsequent measurements are employed to the waveform to detect crackles. Fig-

ure 1.6 shows the general waveform of crackles which arises with an initial deflection and continues with a damped sinusoidal wave along. The duration between the first deflection and the beginning of crackle is defined as initial deflection (IDW). Additionally, the duration from the beginning of crackle to the point which two-cycle completed is the two-cycle duration (2CD). Total duration width (TDW) is the total duration of crackle waveform.

Table 1.1 summarizes the American Thoracic Society (ATS) and Computerized Respiratory Sound Analysis (CORSA) groups suggestions about the above measurements in milliseconds.

Table 1.1. Properties used for the classification of crackles according to ATS and CORSA

	Intensity	**Pitch**	**IDW (ATS)**	**2CD (ATS)**	**2CD (CORSA)**
Course Crackles	Loud	Low	About 1.5 ms	About 10 ms	>10 ms
Fine Crackles	Less loud	Higher	About 0.7 ms	<5 ms	<10 ms

1.2.3.2. Wheezes. Wheezes are continuous adventitious sounds which are superimposed on breath sounds. They play a vital role in the diagnosis and prognosis of respiratory diseases such as obstructive airway pathologies. They have a musical nature because of their narrow frequency bands rather than crackles [10].

The interaction between gas moving through lung airways and airway wall causes oscillation, which generates musical sounds. These musical sounds, wheezes, are often associated with musical instruments that use the wind to generate sound [18]. For example, Forgacs [12] linked the mechanism of wheeze with the toy trumpet. Toy trumpet produce sounds by vibrating reed, and these sounds are characterized by the elasticity and mass of reed. Like in toy trumpets, he stated that the pitch of wheeze is influenced by mass and flexibility of airways, not by the size or length of the lung airway, which was suggested by Laënnec.

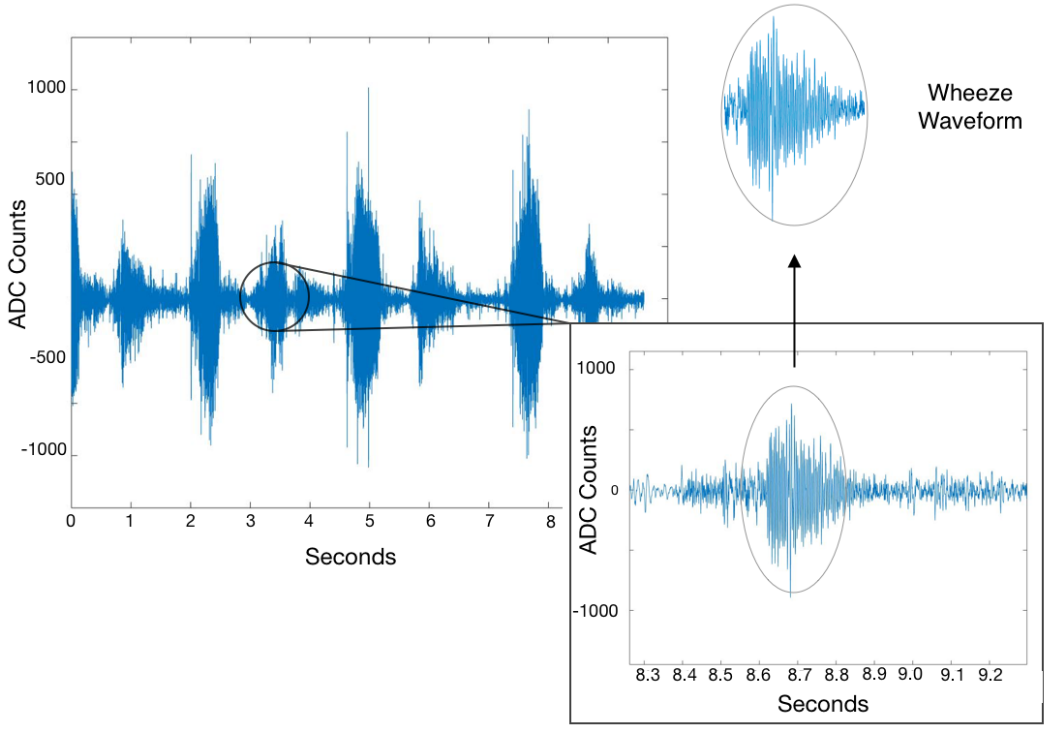

Figure 1.7. Plot of a pathological respiratory sound with wheezes and a wheeze wave

The mathematical model of wheezing was presented by Grotberg and Davis in 1980 [19]. Their model gives a prediction about the oscillation of airways' wall in collapsible tubes when the gas flow velocity is greater than limitation of airways diameter. They suggested that flow is always limited when wheezing occurs; however, a limited flow may not always generate wheezes.

Wheezes can be divided into two categories according to their frequency characteristics. Monophonic wheezes that have a single dominant frequency behaves like a single musical tone, whereas poly-phonic wheezes have more than one musical tone [20].

Wheezes are not always an indication of pulmonary diseases, and they can be detected in healthy subjects' lung sounds through forced expiration [18]. Also, the detection of wheezes is relatively easier than the detection of crackles because wheezes are generally louder than normal breath sounds. They can be heard by auscultation over the larynx and at the subjects' mouths [10].

The word "continuous" is mostly used for wheezes because of their longer duration than crackles. According to the American Thoracic Society (ATS), wheezes are longer than 250 ms, on the other hand, Computerized Respiratory Sound Analysis (CORSA) groups defined that duration of wheezes is longer than 100 ms. Besides, these studies [21, 22] defend that their duration is longer than 50 and 100 ms relatively. Another study [22] also defends that the duration of wheezes is shorter than 250 ms.

Table 1.2. Durations of Wheezes

	Duration(ATS)	**Duration(CORSA)**	**Duration(Taplidou et al. [22])**
Wheeze	$>$ 250 ms	$>$ 100 ms	$>$ 100 ms $<$ 250 ms

The frequency range of wheezes is also not entirely clear. ATS specifies that 400 Hz is the dominant frequency for wheezes [23]; on the other hand, according to definitions of CORSA guidelines, wheezes have a dominant frequency higher than 100 Hz [9]. One study [15] presents that wheezes have a frequency band between 100 to 2500 Hz.

Table 1.3. Frequency Ranges of Wheezes

	DF (ATS)	**DF (CORSA)**	**Frequency Range (Reichert et al. [15])**
Wheeze	About 400 Hz	$>$100 Hz	$>$100 Hz $<$2500 Hz

DF: Dominant Frequency

Many pathologies are interconnected with continuous adventitious sounds, wheezes. Their dominant frequency, numbers, and locations on lungs are crucial clinical features to evaluate the hardness of airway obstruction and to observe its progress. The list which is presented by Meslier and Charbonneau [18] in table 1.4 is the list of diseases associated with wheezes.

1.3. Computerized Wheeze Detection Algorithms

According to the World Health Organization, chronic obstructive pulmonary diseases are the third cause of death all over the world. They are in charge of more than 18% percent of deaths every year [24]. Monitoring this illnesses and early detection of

Table 1.4. Diseases related with wheezes

Asthma
Infections, whooping cough, croup, acute tracheobronchitis
Tracheal stenosis
Foreign body aspiration
Pulmonary oedema
Chronic obstructive pulmonary disease
Bronchorrhoeal states, for example bronchiectasis, chronic bronchitis
Tracheal stenosis

them play an essential role in decreasing the death rate. For this reason, automatic detection of respiratory anomalies such as identification and detection of wheezes has been a notable research of interest. With the help of the electronic stethoscopes, researchers attain a new chance to record respiratory sounds and develop new algorithms to automatically label wheezes in these recorded breath sounds. Different discriminatory features with various classifiers have been used to label abnormal sounds in different studies for years. Some of the features which several studies use are briefly reviewed in the following.

Orjuela-Canon et al. [25] created 13 artificial neural network classifier with Mel frequency cepstral coefficients (MFCC). By using the Leave One Out Cross -Validation method, he achieved to detect all wheeze with a 100 % detection rate. On the other hand, their algorithm detects normal breath sounds with an accuracy of 80%.

Another study [26] also revealed that Mel frequency cepstral coefficients (MFCC) are more discriminative than Linear Predictive Coding features and wavelet transform features. Bahoura et al. found the best accuracy as 94.2% specificity and 97.2% sensitivity by using Gaussian Mixture Model (GMM) with 24 MFCC among Vectorial

Quantification (VQ) and Multilayer Perceptron (MLP).

Oweis et al. [27] announced a 98.6% accuracy by applying 32 averaged power spectrum features to the Artificial Neural Network to classify normal and abnormal respiratory sounds, including events of wheeze events.

Mendes et al. [28] used features from spectrogram space and features usually employed in the context of Music Information Retrieval. Mendes reached the best result as 92.7% sensitivity and 90.9% specificity with the Random Forest Algorithm.

Wisniewski and Zielinski [29] worked with two features: Tonality Index (TI) and Audio Spectral Envelope (ASE). They are widely used technics for compressing sound files with MPEG standards. Wisniewski and Zielinski reported Area Under Curve (AUC) performance of 0.905 for TI and 0.951 for ASE by using the Support Vector Machine (SVM) classifier.

Kurtosis, Renyi entropy, quartile frequency ratios, and Mean Crossing Irregularity (MCI) are applied with Fischer Discriminant Analysis by Aydore et al [30]. They reached the accuracy of 93.5% in detecting wheeze and non-wheeze sound pieces.

Oletic et al. [31] worked on the LPC error ratio to detect wheeze, which is superimposed on normal respiratory sound. They used a simple threshold classifier and reached an accuracy of 90.29% in classifying wheezes.

Also, Liu et al. [32] employed entropy features with a simple threshold classifier to detect wheeze, stridor, crackle, and normal respiratory sound. They achieved 70% accuracy for wheezes and 99 for normal lung sounds on 45 recordings.

Finally, the differentiation of monophonic and polyphonic wheezes is another major topic that was researched by Ulukaya et al. [33]. They used mean crossing irregularity and multiple quartile frequency ratios with several classifiers (Support Vector

Machine, k-Nearest Neighbour and Naive Bayesian). They achieved 75.78% accuracy as the best result by using k-Nearest Neighbour.

1.4. Motivation and Aim

This thesis aims to state the highest discriminative features and to select the best-performed classifier to detect wheezes superimposed on respiratory sound.

The issues originated from the conventional auscultation mentioned in Section 1.1 can be alleviated by the intelligent algorithms which detect the anomalies in respiratory sounds recorded by electronic stethoscopes. Wheezes contain valuable information about lung diseases so that monitoring the occurrence of wheezes in respiratory sounds plays a vital role in diseases mentioned in Table 1.4. On the other hand, in day to day management of these diseases, patients are generally asked to log the experience of wheezes because wheezes that occur at night give diagnostic information, and logging them is essential. Wearable smart stethoscopes can solve this problem and help continuous monitoring of wheezes. Automatic wheeze detection algorithms identify wheezes from recorded respiratory sounds by the help of smart stethoscopes. So, wheeze detection algorithms are a crucial part of the automatic lung sound analysis.

In another study, Renard et al. [34] reviewed 25 different types of features practiced in earlier studies. They presented the performance of the features individually by using a simple linear threshold to classify wheeze and normal respiratory sounds. Their main aim is to find the feature that gives the best-performance in mobile, battery-powered wearable devices. So, they choose simple classifiers that consume less computational load. Today, many IoT devices connect to the servers with high computational power via the internet. So, running algorithms on the IoT devices will no longer be needed. Fast server computers can analyze respiratory sounds and perform complicated classifiers to detect wheezes. Figure 1.9 shows the system schematics of the block diagram of a smart stethoscope system.

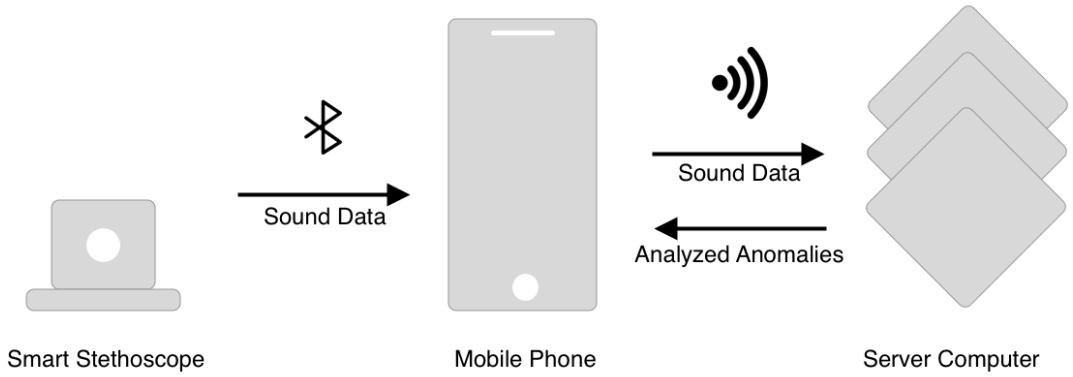

Figure 1.8. The block diagram of a Smart Stethoscope System

In this thesis, more complex classification methods such as Support Vector Machines are used to distinguish wheeze and non-wheeze sounds to increase the performance of accuracy. Moreover, using multiple features are suitable for many applications with the help of server computers with high computational ability. Therefore, the best feature set that includes numerous features is presented with the best classification method. Following properties will be considered to select the best classification method:

- Model accuracy
- Model sensitivity
- Model specificity
- Model F1 score

In this thesis, nine features in Table 1.5 are selected to observe their discriminative abilities on classifying wheezes.

First, features are calculated for every window as detailed in Chapter 2, and relative frequency histograms of features for the wheeze and non-wheeze windows are

Table 1.5. Features which are selected to observe their discriminative abilities on classifying wheezes

Kurtosis
Renyi Entropy
Percentile Frequency Ratios
Mean Crossing Irregularity
Auto Regression Paramaters
Tonality Index
Audio Spectral Envelope
Mel Frequency Cepstral Coefficient
FFT Peak Baseline Difference in dB

shown to distinguish the difference between classes visually.

Second, the following classifiers are applied to the feature set, and their performance metrics are compared.

- Support Vector Machines (SVM)
- K Nearest Neighbour (K-NN)
- Bayesian Classifier with Gaussian Likelihood

Third, the next feature selection techniques are used to state the best set of features that separates wheeze and normal sounds.

- P Values
- Forward Sequential Feature Selection Algorithm
- Backward Sequential Feature Selection Algorithm
- Brute-Force Feature Selection
- Principal Component Analysis (PCA)

- Linear Discriminative Analysis (LDA)
- Computational Time of Each Feature

As the last step, the wheeze detection algorithm for all fifteen seconds of respiratory sounds is developed with the best classifier obtained.This thesis presents the result and details of the detection algorithm.

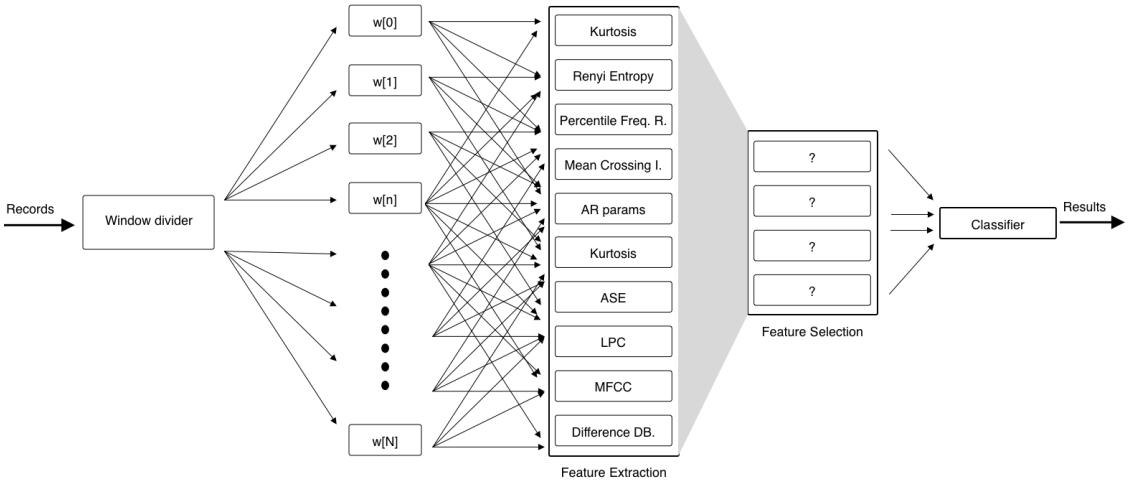

Figure 1.9. The block diagram of wheeze classification and detection

2. DATA

The dataset used in this work contains 34 recordings from 7 subjects (four males and three females). Three subjects have chronic obstructive pulmonary disease (COPD) and the others have Asthma. Furthermore, all the patients were under treatment in the Istanbul Yedikule Teaching Hospital for Chest Diseases and Thoracic Surgery. Their ages were between 33 and 67.

Each respiratory sound recording session has a duration of 15 seconds. The sound signal was sampled at 9600 Hz with 12 bit sample resolution. The system used for recordings was developed at Bogazici University, and the dataset was created in 2012.

The system which is used to record respiratory sounds in the dataset is a fourteen channel device which is capable of acquisition and processing analog sound data. The system uses 14 electret microphones (Sony ECM-44BPT) attached inside Teflon capsules with a conical air cavity. These microphones are located on 14 predefined points on the posterior chest wall. Figure 2.1 shows these points' locations. More detailed information about the system can be found in [7].

Two experts labelled all recordings by auditory confirmation by visually inspecting the time-expanded sound signals. They labelled the total number of 308 wheezes, and they marked the location of wheeze intervals. Because finding the correct start points and endpoints of wheezes is almost impossible with the human sense, 106 ms (512 samples long) before and after of wheeze signals were labelled as undecided area. After labelling wheezes and undecided intervals, the remaining parts of the sounds were labelled as non-wheeze sounds which are normal respiratory sounds. An example

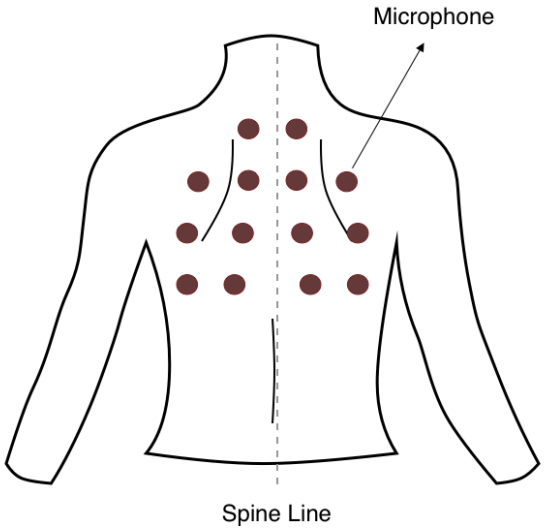

Figure 2.1. Microphone locations on the posterior chest wall

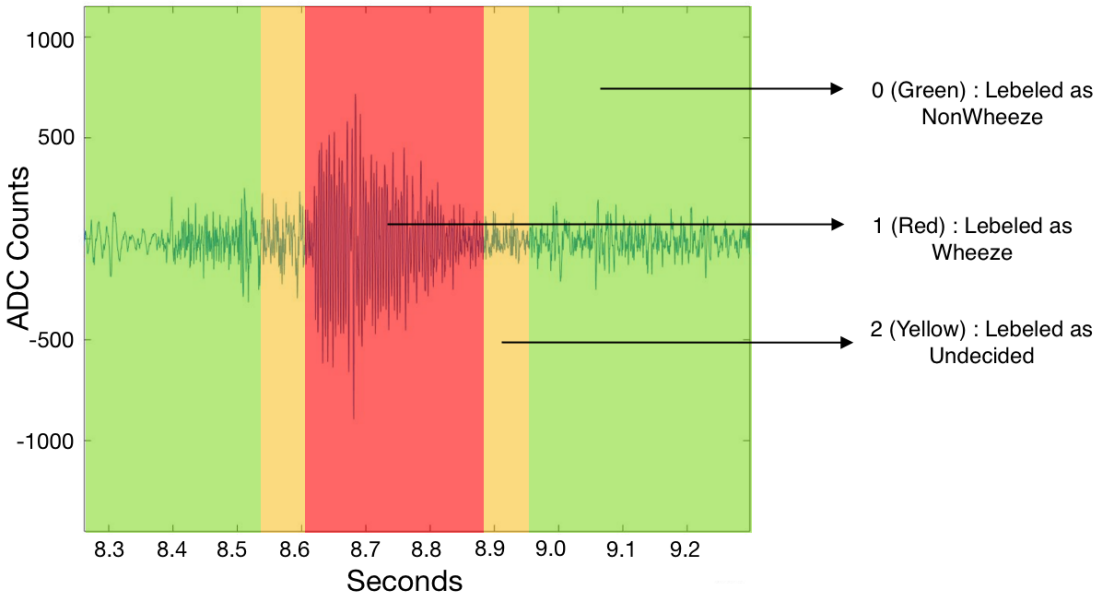

Figure 2.2. Example label of a respiratory sound partition that includes wheeze sound

partition of respiratory sound that includes wheeze is showed in Figure 2.2.

2.1. Preparation

Each respiratory sound recording session which is sampled at 9600 Hz, with 12-bit sample resolution has a duration of 15 seconds. In this study, every recording is divided into shorter windows that contain 512 samples with a 25 per cent overlap.

Windowing operation divides 34 recordings to 12716 windows. Experts labeled 308 wheeze intervals and these intervals consist of 3238 windows. Also, 8280 windows are labelled as non-wheeze. 1198 windows are marked as undecided.

Table 2.1. Properties of wheezes' duration in dataset

Maximum duration	1417 ms
Minimum duration	75 ms
Mean of durations	474ms

3. FEATURES

In this study, nine features are evaluated to classify wheeze and normal respiratory sounds. We selected these nine features both in frequency and time domain because we believe that their discriminative abilities are suitable for the wheeze signal according to researches in Section 1.3 and our experience.

After the feature calculations via MATLAB, every window has 33 dimensional space. Some features like Mel Frequency Cepstral Coefficients (MFCC) and Auto Regression Parameters (AR) account to more than one value. MFCC has 13 parameters and AR has 7 parameters with error term, therefore the dimension of dataset with nine features increased to 33 dimensions.

3.1. Kurtosis

Kurtosis gives a degree of how much the sample distribution fits the normal distribution. It is also a prediction about the shape of the distribution. Any normal distribution has a kurtosis value as 3. If the kurtosis value of the distribution is lower than 3, distribution is more like uniform distribution and less outlier-prone. On the other hand, the sample distributions which have higher kurtosis values are more outlier-prone. The below formula defines the kurtosis for random variable X.

$$k = \frac{E\left(X - \mu\right)^4}{\sigma^4} \tag{3.1}$$

where mean of X is μ, kurtois value is k and σ is the standard deviation.

Every sound sample in respiratory sounds in the time domain has been assumed as a real-valued discrete random variable. Respiratory sound distribution over lungs is non-uniform distribution.Moreover, normal respiratory sounds and wheezes have different distribution characteristics. Because of the wheeze's steady signal type, its

distribution is more similar to the uniform distribution, so its kurtosis value is expected to be lower than the kurtosis value of normal sounds. Figure 3.1 shows the relative frequency histogram of the kurtosis values regarding both classes.

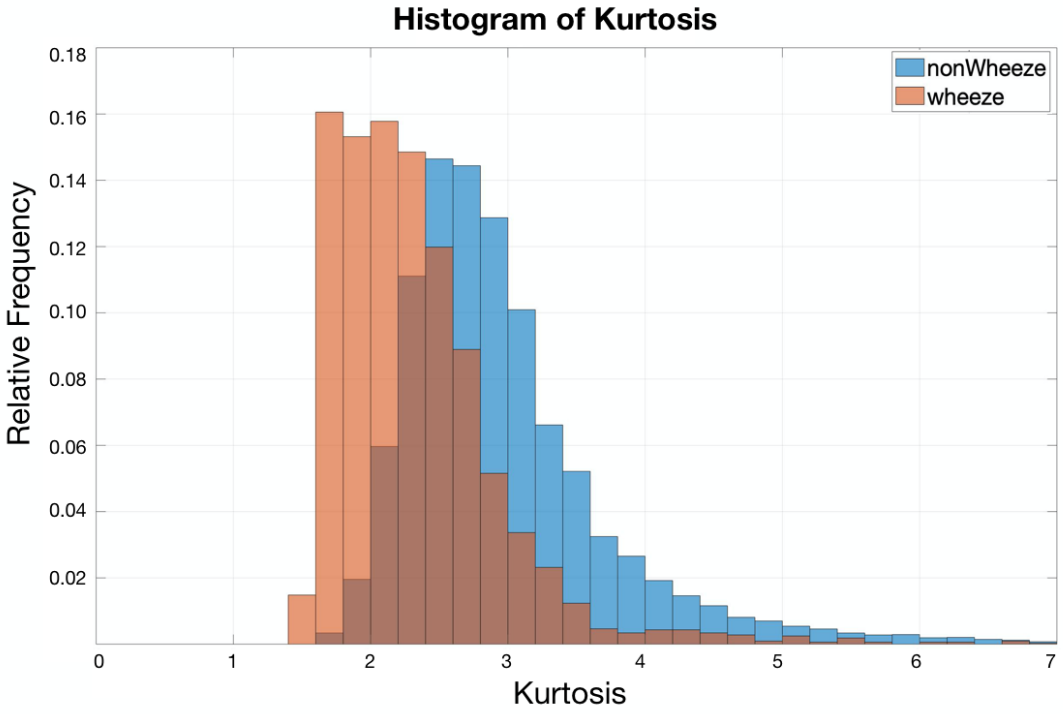

Figure 3.1. Relative frequency histogram of the kurtosis values regarding both classes

As seen in Figure 3.1, windows labeled as normal sounds have kurtosis value mostly distributed near 3. On the other hand, wheeze labeled windows' kurtosis values are distributed around 2. The result is expected because normal respiratory signals have a normal distribution, whereas wheeze signals have a uniform distribution.

3.2. Renyi Entropy

Claud Shannon [35] published the paper "A mathematical theory of communication" in 1948, which has led to the birth of information theory and informational entropy. According to Shannon, the information entropy is the measure of the average level of uncertainty inherent in the random variable's possible outcomes.

Renyi entropy, which is used in this study, is the generalized variant of Shannon entropy. The only difference is that Shannon entropy has an order value of 1. The following formula defines the Renyi entropy of discrete random variable with order α.

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2(\sum_{i=1}^{n} p_i^\alpha) \tag{3.2}$$

where p_i are sound samples for i = 1,...,n. In this thesis, three α values were chosen to classify wheeze and non-wheeze sounds. Figures 3.2, 3.3 and 3.3 show the relative frequency histograms of the Renyi entropy of order 1,2 and 3 regarding both classes, respectively.

Wheezes, as abnormal sounds, cause more uncertainty in sound characteristics in contrast to normal respiratory sounds. The most obvious reason is the high frequency components in their frequency response. Consequently, windows which are labeled as wheeze is expected to have higher entropy values in opposition to windows labeled normal. Figures 3.2, 3.3 and 3.3 have no discriminating differences according to their distributions.

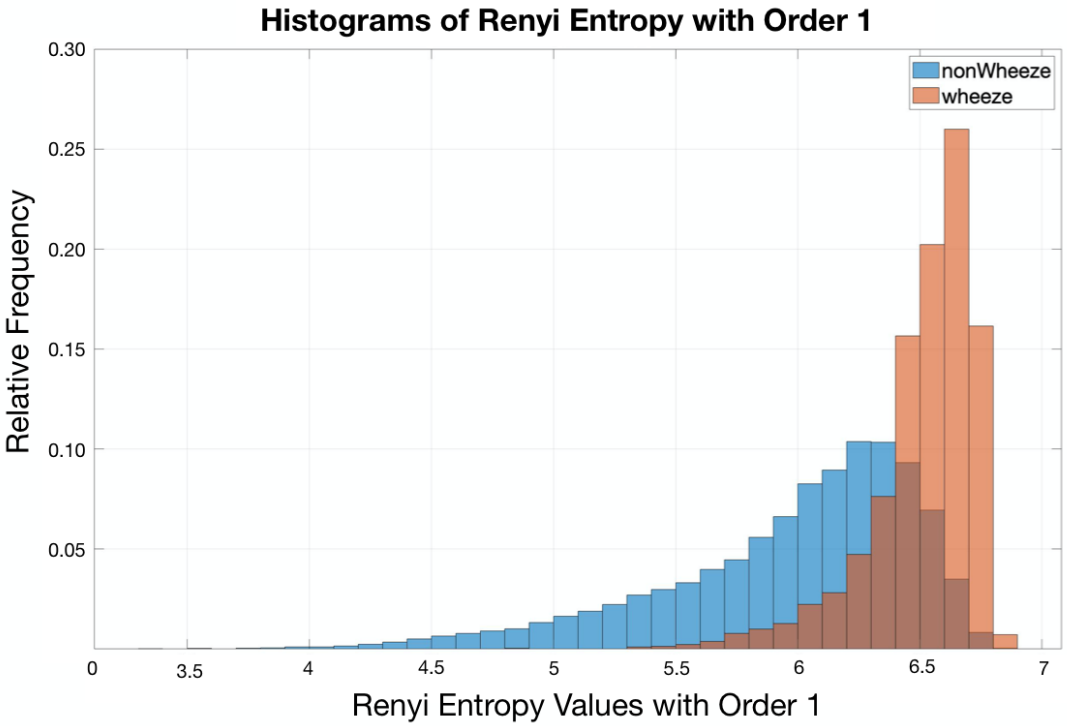

Figure 3.2. Relative frequency histogram of the Renyi Entropy values with order 1

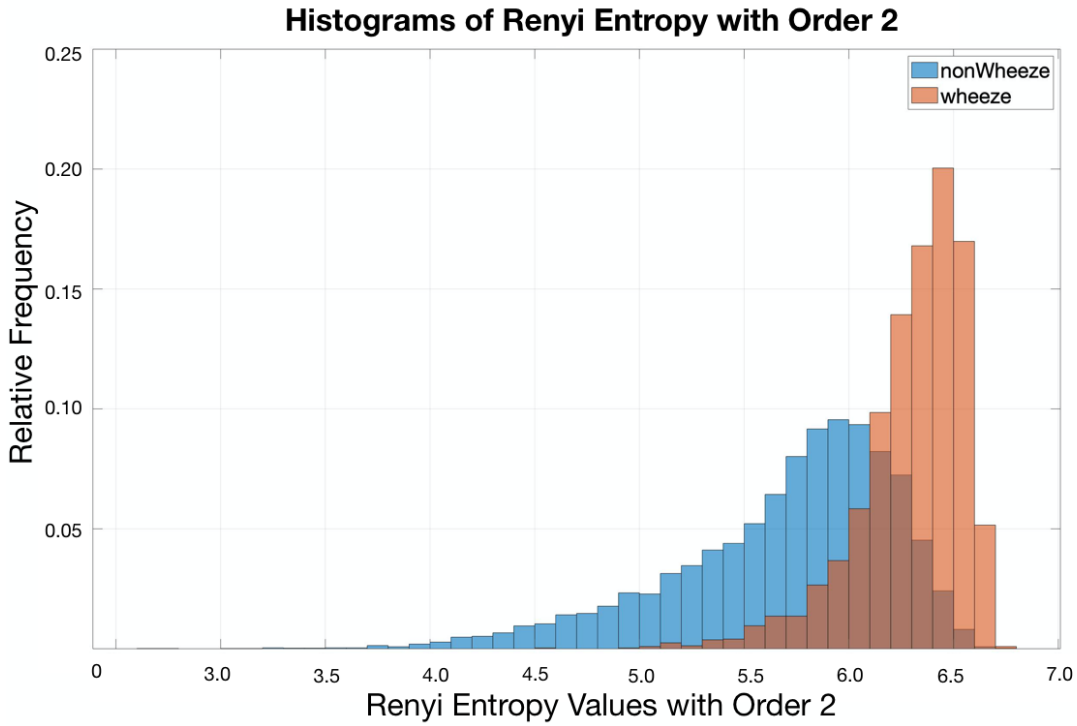

Figure 3.3. Relative frequency histogram of the Renyi Entropy values with order 2

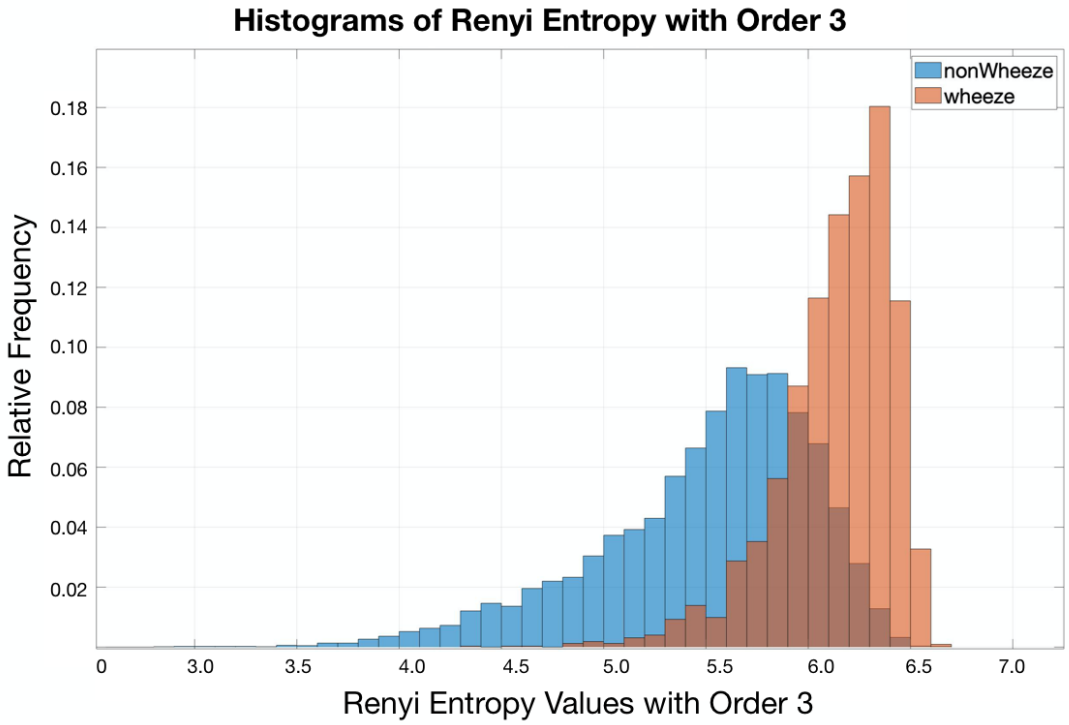

Figure 3.4. Relative frequency histogram of the Renyi Entropy values with order 3

3.3. Percentile Frequency Ratios

The power spectral density shows the power of a signal as a function of frequency. f_{25}, f_{50}, f_{75}, f_{90} are the frequencies which 25%, 50%, 75% and 90% of the total area under the normalized power spectral density function is reached, respectively.

The Welch method is used to estimate the power spectral density of the sound signal. The segment length and overlapping ratio used in the Welch method are 128 and 50%, respectively. Each segment windowed with the Hamming Window, and 256 is the number of Fourier Transform points to calculate power spectral density.

When a single frequency or a group of similar frequencies make a peak in power spectral density function, some percentile frequencies are expected to be close to each other according to peak frequency value. Four frequency ratios, which are f_{25}/f_{50}, f_{25}/f_{90}, f_{50}/f_{75}, and f_{50}/f_{90} are used to find the best discriminative percentile frequency ratio.

Wheezes usually make a peak in the range of high frequencies on its frequency response. On the other hand, normal respiratory sounds usually have a peak in the range of low frequencies. Figure 3.5 shows the plots of normalized power spectral density of wheeze and normal sounds.

Figures 3.6, 3.7, 3.8, and 3.9 show the relative frequency histograms of f_{25}/f_{50}, f_{25}/f_{90}, f_{50}/f_{75}, and f_{50}/f_{90} ratios.

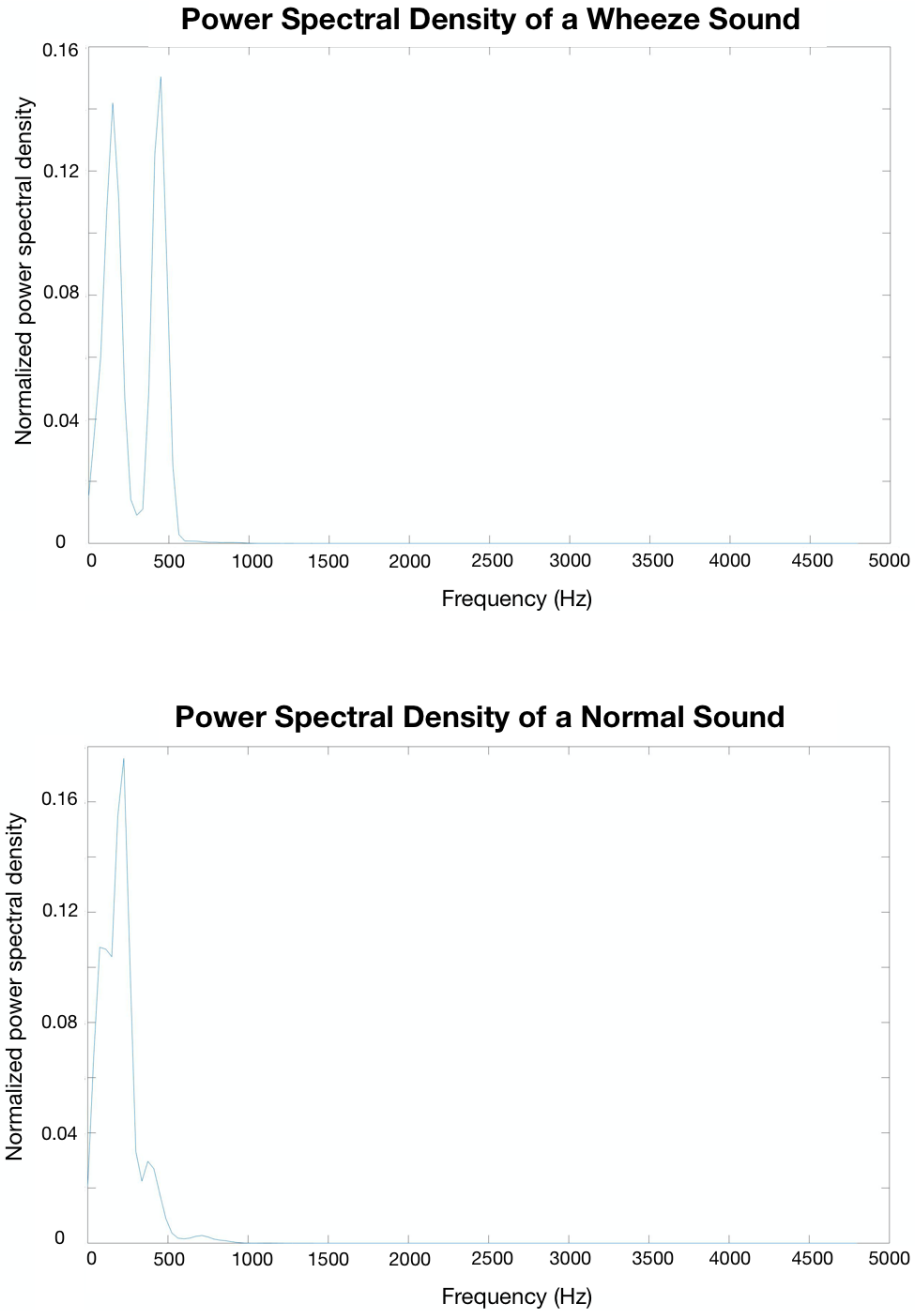

Figure 3.5. Plots of normalized power spectral density of wheeze and normal sounds.

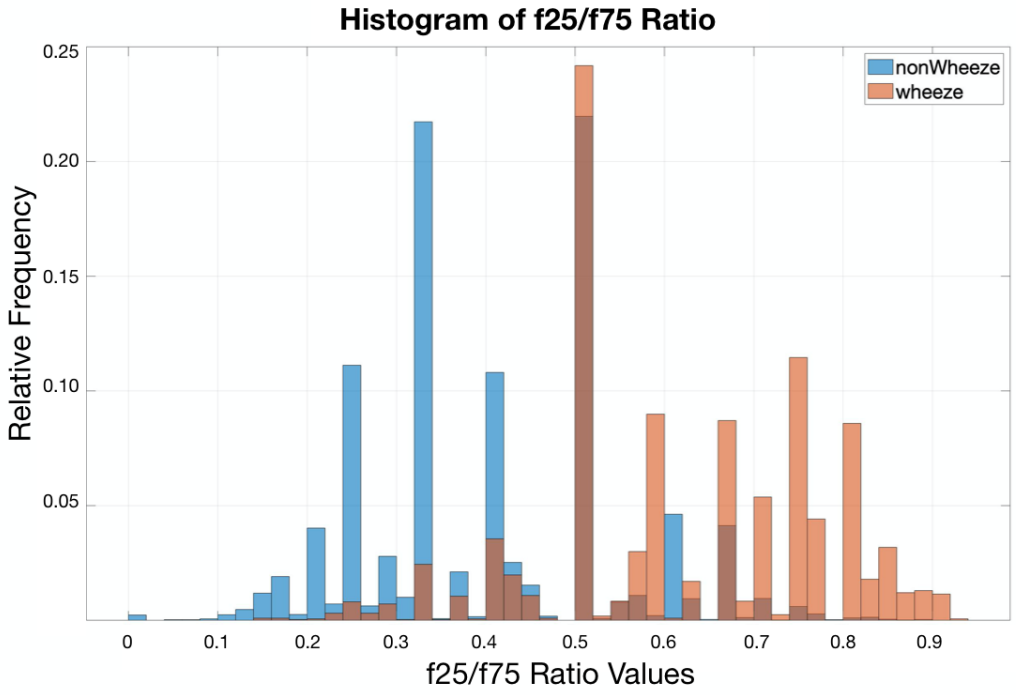

Figure 3.6. Relative frequency histogram of the f_{25}/f_{75} values regarding both classes

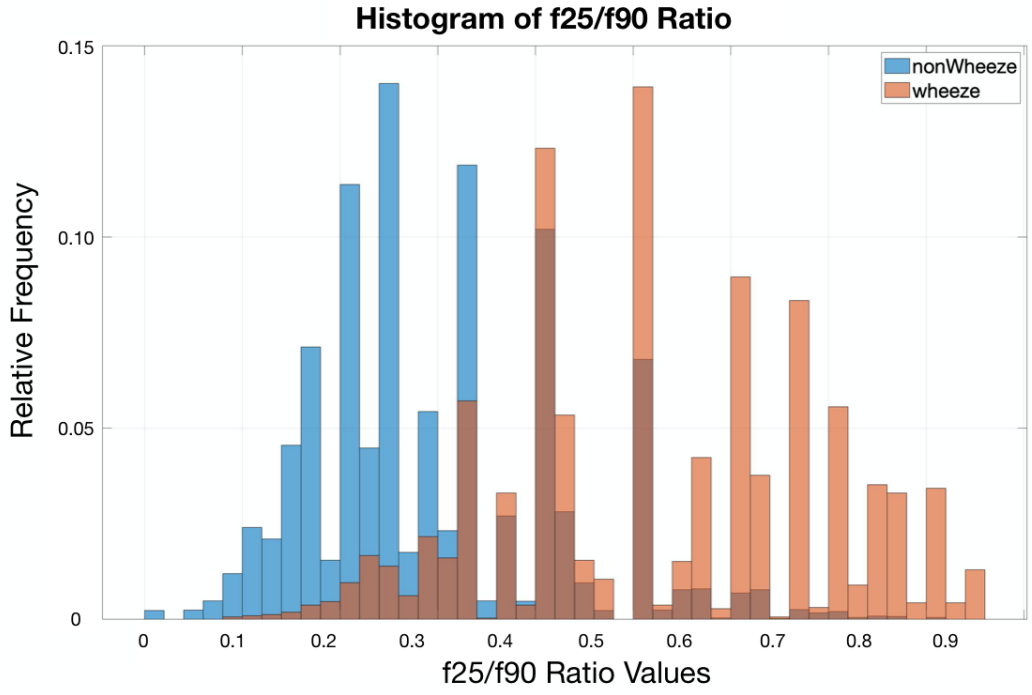

Figure 3.7. Relative frequency histogram of the f_{25}/f_{90} values regarding both classes

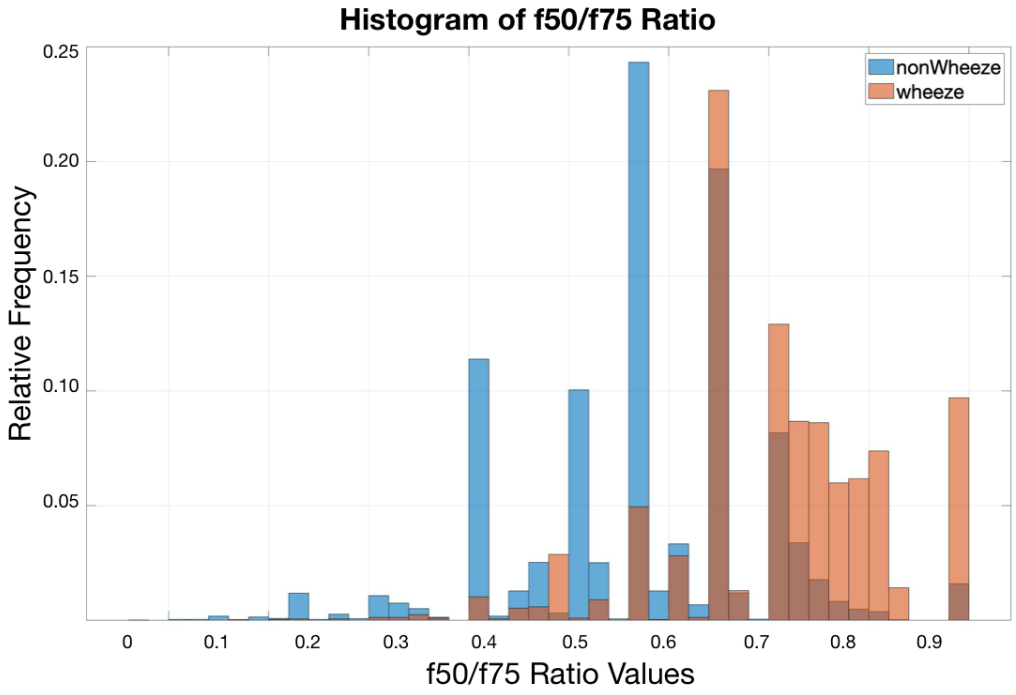

Figure 3.8. Relative frequency histogram of the f_{50}/f_{75} values regarding both classes

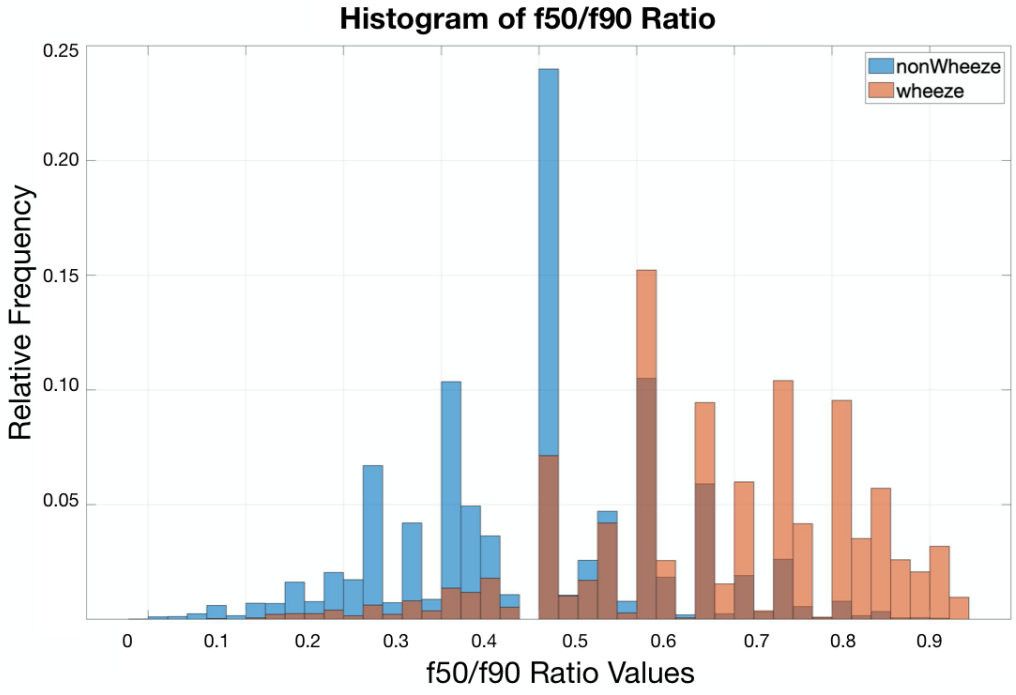

Figure 3.9. Relative frequency histogram of the f_{50}/f_{90} values regarding both classes

3.4. Mean Crossing Irregularity

Zero crossing is a mathematical term defined as the point where the function of the signal changes its sign by crossing the axis (zero value). It is widely used in acoustic, sound - image processing, and electronics. In this study, the regularity level of the intervals between zero crossings can be a clue of the occurrence of wheezes. Mean crossing irregularity is defined as following equation.

$$MCI = \frac{\sqrt{Var\,(X)}}{E(X)} \tag{3.3}$$

where X is the interval between two sequential zero crossing in signal.

Wheeze and non-wheeze signals have different mean crossing behaviour because wheeze waveform has more regular oscillations than non-wheeze waveform [30]. Figure 3.10 shows the relative frequency histograms of the MCI values regarding both wheeze and non-wheeze signals.

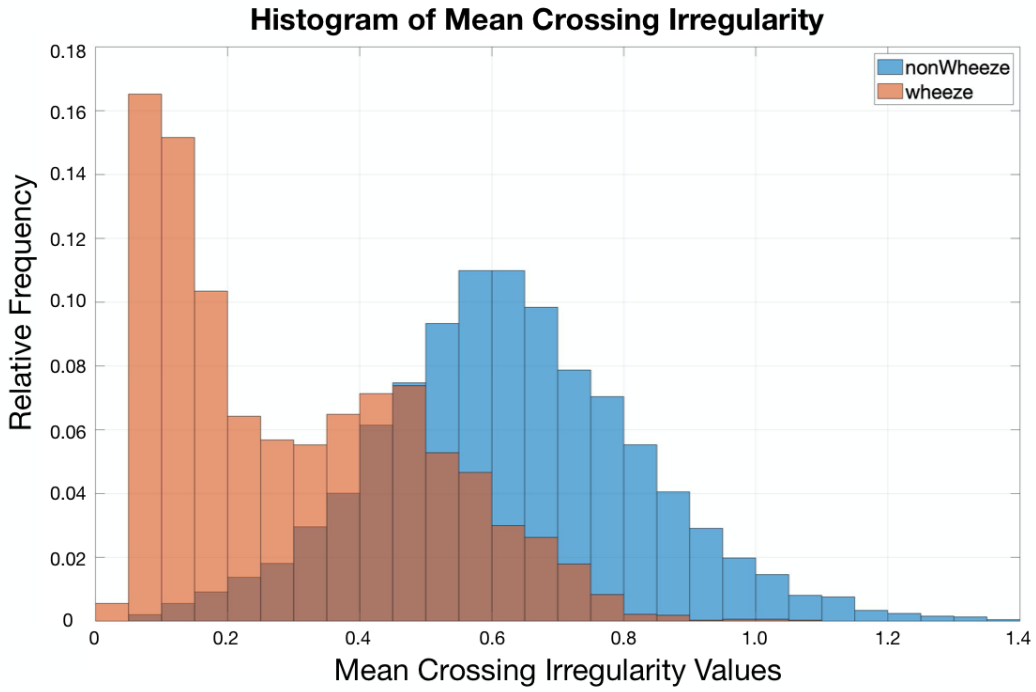

Figure 3.10. Relative frequency histogram of the Mean Crossing Irregularity values regarding both classes

3.5. Autoregressive Model (AR) parameters

The autoregressive model is widely used in signal processing, statistic, and econometrics. It estimates the current sample of a sound signal by creating the linear function of previous sound samples. The autoregressive model gives an accurate estimation of the speech signal, and it is widely used in the transmission of these signals. It is also used in stock market prediction, channel prediction, and many other applications. The following equation is the fundamental equation behind the autoregressive model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t \tag{3.4}$$

where ϵ_t is the error of model and $\phi_1, \phi_2....\phi_p$ are the AR coefficients. The number of parameters, p, in 3.4 represents the order of AR model. In this thesis, 6'th order AR paremeters are used [36].

AR model, solving by Yule-Walker equations method, is a discriminative technic when the sound signal within the window is stationary. So, the respiratory signals can be represented well with AR Model because of their stationary structures. The used algorithm in estimating AR parameters is the Autocorrelation Levinson Durbin method just like in Liner Predictive Coding (LPC) paramaters' calculation. AR parameters are same as the LPC parameters when Yule-Walker equations and Autocorrelation Levinson Durbin method is used.

Coefficients of AR model can exceptionally provide information about the waveform of stochastic concern. Figure 3.11, Figure 3.12, Figure 3.13, Figure 3.14, Figure 3.15, Figure 3.16 and, Figure 3.17 shows the relative frequency histograms all six AR coefficients and error term.

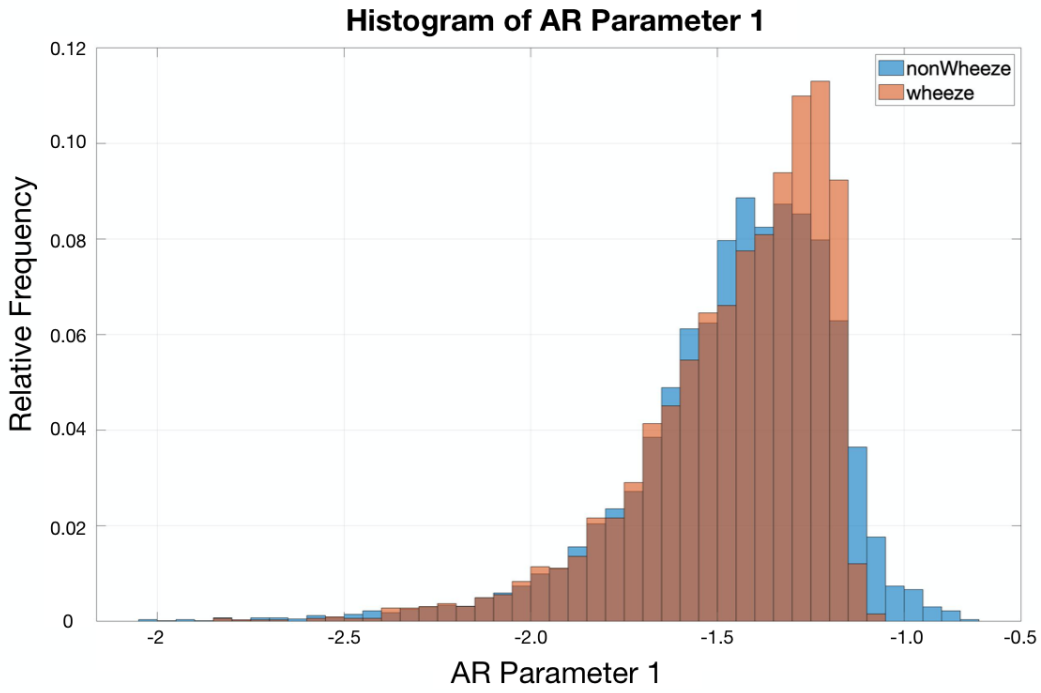

Figure 3.11. Relative frequency histogram of AR 1 values regarding both classes

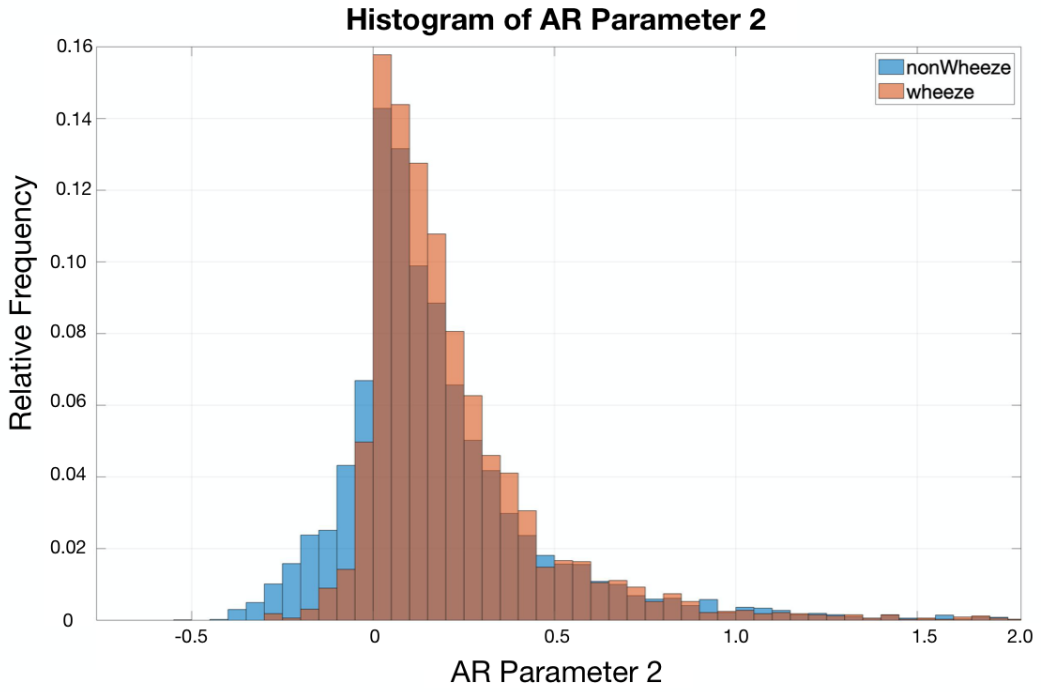

Figure 3.12. Relative frequency histogram of AR 2 values regarding both classes

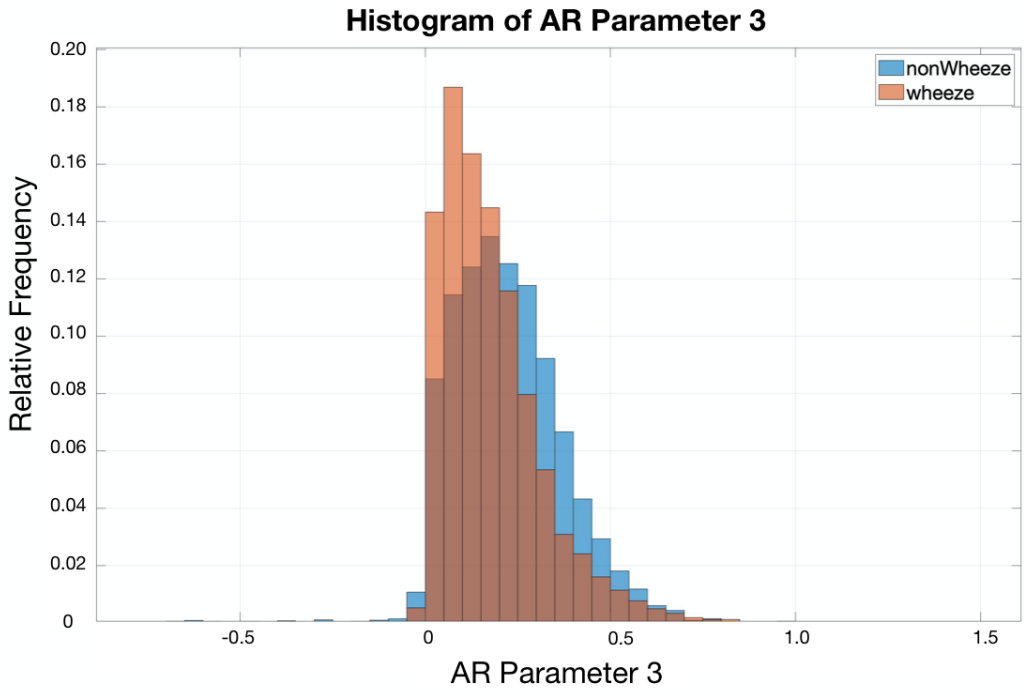

Figure 3.13. Relative frequency histogram of AR 3 values regarding both classes

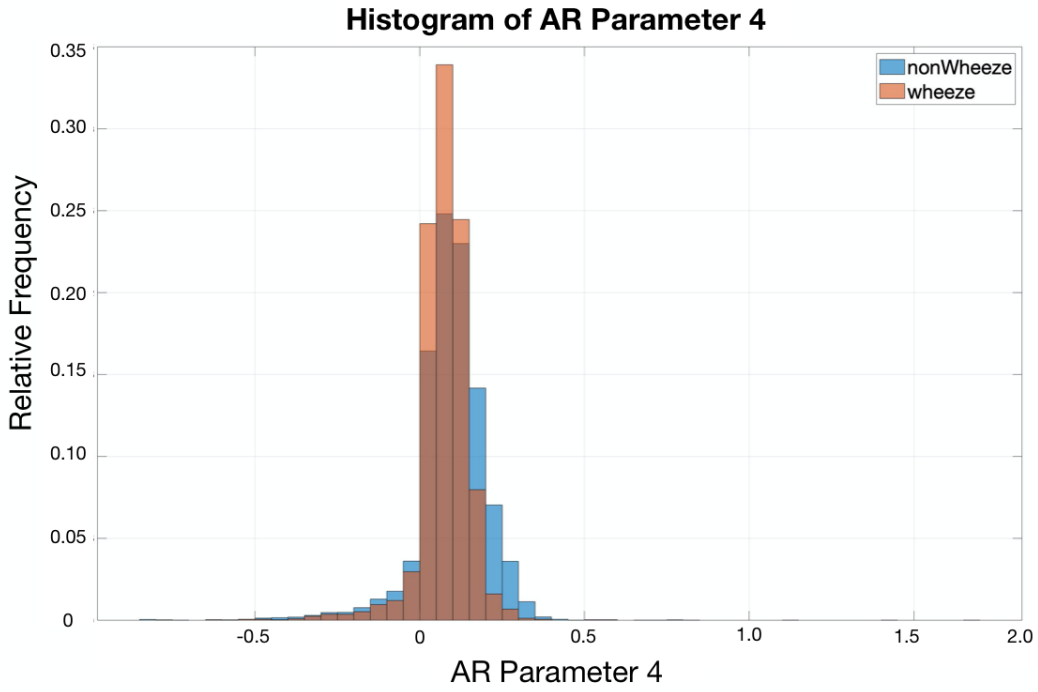

Figure 3.14. Relative frequency histogram of AR 4 values regarding both classes

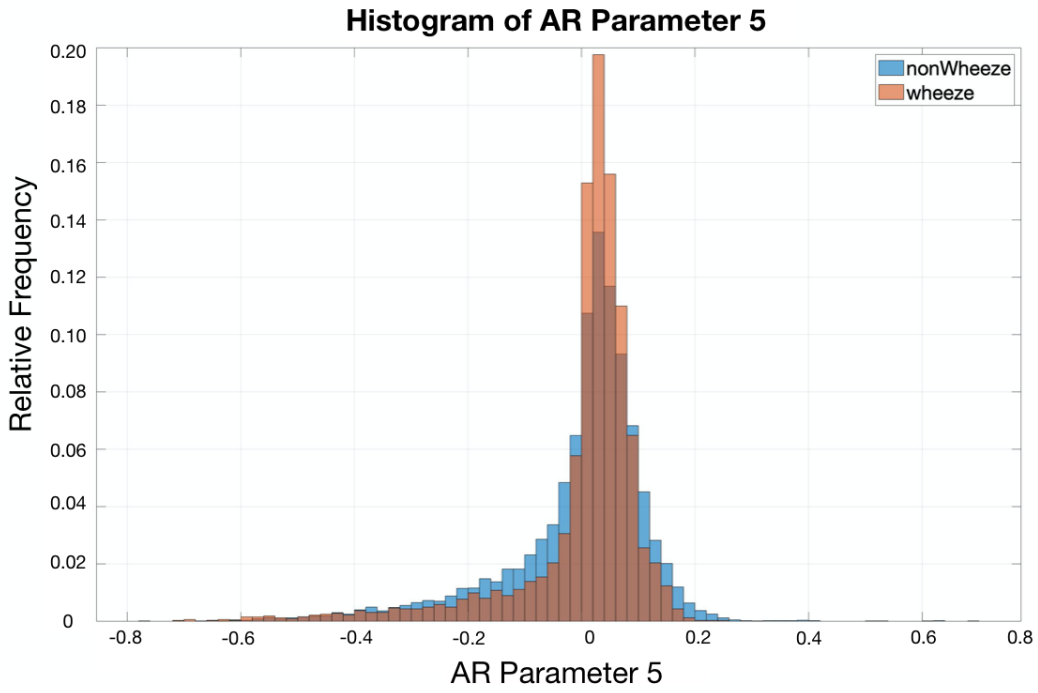

Figure 3.15. Relative frequency histogram of AR 5 values regarding both classes

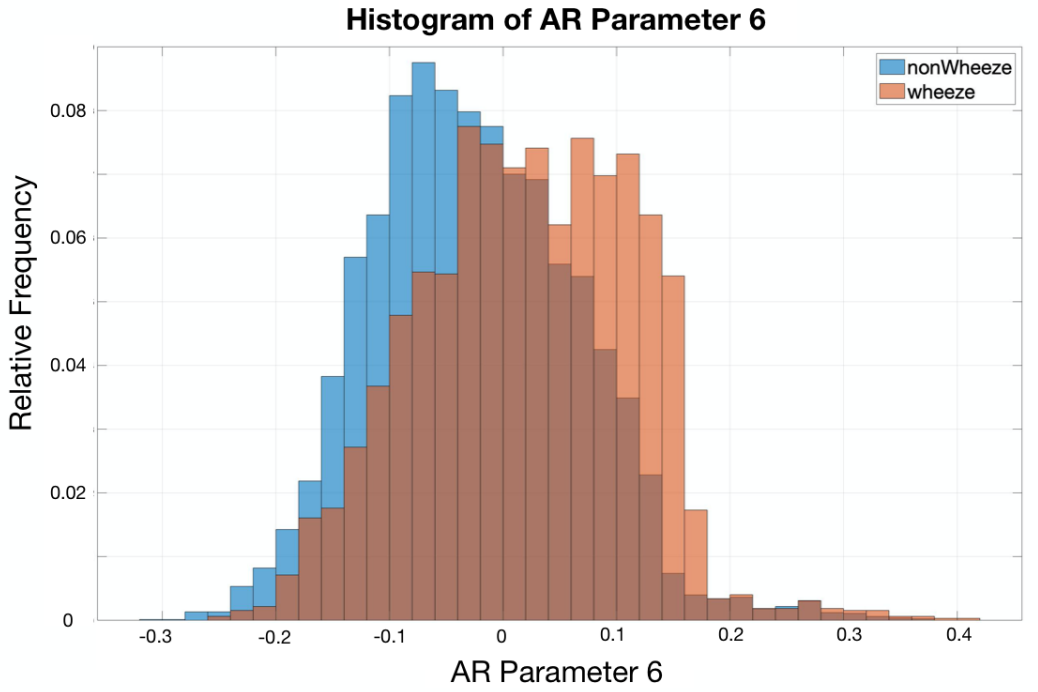

Figure 3.16. Relative frequency histogram of AR 6 values regarding both classes

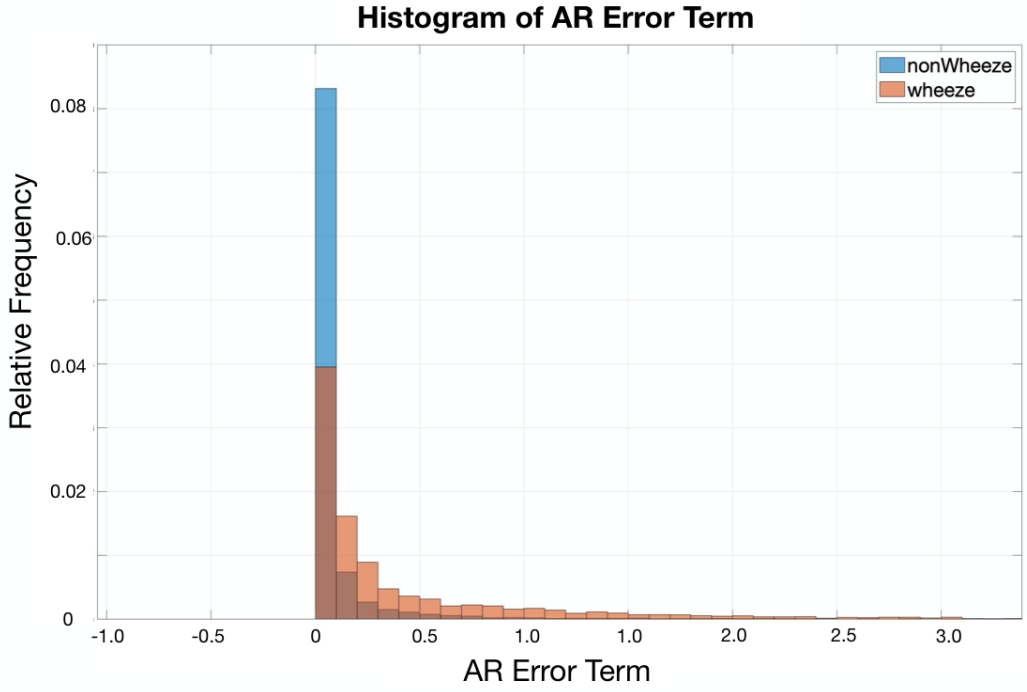

Figure 3.17. Relative frequency histogram of AR Error Term values regarding both classes

3.6. Mel Frequency Cepstral Coefficients

Mel frequency cepstral coefficients (MFCC) are the representation of speech signals according to the hearing ability of human being. MFCC uses the Discrete Fourier Transform (DFT) with a different approach. The frequency bands are placed logarithmically according to Mel scale in DFT to compute MFCC so that sounds can be represented more like a human hearing auditory model.

First, the Discrete Fourier Transform is calculated for every respiratory sound for computation of MFCC. Then, thirteen overlapping triangular filters are applied to the spectrum according to Mel Frequency scale. The logarithmic equation in 3.5 is used to convert the energy of each Mel frequency bands to the logarithmic scale. Finally, the amplitudes of discrete Cosine Transform of Mel frequencies in logarithmic scale give the Mel frequency coefficients.

$$M(f) = 1125 \ln(1 + f/700) \tag{3.5}$$

Even though Mel Frequency Cepstral Coefficients (MFCC) are developed mainly for speech recognition, it also used to to classify abnormal sounds superimposed on respiratory sounds [25]. Figure 3.18 - 3.31 shows the relative frequency histograms of fourteen Mel Frequency Cepstral coefficients.

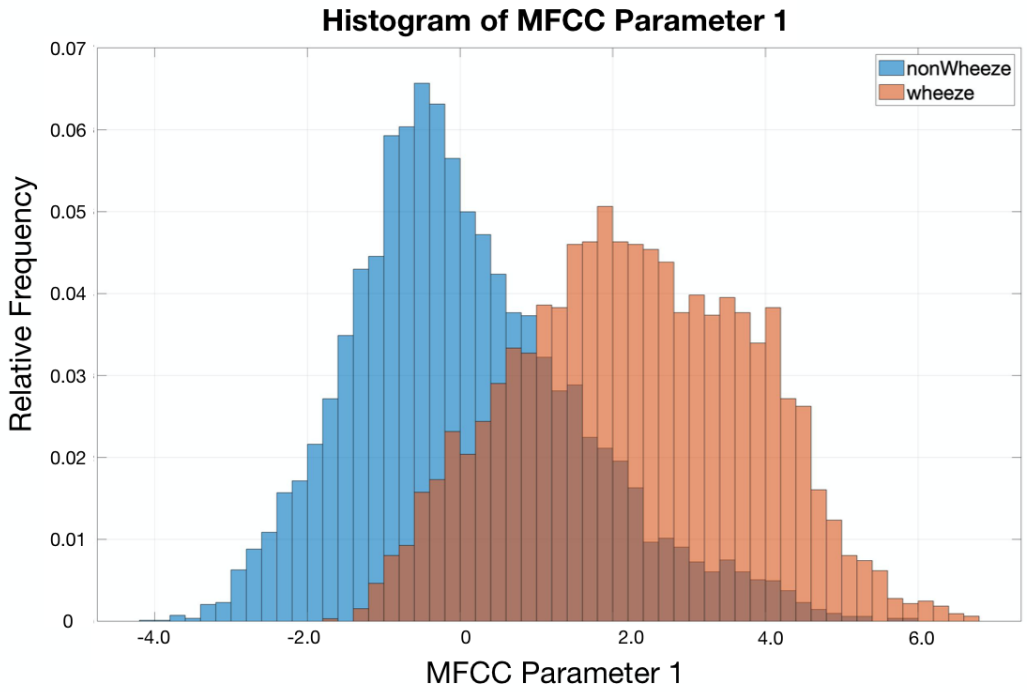

Figure 3.18. Relative frequency histogram of MFCC 1 values regarding both classes

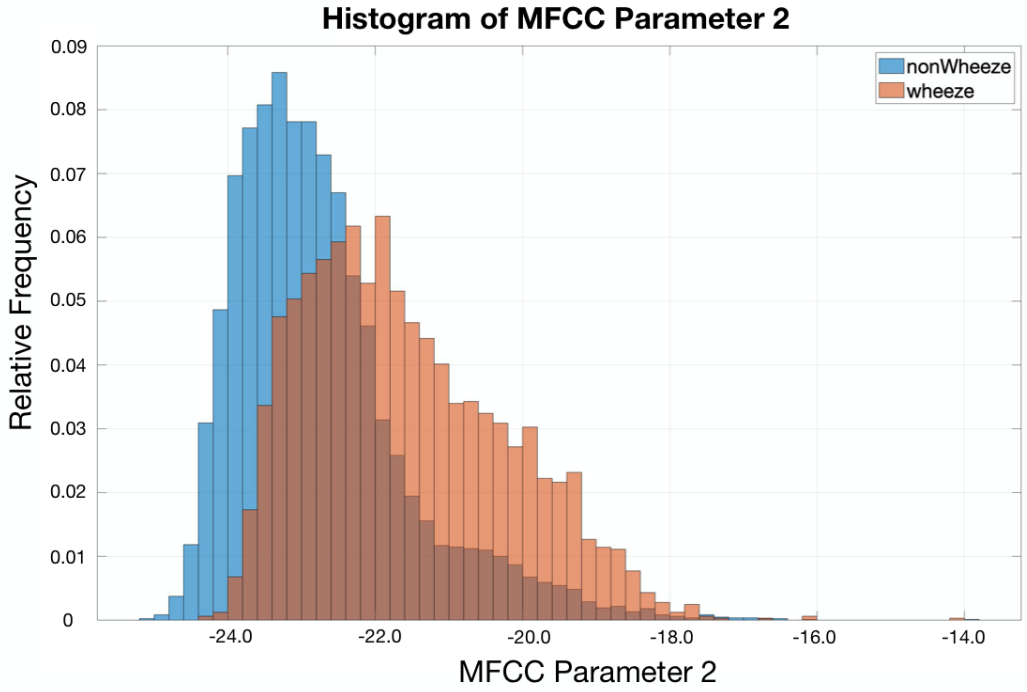

Figure 3.19. Relative frequency histogram of MFCC 2 values regarding both classes

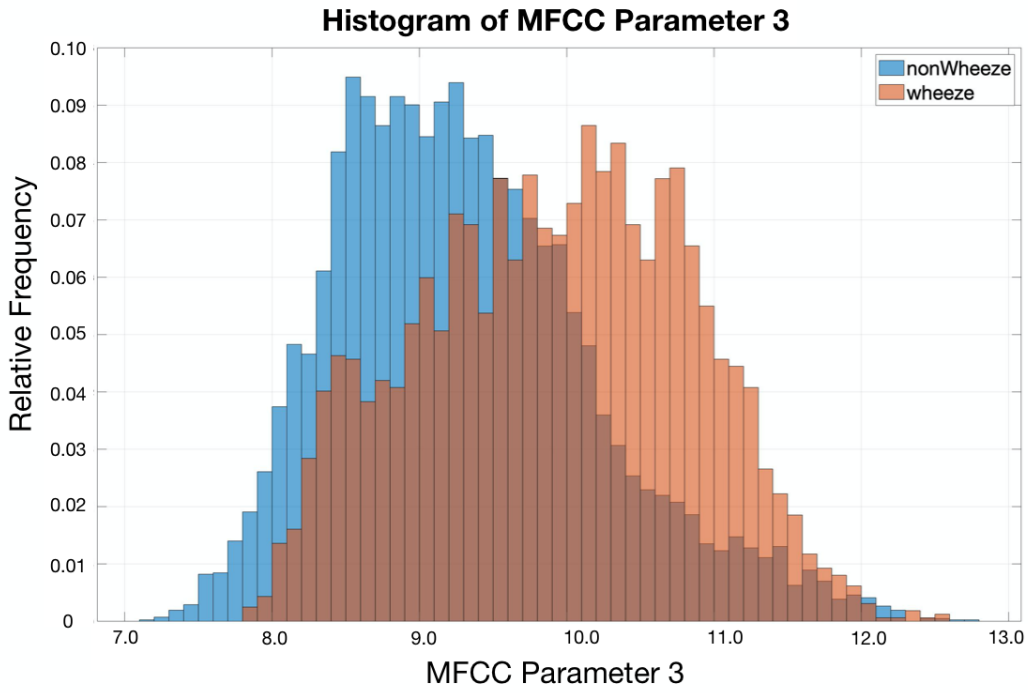

Figure 3.20. Relative frequency histogram of MFCC 3 values regarding both classes

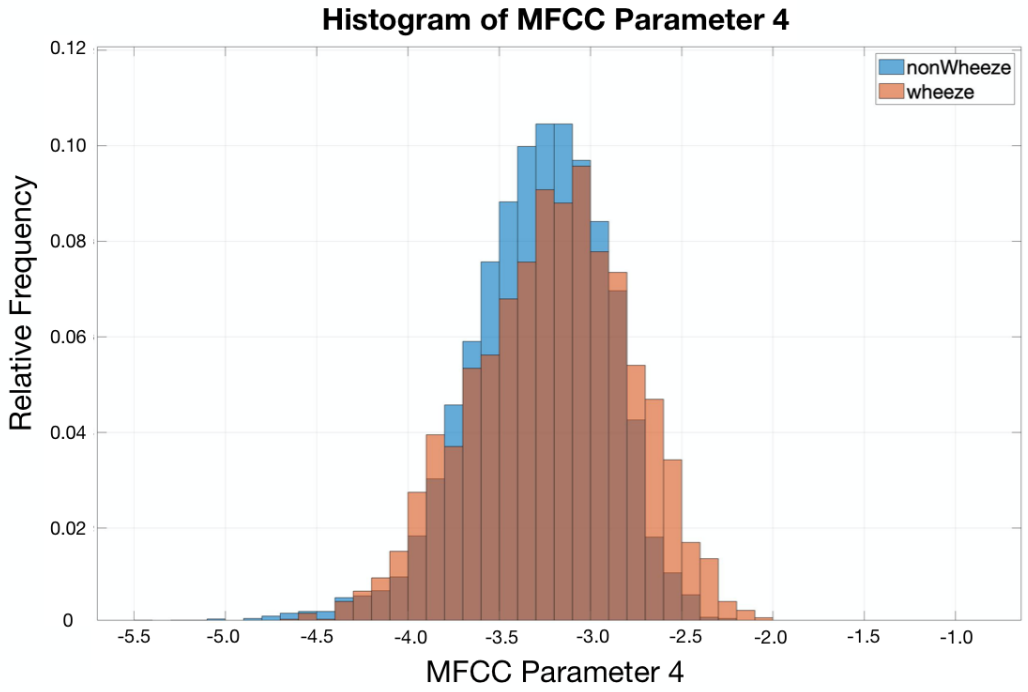

Figure 3.21. Relative frequency histogram of MFCC 4 values regarding both classes

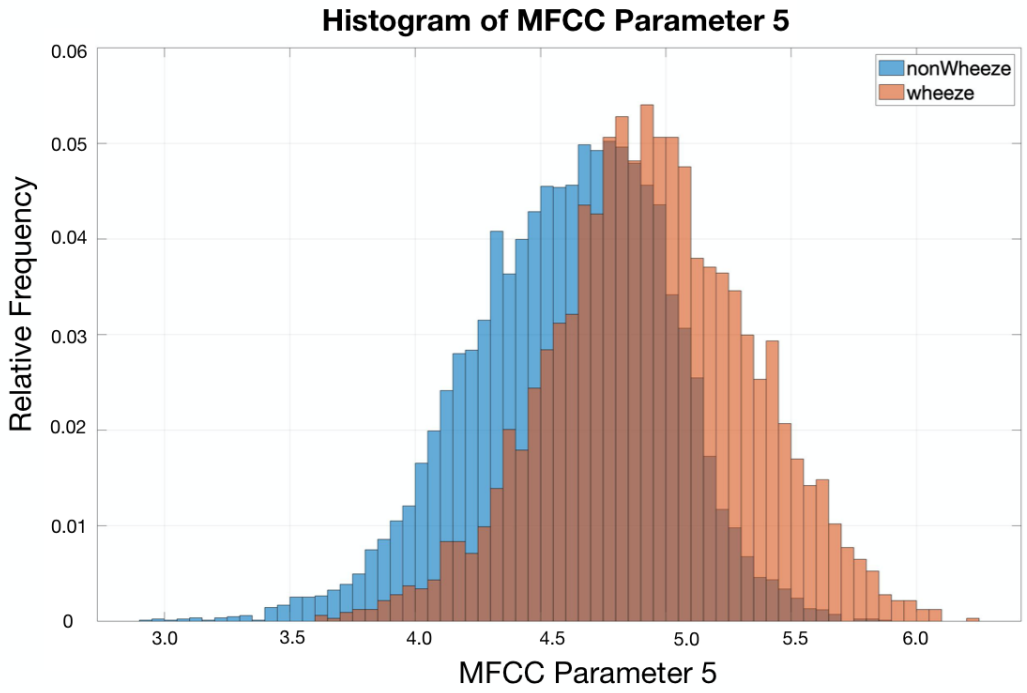

Figure 3.22. Relative frequency histogram of MFCC 5 values regarding both classes

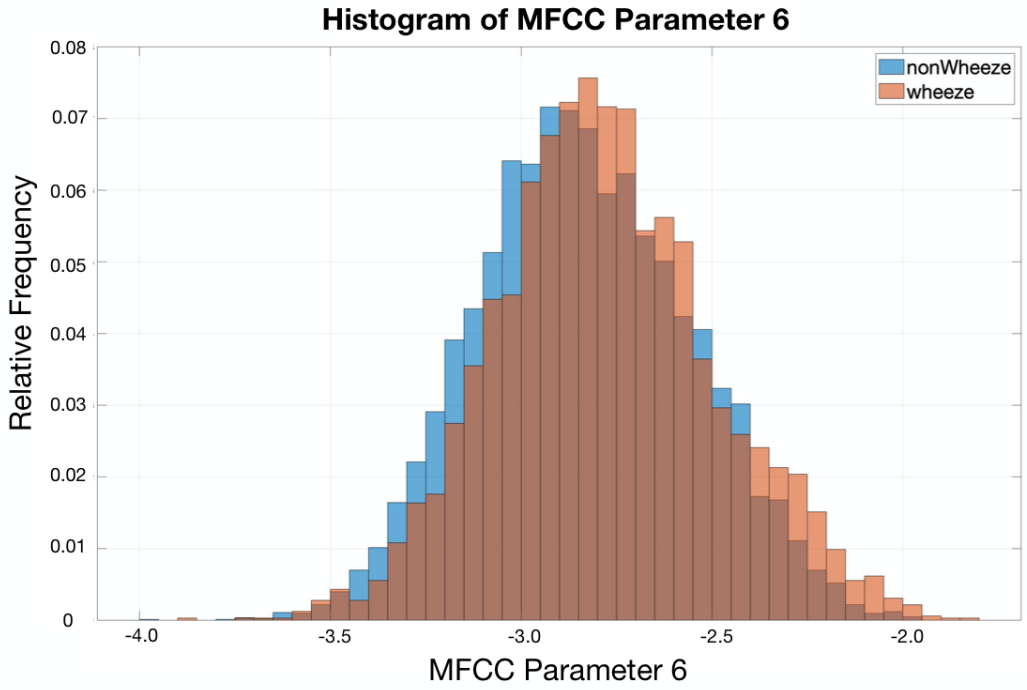

Figure 3.23. Relative frequency histogram of MFCC 6 values regarding both classes

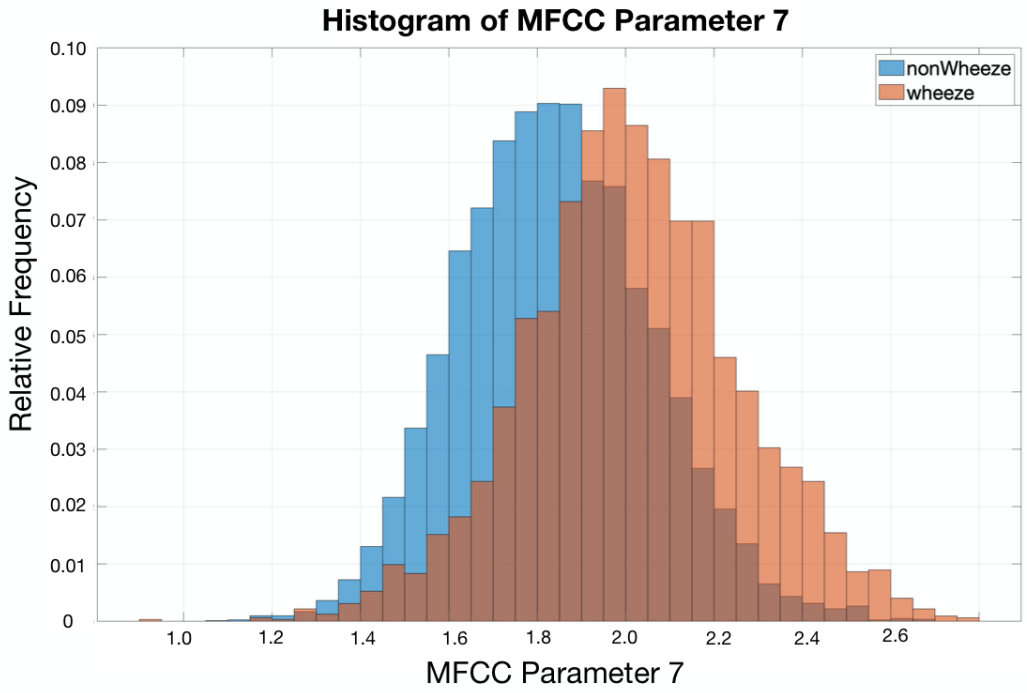

Figure 3.24. Relative frequency histogram of MFCC 7 values regarding both classes

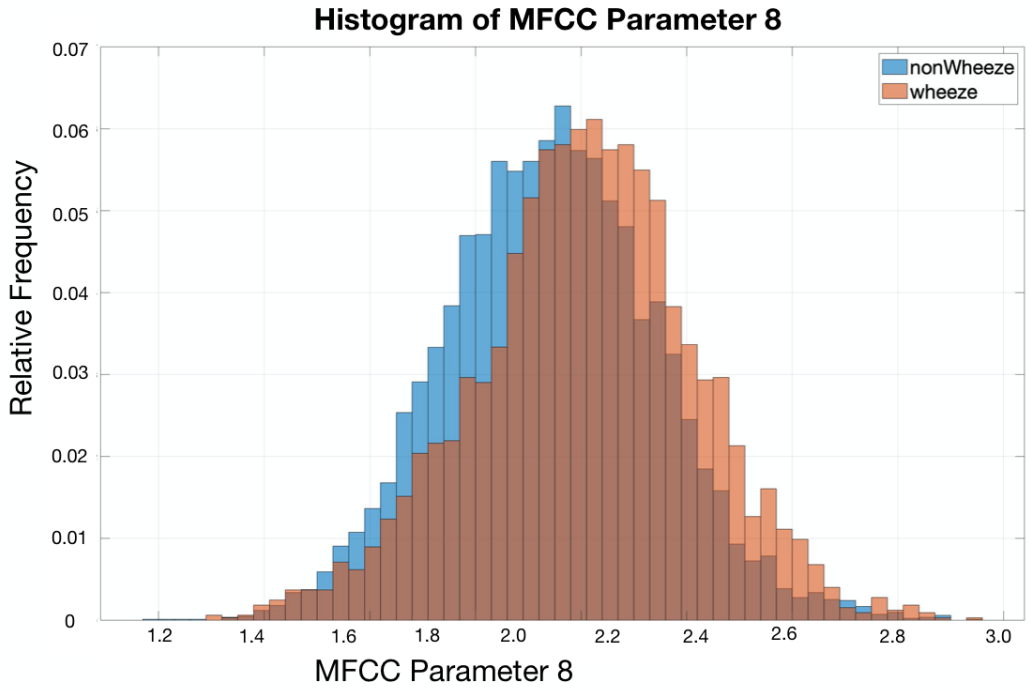

Figure 3.25. Relative frequency histogram of MFCC 8 values regarding both classes

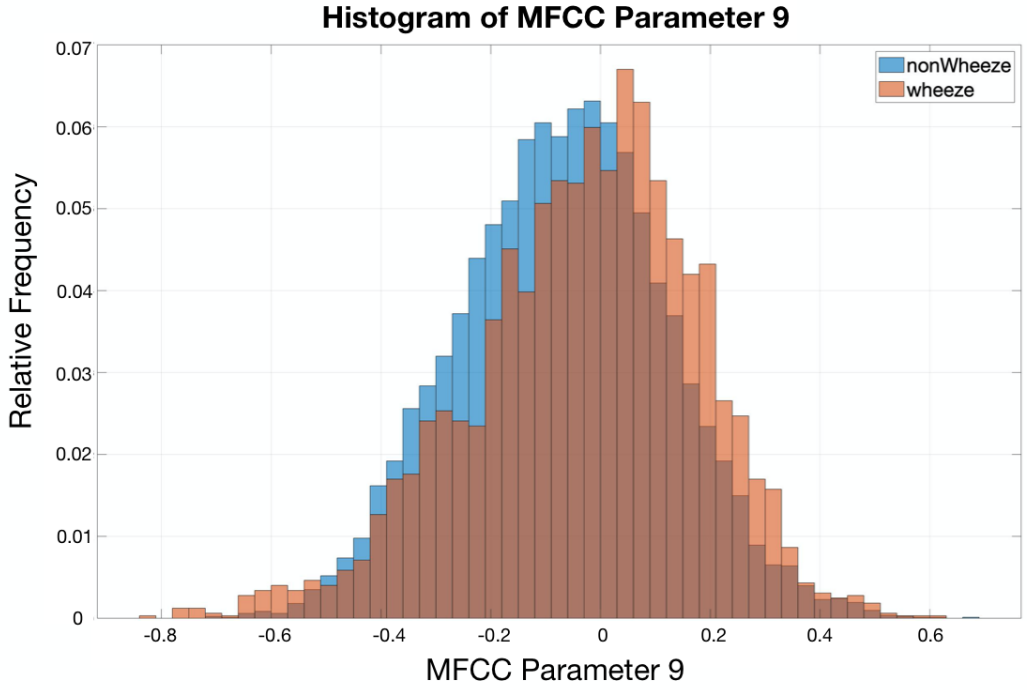

Figure 3.26. Relative frequency histogram of MFCC 9 values regarding both classes

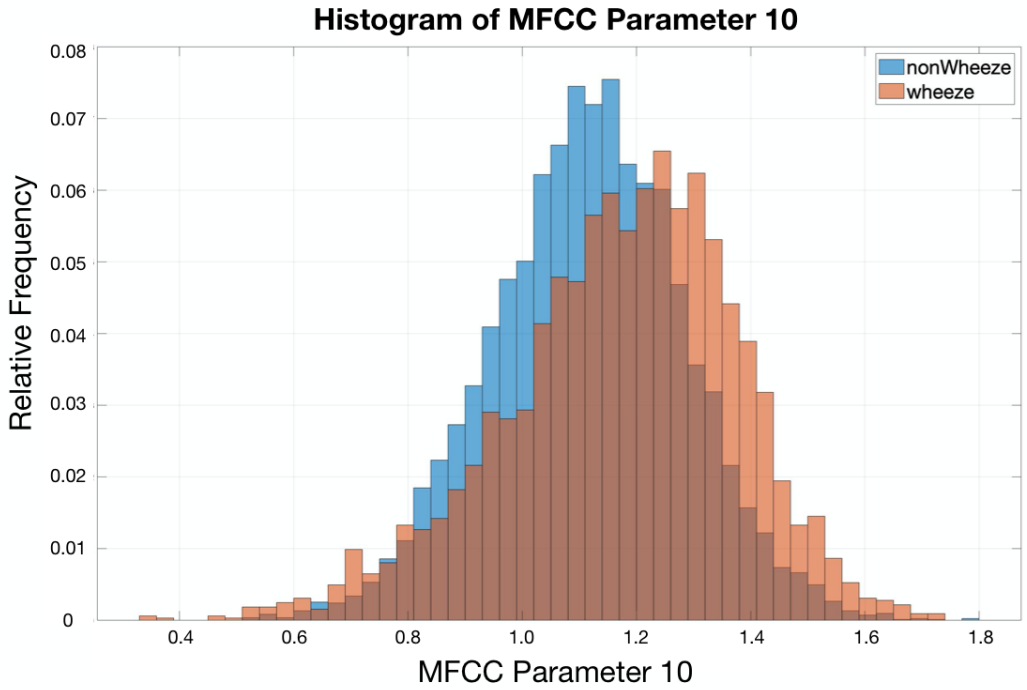

Figure 3.27. Relative frequency histogram of MFCC 10 values regarding both classes

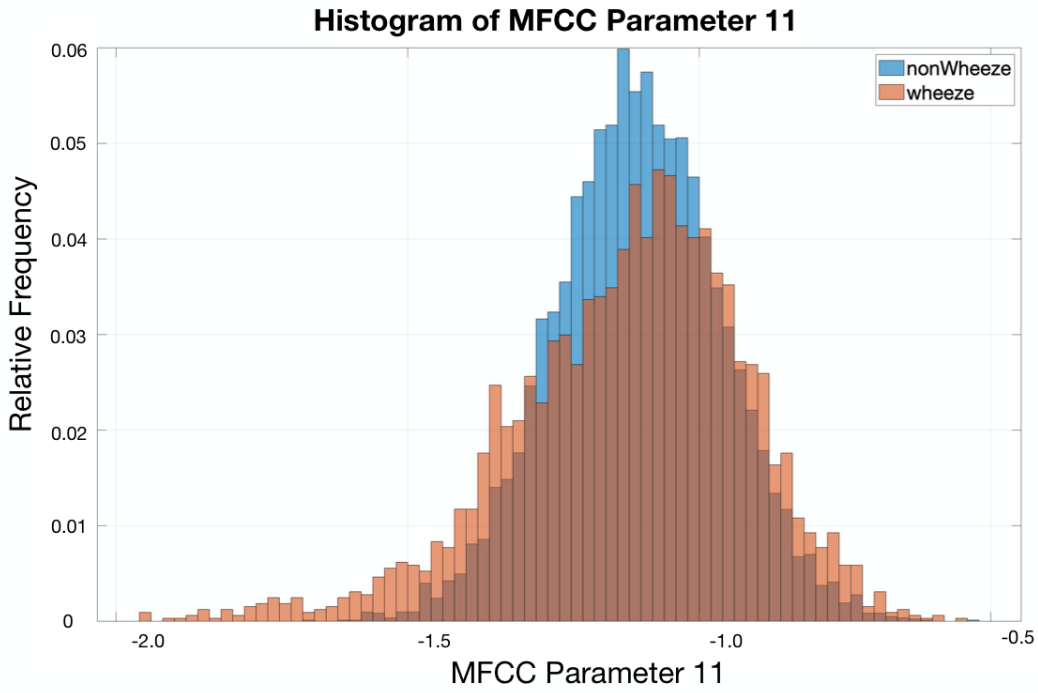

Figure 3.28. Relative frequency histogram of MFCC 11 values regarding both classes

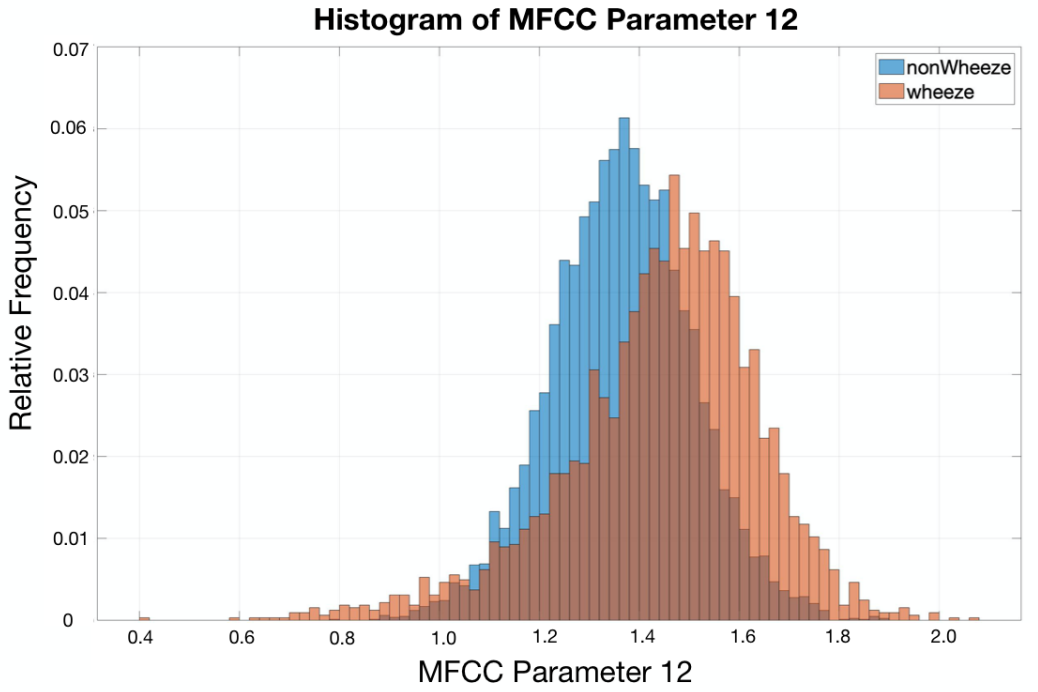

Figure 3.29. Relative frequency histogram of MFCC 12 values regarding both classes

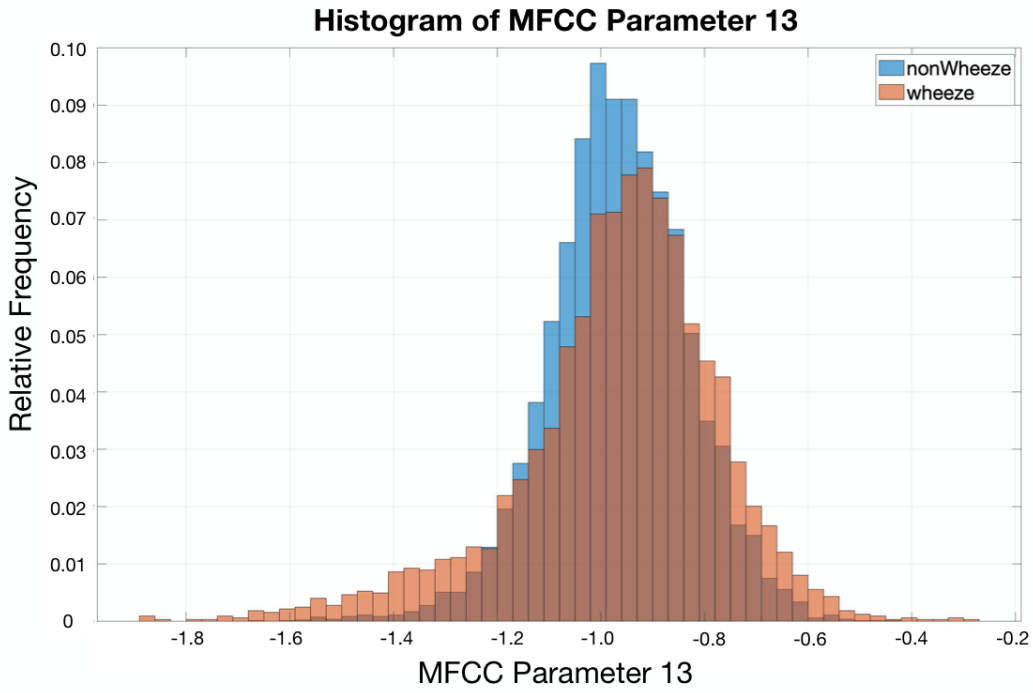

Figure 3.30. Relative frequency histogram of MFCC 13 values regarding both classes

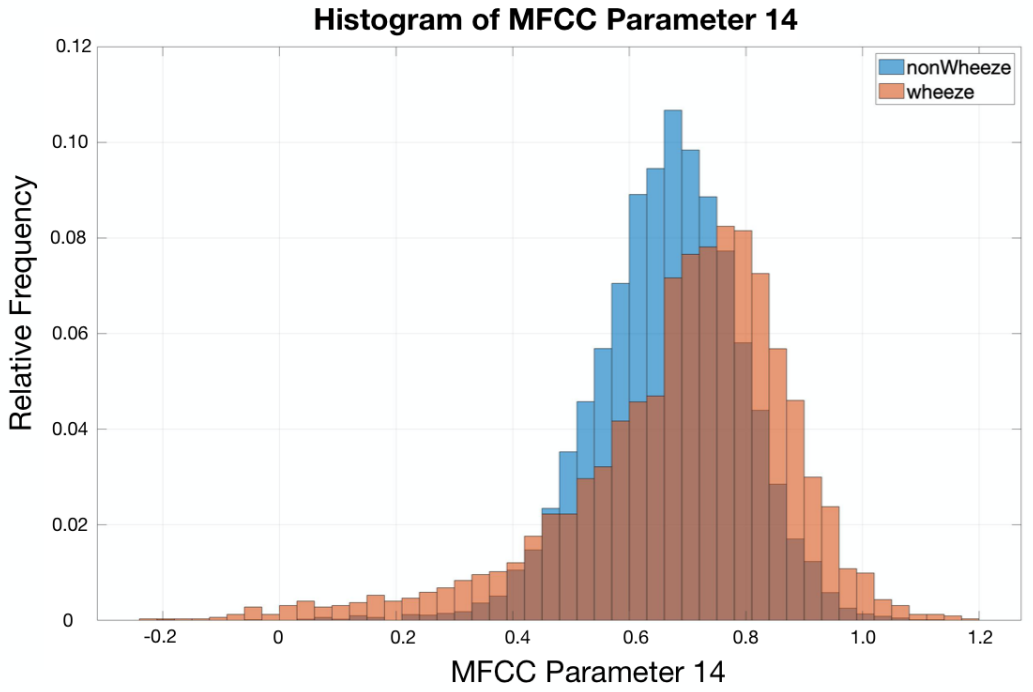

Figure 3.31. Relative frequency histogram of MFCC 14 values regarding both classes

3.7. Audio Spectral Envelope

Spectral envelope of a signal gives valuable information about the spectral content of the signal. It is an envelope curve in the plane of the Fourier magnitude spectrum. In these studies, audio spectral envelope is first introduced to classify wheeze and non-wheeze respiratory signals. The Audio Spectral Envelope (ASE) is a technique which is used widely in ISO/IEC MPEG-7 standard to compress the audio signal, and it is a primary low-level descriptor of the energy of the signal.

To compute of Audio spectral envelope, the following steps are followed.

(i) Discrete Fourier Transform (DFT) is applied every window with 256-point.

$$X_m(k) = \sum_{n=0}^{N} x_m(n) w(n) e^{-j\frac{2\pi}{N}kn} \tag{3.6}$$

where N = 512 (window size) and k is $k_{min} = 4$ and $k_{max} = 38$ for m'th window.

(ii) The sampling frequency of the respiratory signals in the database is 9600 Hz. With 256 points DFT, frequency resolution becomes 9600 Hz /256 = $37,5$ Hz. For wheeze detection, the frequency range of concern is 100 Hz to 1400 Hz so that points between $k_{min} = 4$ and $k_{max} = 38$ in DFT covers the all frequency range.

(iii) Power of m'th window is calculated as $P_m(k)$ for every k bin.

(iv) Averages of spectra fragments of every window are computed in time. In this study, L_{ASE} is chosen as 13. It means that spectra fragments averaged for last 13 signal frames.

(v) Every spectra windows are normalized into the rang [0,1].

(vi) The spectrum fluctuation is computed finally.

Because of the different spectral characteristic of a wheeze, audio spectral envelope is expected to be a discriminative feature when classifying wheeze and non-wheeze sounds. Figure 3.33 shows the relative frequency histograms of the ASE according to classes.

repeat $k_{min} = 4$ to $k_{max} = 38$

$P_m(k) =\mid X_m(k) \mid^2$ power calculation

$ASE_m(k) = \frac{1}{L_{ASE}} \sum_{l=m-L_{ASE}+1}^{m} P_l(k)$ mean

$ASE_m^{norm}(k) = \frac{ASE_m(k) - min_k(ASE_m(k))}{max_k ASE_m(k) - min_k(ASE_m(k))}$ norm

end

$fluctASE_m = \sum_{k=k_{min}+1}^{k_{max}} \mid ASE_m^{norm}(k) - ASE_m^{norm}(k-1) \mid$ **final**

Figure 3.32. Audio Spectral Envelope Algorithm

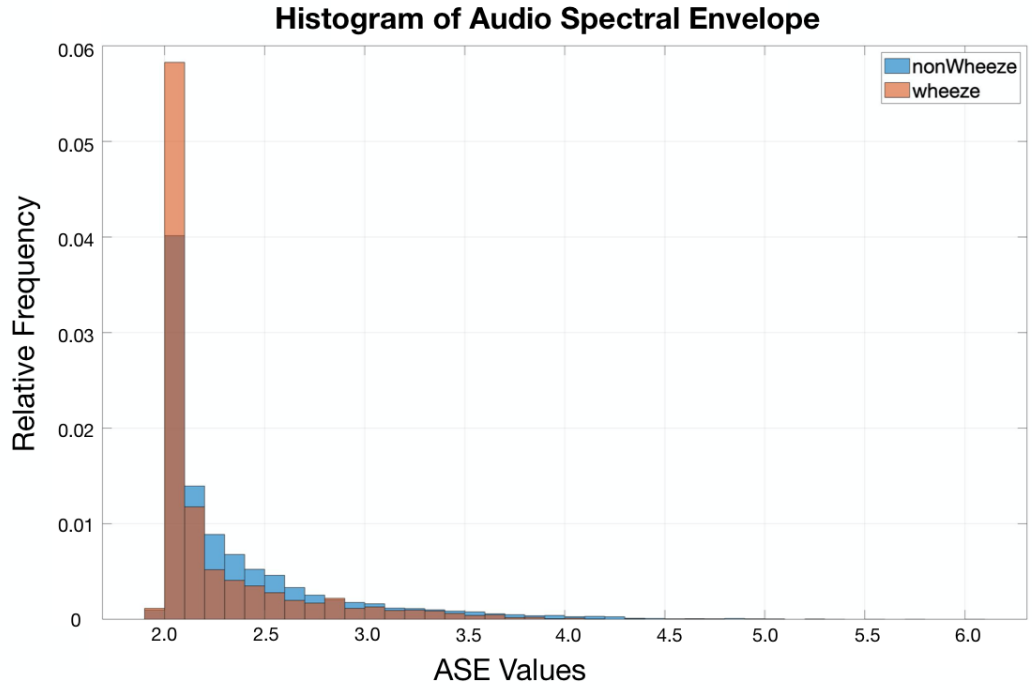

Figure 3.33. Relative frequency histograms of the Audio Spectral Envelope values regarding both classes

3.8. Tonality Index

The tonality index (TI) is widely used audio feature to compress audio files in MPEG-audio standard. In a MP3 encoder, it switches the working decision of second filter bank in the psychoacoustic model no 2. To compress sounds in MP3 encoder, the acoustic signal first divided into 32 frequency subbands. Then, according to whether a subband is tonal or not, signals additionally are divided into 18 sub-channels for better resolution. [37]

Tonality index is a likelihood measure which decides if a sound signal segment is more tone-like or noise-like. This index estimates the current spectral components by creating a function of two previous spectral components.To calculate tonality index, following steps are applied. Also, The tonality index algorithm is depicted in Table 3.8.

(i) Discrete Fourier Transform (DFT) is applied every window with 256-point.

$$X_m(k) = \sum_{n=0}^{N} x_m(n) w(n) e^{-j\frac{2\pi}{N}kn} \tag{3.7}$$

where N = 512 (window size) and k is $k_{min} = 4$ and $k_{max} = 38$ for m'th window.

(ii) The sampling frequency of the respiratory signals in the database is 9600 Hz. With 256 points DFT, frequency resolution becomes 9600 Hz /256 = $37,5$ Hz. For wheeze detection, the frequency range of concern is 100 Hz to 1400 Hz so that points between $k_{min} = 4$ and $k_{max} = 38$ in DFT covers the all frequency range.

(iii) Magnitude and phase of k'th bin are calculated as $r_m(k)$ and $\phi_m(k)$ for m'th window.

(iv) Prediction of magnitude and phase by using 2 previous windows are calculated as $\overline{r_m}(k)$ and $\overline{\phi_m}(k)$ for m'th window.

(v) Spectral unpredictability is calculated for every bin k'th as $c_m(k)$.

(vi) Total energy and weighted spectral unpredictability are computed for m'th window.

(vii) Averages of energy and unpredictability of every window are computed by using seven previous windows. In this study, L_{TI} is chosen as 7.

(viii) The Tonality Index is computed finally.

repeat for $k_{min} = 4$ to $k_{max} = 38$

$r_m(k) =\mid X_m(k) \mid$

$\phi_m(k) = angle(X_m(k))$

$\overline{r_m}(k) = 2r_{m-1}(k) - r_{m-2}(k)$ predict magnitude

$\overline{\phi_m}(k) = 2\phi_{m-1}(k) - \phi_{m-2}(k)$ predict phase

$A_m(k) = r_m(k)cos(\phi_m(k)) - \overline{r_m}(k)cos(\overline{\phi_m}(k))$

$B_m(k) = r_m(k)sin(\phi_m(k)) - \overline{r_m}(k)sin(\overline{\phi_m}(k))$

$c_m(k) = \frac{\sqrt{A_m(k)^2+B_m(k)^2}}{r_m(k)+\mid\overline{r_m}(k)\mid}$ spectral unpredictability

end

$e_m = \sum_{k=k_{min}}^{k_{max}} r_m^2(k)$ energy

$c_m^{(w)} = \sum_{k=k_{min}}^{k_{max}} c_m(k)r_m^2(k)$ weighted spectral unpredictability

$\overline{e_m} = \frac{1}{L_{TI}} \sum_{l=m-L_{TI}+1}^{m} e_l$ energy mean

$\overline{c_m^{(w)}} = \frac{1}{L_{TI}} \sum_{l=m-L_{TI}+1}^{m} c_l^{(w)}$ unpredictability mean

$TI_m = log_{10}(\frac{\overline{c_m^{(w)}}}{\overline{e_m}})$ final

Figure 3.34. Tonality Index Algorithm

The sound of wheeze is a musical, which is described in Chapter 1. So, wheezes are more likely to have tonal components rather than normal respiratory sounds. Figure 3.35 shows the relative frequency histograms of the TI according to classes.

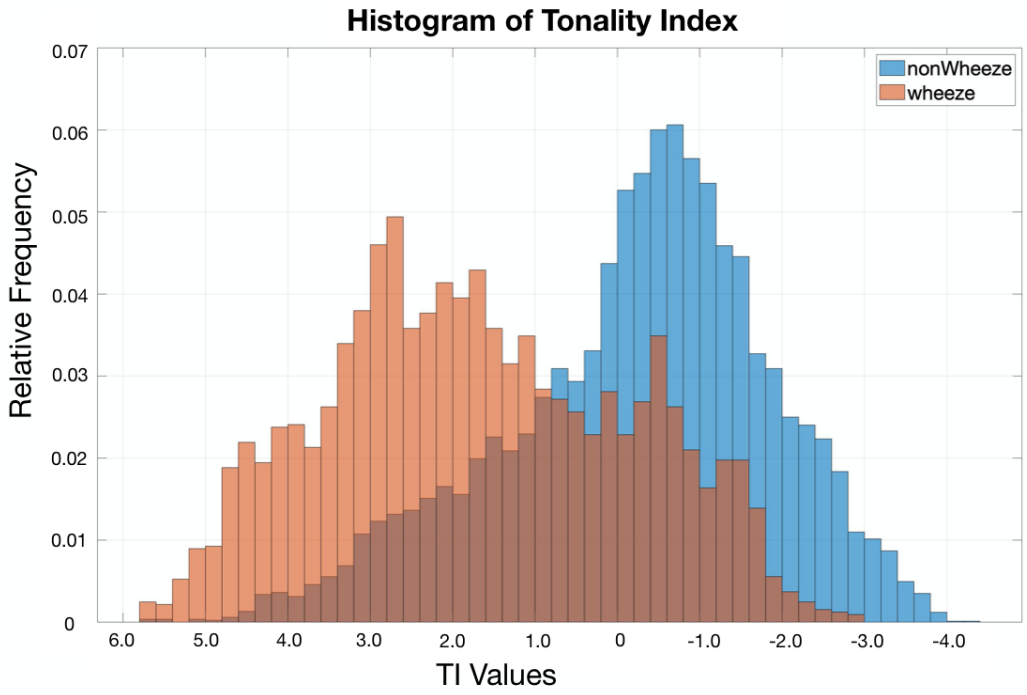

Figure 3.35. Relative frequency histograms of the Tonality Index values regarding both classes

3.9. FFT Peak-Baseline Difference in dB

The feature, FFT peak-baseline difference in dB (FPBD), was first introduced by Ipek Sen in 2005 [7]. Wheezes have dominant frequencies and make noticable peak around its dominant frequency on Fourier magnitude spectrum plane. FPBD proposes a method that uses the discrete Fourier transform of sound window and its polynomial fit to detect an unusual peak around the dominant frequency. The order of polynomial of fit is chosen as 10 in this thesis.

The next steps are followed to calculate FPBD.

(i) Discrete Fourier Transform (FPBD) is applied every window with 512-point.
(ii) Tenth order polynomial fit is calculated.
(iii) Local maxima of polynomial fit is found.
(iv) FPBD is the maximum difference around ($\pm$100 points) local maxima between DFT and its polynomial fit.

The FPBD values are expected to separate the wheeze and non-wheeze sounds because of the spectral characteristic of wheeze. Figure 3.36 shows the relative frequency histograms of the FPBD according to classes.

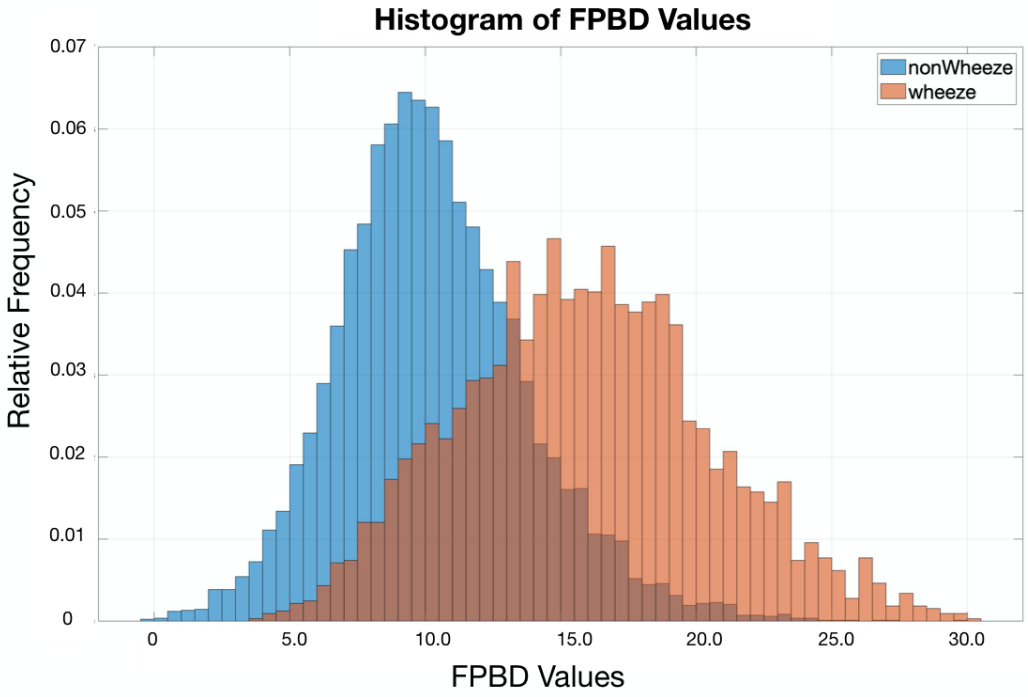

Figure 3.36. Relative frequency histograms of FFT peak baseline difference in dB regarding both classes

4. CLASSIFIERS

In chapter 3, nine features are calculated to compare their discriminatory abilities. Because some features contain more than one parameter, the feature vector consists of 33 columns. There are three classifiers used in this thesis. The comparison of these classifiers is made with the following metrics.

4.1. Result Metrics

There are four metrics in this thesis to evaluate the discriminatory ability of the classifiers. These metrics are accuracy, sensitivity, specificity, and F1 score. They are described as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FN+TN+FP}} \tag{4.1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{4.2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{4.3}$$

$$\text{F1} = \frac{2\,\text{TP}}{2\,\text{TP+FP+FN}} \tag{4.4}$$

In the above equations, true Positive (TP) means the number of correctly labeled wheezes. True negative (TN) describes the number of correctly marked normal respi-

ratory sounds. On the other hand, false positive (FP) means the number of normal sounds labeled as wheezes incorrectly. Contrary to FP, FN is the number of wheezes mislabeled as normal respiratory sounds.

Accuracy is the ratio of how much the system labels samples correctly. Sensitivity is the capability to label wheezes correctly among all wheezes; besides that, specificity is the ratio of how well classifier labels normal respiratory sounds correctly among all normal sounds.

F1 score represents the harmonic mean of the sensitivity and the positive predicted value (PPV). It is generally used to define the trade-off between sensitivity and PPV. Its value changes between zero to one. F1 score is usually used to compare two machine learning algorithm. Following equation shows the positive predicted value equation.

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{4.5}$$

The training data set is used with the 20-fold cross-validation to determine the accurate performance of the classifiers. Because of the limited number of training data, cross-validation increases the performance of the classifier by training more samples.

4.2. Support Vector Machines (SVM)

4.2.1. Description

Support vector machines (SVM) is supervised machine learning method widely used in classification and regression. It tries to find the best hyperplane, which separates the data points belongs to different classes. SVM is also named as a maxiumum margin classifier because it minimizes the classification error iteratively and maximizes the geometric margin at the end.

A hyperplane is a subspace of dimension n-1 with n-dimensional ambient space and the generalized equation of hyperplane is given in Equation 4.6:

$$w \cdot x + b = 0 \tag{4.6}$$

where w is p-dimensional hyperplane vector and b is bias term.

Given a data set $D = \{(x_i, y_i) \mid x_i \in R^n, y_i \in \{-1, 1\}\}_{i=1}^m$ for any $i = 1, ..., m$ where m is number of samples, two parallel hyperplanes are needed and created by adding a offset to b in Equation 4.6 to increase margin. Hyperplanes are defined as:

$$\begin{aligned} w \cdot x + b &= 1 \\ w \cdot x + b &= -1 \end{aligned} \tag{4.7}$$

Two hyperplanes must be located according to two basic rules to classify linearly separable training data: There must be no data points inside these hyperplanes, and the distance between hyperplanes is maximized. The distance between hyperplanes is $\frac{2}{|w|}$ so that $\mid w \mid$must be minimized.

Data points on hyperplanes are the only points that affect the distance between hyperplanes so that they are called support vectors in SVMs. These support vectors satisfy the following equation.

$$y_j[w^T \cdot x_j + b] = 1 \tag{4.8}$$

The optimization problem in soft margin SVM eventually becomes:

$$\underset{w,b,\zeta}{\text{minimize}} \quad \frac{1}{2}w^t w + \mathrm{C}\sum_{i=1}^{m}\zeta_i \tag{4.9a}$$

subject to

$$y_i(w\phi(x_i + b)) + \zeta_i - 1 \geq 0, \tag{4.9b}$$

$$\zeta \geq 0 \tag{4.9c}$$

for any $i = 1, ..., m$ where m is number of samples, ζ is regularization parameter which makes SVM do error. C is called box constraint. The optimization problem can be solved by using Wolfe's dual Lagrangian function. The Wolfe's dual problem of this optimization problem is:

$$\underset{\alpha}{\text{maximize}} \quad \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \tag{4.10a}$$

subject to

$$C \geq \alpha_i \geq 0, \tag{4.10b}$$

$$\sum_{j=1}^{m}\alpha_i y_i \tag{4.10c}$$

for any $i = 1, ..., m$ where m is number of samples. K is the kernel function which calculates the dot products of vectors in other space. The final hypothesis function is defined in following equation:

$$h(x_i) = sign(\sum_{j_i=1}^{S}\alpha_j y_j K(x_j, x_i) + b) \tag{4.11}$$

Kernels are special dot product functions used in SVM to map data in higher dimensions.There are many different kernel functions used in SVM. The most widely used ones are the linear kernel, polynomial kernel, and radial basis kernel (RBF). In this thesis, SVM with these three types of kernels are trained to classify wheeze and

normal sounds.

The linear kernel in $K(x, \acute{x}) = x^T \acute{x} + \gamma$ is the easiest kernel given by inner dot product of samples with parameter γ.

The polynomial kernel in $K(x, \acute{x}) = (\gamma x^T \acute{x} + 1)^d$ maps samples to higher space to separate them with degree parameter d and scale parameter γ.

The Gaussian kernel or Radial Basis Function kernel in $K(x, \acute{x}) = exp(-\gamma \parallel x - \acute{x} \parallel^2)$ also maps samples to higher space. It uses fewer hyperparameters, unlike polynomial kernel.γ is the only paramater in RBF kernel.

The other crucial parameter which controls the SVM algorithm is box parameter C. The parameter is used to control SVM on how to respond to errors. Higher C values increase the sensitivity of SVM to errors, which means that the classifier chooses a smaller margin hyperplane to improve accuracy in the train set. On the other hand, lower C parameters make classifier numb to misclassification.

4.2.2. Result

Three SVM classifiers with linear, polynomial and RBF kernels are trained with 20-fold cross validation with all features.

In linear SVM, C parameter is choosen as 1 after parameter tuning. The final accuracy is 88.6%. Table 4.1 gives the details.

Table 4.1. Performance of SVM with Linear Kernel

Metrics	**Results**
Accuracy	88.62%
Sensitivity	82.49%
Specificity	90.64%
F1 Ratio	0.78
Training time	12.11 seconds with 20-fold cross validation

In polynomial SVM, polynomial kernel is used with order 3. C parameter is choosen as 1 and γ parameter is decided as 4.05 after parameter tuning. The final accuracy is 91.7%.

Table 4.2. Performance of SVM with Polynomial Kernel

Metrics	**Results**
Accuracy	89.51%
Sensitivity	84.45%
Specificity	91.16%
F1 Ratio	0.82
Training time	480.33 seconds with 20-fold cross validation

Lastly, C parameter is choosen as 1 and γ parameter is decided as 4.12 for SVM with Radial Basis Function (RBF) kernel. The final accuracy is 92.66%.

As a result, SVM with RBF kernel separates normal sounds and wheezes best with 92.6 percent accuracy. Also, polynomial SVM performs significant accuracy, with 91.7%. But, polynomial SVM's training time is 430.6 seconds and much higher than SVM with RBF kernel, so that the SVM with RBF kernel is chosen for feature selection. The details of the performance of the chosen SVM classifier is shown in Table 4.3 and Figure 4.1

Table 4.3. Performance of SVM with RBF Kernel

Metrics	**Results**
Accuracy	92.66%
Sensivity	89.92%
Specificity	93.62%
F1 Ratio	0.86
Training time	72.33 seconds with 20-fold cross validation

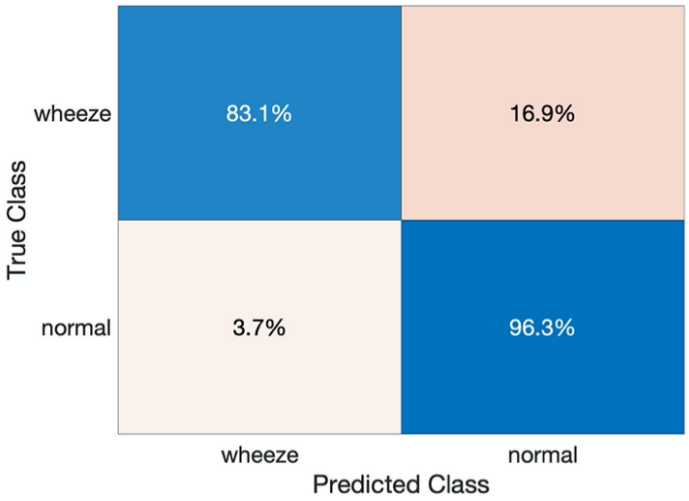

Figure 4.1. Confusion Matrix of SVM with RBF Kernel

4.3. k-Nearest Neighbour (k-NN)

4.3.1. Description

k-nearest neighbor(k-NN) is a well-known supervised nonparametric machine learning algorithm that is widely used in classification and regression. Despite its simplicity, it is still popular and has high accuracy rates.

First, k is defined by user. After k is selected, the algorithm assigns the sample in the test set to class according to the majority of its number of k nearest neighbors in the training set. The distance, Euclidean or Mahalanobis, between samples defines the nearest neighborhood rule. The parameter k should not be multiple of the number of classes because the final class selection can stay in uncertainty.

The Euclidean distance is a popular distance method in k nearest neighbor. Euclidian distance for two samples, x_1 and x_2 with m dimension, is:

$$dist(A, B) = \sqrt{\frac{\sum_{i=1}^{m}(x_{1i} - x_{2i})}{m}} \tag{4.12}$$

A significant disadvantage of k-NN is that distance between all samples in the training set must be calculated when a new sample is added.

4.3.2. Result

Since deciding the right k parameter is the backbone of the k-NN algorithm, k-NN classifiers with k parameters 1 to 100 are trained with 20-fold cross validation and their accuracies are plotted in Figure 4.2. As seen in Figure 4.2, when k is chosen as 9, maximum accuracy is achieved. Then the k-NN classifier is trained with 20-fold cross-validation and gives the best accuracy as 91.43% with 20-fold cross-validation. Also, the distance metric is chosen as Euclidian. The details of the performance of

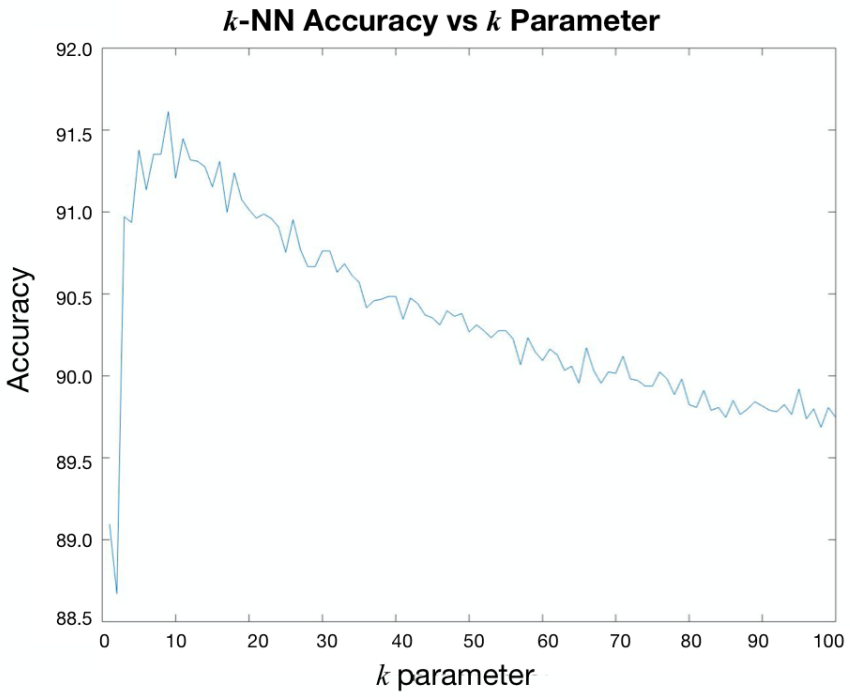

Figure 4.2. k-NN accuracy vs k parameters

k-NN classifier with 9 number of neighbors and Euclidean distance are shown in Table 4.4 and Figure 4.3.

Table 4.4. Performance of Euclidian k-NN with 9 number of neighbors

Metrics	**Results**
Accuracy	91.43%
Sensivity	88.1%
Specificity	92.5%
F1 Ratio	0.84
Training time	4.20 seconds for 20-fold cross validation

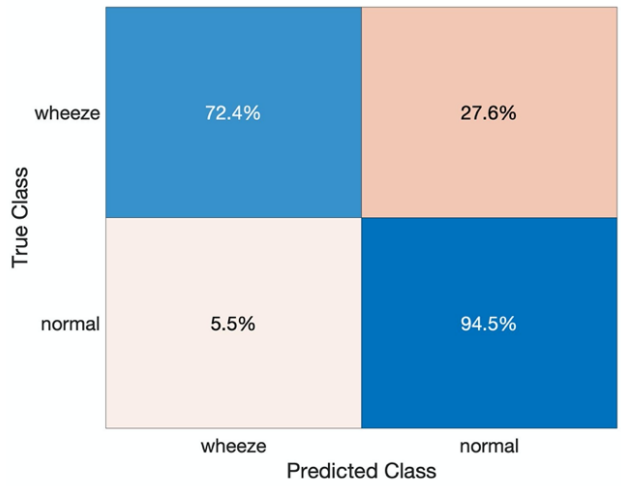

Figure 4.3. Confusion Matrix of K-NN

4.4. Bayesian Classifier with Gaussian Likelihood

4.4.1. Description

Maximum likelihood estimation is a method that estimates the probability distribution parameters by maximizing the likelihood function of the distribution. Because of the high dimensionality of the dataset, multivariate gaussian distribution is used for the modeling feature set. The probability density function of multivariate gaussian distribution with d-dimensional $\boldsymbol{x}$:

$$p(\boldsymbol{x}|\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{4.13}$$

where statistic $\boldsymbol{\Sigma}$ is covariance matrix and statistic $\boldsymbol{\mu}$ is the mean vector. The probability density function (pdf) in Equation 4.13 becomes the class likelihood function as following equation:

$$p(\boldsymbol{x}|C_i) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right) \tag{4.14}$$

Then we need posterior probability $p(C_i|\boldsymbol{x})$ to compare the probability of class for a given new $\boldsymbol{x}$. Bayesian theorem helps at this point.

$$p(C_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|C_i)p(C_i)}{p(\boldsymbol{x})} \tag{4.15}$$

where $p(\boldsymbol{x})$ is the evidence and its value is same for every class. Then, decision rule becomes for the C_i:

$$\begin{aligned} &\text{If } p(C_{wheeze}|\boldsymbol{x}) > p(C_{nonWheeze}|\boldsymbol{x}) \text{ then } \boldsymbol{x} \text{ belongs to wheeze} \\ &\text{If } p(C_{nonWheeze}|\boldsymbol{x}) > p(C_{wheeze}|\boldsymbol{x}) \text{ then } \boldsymbol{x} \text{ belongs to normal sounds} \end{aligned} \tag{4.16}$$

The classification algorithm first finds the parameters of the Gaussian distribution, which are the mean vector and covariance matrix of each classes. Then it calculates the posterior probability of the new sample for every class using Bayesian theorem. Then the algorithm assigns the new sample to class of the highest posterior probability.

4.4.2. Result

The Bayesian classifier has the lowest accuracy rate among three classifier trained in this thesis. The details of the performance of Bayesian classifier are shown in Table 4.5 and Figure 4.4.

Table 4.5. Performance of Bayesian Classifier

Metrics	**Results**
Accuracy	87.31%
Sensitivity	88.32%
Specificity	90.57%
F1 Ratio	0.83
Training time	0.38 seconds 20-fold cross validation

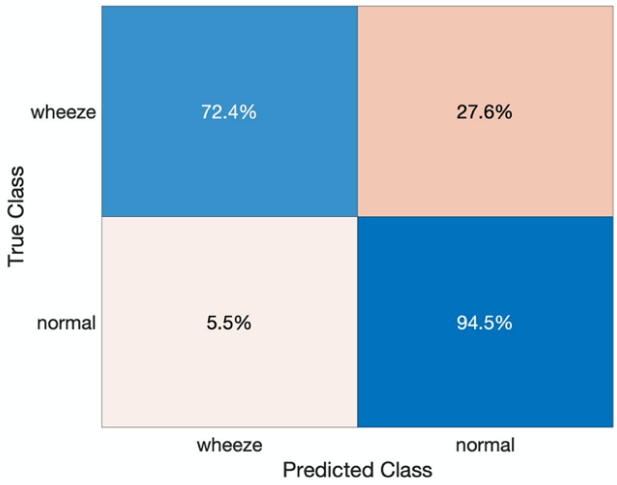

Figure 4.4. Confusion Matrix of Bayesian Classifier

4.5. Result: Best Classifier

When the performance metrics of all classifiers are compared, SVM with RBF kernel has the best performance. Its accuracy is 92.66 and F1 ratio is 0.86. In Figure 4.5, all performance metrics are given for all classifiers.

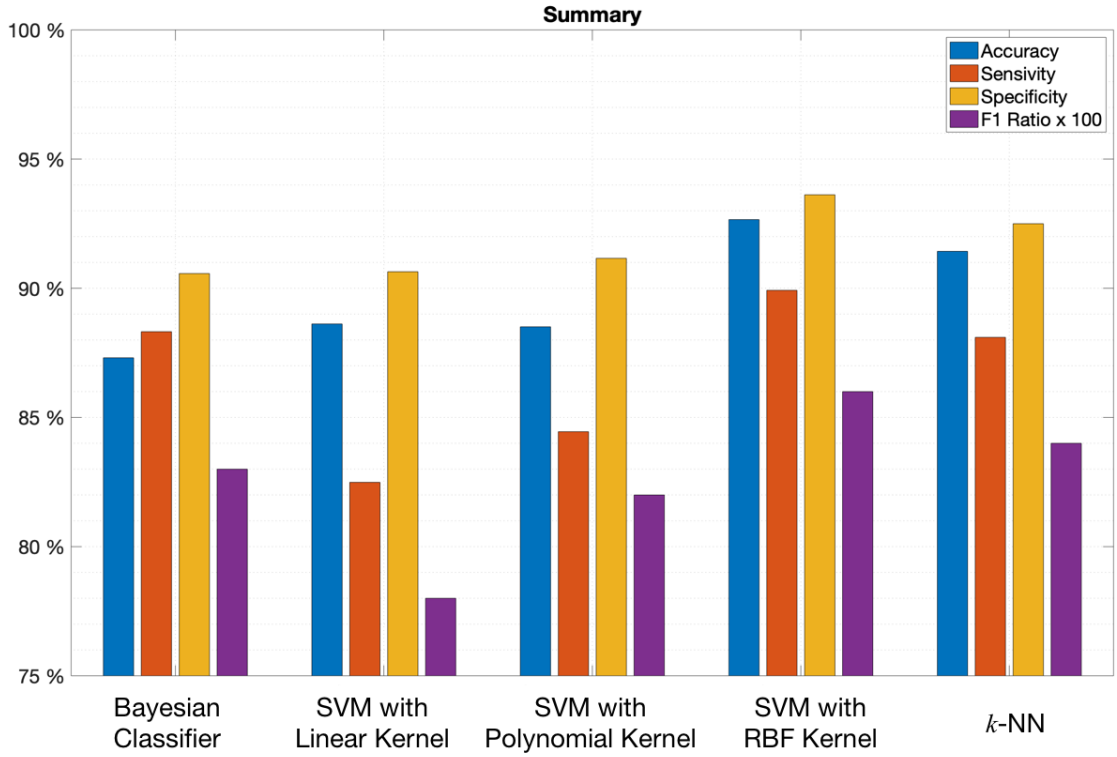

Figure 4.5. Performance metrics of all classifiers

5. FEATURE SELECTION

The training set consists of 33 columns with nine features. In this chapter, the main aim is to reduce the number of features to find features with the best discriminatory ability. To do this, SVM with RBF Kernel is chosen because it gives the best performance among the three classifiers with the 92.66 percent accuracy. Also, k-NN needs adaptive k parameter optimization when the feature set is changed iteratively. In the first section, the P values of the two-sample t-test are presented. In the second and third sections, two popular feature selection algorithm, forward feature selection, and backward feature selection are employed with the SVM with RBF Kernel.

5.1. P Values of Each Feature

Two-sample t-test is a test for determining whether the two samples belong to the same population or not by testing equality of their sample means. The test statistic of two-sample t-test with unequal variance is:

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} \tag{5.1}$$

where x and y are two sample set of wheezes and normal respiratory sounds. Also, n and m are the sample sizes.s_x and s_y are the sample standard deviations which are calculated with following equations:

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i - \overline{x}}{n-1}} \tag{5.2}$$

The null hypothesis for this test is that means of the wheeze samples and normal samples are equal for every feature. P value is the probability value under the null hypothesis. Small P values indicate that the rejection of the null hypothesis is the

right choice. In this test, the features with lower P values are more discriminative than the features with higher P values. Table 5.1 shows the P values for each features.

Table 5.1. Two-sample t-test P values of each feature in ascending order

Features	**P Values**	**Features**	**P Values**	**Features**	**P Values**
f_{25}/f_{75} Ratio	0	Renyi alpha 3	0	AR 4	5.7308e-31
f_{25}/f_{90} Ratio	0	MFCC 5	8.7416e-312	AR 5	7.0106e-31
f_{50}/f_{75} Ratio	0	MFCC 7	1.9013e-247	MFCC 6	2.6988e-22
f_{50}/f_{90} Ratio	0	Kurtosis	6.1195e-208	MFCC 9	2.8669e-20
Mean Cross. Irr.	0	MFCC 3	7.9878e-208	MFCC 4	8.4801e-20
FPBD	0	AR 6	3.0064e-120	MFCC 14	1.01e-07
Tonality Index	0	AR Error	9.7557e-91	MFCC 11	4.839e-05
MFCC 1	0	MFCC 12	2.8626e-89	AR 2	0.0017216
MFCC 2	0	MFCC 8	1.5978e-49	AR 5	0.032718
Renyi alpha 1	0	ASE	1.2393e-45	MFCC 13	0.69849
Renyi alpha 2	0	MFCC 10	4.5168e-44	AR 1	0.73861

5.2. Forward Sequential Feature Selection Algorithm

Forward sequential feature selection is a well-known method to reduce data dimensionality. Reducing the number of features, decreases the computational time and increases simplicity. Forward feature selection starts with having no feature in the data set, and iteratively adds new features with the best accuracy to set. Figure 5.1 shows the accuracy graph versus the newly added features to the training set. Also, Table 5.3 shows the change in F1 ratios of the model when the most discriminative feature added.

As seen in Table 5.2 and Figure 5.1, Mel frequency cepstral coefficients have the most discriminatory ability to distinguish wheezes among normal respiratory sounds with the accuracy of 92.04%. When FPDB feature is included in the feature set, accu-

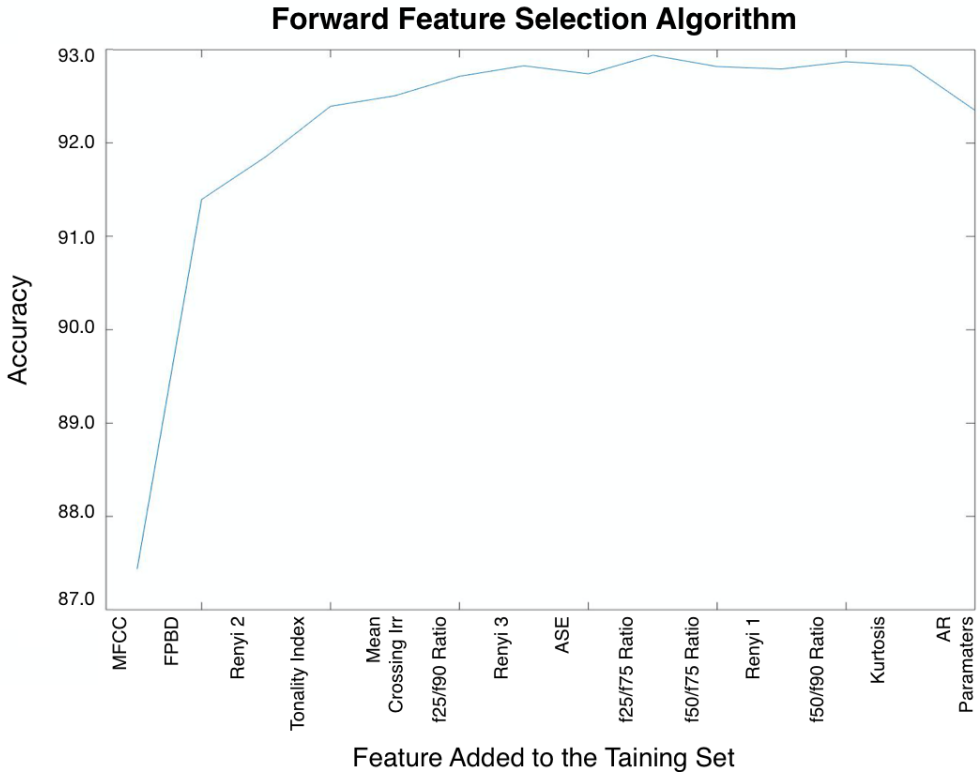

Figure 5.1. Accuracy graph of the model when the most discriminative feature added

Table 5.2. A change in accuracy of the model when the most discriminative feature added

Features	**Accuracy**	**Features**	**Accuracy**
MFCC	87.44%	Audio Spectral Envelope	92.74%
FPBD	91.40%	f_{25}/f_{75} Ratio	92.94%
Renyi alpha 2	91.86%	f_{50}/f_{75} Ratio	92.82%
Tonality Index	92.39%	Renyi alpha 1	92.79%
Mean Cross. Irr.	92.51%	f_{50}/f_{90} Ratio	92.87%
f_{25}/f_{90} Ratio	92.72%	Kurtosis	92.83%
Renyi alpha 3	92.83%	AR Parameters	92.35%

Table 5.3. A change in F1 ratios of the model when the most discriminative feature added to feature set

Features	**F1 Ratio**	**Features**	**F1 Ratio**
MFCC	0.787	f_{50}/f_{75} Ratio	0.859
FPBD	0.828	Renyi alpha 2	0.859
Renyi alpha 3	0.846	f_{25}/f_{75} Ratio	0.859
Tonality Index	0.851	f_{25}/f_{90} Ratio	0.858
Mean Cross. Irr.	0.855	Audio Spectral Envelope	0.858
Kurtosis	0.859	13. Renyi alpha 1	0.860
f_{50}/f_{90} Ratio	0.863	14. AR Parameters	0.851

racy ascends to about 91.5%, which is only one point smaller than optimum accuracy of 92.6%. After the Renyi Entropy with order 2 and Tonality Index are added, accuracy rises to about 92.4%.

5.3. Backward Sequential Feature Selection Algorithm

Backward feature selection is another method to reduce data dimensionality. It starts with all features in the set. Then it eliminates the least significant feature one by one. Each elimination must improve the accuracy of the classifier with a remaining feature set. Figure 5.2 shows the accuracy graph versus eliminated features from the training set. Also, Table 5.5 shows the change in F1 ratios of the model when the most discriminative feature added.

As seen in Table 5.4 and Figure 5.2, Ar parameters, the Renyi Entropy with order 3, and the Renyi Entropy with order 1 and Kurtosis are the least significant features when detecting wheezes. AR parameters' frequency histograms give a clue about their inconsistency in Chapter 3. Also, Renyi Entropies with order 1 and 3 are a surprise when their histograms are evaluated first. Nevertheless, when the table is

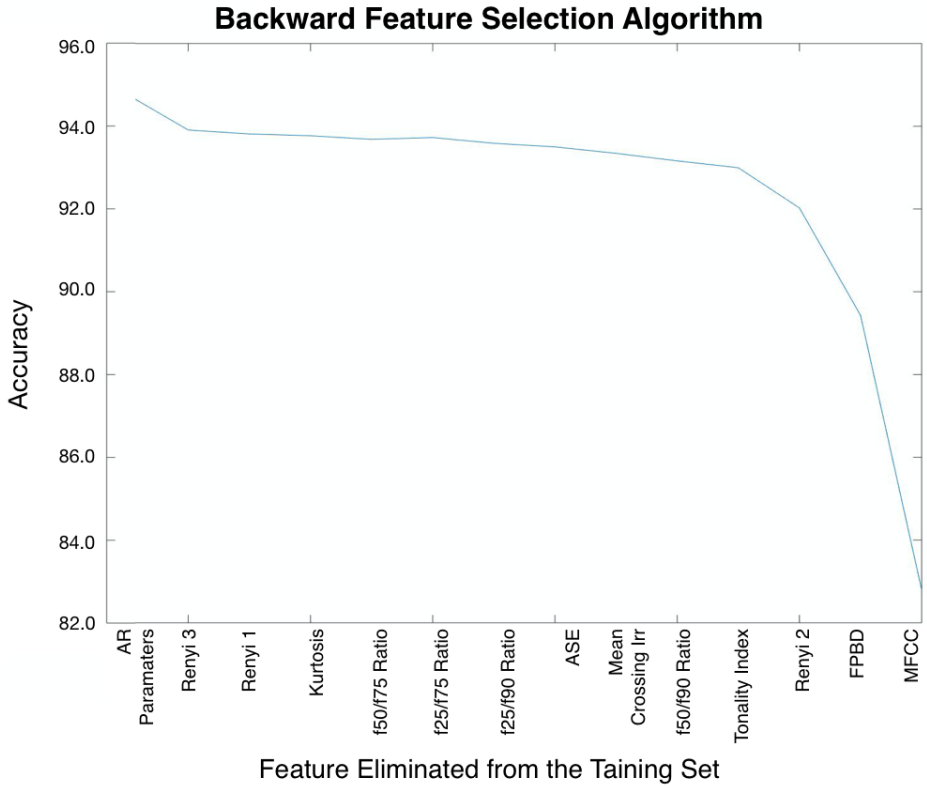

Figure 5.2. Accuracy graph of the model when the least discriminative feature eliminated

Table 5.4. A change in accuracy of the model when the least discriminative feature is eliminated according to accuracy

Features	**Accuracy**	**Features**	**Accuracy**
AR Parameters	92.77%	Audio Spectral Envelope	91.50%
Renyi alpha 3	91.91%	Mean Cross. Irr.	91.34%
Renyi alpha 1	91.81%	f_{50}/f_{90} Ratio	91.16%
Kurtosis	91.77%	Tonality Index	91.00%
f_{50}/f_{75} Ratio	91.68%	Renyi alpha 2	90.02%
f_{25}/f_{75} Ratio	91.73%	FPBD	87.43%
f_{25}/f_{90} Ratio	91.59%	MFCC	86.12%

Table 5.5. A change in F1 ratios of the model when the least discriminative feature eliminated from feature set

Features	**F1 Ratio**	**Features**	**F1 Ratio**
AR Parameters	0.869	f_{50}/f_{90} Ratio	0.842
Renyi alpha 3	0.851	Mean Crossing Irregularity	0.838
f_{25}/f_{90} Ratio	0.848	Tonality Index	0.835
f_{50}/f_{75} Ratio	0.848	Kurtosis	0.834
f_{25}/f_{75} Ratio	0.847	Renyi alpha 2	0.811
Renyi alpha 1	0.844	FPBD	0.791
Audio Spectral Envelope	0.844	MFCC	0.786

investigated carefully, Renyi Entropy with order 2 is at the end of the list, and it is the third discriminative feature. This situation is understandable because the correlation among Renyi entropies with different alpha is high.

5.4. Brute-Force Feature Selection

In this thesis, we also decided to train the SVM classifier with all combination of feature set (1 to 5 sized) to detect the best features. This method took too much time and computational power. However, it is needed to understand the nature of the features used for classification. Table 5.6 shows the best combinations of features per size and their accuracies.

As seen in table 5.6, Mel frequency cepstral coefficient is the most selective feature followed by f_{25}/f_{90} Ratio and AR parameters. AR parameters are the third discriminative feature, among others; nevertheless, it is not seen in the best feature sets of more than one number. Although confusing at first glance, since MFCC and AR parameters are highly correlated; this result is to be expected.

Table 5.6. Best sets of features with different sizes and their accuracies with descending order

Features	**Ranking**	**Best set of features**	**Accuracy**	**F1**
1	Best	MFCC	88.71%	0.78
1	Second	f25/f90 Ratio	82.35%	0.59
1	Third	AR Params.	82.13%	0.60
2	Best	MFCC, FPBD	91.47%	0.83
2	Second	MFCC, Mean C. Irr.	91.01%	0.82
2	Third	MFCC, Renyi 2	90.31%	0.82
3	Best	MFCC, FPBD, Renyi 1	91.80%	0.84
3	Second	MFCC, FPBD, Renyi 2	91.75%	0.85
3	Third	MFCC, FPBD, Mean C. Irr.	91.62%	0.84
4	Best	MFCC, FPBD, Renyi 3, TI	92.18%	0.85
4	Second	MFCC, FPBD, Renyi 3, Mean C. Irr	92.14%	0.85
4	Third	MFCC, FPBD, Renyi 2, f25/f90	92.10%	0.85
5	Best	MFCC, FPBD, Renyi 2, Mean C. Irr, TI	92.61%	0.86
5	Second	MFCC, FPBD, TI, Mean C. Irr, f25/f90	92.41%	0.86
5	Third	MFCC, FPBD, Renyi 1, Mean C. Irr, f25/f90	92.38%	0.86

The feature set with MFCC and FPBD increases the accuracy from 88.71 % to 91.47% percent. FFT peak baseline difference in dB feature is the best fellow of MFCC to detect wheezes. Then Renyi entropy with order 3 and Tonality Index must be added to the feature set to increase the accuracy.

5.5. Correlations Plot

Correlation measures how much two variables depend on each other linearly. It gives both direction and intensity of the linear relationship. In this thesis, the correlation plot of the features is useful to understand which features are dependent on each other. For example, Figure 5.3 shows that Renyi entropies have high correlation values when they are compared with each other. The meaning of these high correlation values and scatter plots is that one of the Renyi entropy values is sufficient to classify wheezes superimposed on normal respiratory sounds.

To decide the optimum feature set, the correlation values and scatter plots of nine features are shown in Figure 5.3. Because MFCC and AR features have more than one variable, Principal Component Analysis (PCA) is applied to those features, and the first column of score matrix of each one are used for calculation of correlation and presentation of plots.

As we observe the correlation plot in Figure 5.3, we can conclude that f_{25}/f_{90} is highly correlated with f_{25}/f_{75} Ratio, mean crossing irregularity is highly correlated with f_{50}/f_{90} Ratio, tonality index is highly correlated with the MFCC values with PCA applied and Renyi values are highly correlated with each other.

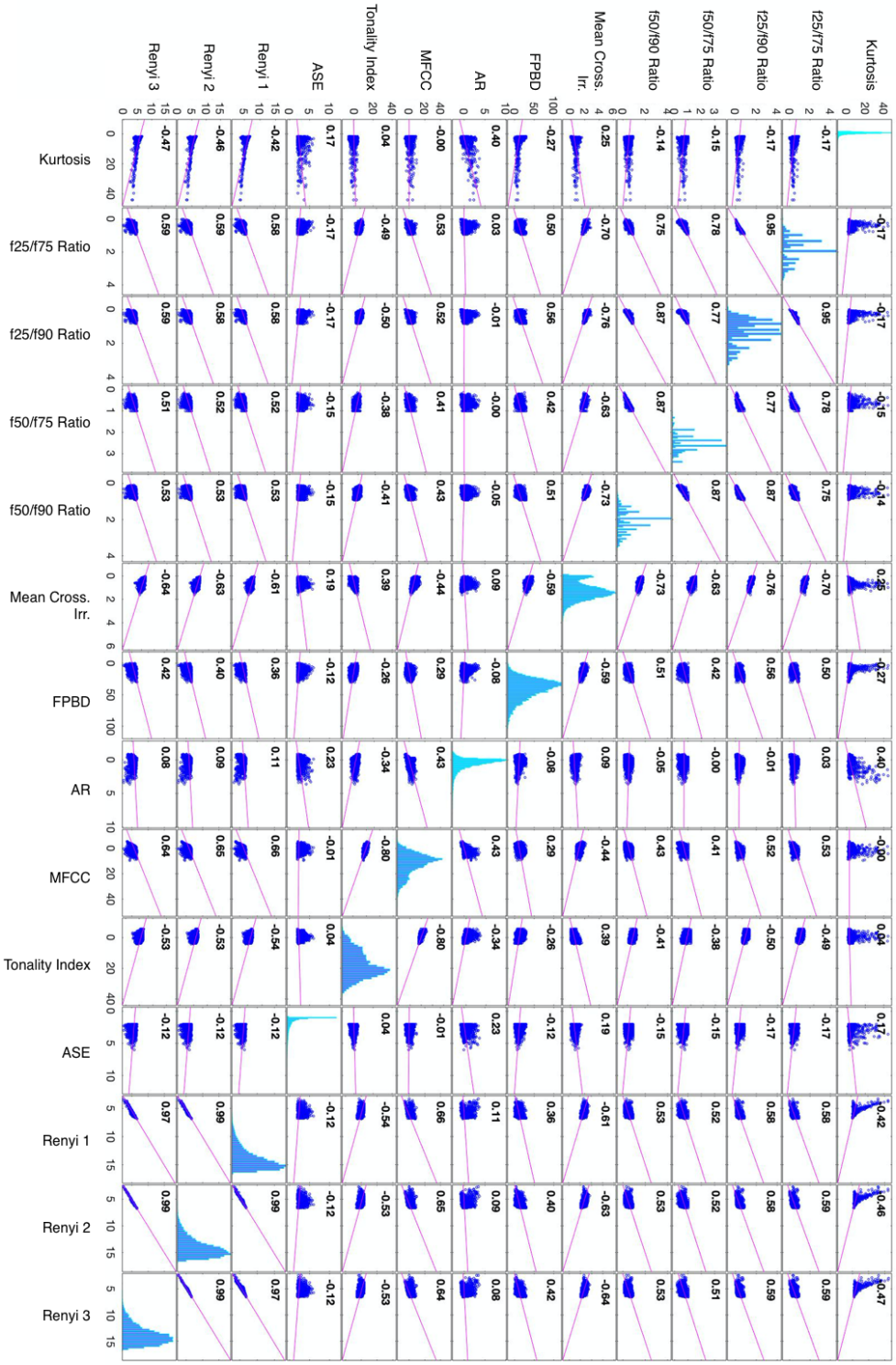

Figure 5.3. Correlation Matrix of Features

5.6. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an unsupervised dimensionality reduction method that is based on the variance of the dataset. This method is used for feature extraction by maximizing the variance with minimum loss of information. The projection of sample $\boldsymbol{x}$ on the direction of w is

$$z = w^T \boldsymbol{x} \tag{5.3}$$

In PCA, principal component w must be orthogonal to each other and variance of score matrix z must be maximized. Given that

$$\begin{aligned} E[w^T \boldsymbol{x}] &= w^T E[x] = w^T \mu & (5.4)\\ Var(w^T \boldsymbol{x}) &= E[(w^T \boldsymbol{x} - w^T \mu)^2] = E[(w^T \boldsymbol{x} - w^T \mu)(w^T \boldsymbol{x} - w^T \mu)] & (5.5)\\ &= E[(w^T (\boldsymbol{x} - \mu)(\boldsymbol{x} - \mu)^T w] = w^T E[(\boldsymbol{x} - \mu)(\boldsymbol{x} - \mu)^T] w & (5.6)\\ &= w^T \Sigma w & (5.7) \end{aligned}$$

where $Cov(x) = \Sigma$ and Σ is the covariance matrix of sample matrix. The optimization problem becomes:

$$\begin{aligned} \underset{w}{\text{maximize}} \quad & w^T \Sigma w & (5.8a)\\ \text{subject to} \quad & w^T w = I & (5.8b) \end{aligned}$$

We can solve this problem as a Lagrange problem so that:

$$\max_w w^T \Sigma w - \alpha(w^T w - 1) \tag{5.9}$$

After the taking derivative of Equation 5.5 and setting it as zero, we have:

$$2\Sigma w - 2\alpha w = 0 \tag{5.10}$$

$$\Sigma w = \alpha w \tag{5.11}$$

When w is chosen as eigenvector of Σ and corresponding eigenvalue α, (5.11) will be true. Moreover, maximum value of α must be chosen to maximize the variance because:

$$w^T \Sigma w = \alpha w^T w = \alpha \tag{5.12}$$

After the PCA is applied to dataset with nine features, the dimension of feature space is reduced from 33 to 1. The wheeze and non-wheeze samples in a one-dimensional feature set are fitted to a normal distribution, and Figure 5.4 shows the distribution plots of these samples. The corresponding eigenvalue of the first eigenvector is 22.8 and the proportion of variance explained is 62.8%. Then maximum likelihood classifier in Section 4.4 is applied to samples and Table 5.7 shows the results.

Table 5.7. Performance of Bayesian Classifier on one dimensional feature set after PCA

Metrics	**Results**
Accuracy	83.14%
Sensitivity	93.08%
Specificity	57.37%
F1 Ratio	0.89
Training time	0.43 seconds 5-fold cross validation

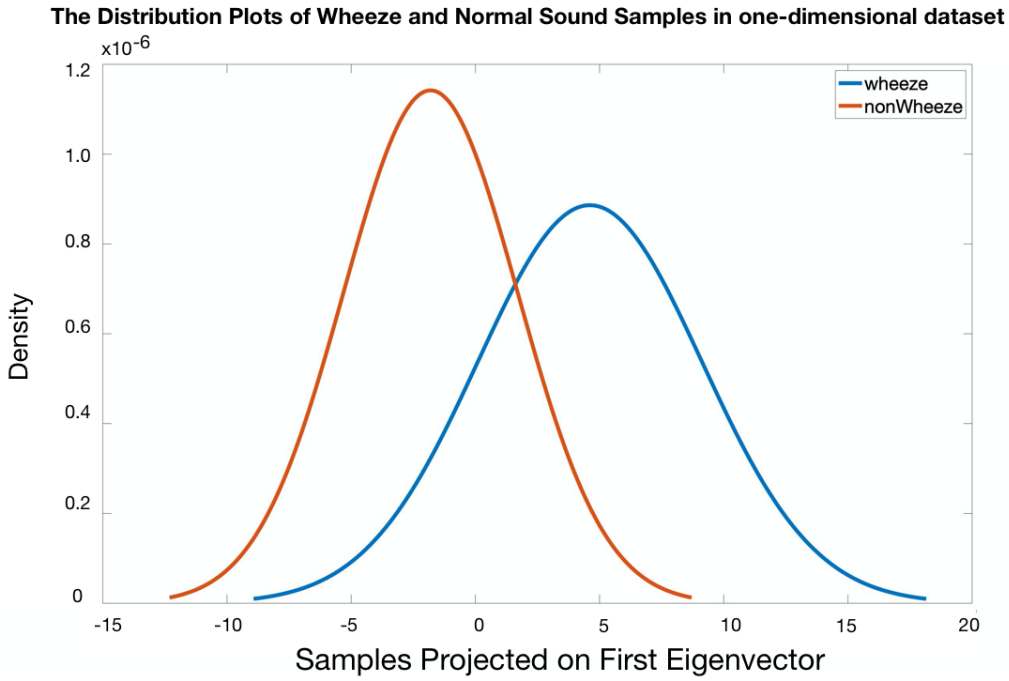

Figure 5.4. The plot of distributions of wheeze and non-wheeze samples in one dimensional feature set after PCA

This time, the PCA is applied again to reduce the dimension of feature space from 33 to 2. The plot of wheeze and non-wheeze samples in a two-dimensional feature set is shown in Figure 5.5. The corresponding eigenvalues of the first and second eigenvectors are 22.8 and 4.2 respectively, and the proportion of variance explained is 81.9%. Then support vector machine with RBF Kernel in Section 4.2 is applied to samples and Table 5.8 shows the results.

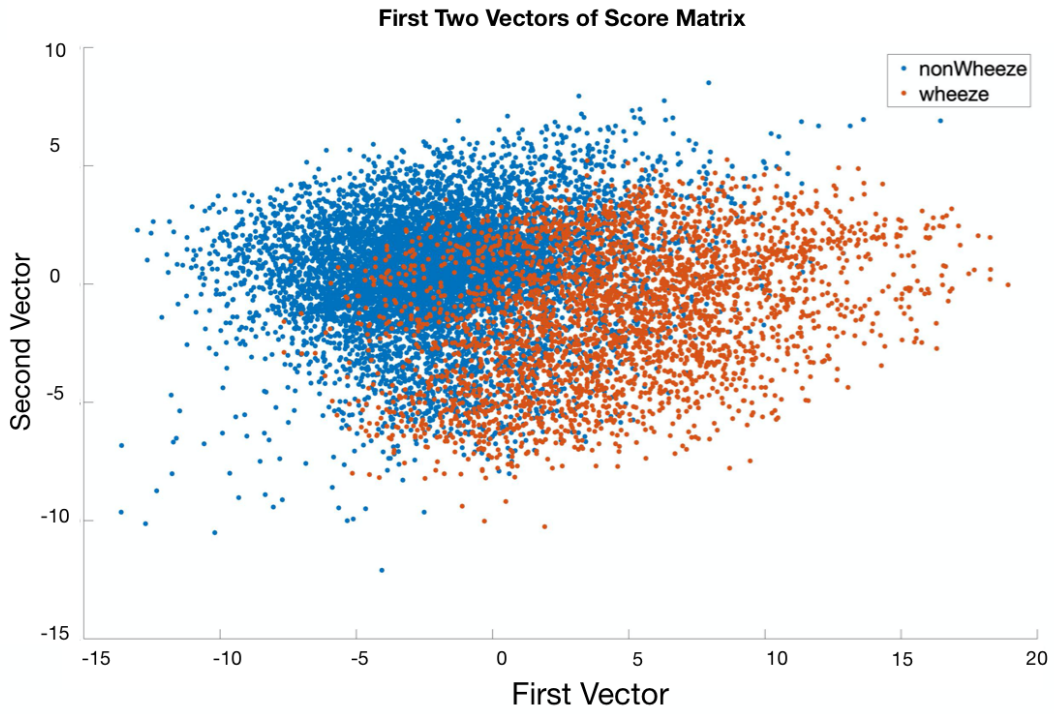

Figure 5.5. The plot of first two-column of the Score Matrix

Table 5.8. Performance of SVM with RBF Classifier on two dimensional feature set after PCA

Metrics	**Results**
Accuracy	82.36%
Sensitivity	72.79%
Specificity	85.22%
F1 Ratio	0.65
Training time	5.4 seconds 5-fold cross validation

Last, the PCA is applied again to reduce the dimension of feature space from 33 to 3. The plot of wheeze and non-wheeze samples in a three-dimensional feature set is shown in Figure 5.6. The corresponding eigenvalues of the first second and third eigenvectors are 22.8, 4.2 and 1.5 respectively, and the proportion of variance explained is 91.83 %. Then support vector machine with RBF Kernel in Section 4.2 is applied to samples and Table 5.9 shows the results.

Table 5.9. Performance of SVM with RBF Classifier on three dimensional feature set after PCA

Metrics	**Results**
Accuracy	85.55%
Sensitivity	79.06%
Specificity	87.54%
F1 Ratio	0.72
Training time	6.7 seconds 5-fold cross validation

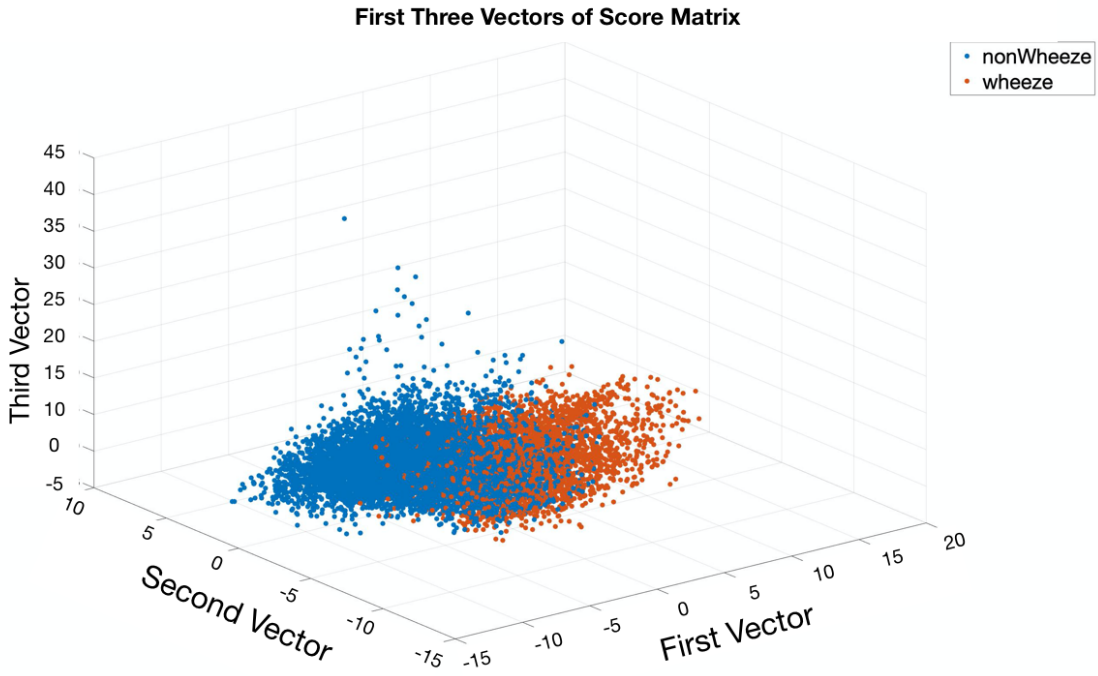

Figure 5.6. The plot of first three-column of the Score Matrix

5.7. Linear Discriminant Analysis (LDA)

In comparison to the PCA, linear discriminant analysis is another method to reduce the dimensionality of classification problems. For the two-class, wheeze, and normal, LDA seeks to find a direction vector. When the feature sets are projected onto w, the samples of two classes are as well-separated as possible.

$$z = w^T \boldsymbol{x} \tag{5.13}$$

where $\boldsymbol{x}$ is the feature set and w is direction vector. Note that, in this case, $\boldsymbol{x} \in R^{33}$ and $z \in R^1$. So, LDA will reduce dimensionality 33 to 1. Note that, μ_n and μ_w are the means of normal and wheeze samples, respectively. Then scatter of normal and wheeze samples are:

$$s_w^2 = (w^T \boldsymbol{x}_w - w^T \mu_w)^T (w^T \boldsymbol{x}_w - w^T \mu_w) \tag{5.14}$$

$$s_n^2 = (w^T \boldsymbol{x}_n - w^T \mu_n)^T (w^T \boldsymbol{x}_n - w^T \mu_n) \tag{5.15}$$

The optimization problem for LDA beacomes:

$$\max_w \frac{(w^T \mu_w - w^T \mu_n)^2}{s_w^2 + s_n^2} \tag{5.16}$$

Then numerator of maximization equation is:

$$\begin{aligned} (w^T \mu_w - w^T \mu_n)^2 &= w^T (\mu_w - \mu_n)(\mu_w - \mu_n)^T w \\ &= w^T S_B w \end{aligned} \tag{5.17}$$

where $S_{BC} = (\mu_w - \mu_n)(\mu_w - \mu_n)^T$ is the between-class scatter matrix. Also denominator is:

$$\begin{aligned}(s_w^2 + s_n^2) &= w^T(x_w - \mu_w)(x_w - \mu_w)^T w + w^T(x_n - \mu_n)(x_n - \mu_n)^T w \\ &= w^T S_w w + w^T S_n w \\ &= w^T(S_w + S_n)w\end{aligned} \tag{5.18}$$

where $S_{WC} = (S_w + S_n)$ is the within-class scatter matrix. To maximize Equation 5.14, the derivative of optimization equation with respect to w is taken and set to zero.

$$\frac{w^T(\mu_w - \mu_n)}{w^T S_{WC} w}(2(\mu_w - \mu_n) - \frac{w^T(\mu_w - \mu_n)}{w^T S_{WC} w})S_{WC} w = 0 \tag{5.19}$$

where $\frac{w^T(\mu_w - \mu_n)}{w^T S_{WC} w}$ is a constant. Then, w can be obtained with

$$w = c\, S_{WC}^{-1}(\mu_w - \mu_n) \tag{5.20}$$

where $S_{WC} = (S_w + S_n) = (\boldsymbol{x}_w - \mu_w)(\boldsymbol{x}_w - \mu_n)^T + (\boldsymbol{x}_n - \mu_n)(\boldsymbol{x}_n - \mu_n)^T$. Because of the importance of direction, c can be taken as 1.

After LDA is applied to 14 features dataset, the dimension of feature space is reduced from 33 to 1. The wheeze and non-wheeze samples in a one-dimensional feature set are fitted to a normal distribution, and Figure 5.7 shows the distribution plots of these samples. Then maximum likelihood classifier in Section 4.4 is applied to samples and Table 5.10 shows the results.

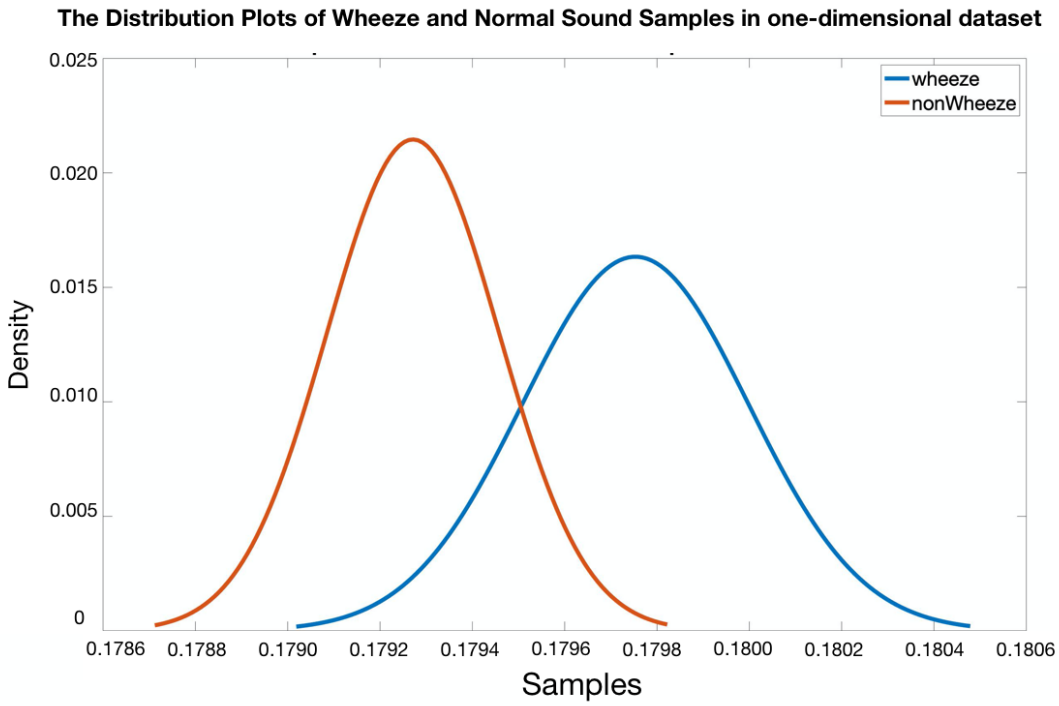

Figure 5.7. The plot of distributions of wheeze and non-wheeze samples in one dimensional feature set after LDA

Table 5.10. Performance of Bayesian Classifier on one dimensional feature set after LDA

Metrics	**Results**
Accuracy	87.87%
Sensitivity	93.91%
Specificity	74.22%
F1 Ratio	0.92
Training time	5.4 seconds 5-fold cross validation

5.8. Computational Time Consumption of Each Feature

The time consumption of the calculation of each feature is shown in the table 5.11. To make time consumption analysis, an Apple MacBook Pro (2018) is used with 2.3 GHz Quad-Core Intel Core i5 processor and 8 GB 2133 MHz RAM.

Every feature calculation is made with a 15 seconds respiratory sound. This respiratory sound is windowed with 25%, 374 windows, overlap before making the feature calculation.

Table 5.11. Computational time consumption of each feature

Feature	**Time**	**Feature**	**Time**
1. Mean Crossing Irregularity	20 ms	8. f_{25}/f_{90} Ratio	728 ms
2. Renyi Entropy with alpha 1	34 ms	9. f_{50}/f_{75} Ratio	799 ms
3. Renyi Entropy with alpha 2	36 ms	10. f_{50}/f_{90} Ratio	813 ms
4. Kurtosis	41 ms	11. f_{25}/f_{75} Ratio	1216 ms
5. Renyi Entropy with alpha 3	43 ms	12. MFCC	2112 ms
6. AR Parameters	100 ms	13. FFT Peak Baseline Diff. in DB	2309 ms
7. Tonality Index	108 ms	14. Audio Spectral Envelope	3525 ms

5.9. Result: Best Feature Set

In this chapter, various feature selection algorithms are applied our data set to select best features. We compare their computational time and discriminatory abilities. We concluded that:

(i) Mel Frequency Cepstral Coefficients (MFCC)
(ii) FFT peak baseline difference in dB (FPBD)
(iii) Renyi Entropy with alpha 2 (Renyi 2)
(iv) Mean Crossing Irregularity (MCI)

compose the best feature set with final accuracy of **92,18** and **0,86** F1 ratio. (SVM with RBF Kernel, 20 - fold cross validation, Kernel scale: 2.50 and C parameter 1)

6. DETECTION

In the previous chapter, the best features for classifying wheeze and normal sounds are decided with various feature selection techniques. Consequently, in this chapter, the ultimate aim of the thesis, detection algorithm, is presented. With the 92 percent accuracy rate, the SVM with RBF Kernel model is employed to predict the wheezes among normal respiratory sounds with the feature set contains MFCC, FPBD, Renyi 2 and MCI.

6.1. Windowing and Feature Calculation

First, fifteen seconds of respiratory sound is divided into 374 windows that contain 512 samples with a 25 percent overlap. Then, each feature is calculated for every window. The training vector has 374 rows, number of windows, and four columns, number of features. Figure 6.1 shows a 15 seconds respiratory sound with wheeze labels in green rectangular.

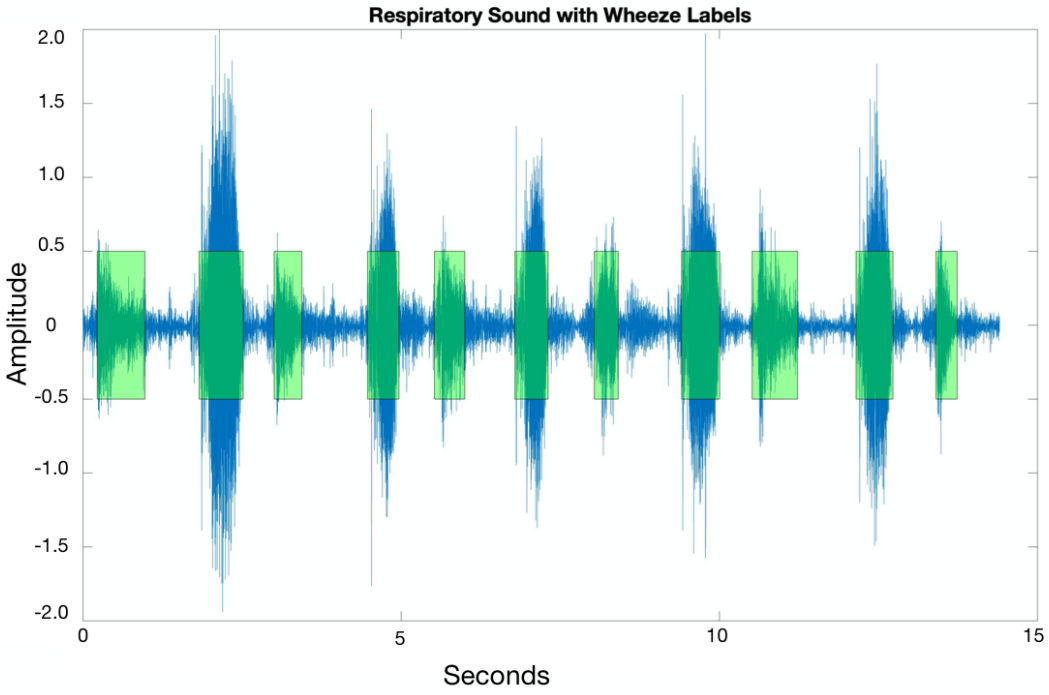

Figure 6.1. Respiratory sound with wheezes (green) labeled by expert

6.2. Prediction

After windowing and feature calculation, the trained SVM model predicts the labels of windows as wheeze and non-wheeze. Figure 6.2 shows the 15 seconds long respiratory sound and the windows, which are labeled as wheeze with the red-colored rectangles.

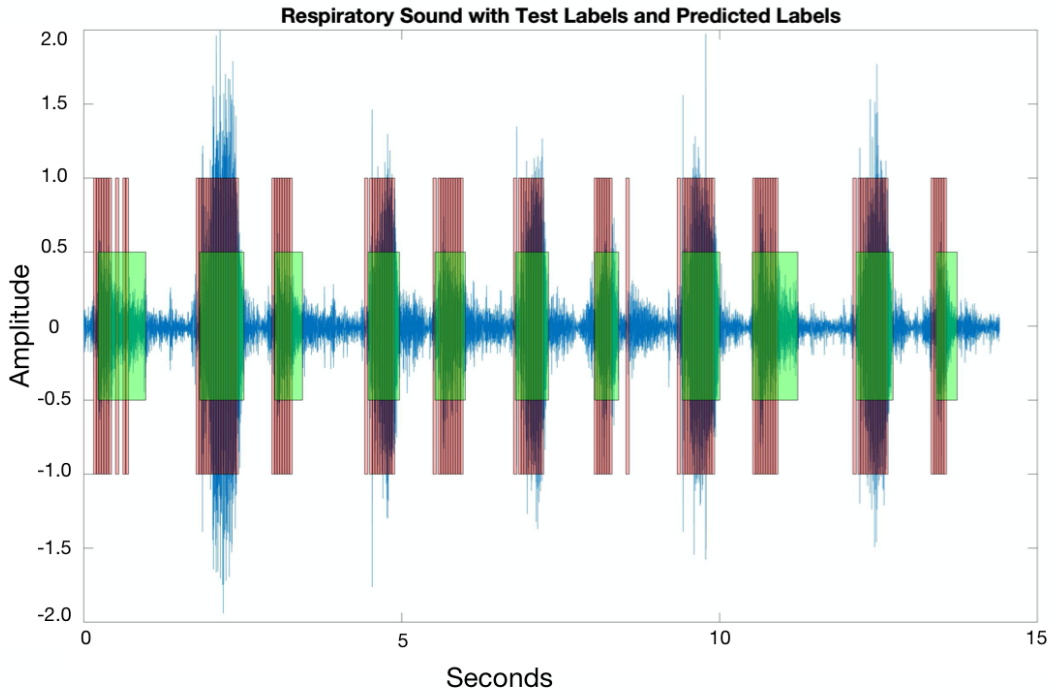

Figure 6.2. Respiratory sound with test windows (green) and predicted wheeze windows (red)

6.3. Wheeze decision

The determination of the start and end points of wheeze superimposed on respiratory sounds is one of the most significant parts of this thesis.

As seen in Figure 6.2, many windows predicted as wheeze are formed as groups. Clusters of wheezes show the probable interval of predicted wheezes. However, some-

times false negative windows (wheeze window but labeled as normal) remain between these clusters. To correct these windows, an algorithm is developed. The algorithm relabels these as wheeze if they are between two true positive (wheeze window labeled correctly as wheeze) windows. The number of false negative windows between true positive windows can be a maximum of two. The number of 2 is selected empirically.

Also, the algorithm makes one bigger window from clusters if the windows inside clusters are strictly bounded to each other. After this operation, the algorithm calculates the duration of bigger sized windows and removes them if their duration is smaller than 100 ms. The value of 100 ms is the minimum time duration of a wheeze can remain according to Computerized Respiratory Sound Analysis (CORSA). Figure 6.3 shows wheezes labeled by expert (green) and wheezes labeled by algorithm (yellow) superimposed on the 15 seconds long respiratory sound.

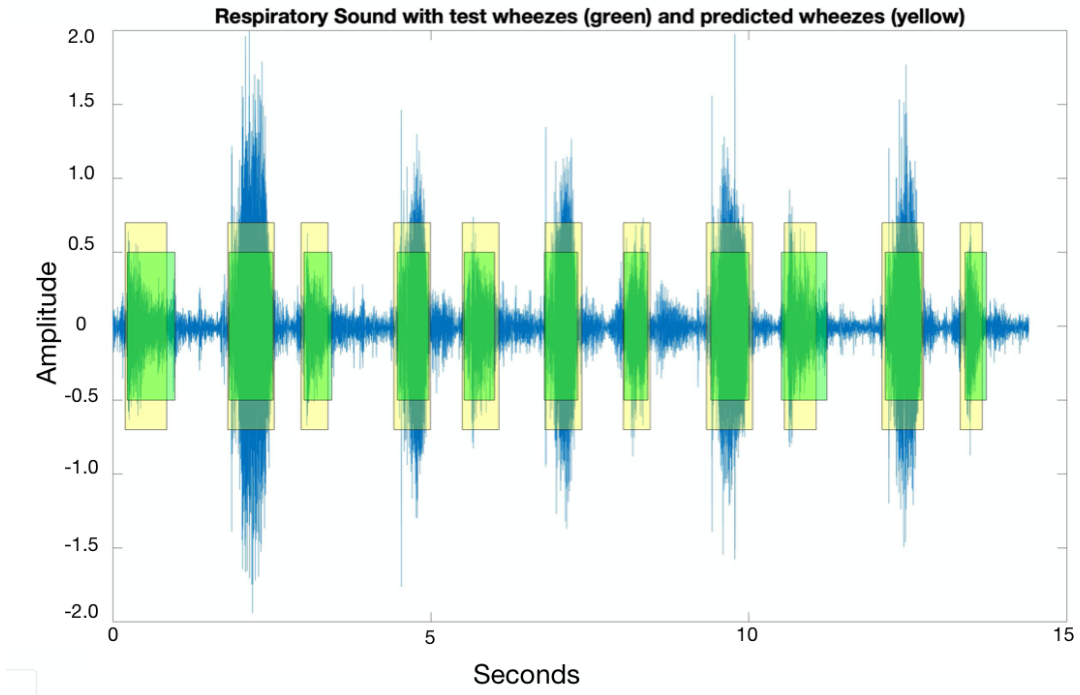

Figure 6.3. Respiratory sound with test windows (green) and windows predicted as wheeze (yellow)

6.4. Result

The final aim of the thesis is accomplished by presenting the detection algorithm. The algorithm is fast and straightforward. Its accuracy mostly depends on the classifier's results. However, there is no known metric to measure the exact accuracy of the detection algorithm. In this thesis, we present a new metric named YCI value. YCI value is calculated with the following equation.

$$YCI = \sqrt{\frac{(A_{detection} \cap A_{test})^2}{A_{detection} A_{test}}} \tag{6.1}$$

where $A_{detection}$ is the interval indices of detected wheeze, and A_{test} is the interval indices of test wheeze labeled by expert. The YCI value ranges typically from 0 to 1. 1 is the optimum value, which means that all the wheezes are detected correctly. 0 is the worst value, which means that none of the wheezes are detected correctly. The YCI value of the respiratory sounds in Figure 6.3 is the 0.90.

Table 6.1. YCI values of all recordings

Recordings	**YCI**	**Recordings**	**YCI**	**Recordings**	**YCI**
Rec 1	0.90	Rec 13	0.63	Rec 25	0.79
Rec 2	0.87	Rec 14	0.88	Rec 26	0.81
Rec 3	0.84	Rec 15	0.89	Rec 27	0.76
Rec 4	0.88	Rec 16	0.73	Rec 28	0.78
Rec 5	0.85	Rec 17	0.89	Rec 29	0.84
Rec 6	0.87	Rec 18	0.78	Rec 30	0.83
Rec 7	0.87	Rec 19	0.91	Rec 31	0.87
Rec 8	0.82	Rec 20	0.93	Rec 32	0.86
Rec 9	0.77	Rec 21	0.79	Rec 33	0.90
Rec 10	0.91	Rec 22	0.85	Rec 34	0.88
Rec 11	0.73	Rec 23	0.90	Average	0,84∓0,04
Rec 12	0.89	Rec 24	0.78		

7. CONCLUSIONS

Analyzing respiratory sounds and detecting anomalies in them with intelligent algorithms has opened a new era for auscultation that has 250 years of history. Furthermore, wheezes as abnormal sounds are one of the hot topics currently being researched by many researchers. In this thesis, we presented a new intelligent algorithm to detect wheezes that are superimposed on vesicular sounds.

The machine learning algorithm we found quite successful in detecting wheeze is the Gaussian SVM classifier that was trained with nine sound features used in other researches. Gaussian SVM among five classifiers gave the best results with the parameters determined in this study. It reached a 92.66% accuracy rate. The methodology followed, details of classifiers and the other result metrics achieved are presented and documented in Chapter 4.

The number of sound features used in this study is nine, and they were chosen among the most known and researched features so far. The features are documented and their frequency histograms according to classes are presented in Chapter 3. Also, the algorithms of complicated features are noted in this chapter.

In this work, one of the features is first named and detailed for the classification of wheezes for the first time. It was named as 'FFT peak baseline difference in dB' and shortened it as FPBD. Moreover, this feature has been concluded as the member of the final feature set that distinguishes wheezes best among all nine features with various selection algorithms. The features which form the best discriminatory feature set are Mell Frequency Cepstral Coefficient (MFCC), FFT peak baseline difference in dB (FPBD), Renyi Entropy with alpha 2 (Renyi 2), and Mean Crossing Irregularity (MCI).

These four features are the most discriminatory features with Gaussian SVM ac-

cording to both F1 ratio and accuracy. This result is corrected by forward sequential feature selection, backward sequential feature selection, and brute-force feature selection techniques. In backward sequential feature selection, kurtosis takes the place of tonality index according to F1 ratios but this can be understandable because correlation between kurtosis and tonality index is high.

In this study, PCA and LDA are applied to the feature set for observing the effects of dimensionality reduction. After the dimensionality reduction, different types of classifiers are used to classify wheezes. But, PCA has no major advantage. On the other hand, after LDA is applied, the Bayesian classifier gives the highest F1 ratio of 0.92. This result is surprisingly interesting.

Also, for the detection algorithm developed in the thesis, a new metric was introduced and defined. it was named as YCI Value. It ranges from 0 to 1, and it is the metric of the measure of how much wheezes are correctly detected by a detection algorithm. The details of the detection algorithm and YCI value are noted in chapter 6. Average of YCI values for all the recordings in the dataset is 0.84.

REFERENCES

1. Welsby, P. D., G. Parry and D. Smith, "The stethoscope: Some preliminary investigations", *Postgraduate Medical Journal*, Vol. 79, No. 938, pp. 695–698, 2003.

2. Sciencemuseumgroup.org.uk, "Laennec's stethoscope — Science Museum Group Collection", 2020, `https://collection.sciencemuseumgroup.org.uk/laennecs-stethoscope` `March 2020`.

3. Pasterkamp, H., S. S. Kraman and G. R. Wodicka, "Respiratory Sounds", *American Journal of Respiratory and Critical Care Medicine*, Vol. 156, No. 3, pp. 974–987, 1997.

4. Vannuccini, L., J. E. Earis, P. Helistö, B. M. Cheetham, M. Rossi, A. R. Sovijärvi and J. Vanderschoot, "Capturing and preprocessing of respiratory sounds", *European Respiratory Review*, Vol. 10, No. 77, pp. 616–620, 2000.

5. Mangione, S. and L. Z. Nieman, "Pulmonary auscultatory skills during training in internal medicine and family practice", *American Journal of Respiratory and Critical Care Medicine*, Vol. 159, No. 4, pp. 1119–1124, 1999.

6. Chamberlain, D., J. Mofor, R. Fletcher and R. Kodgule, "Mobile stethoscope and signal processing algorithms for pulmonary screening and diagnostics", *Proceedings of the 5th IEEE Global Humanitarian Technology Conference GHTC 2015*, pp. 385–392, Seattle, 2015.

7. Sen, I. and Y. P. Kahya, "A multi-channel device for respiratory sound data acquisition and transient detection", *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, Vol. 7, pp. 6658–6661, Shanghai, 2005.

8. Suarez, S. E., M. E. Brookfield, E. J. Catlos and D. F. Stöckli, "A U-Pb zircon age constraint on the oldest-recorded air-breathing land animal", *PLOS ONE*, Vol. 12, No. 6, p. e0179262, 2017.

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

10. Marques, A. and A. Oliveira, "Normal versus adventitious respiratory sounds", *Breath Sounds From Basic Science to Clinical Practice*, pp. 181–206, first edn., Springer International Publishing, New York, 2018.

11. Pasterkamp, H. and I. Sanchez, "Effect of gas density on respiratory sounds", *American Journal of Respiratory and Critical Care Medicine*, Vol. 153, No. 3, pp. 1087–1092, 1996.

12. Forgacs, P., "Crackles and wheezes.", *Lancet*, Vol. 2, No. 7508, pp. 203–205, 1967.

13. Holford, S. K., "Discrete lung sounds: Crackles (rales) as stress-relaxation quadrupoles", *Journal of the Acoustical Society of America*, Vol. 73, No. 3, pp. 1036–1046, 1983.

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

15. 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, p. CCRPM.S530, 2008.

16. Workum, P., S. K. Holford, E. A. Delbono and R. L. Murphy, "The prevalence and character of crackles (rales) in young women without significant lung disease", *American Review of Respiratory Disease*, Vol. 126, No. 5, pp. 921–923, 1982.

17. Thacker, R. E. and S. S. Kraman, "The prevalence of auscultatory crackles in subjects without lung disease", *Chest*, Vol. 81, No. 6, pp. 672–674, 1982.

18. Meslier, N., G. Charbonneau and J. L. Racineux, "Wheezes", *The European respiratory journal*, Vol. 8, No. 11, pp. 1942–8, 1995.

19. Grotberg, J. B. and S. H. Davis, "Fluid-dynamic flapping of a collapsible channel: Sound generation and flow limitation", *Journal of Biomechanics*, Vol. 13, No. 3, pp. 219–230, 1980.

20. Dokur, Z., "Respiratory sound classification by using an incremental supervised neural network", *Pattern Analysis and Applications*, Vol. 12, No. 4, pp. 309–319, 2009.

21. Yi, G. A., "A software toolkit for acoustic respiratory analysis", *MIT Computer Sound and Artificial Intelligence Laboratory*, Vol. 1, No. 3, pp. 215–216, 2004.

22. Charleston-Villalobos S.*, A. T. G.-C. R., Aljama-Corrales, "On applying continuous wavelet transform in wheeze analysis Analysis of Simulated Heart Sounds by Intrinsic Mode Functions", *Engineering in Medicine and Biology Society*, Vol. 5, No. 32, pp. 2848–2851, 2006.

23. Society, A. T., "Updated nomenclature for membership reaction", *ATS NEWS*, Vol. 3, pp. 5–6, 1977.

24. *World Health Organization — Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach*, Tech. rep., 2012.

25. Orjuela-Cañón, A. D., D. F. Gómez-Cajas and R. Jiménez-Moreno, "Artificial neural networks for acoustic lung signals classification", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 8827, pp. 214–221, Cham, 2014.

26. Bahoura, M., "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes", *Computers in Biology and Medicine*, Vol. 39, No. 9, pp. 824–843, 2009.

27. Oweis, R. J., E. W. Abdulhay, A. Khayal and A. Awad, "An alternative respiratory sounds classification system utilizing artificial neural networks", *Biomedical Journal*, Vol. 38, No. 2, pp. 153–161, 2015.

28. Mendes, L., I. M. Vogiatzis, E. Perantoni, E. Kaimakamis, I. Chouvarda, N. Maglaveras, V. Tsara, C. Teixeira, P. Carvalho, J. Henriques and R. P. Paiva, "Detection of wheezes using their signature in the spectrogram space and musical features", *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Vol. 2343, pp. 5581–5584, Milan, 2015.

29. Wisniewski, M. and T. P. Zieliński, "Joint application of audio spectral envelope and tonality index in an E-asthma monitoring system", *IEEE Journal of Biomedical and Health Informatics*, Vol. 19, No. 3, pp. 1009–1018, 2015.

30. Aydore, S., I. Sen, Y. P. Kahya and M. Kivanc Mihcak, "Classification of respiratory signals by linear analysis", *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, pp. 2617–2620, Minneapolis, 2009.

31. Oletic, D., B. Arsenali and V. Bilas, *Towards continuous wheeze detection body sensor node as a core of asthma monitoring system*, New York, first edn., 2012.

32. Liu, X., W. Ser, J. Zhang and D. Y. T. Goh, "Detection of adventitious lung sounds using entropy features and a 2-D threshold setting", *10th International Conference on Information, Communications and Signal Processing*, Shanghai, 2015.

33. Ulukaya, S., I. Sen and Y. P. Kahya, "Feature extraction using time-frequency anal-

ysis for monophonic-polyphonic wheeze discrimination", *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Vol. 987, pp. 5412–5415, Milan, 2015.

34. Adhi Pramono, R. X., S. A. Imtiaz and E. Rodriguez-Villegas, "Evaluation of features for classification of wheezes and normal respiratory sounds", *PLoS ONE*, Vol. 14, No. 3, p. e0213659, 2019.

35. Shannon, C. E., "A Mathematical Theory of Communication", *Bell System Technical Journal*, Vol. 27, No. 3, pp. 379–423, 1948.

36. Kahya, Y. P., M. Yeginer and B. Bilgic, "Classifying respiratory sounds with different feature sets", *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pp. 2856–2859, Hong Kong, 2006.

37. Johnston, J. D., "Estimation Of Perceptual Entropy Using Noise Masking Criteria", *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 2524–2527, New York, 1988.