DESIGN OF A SYSTEM FOR DIAGNOSIS OF LUNG DISEASES USING PULMONARY SOUNDS

by

Erman Kömürcü

B.S., Electrical Electronic Engineering, Boğaziçi University, 2014

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 Electronic Engineering
Boğaziçi University
2019

ACKNOWLEDGEMENTS

I would first like to express my gratitude to my thesis supervisor Prof. Dr. Yasemin Kahya for her guidance and support throughout this study. Without her advice, this thesis could not have been successfully completed. Besides, I would like to thank my evaluation committee members Prof. Dr. Oğuzhan Çiçekoğlu and Asst. Prof. İpek Şen for their valuable feedback.

I would also like to thank Electrosalus biomedical company members for data sharing and their supports. I would also like to acknowledge Gunay Turan as the second reader of this thesis, and I am gratefully indebted to his for his very valuable comments on this thesis and his support during .mythesis work

Finally, I would also like to express my gratitude to my family for providing me with unfailing support and continuous encouragement throughout my years of study. Thank you.

ABSTRACT

DESIGN OF A SYSTEM FOR DIAGNOSIS OF LUNG DISEASES USING PULMONARY SOUNDS

Auscultation is a common method among physicians to detect adventitious sounds which are mainly wheezes and crackles, and are important indicators of a pathological condition. According to the presence of these adventitious sounds doctors diagnose a patient as healthy or pathological and may require further examinations. However, auscultation is highly subjective method and there is always a risk of missing adventitious sounds or misevaluating them. In this thesis, a new system is designed and implemented to reduce this risk by removing subjectiveness of the method using computerized techniques. Another advantage of this system is the elimination of unnecessary examinations and the saving of time for both doctors and patients. This system is intended to be used for detection of pathological conditions in recorded sounds rather than exact diagnosis of respiratory diseases.

This system is designed to determine the presence of any respiratory disorder in a subject. For this purpose, the system analyzes recorded pulmonary sounds with digital signal processing tools and machine learning techniques. It consists of three different detection algorithms and a final classifier to combine outputs of all algorithms to produce a result. These algorithms are wheeze detection algorithm, crackle detection algorithm, and disease detection algorithm.

ÖZET

SOLUNUM SESLERİNİ KULLANARAK AKCİĞER HASTALIKLARINI TESPİT EDEN ELEKTRONİK SİSTEM TASARIMI

Oskültasyon, hekimler arasında, çoğunlukla hırıltı ve çatırtıdan oluşan ve patolojik bir durumun önemli göstergesi sayılan normal olmayan solunum seslerini saptamak için kullanılan yaygın bir yöntemdir. Bu normal olmayan solunum seslerinin varlığına göre doktorlar bir hastayı sağlıklı veya patolojik olarak teşhis eder ve duruma göre ileri tetkiklere ihtiyaç duyabilirler. Bununla birlikte, oskültasyon son derece öznel bir yöntemdir ve her zaman normal olmayan solunum seslerini kaçırma veya bunları yanlış değerlendirme riski vardır. Bu tezde, bazı dijital teknikler kullanılarak yöntemin öznelliğini ortadan kaldırmak ve yanlış değerlendirme riskini azaltmak için yeni bir sistem tasarlanmış ve uygulanmıştır. Bu sistemin bir başka avantajı da gereksiz muayenelerin ortadan kaldırılması sayesinde hem doktorlar hem de hastalar için zaman tasarrufu sağlamaktır. Bu sistem, akciğer rahatsızlıklarının kesin teşhisi yerine, akciğerler ile ilgili herhangi bir patolojik durumun tespiti için kullanılır.

Bu sistem, bir denekte varolan herhangi bir akciğer rahatsızlığının varlığını belirlemek için tasarlanmıştır. Bu amaçla, sistem kaydedilmiş solunum seslerini dijital sinyal işleme araçları ve makine öğrenme teknikleri ile analiz eder. Bu sistem, bir sonuç üretmek için üç farklı algılama algoritmasından ve bu algoritmaların çıktılarını birleştiren son bir sınıflandırıcıdan oluşur. Bu algoritmalar, hırıltı algılama algoritması, çatırtı algılama algoritması ve hastalık algılama algoritmasıdır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a	Dilation parameter
a_k	Auto-Regressive model coefficient
$\mathrm{a}^m(k)$	Model parameter
b	Bias variable or translation parameter
C	Upper bond
d	Input count
e	Vector of all ones or prediction error
$\hat{e}$	Reconstructed prediction error
$e^m(n)$	White noise
E^t	Cross entropy
$E(X)$	Expected value
$H_\alpha(X)$	Renyi entropy
k	Kurtosis value
$K(x_i, x_j)$	Kernel function
m	Window index
n	Sample count
N	Window size
p	Dimension of vector space
p_i	Series of events
Q	Positive semi-definite matrix
r^t	Label of instance
s	Respiratory sound record
$\hat{s}$	Predicted value
t	Instance index
v	Weight of last layer
w	Normal vector of a hyperplane or connection weight
w	Weight vector
W	Continuous wavelet transform

$w[m]$	Discrete window function
$w(\tau)$	Continuous window function
x	Input vector
X	Real-valued random variable
x_i	Data vector
$x^{'}$	Teager's energy function
y	Output of a perceptron
y_i	Output vector
z	Output of a perceptron in hidden layer
α	Order of Renyi entropy
γ	Gamma parameter
Δw	Weight update rule
ζ_i	Slack parameter
η	Learning rate
λ	Compromise value between large margin size
μ	Mean value
σ	Standard deviation
τ	Time index
χ_0	Decomposition subband
$\phi(x)$	Mapping function
ψ	Mother wavelet
ψ^*	Complex conjugate of the mother wavelet
Ψ	Fourier transform of the mother mother wavelet
ω	Angular frequency

LIST OF ACRONYMS/ABBREVIATIONS

2CD	Two Cycle Duration
2D	Two Dimensional
ADC	Analog to Digital Converter
APF	All-Pole Filter
AR	Auto-Regressive
ATS	American Thoracic Society
CNN	Convolutional Neural Network
DWT	Discrete Wavelet Transform
FN	False Negative
FP	False Positive
GPU	Graphical Processing Unit
HPF	High Pass Filter
IDW	Initial Deflection Width
IDWT	Inverse Discrete Wavelet Transform
LAL	Lung Acoustic Laboratory
LPF	Low Pass Filter
MCU	Microcontroller
MLP	Multilayer Perceptron
PEF	Prediction Error Filter
PSD	Power Spectral Density
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
TFR	Time-Frequency Representation
TN	True Negative
TP	True Positive

1. INTRODUCTION

1.1. Background

Auscultation is a widely used method in the diagnosis of the respiratory disorders. It is a noninvasive method and performed mostly using a stethoscope. Pulmonary sounds are considered as valuable sources for diagnosing respiratory diseases. Using stethoscope for diagnosing pulmonary diseases has some shortcomings due to its inherent properties which reduce the value of the diagnosis. First, it is subjective since it depends strongly on the hearing capability and experience of the physician. Moreover due to the stethoscope's limited frequency response, it acts like a low pass filter with 112 Hz cut-off frequency [1]. However, the spectrum of respiratory sounds contains frequencies up to 2000 Hz for normal breath sounds and 6000 Hz for some pathological cases [2].

With the advancements in computer technology and signal processing techniques, overcoming these drawbacks is possible. Auscultation subjectivity can be decreased by parameterizing the sound data and evaluating these parameters with advanced techniques in machine learning. This process can be summarized as follows. The first step is capturing the respiratory sound with sound transducers. Then the captured respiratory sound signal is passed through an analog circuitry for filtering and preprocessing. After that, the signal is digitized, and various digital signal processing techniques are applied to it. Then recorded signals are fed into some machine learning algorithms to increase the diagnostic value of auscultation. These processed digital respiratory sound data are stored for building a database where they can be reached and processed later. There are various studies on analysis of pulmonary sound and some of them are given in the review articles [1, 3, 4].

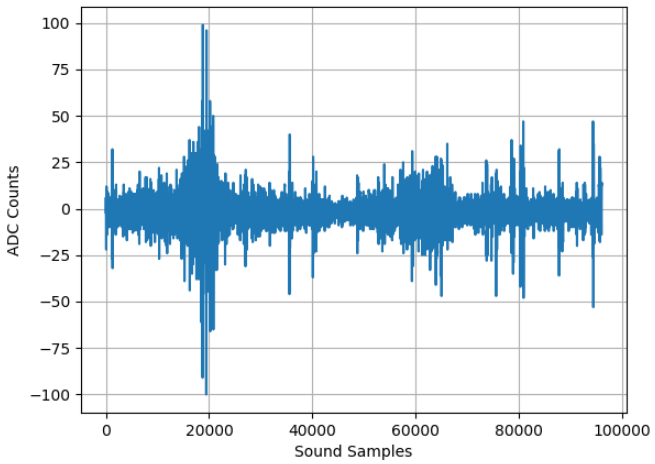
1.2. Respiratory Sounds

It is believed that air turbulence in the airways of the lungs induces pulmonary sounds, which are also called breath sounds or respiratory sounds [5]. Respiratory sounds consist of components occurring during inspiratory and expiratory phases of the breath circle. It is a well established belief that its inspiratory component is generated primarily within the lobar and segmental airways, whereas the expiratory component comes from more proximal locations [1].

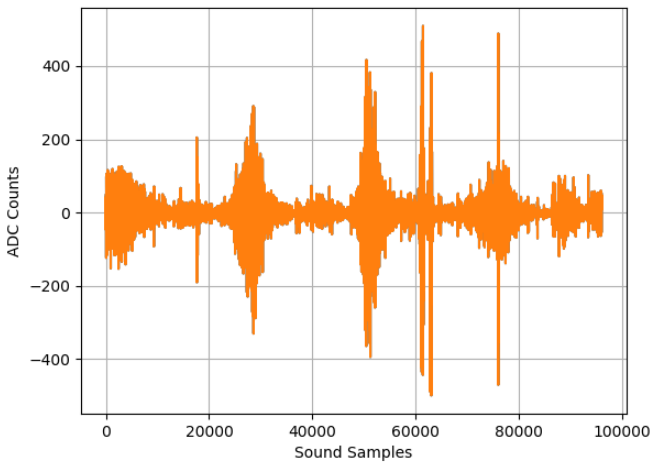
Normal respiratory sound, which is recorded from the chest wall of healthy human subjects, has a low-frequency noisy sound characteristic in inspiratory phase and hardly audible in expiratory phase [6]. If air flows through the airways without any obstruction, normal breath sounds occur. However, in case of existence of any obstructions such as constriction, fluid, or hyperexpansion, abnormal breath sound will occur. The change of lung structure due to some diseases affects the waveform of sounds heard over the chest wall and may cause abnormal breath sounds [1].

Pulmonary sounds include both vesicular sounds and adventitious sounds. Vesicular sound is normal breath sound detected over the chest wall and the most common sound heard in the absence of respiratory diseases. It has a soft and low-pitched characteristic and is louder during inhalation. The location of auscultation and pathological conditions may affect vesicular sound's amplitude and spectrum [7]. Adventitious sounds are additional abnormal breath sounds which are superimposed on normal breath sounds. They usually indicate various respiratory diseases. They comprise of both continuous (like wheezes) and discontinuous (such as crackles) sounds.

There are more adventitious sounds besides wheezes and crackles. However, for analyzing sound data recorded from patients' chest wall, wheezes and crackles are the most promising phenomena to help diagnosis of respiratory disorders [8]. In this thesis, crackle and wheeze have been used as adventitious sounds for detection of respiratory diseases. Raw lung sound signal for both healthy and pathological subjects can be seen in Figure 1.1. Crackles are discontinuous, non-musical, brief adventitious sounds that



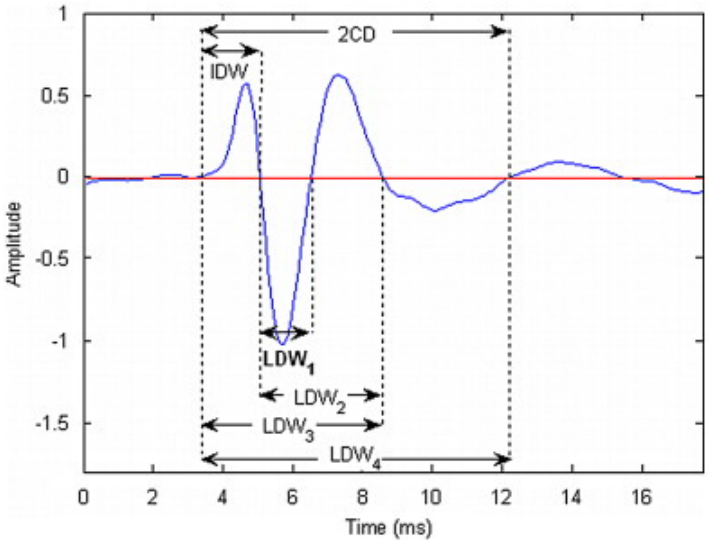
(a)



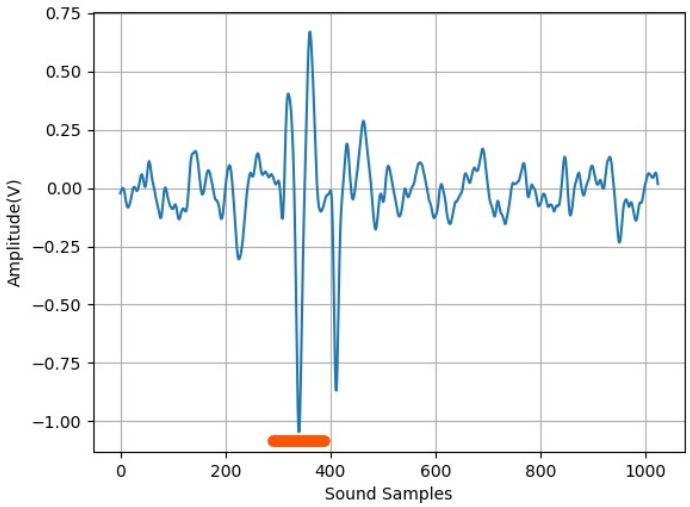
(b)

Figure 1.1. Lung sounds: (a) Normal lung sound, (b) Pathological lung sound

are heard during both inhalation and exhalation. In which breathe cycle crackles are mostly concentrated is an indicator of a disease type. They have a short duration of less than 20 ms. They occur in pathological conditions and are superimposed on normal breath sounds. They have explosive, intermittent and transient characteristics and appear frequently in cardio-respiratory diseases [1, 5]. An example of crackle waveform is depicted in Figure 1.2(a) [9] and Figure 1.2(b). Murphy et al. have suggested criteria for crackle waveforms [10]. There are two types of crackles, mainly coarse and fine



(a)



(b)

Figure 1.2. (a) Crackle waveform, (b) Crackle waveform in a lung sound

crackles but in this thesis, types of crackles are not investigated. Detection of crackles within a respiratory cycle indicates the presence of a respiratory disorder. Wheezes are another type of adventitious sounds which are continuous and musical. They usually appear in expiratory phase but sometimes they may be heard during inspiration. Wheezes are generally caused by obstructions in the airways. They are induced when air flows through airways narrowed by secretions, foreign bodies, or obstructive lesions.

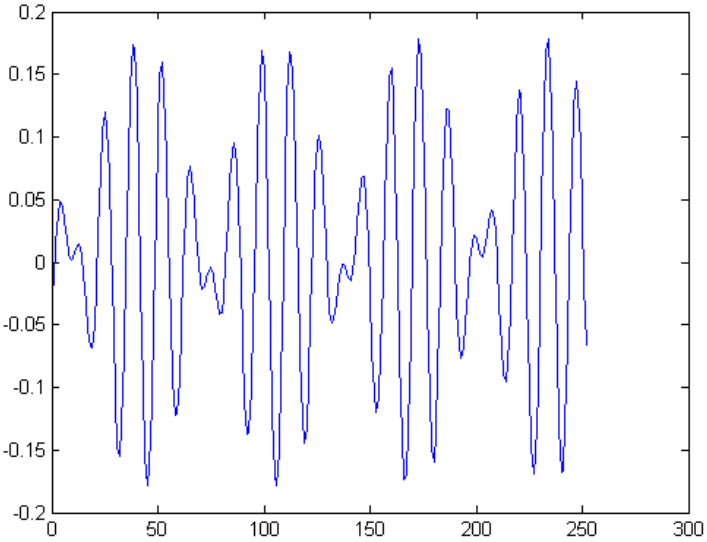


Figure 1.3. Wheeze waveform

They last longer than crackles with duration of generally longer than 100 ms and have periodic waveforms with a dominant frequency generally over 100 Hz [11]. Wheezes may contain either single frequency or several frequencies. Determining the number of wheezes in a record helps us to detect presence of a respiratory disease. Expanded wheeze waveform can be seen in Figure 1.3 [12].

Some subjects who do not have any adventitious sounds in their records might have a respiratory disease. In this case, characteristics of lung sound acquired from those subjects differ from healthy subjects. Because of this, in this thesis, unlabeled respiratory sound data are also examined for detection of any respiratory disease.

1.3. Motivation and Aim

The main purpose of the thesis is to construct a system to analyze recorded respiratory sounds with digital signal processing tools and machine learning techniques. This system is intended to be used for detection of pathological conditions in a record rather than obtaining exact diagnosis of respiratory diseases.

Figure 1.4. Smart Stethoscope

Adventitious sounds play an important role in auscultation. Especially, wheezes and crackles are commonly used to determine diseases. In this thesis, two different databases are used for crackle and wheeze detection separately. Both of these databases consist of lung sound records from thirteen different subjects with various respiratory disorders. The sounds were acquired from the posterior chest wall of each subject. These records have been checked by a specialist to determine locations of wheezes and crackles.

The algorithm which was used to detect presence of respiratory disorders in a record was applied on records that were acquired by Electrosalus's single channel electronic stethoscope in Yedikule Chest Diseases and Thoracic Surgery Training and Research Hospital under the guidance of a physician. This device is small and portable and it has internal analog filters to allow us to process sound signal before recording it. This stethoscope is seen in Figure 1.4.

The proposed disease detection system consists of three different independent units and a fourth unit which is used to merge the outputs of the three units with a few additional features to produce the final result. These three independent parts are wheeze detection, crackle detection and disease detection and the last part is classifi-

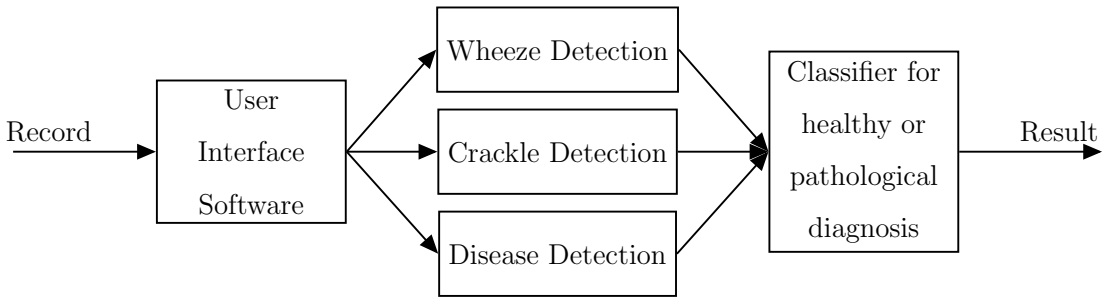


Figure 1.5. Block diagram for disease detection

cation part. A block diagram of the system is depicted in Figure 1.3. First, recorded respiratory sound data from a healthy or pathological subject is fed into the system on a computer. Then, this data record is processed by the three independent parts separately. In the wheeze detection part, record is searched for the presence of a wheeze and then if wheezes exist in the record then they are counted and number of wheezes is given as output to the last part. In the crackle detection unit, number of crackles is detected by using some digital signal processing techniques and machine learning algorithms and obtained result is sent to the last unit. In the disease detection unit, the same record is processed for detecting any disease sign and its result is also fed to the last unit. In the last unit, which merges the results of three independent algorithms, all results coming from previous units are combined with Auto-Regressive (AR) parameters and percentile frequencies to construct feature set of the Support Vector Machine (SVM) to obtain the final result.

Wheeze detection in this study consists of window separation, feature extraction and employing an SVM classifier. The records are divided into small windows and then a wheeze detection algorithm is used to detect wheezes in each window. Kurtosis, Renyi Entropy, Percentile Frequencies, and Zero Crossing Numbers are used in the feature vector for classification. In Chapter 3, this detection algorithm and its results are explained in detail.

For crackle detection, first, discrete wavelet decomposition is applied to records to localize crackles in time and frequency. Then, decomposed signal bands are thresholded

and reconstructed. After that, these signals are fed to a Multilayer Perceptron (MLP) network. The MLP network gives a label for the processed window which indicates whether a crackle exists in the window or not. In Chapter 4, this algorithm is presented in detail.

For detecting the presence of a respiratory disorder in a record, Short Time Fourier Transform (STFT) is applied to the whole signal and then results are fed to Convolutional Neural Network (CNN). For developing and training machine learning model for this algorithm, Tensorflow which is an open source machine learning software library officially supported by Google was preferred. It is mostly used in neural network applications and can utilize Graphical Processing Unit (GPU) to shorten training duration significantly. Since this process requires heavy computation, Tensorflow is used and runs on GPU. Details of this process are explained in Chapter 5.

For the final diagnosis of healthy vs. pathological classification, records are divided into smaller windows and then Auto-Regressive (AR) parameters and percentile frequencies are calculated for each window. A Support Vector Machine (SVM) is used to evaluate the record after gathering parameters from CNN, wheeze detection algorithm, crackle detection algorithm, AR parameters and percentile frequencies. This is the final result showing the presence of respiratory disorders. In Chapter 6, the final classifier is described and results are given.

This system is designed for determining the presence of any respiratory disorder in a subject. If the subject has any respiratory disorder then the system will detect it and lead the subject to physical examination for diagnosing the disorder. This will eliminate unnecessary examinations and save time for both doctors and patients.

2. DATA

Three different databases were used for training and validation of the algorithms. Since each database has specific characteristics, they have been used separately in different detection algorithms. Wheeze detection algorithm uses the wheeze database where each wheeze in records was marked by a specialist. Crackle detection algorithm was applied to crackle database. In this database, location of each crackle was determined and noted by a specialist. For the disease detection algorithm, each record in the Electrosalus's database has been labeled by a specialist as healthy or pathological. The three different sets of data used are as follows:

(i) Wheeze Data Set
(ii) Crackle Data Set
(iii) Electrosalus's Data Set

2.1. Wheeze Data

For wheeze detection, a database consisting of 98 lung sound records of different subjects has been used. This database was constructed by Bogazici University Lung Acoustics Laboratory (LAL). A detailed description of the acquisition system may be found in [13]. These records include data from both healthy and pathological subjects. There are 246 different wheezes in the records which were detected by two different specialists. Properties of the respiratory sound acquisition system may be summarized as follows: Frequency sampling rate is 9600 Hz, resolution is 12-bits, and each record lasts 15 second long. Therefore each record consists of 144400 sample points. These lung sounds have been acquired from 14 different locations on the chest wall.

2.2. Crackle Data

For crackle detection, a different database has been used. Like wheeze database, it has been also constructed by Bogazici University LAL. The same techniques as in

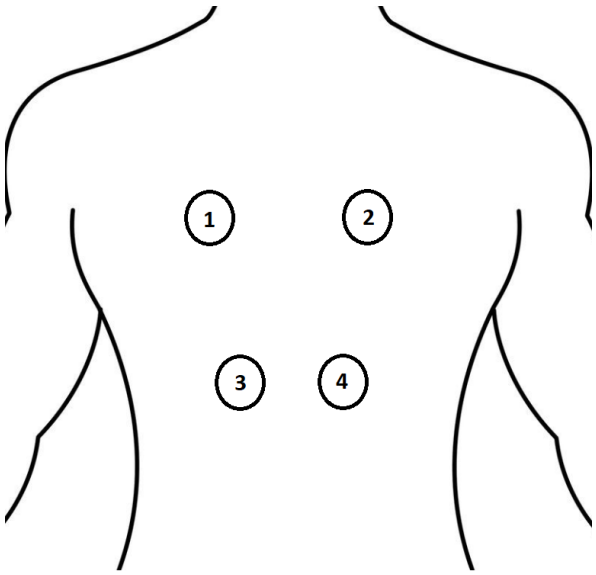


Figure 2.1. Recording locations on the posterior chest wall

previous section have been used for gathering this database. In this database, there are 56 different records, which contain total 998 crackles.

2.3. Electrosalus's Data

This database was constructed by Electrosalus Bio-medical Company under the guidance of a physician specialized in the pulmonary medicine at Yedikule Teaching Hospital for Chest Diseases and Thoracic Surgery. The data acquisition procedure had the approval of the 2nd Ethical Committee on Clinical Research of Istanbul in compliance with Declaration of Helsinki. Lung sounds were captured by the mobile electronic stethoscope of the company.

This database includes 40 records of 8 different patients and 24 records of 14 different healthy subjects. Four different locations on the chest wall have been determined by a physician specialized in pulmonary medicine for lung sound recording and they are represented in Figure 2.1. These regions were used when recording a subject's lung sound.

Some of the basic hardware parts of the stethoscope are an electret microphone, a microcontroller (MCU), a few analog filters, and a Bluetooth device. The electret microphone located inside of a teflon capsule for capturing respiratory sound is used in the stethoscope to convert the sound to an electric signal. The analog filters inside the stethoscope are used to filter out the high frequencies which exist in the recorded signal and to eliminate ambient noise from the signal on the mobile device. The MCU is employed for controlling the recording process and communication, and the Bluetooth device is used to communicate with the external interface.

Characteristics of the stethoscope can be summarized as follows. An 8th order Butterworth Low Pass Filter (LPF) with 3600 Hz cut-off frequency is used to filter out high frequencies that are not useful for diagnosis and a 6th order Bessel High Pass Filter (HPF) with 80 Hz cut-off frequency is used to filter out heart sounds and friction noise. Analog signal is converted to digital signal using a 10 bit Analog to Digital Converter (ADC) and its sampling rate was set to 9600 Hz.

A mobile application is used to receive data from the stethoscope and send them over the internet to the server, where the records are stored and analyzed.

3. WHEEZE DETECTION

The detection of a wheeze in the respiratory cycle has great importance in the diagnosis of obstructive airways pathologies [14]. Presence of a wheeze is an indication of a pathological condition thus its detection is important to diagnose whether a subject is healthy or pathological. In order to identify wheezes within a respiratory sound signal, a feature set based on the article by Aydore et al. is used [15].

3.1. Methodology

In [15] , four features which were defined for detection of wheezes are Kurtosis, Renyi Entropy, f_{50}/f_{90} ratio, and Mean Crossing Irregularity. In this thesis, instead of Mean Crossing Irregularity, Zero Crossing is used to increase the performance of the algorithm.

Block diagram for the wheeze detection algorithm is depicted in Figure 3.1 and applied steps of the algorithm are described as follows:

(i) Lung sound records from the wheeze database are divided into smaller window sizes of 120 ms samples with a 25 percent overlap.
(ii) Each window is labeled as either wheeze or non-wheeze.
(iii) Four features are calculated in Python for each window.
(iv) Parameters are fed to the SVM and final results are obtained.

The features and SVM are explained in the following sections.

3.1.1. Kurtosis

Kurtosis is a tail weight of a probability distribution of a real-valued random variable. It gives us an idea about the shape of a probability distribution. Any univariate normal distribution has the Kurtosis value of 3 and if it is less than 3 then it

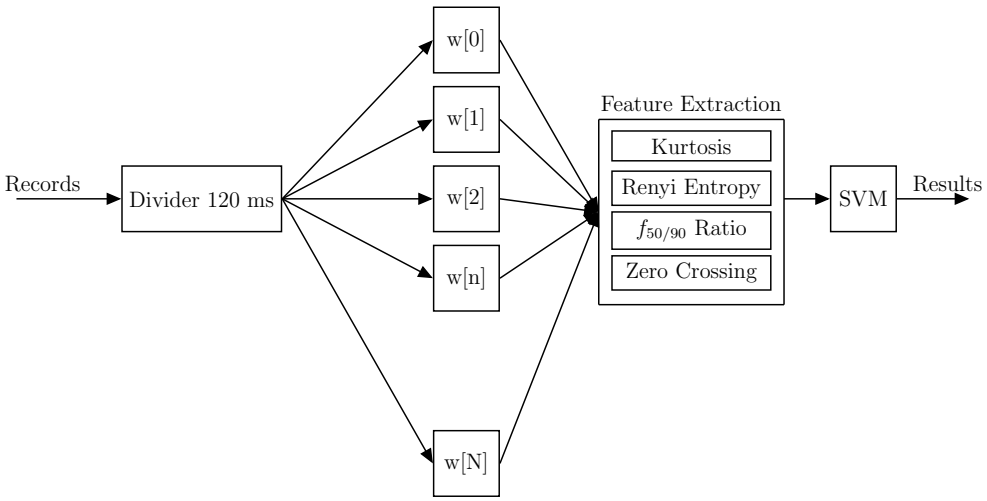


Figure 3.1. Block diagram for wheeze detection

will be similar to uniform distribution with fewer and less extreme outliers and if it is greater than 3 then the tails of the distribution will approach to zero more slowly than a Gaussian distribution with more outliers. Kurtosis is defined as the formula below.

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

where k is kurtosis value, X is real-valued random variable, μ is mean of X, and σ^2 is variance of X, and E(.) is expected value.

Lung sounds in time domain have been assumed as a probability mass function of a real-valued discrete random variable. It is expected that non-wheeze signal and wheeze signal have different Kurtosis values. Wheeze signals are similar to a uniform distribution, whereas non-wheeze signals resemble a normal distribution. In the thesis, Kurtosis value of each window has been calculated in Python using Equation (3.1). Histogram of the calculated Kurtosis values of each window can be seen in Figure 3.2. This figure is normalized to see the difference between non-wheeze and wheeze windows easily. It can be seen that there is a slight separation between non-wheeze windows and wheeze windows. In Figure 3.2, wheeze windows mostly have Kurtosis

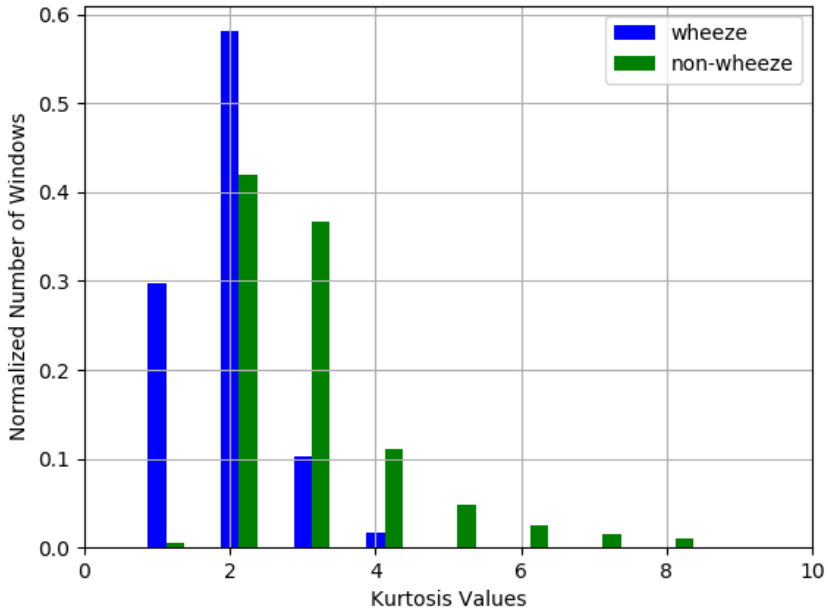


Figure 3.2. Histogram of Kurtosis Values

values lower than 3. Contrary to that, non-wheeze windows usually have Kurtosis values higher than 3.

3.1.2. Renyi Entropy

Rényi entropy is the generalized form of Shannon entropy with an order greater than 0. It is used as measures of uncertainties. The Renyi entropy of a discrete random variable X is defined by

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

where α is the order of Renyi Entropy, p_i is series of events for $i = 1, ..., n$. In this thesis, Renyi's entropy with $\alpha = 2$ is used for calculations.

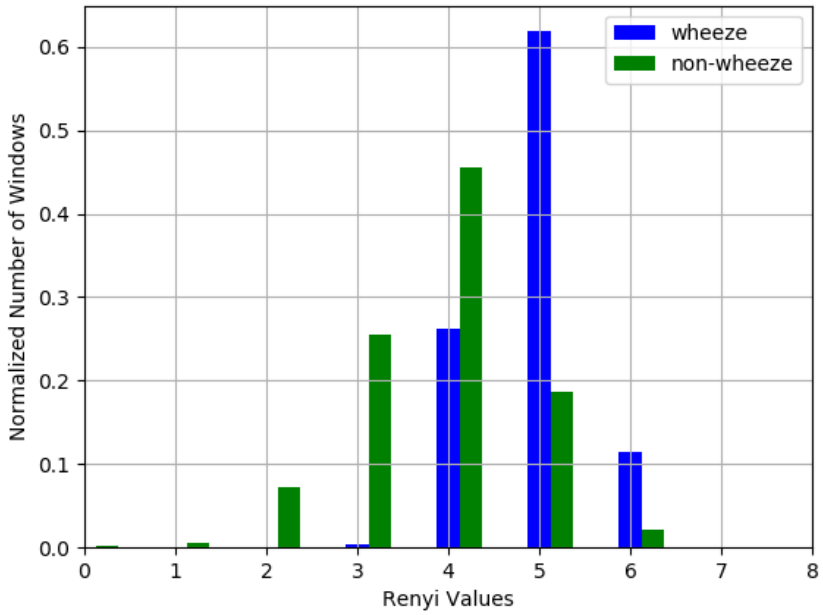


Figure 3.3. Histogram of Renyi Entropy

Presence of a wheeze on a lung sound record causes a deviation from normal state, which leads uncertainties in the record. Because of this, it is expected that Rényi entropy of a window containing a wheeze should be higher than a window which does not contain any wheeze. Renyi entropy for each window has been calculated in the time domain and normalized histogram of computed Renyi entropy for those windows is shown in Figure 3.3. In the normalized histogram, Renyi entropy of the wheeze signals is generally higher than 4 whereas most of the non-wheeze signals are concentrated on the left side of Renyi value of 4. This separation is used as a feature for distinction between non-wheeze and wheeze signals.

3.1.3. f_{50}/f_{90} Ratio

The power spectral density (PSD) describes power distribution of the signal in the frequency domain. f_{50} and f_{90} cover the area under PSD function in the range of 0 to 50% and 0 to 90% respectively.

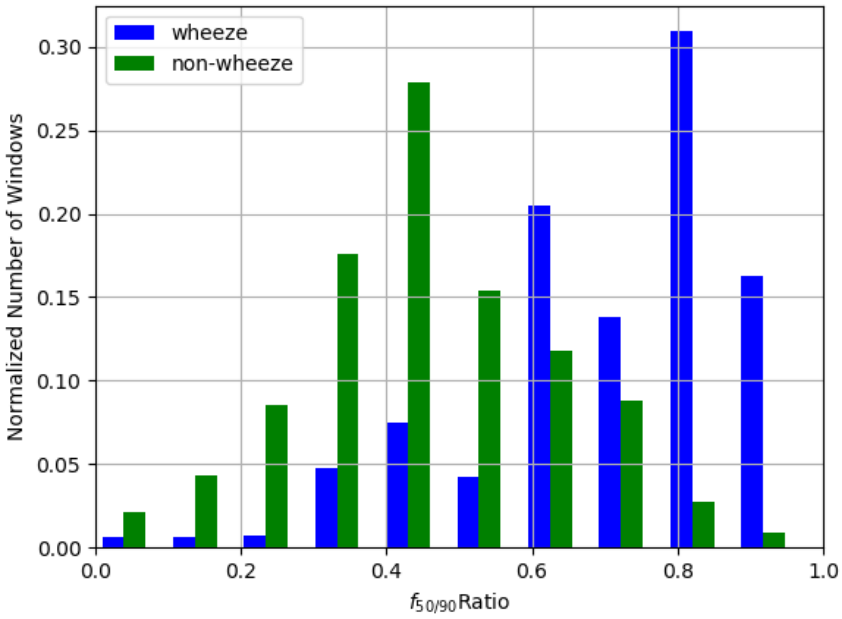


Figure 3.4. Histogram of f_{50}/f_{90} Ratio

PSD was calculated for each window using Welch method where window length and number of Fourier Transform points are 256, overlapping ratio is 50% and Hamming window function is used.

It is expected that if a window includes wheeze then its spectrum is concentrated along a single frequency. Therefore, f_{50} and f_{90} ratio of the window should be close to 1. Otherwise, it is close to 0.5. In Figure 3.4, normalized histogram of f_{50}/f_{90} ratio shows the difference between the non-wheeze and wheeze windows. In the histogram, wheeze windows have higher f_{50}/f_{90} ratio than non-wheeze windows have.

3.1.4. Zero Crossing

Zero crossing is a point in the graph where a signal changes its sign. In other words, it is a point on the signal where it intercepts the x-axis. Counting zero crossing is especially used in speech processing to determine the frequency of a waveform. It is expected that if a wheeze exists in a window, that window has higher zero crossing

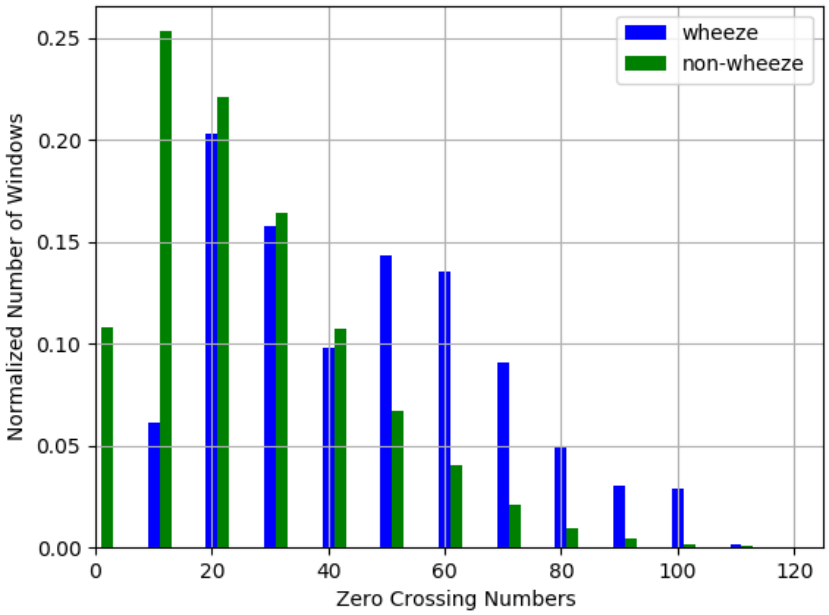


Figure 3.5. Histogram of Zero Crossing

count than a non-wheeze window has since the waveform of wheeze in the time domain causes this situation. Figure 3.5 represents this phenomenon, where wheeze signals tend to have higher zero crossing count than non-wheeze signals have.

3.1.5. SVM

SVM is a supervised learning model and is used to find a separating hyperplane which divides a data-set into classes. Support vectors are data points in a given data-set that lie closest to the hyperplane. This hyperplane is constructed using an iterative training algorithm that tries to minimize a predefined error function and is also called decision surface. Training algorithm also tries to maximize the margin between decision surface.

For given data vectors $x_i \in \mathbb{R}^p, i = 1, \ldots, n$. And for two classes, an output vector y, where $y_i \in \{+1, -1\}^n$, SVMs should minimize

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0, 1 - y_i(w \cdot x_i - b)\right)\right] + \lambda\|w\|^2 \tag{3.3}$$

where w is the normal vector of the hyperplane, b is bias, and λ is a compromise between large margin-size and ensuring that the x_i lie on the correct side of the margin.

Equation (3.3) can be minimized in the following way:

For each $i \in \{1, \ldots, n\}$, a new slack parameter can be introduced as following way $\zeta_i = \max(0, 1 - y_i(w \cdot x_i - b))$. Note that ζ_i is the smallest non-negative number that satisfies $y_i(w \cdot x_i - b) \geq 1 - \zeta_i$. Then Equation (3.3) can be converted to the following primal problem.

$$\min_{w,b,\zeta} \frac{1}{2}w^T w + C\sum_{i=1}^{n}\zeta_i \tag{3.4}$$

Equation (3.4) is subject to

$$y_i(w.x_i + b) \geq 1 - \zeta_i,\ \zeta_i \geq 0,\ i = 1, ..., n \tag{3.5}$$

The above problem can be simplified by solving for the Lagragian dual and it takes the form as follows

$$\min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \tag{3.6}$$

Equation (3.6) is subject to

$$y^T\alpha = 0\ and\ 0 \leq \alpha_i \leq C,\ i = 1, ..., n. \tag{3.7}$$

where e is the vector of all ones, $C > 0$ is the upper bound, Q is an n by n positive semidefinite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, where $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is the kernel. The function ϕ is used to map training vectors implicitly into a higher dimensional space.

The final decision function for two classes is obtained as

$$\text{sgn}(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) - b) \tag{3.8}$$

In this thesis, SVM with Radial Basis Function (RBF) kernel is employed from skylearn python library which uses internally libsvm library for all computations. RBF has the following kernel function $\exp(-\gamma \|x - x'\|^2)$. Here γ called gamma in the python library must be greater than 0.

When training an SVM with RBF kernel, proper choice of regularization parameter termed as C and gamma should be taken into account. They affect the SVM's performance. C determines the trade-off between misclassification of training data points and simplicity of the system. Large values of C increase the training classification accuracy, while low values of C make the hyperplane surface more smooth, which increases the training errors. Gamma determines the influence of a single training point.

In the thesis, in order to optimize gamma and C values, sklearn's GridSearchCV optimization is used. Optimized C and gamma values were 20 and 0.01 respectively. The other SVM parameters are optimized for getting best results during training and validation. The degree of the system is determined as 5 and tolerance was 0.1

Table 3.1. Confusion Matrix of Wheeze Detection Algorithm

	Predicted 0s	Predicted 1s
Actual 0s	2791	293
Actual 1s	35	187

3.2. Results

Since the number of samples to train the SVM is relatively small, leave-one-out method is used to increase the performance of the algorithm. The SVM gives the direct label of the record and confusion matrix constructed from the output of the SVM is given in Table 3.1. In the table, 0s indicate that there is no wheeze in inspected window and 1s indicate that a wheeze exists in inspected window and numbers are window counts. The algorithm can be evaluated using following criteria:

$$Accuracy = \frac{TP + TN}{Total\ number\ of\ samples} \tag{3.9}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.11}$$

where TP is true positive which predicted label is correct for pathological sample, FN is false negative which predicted label is incorrect for a pathological sample, TN is true negative which predicted label is correct for a healthy sample, and FP is false positive which predicted label is incorrect for a pathological sample.

Accuracy determines how accurate the system is for detecting wheeze in a sample. Sensitivity refers to the algorithm's capability of correct identification of wheeze sample among others. Specificity is related to the algorithm's ability to correctly detect non-wheeze samples. During training, it is observed that specificity is high but sensitivity

is low since database contains a lot of non-wheeze records. A test with high specificity but low sensitivity causes to mislabel subjects who are pathological as healthy. In order to increase the sensitivity of the system class weights were adjusted. The accuracy of the system is calculated as 90%, sensitivity is 84% and specificity is 90%.

4. CRACKLE DETECTION

Crackles correspond to short explosive breath sounds and are usually associated with pulmonary disorders such as lung infection, pneumonia, pulmonary oedema [16]. Their presence is an important sign of pulmonary affections and disorders but it is not always an indication of pathological pulmonary process [16]. Crackles have a typical waveform which begins with a wide deflection followed by a long and damped sinusoidal wave [17]. A typical crackle's highest peak is greater than its background's amplitude by at least a factor of two [18]. Crackles are classified into two different types which are fine crackle and coarse crackle [19]. Initial deflection width (IDW) and two cyle duration (2CD), which are presented in Figure 4.1 [20], are used for determining type of a crackle. According to American Thoracic Society (ATS), fine crackles have 0.7 ms mean duration for IDW and 5 ms mean duration for 2CD, and coarse crackles have 1.5 ms mean duration for IDW and 10 ms mean duration for 2CD [21]. A typical crackle waveform is displayed in Figure 4.1 [20]. This typical waveform in a lung sound signal can be detected by computerized tools. In this thesis, an algorithm based on the article [22] has been applied to the records from crackle database to detect crackles in lung sound.

4.1. Methodology

In the thesis, two different approaches are used for detection of crackles in lung sound signals. The first method is analyzing the records as whole whereas the other method is analyzing small windows extracted from the whole record. Details of the steps are explained in the following subsections. This part is implemented using python's pylab, pywt and Tensorflow libraries.

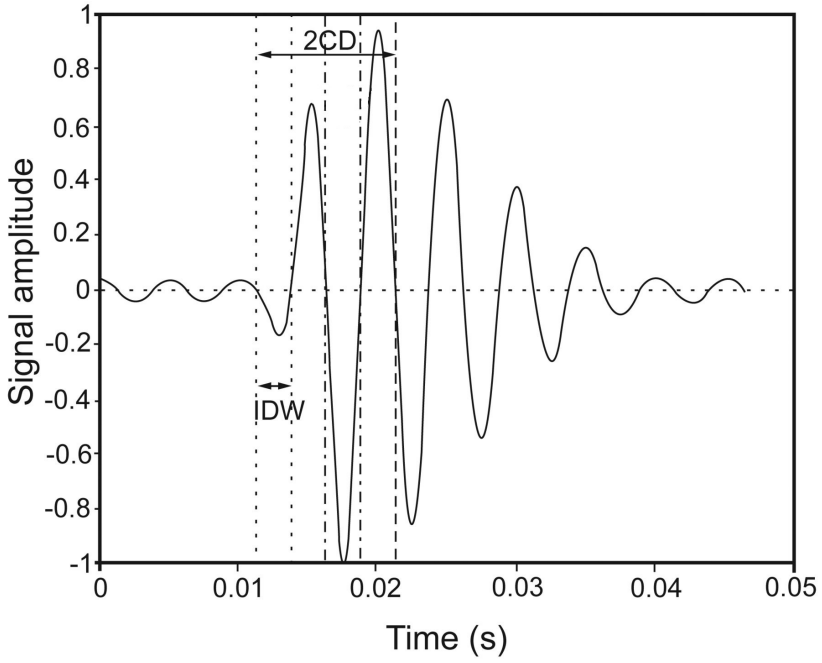


Figure 4.1. A typical crackle waveform

Applied steps for the first method are described as follows:

(i) AR parameters are calculated for the whole record.

(ii) Prediction Error Filter (PEF) with calculated AR parameters are applied to the signal.

(iii) Discrete Wavelet Transform (DWT) is applied to the signal coming from previous step to obtain decomposition bands.

(iv) Background suppression is applied to these decomposition bands.

(v) Inverse Discrete Wavelet Transform (IDWT) is applied to the combination of selected subbands.

(vi) Reconstructed signal is divided into smaller windows size of 50 ms samples with a 40 percent overlap.

(vii) Each window is labeled as either crackle or non-crackle based on their contents.

(viii) This signal is fed to an MLP.

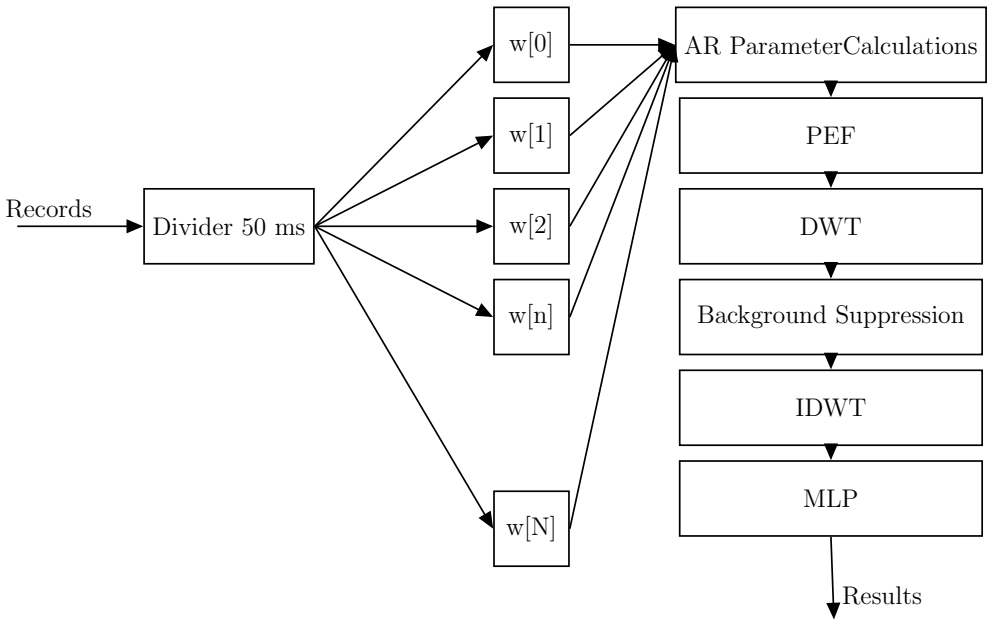


Figure 4.2. Block Diagram of Crackle Detection

Block diagram for second method is illustrated in Figure 4.1 and applied steps for the second method are described as follows:

(i) Lung sound records from the crackle database are divided into smaller windows size of 50 ms with a 40 percent overlap.
(ii) Each window is labeled as either crackle or non-crackle.
(iii) AR parameters are calculated for each window.
(iv) PEF with AR parameters is applied to the windows.
(v) DWT is applied separately to each window.
(vi) Background suppression is applied to produced subbands.
(vii) The signal is reconstructed by applying IDWT to the selected decomposition bands.
(viii) The signal is fed to an MLP.

4.1.1. AR parameters

It is assumed that lung sound is stationary and can be modeled by AR model [23]. If crackles exist in a sound portion then there will be deviations from the AR process.

Detecting these deviations helps us to detect crackles in that sound portion. AR model for each window of the record $s(n)$ is defined as

$$s^m(n) = -\sum_{k=1}^{p} a^m(k)s^m(n-k) + e^m(n),\ n = 1, 2, ..., N \tag{4.1}$$

where m is window index, N is window size, $a^m(1), ..., a^m(p)$ are parameters of the model and $e^m(n)$ is white noise.

Since 6th order AR model's ability of representing lung sound is relatively high [24], model order was chosen as 6 for representing the lung sound. The AR coefficients were computed automatically using Yule-Walker equations in python.

4.1.2. Predicted Error Filter

In order to suppress the background sound while preserving the transients in the signal, an all-zero PEF is applied to the signal. The output of the PEF is the prediction error, $e(n)$, between respiratory sound value $s(n)$ at time n and its predicted value $\hat{s}(n)$,

$$e(n) = s(n) - \hat{s}(n) \tag{4.2}$$

where $\hat{s}(n)$ is calculated with AR coefficients obtained from the first step,

$$\hat{s}(n) = -\sum_{k=1}^{P} a_k s(n-k) \tag{4.3}$$

where p is AR model order and a_k is AR model coefficient.

After that point, the stationary background is omitted from the signal while transients preserve their appearance in the record.

4.1.3. Wavelet Analysis

In signal processing field, wavelet analysis is generally used in analysis of transient signals. It has been used in many application areas such as bio-medical, speech, transient signal analysis, and communication systems due to its effectiveness in localizing a signal in time and frequency.

Wavelets are mathematical functions that are useful for representing signals or other functions and they are generated from a mother wavelet by dilation and translation operations. Like Fourier analysis, Wavelet analysis expands functions using a set of basis functions. Contrary to Fourier analysis, Wavelet analysis does not use trigonometric polynomials as basis functions. Instead, it uses wavelets as basis functions.

The Wavelet transform is a tool used for representing signals at different scales [25]. It has been used for analyzing transient events by decomposing the signal into different frequency subbands using a single prototype function, mother wavelet [26]. The mother wavelet, ψ, is a complex-valued function and satisfies the following conditions:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty. \tag{4.4}$$

$$c_\psi = 2\pi \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} dt < \infty. \tag{4.5}$$

where Ψ is Fourier transform of the mother wavelet and c_ψ is a substitution parameter which is used later.

Continuous wavelet transform of a real-valued signal $x(t)$ using mother wavelet that satisfies above conditions, $\psi(t)$, is defined as follow:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi^*(\frac{t-b}{a}) x(t) dt, \tag{4.6}$$

where ψ^* is the complex conjugate of ψ, and it is defined for $a \in R^+, b \in R$. The parameters a and b correspond to dilation and translation, respectively.

$\psi_{a,b}(t)$ can be defined as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}), \tag{4.7}$$

where the function is re-scaled by a and shifted by b, and Equation (4.6) can be rewritten in the form of

$$W(a,b) = \int_{-\infty}^{\infty} \psi^*_{a,b}(t)x(t)dt. \tag{4.8}$$

The signal $x(t)$ can be reconstructed from its wavelet transform $W(a,b)$ using following inverse formula:

$$x(t) = \frac{1}{c_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(a,b)\psi_{a,b}(t)\frac{1}{a^2}dadb \tag{4.9}$$

In the discrete case, discrete values are used for both dilation and translation parameters. They are represented as $a = a_0^m$ and $b = nb_0$ and the discrete mother wavelets are defined as

$$\psi_{m,n}(t) = a_0^{-m/2}\psi(\frac{t-nb_0}{a_0^m}), \tag{4.10}$$

where m and n are integer values. The DWT and IDWT are defined as follows:

$$W_{m,n} = \int_{-\infty}^{\infty} \psi^*_{m,n}(t)x(t)dt \tag{4.11}$$

$$x(t) = k_\psi \sum_m \sum_n W_{m,n}\psi_{m,n}(t) \tag{4.12}$$

where k_ψ donates a constant for normalization.

In the most common situation, dyadic scale analysis is applied by setting $a = 2^m$ and $b = n2^m$, which leads to orthogonal set of child wavelets to be obtained.

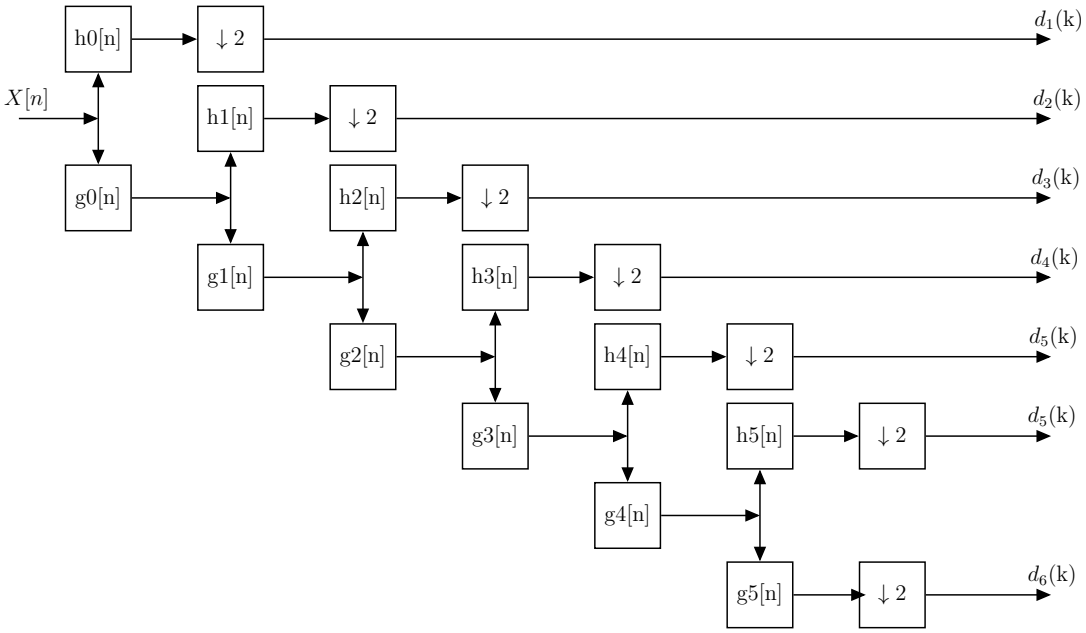


Figure 4.3. Discrete Decomposition

In the DWT, original signal, $x(n)$, is decomposed into low and high frequencies by passing it through a series of filters. First, the signal, $x(n)$, is passed through a low pass filter f_L and a high-pass filter f_H simultaneously. After the filtering, half the frequencies of the signal have been filtered out. Therefore down-sampling is applied to outputs of the filters to discard half the samples according to Nyquist's rule. The outputs of high-pass filter and low-pass filter give the detail coefficients and approximation coefficients respectively.

The output of the LPF, f_L in Figure 4.3 is then down-sampled by 2 and further processed by passing it again through a new LPF f_L and a new HPF f_H with half the cut-off frequency of the previous one. This process is repeated many times to increase the frequency resolution until it reaches the predetermined decomposition level. The decomposition of the signal, $x(n)$ into M subbands, can be described by the following formula:

$$DWT\{x(n)\} = \{\chi_0, \chi_1, \chi_2, ..., \chi_M\} \tag{4.13}$$

Table 4.1. Decomposition Levels

Level	Frequencies(n)	Sample Frequencies Hz	# of Samples
6	0 *to* $f_n/64$	0 to 75	32
	$f_n/64$ *to* $f_n/32$	75 to 150	32
5	$f_n/32$ *to* $f_n/16$	150 to 300	64
4	$f_n/16$ *to* $f_n/8$	300 to 600	128
3	$f_n/8$ *to* $f_n/4$	600 to 1200	256
2	$f_n/4$ *to* $f_n/2$	1200 to 2400	512
1	$f_n/2$ *to* f_n	2400 to 4800	1024

At each level in Figure 4.3 the signal is decomposed into low and high frequencies. Due to the decomposition process, the input signal must have a sample count as a multiple of 2^n where n is the number of levels.

In Table 4.1, a signal with 1024 samples, frequency range of 0 to f_n(4800Hz) and 6 levels of decomposition, and 7 output scales are represented.

The signal, $x(n)$, can be reconstructed from the decomposition subbands, χ_m by applying IDWT, i.e.

$$x(n) = IDWT\{\chi_0, \chi_1, \chi_2, ..., \chi_M\} \tag{4.14}$$

The rationale behind using wavelet transform is that transients can be filtered at different levels and background suppression can be applied to each decomposition bands to increase the transient to background ratio.

In this thesis, Daubechies 6 wavelet is used for decomposition due to its significant resemblance to a crackle waveform [27]. A typical Daubechies 6 wavelet is shown in Figure 4.4. The decomposition level is selected as 6 and therefore, 7 decomposi-

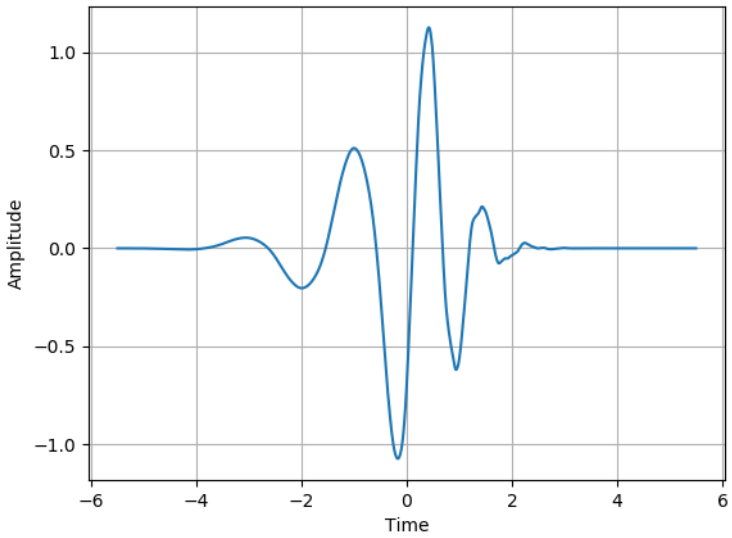


Figure 4.4. Daubechies 6 Waveform

tion subbands $(d_0, d_1, ..., d_7)$ are produced using python's pywt library. In Table 4.1, frequency range of the decomposition subbands can be seen.

4.1.4. Background Suppression

In order to increase the ratio of transients to the background, Teager's Energy Operator is applied to each decomposition band. It is used to attenuate background respiratory sound. For discrete signal $x(n)$, Teager's energy function x', is defined as

$$x'(n) = x^2(n) - x(n+1)x(x-1) \tag{4.15}$$

Teager's energy function for each m wavelet component can be calculated as follows

$$\chi'_m(k) = \chi^2_m(k) - \chi_m(k+1)\chi_m(k-1) \tag{4.16}$$

After using Teager's Energy Operator, steep changes and peaks are amplified whereas background sound is suppressed. Here, it is assumed that after applying Teager's Energy Operation, remaining energy function can be separable into background sound and transient. The next step is employing Otsu Thresholding to determine threshold values T_m to set the background to zero. Otsu's method is used to separate given data into two different classes.

The threshold values, T_m, are empirically calculated in python and applied as follows

$$\chi''_m(k) = \left\{ \begin{array}{cc} \chi'_m(k) & \text{if } \mid \chi'_m(k)\mid > T_m \\ 0 & \text{otherwise} \end{array} \right\} \tag{4.17}$$

4.1.5. Reconstruction

In this step, IDWT is applied to selected thresholded energy functions, χ''_m, $m = 1, 2, 3, 4$ to construct a new signal, $\hat{e}(n)$, with background suppression. Here, $\chi''_0, \chi''_5, \chi''_6$ are ignored since they do not have any significant contribution to crackle waveform. The constructed signal highlights potential locations of transients in the original signal. $\hat{e}(n)$ can be symbolically defined as follows:

$$\hat{e}(n) = IDWT\{\chi''_1, \chi''_2, \chi''_3, \chi''_4\} \tag{4.18}$$

$\hat{e}(n)$ is further processed by passing it through an all-pole filter (APF) with the same parameters as the PEF to recover the original shape of the transients. The APF works as inverse of PEF, and obtained new signal $\hat{x}(n)$, from $\hat{e}(n)$ is shown as follows:

$$\hat{x}(n) = -\sum_{k=1}^{P} a_k \hat{x}(n-k) + \hat{e}(n) \tag{4.19}$$

After that point, $\hat{x}(n)$ is assumed to preserve its transients while its background is removed. A respiratory sound signal and applied steps to it are illustrated in Figure 4.5.

4.1.6. MLP

MLP is a feedforward artificial neural network structure consisting of at least one hidden layer of perceptrons which are basic processing elements of the structure [28]. Each perceptron has the same nonlinear activation function to process its inputs coming from the inputs of the structure or outputs of the other perceptrons in the previous layer. A perceptron, which has nonlinear sigmoid activation function, can be modeled as follows:

$$y = \frac{1}{1 + exp\left[\sum_{j=1}^{d} w_j x_j + b\right]} \tag{4.20}$$

where y is the output of the perceptron, d is the number of inputs, w is connection weight and b is bias. This equation can be written as a dot product.

$$o = \mathbf{w}^T \mathbf{x} \tag{4.21}$$

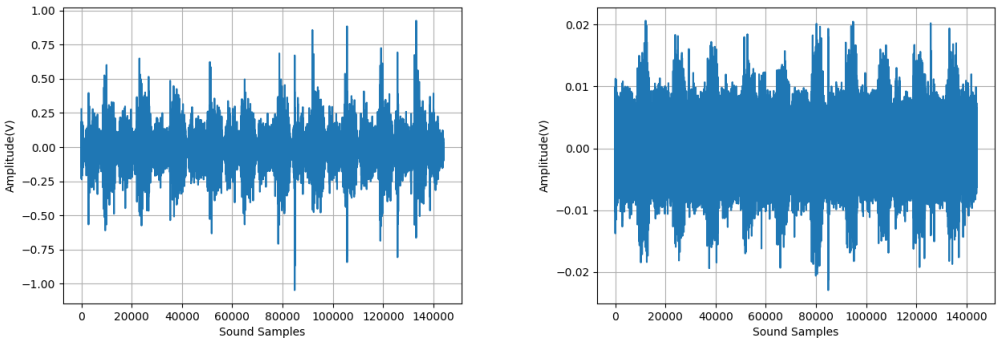
$$y = \frac{1}{1 + exp[-o]} \tag{4.22}$$

where $\mathbf{w} = [w_0, w_1, ..., w_d]^T$ and $\mathbf{x} = [x_0, x_1, ...x_d]^T$ are weight vector and input vector respectively with bias weight and input. When training a perceptron for two classes case, cross entropy is calculated as follows:

$$E^t(\mathbf{w}|\mathbf{x^t}, r^t) = 1/2(r^t - y^t)^2 = 1/2\left[r^t - (\mathbf{w^T x^t})\right]^2 \tag{4.23}$$

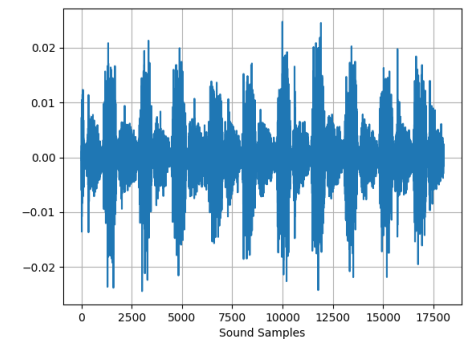
where t is instance index, $r^t = 1$ if $x^t \in C_1$ and $r^t = 0$ if $x^t \in C_2$. When the activation function is sigmoid, Equation (4.23) can be written as

$$E^t(\mathbf{w}|\mathbf{x^t}, r^t) = -r^t \log y^t - (1 - r^t) \log(1 - y^t) \tag{4.24}$$

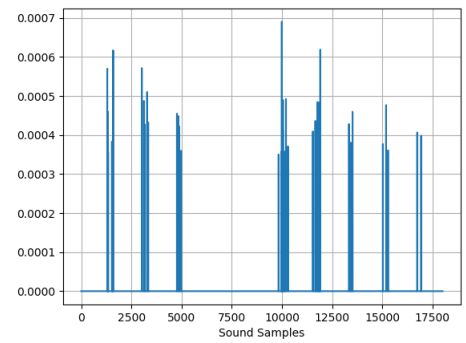


(a)

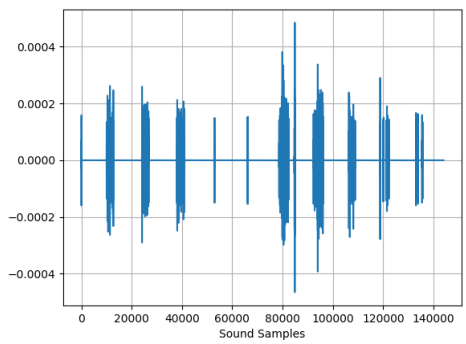
(b)



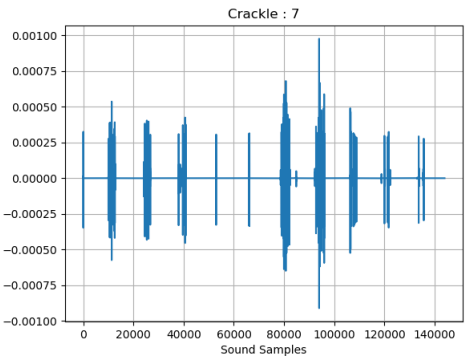
(c)



(d)



(e)



(f)

Figure 4.5. Crackle detection algorithm steps: (a) a respiratory sound signal; (b) the error signal of the previous signal; (c) DWT of the error signal for decomposition subband-3, χ_3; (d) Teager operated and thresholded signal, χ_3''; (e) reconstructed signal from selected subbands; and, (f) output signal of APF.

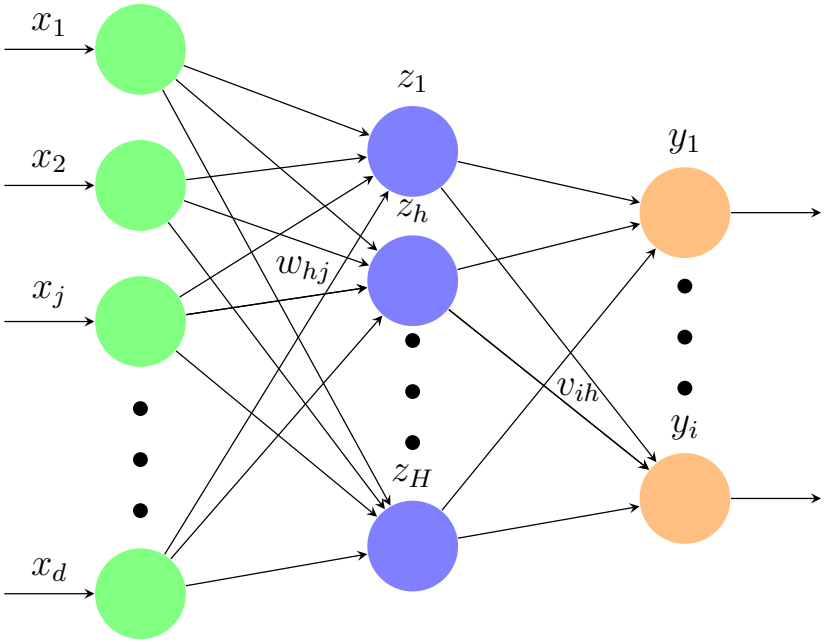


Figure 4.6. 3 Layer MLP structure

The cross-entropy is minimized iteratively using gradient descent and weight update rule for $j = 0, ..., d$ is obtained as

$$\Delta w_j^t = \eta(r^t - y^t)x_j^t \tag{4.25}$$

where η denotes learning rate.

MLP is a supervised learning method which is used for nonlinear classification and regression problems. Hornik et al. [29] have proved that an MLP with one hidden layer can learn any nonlinear function of the input. Structure of 3 layer MLP is shown in Figure 4.6. MLP utilizes back propagation algorithm for training the network. In order to adjust weight of the perceptrons, the algorithm computes the gradients of the error function. Error function over the whole sample for 2 classes case is calculated as

$$E(\mathbf{W}, \mathbf{v}|X) = -\sum_t r^t log y^t + (1 - r^t)\log(1 - y^t) \tag{4.26}$$

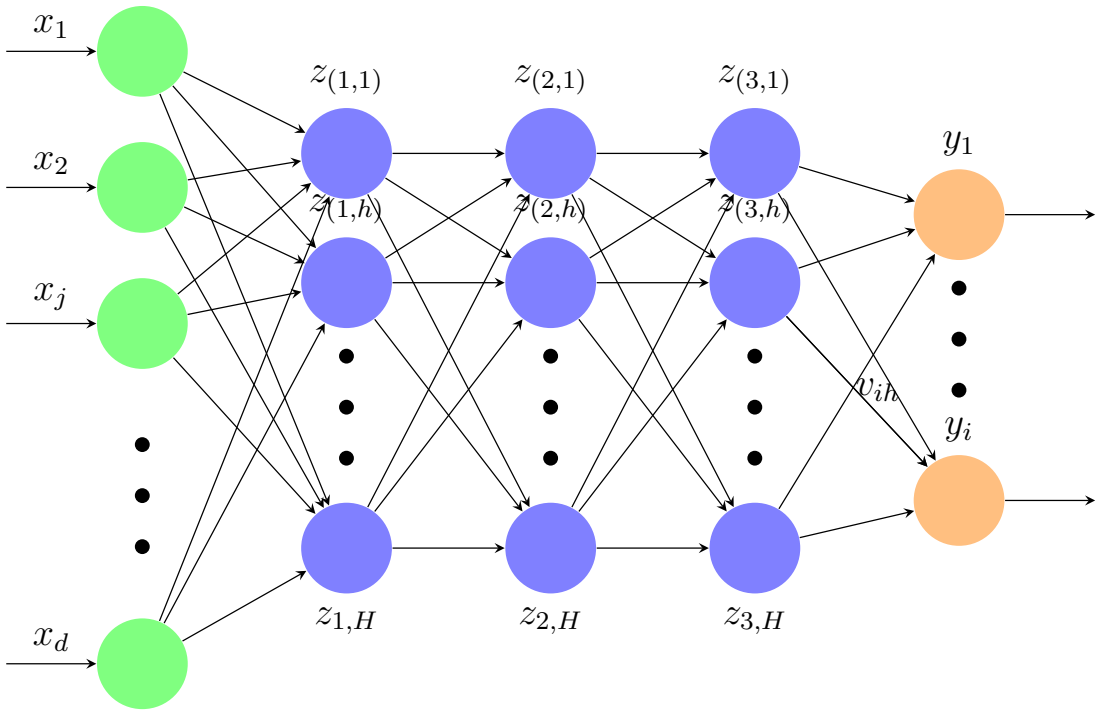


Figure 4.7. 5 Layer MLP structure used in the algorithm

and weight update rule for weights on the last layer is

$$\Delta v_h = \eta \sum_t (r^t - y^t) z_h^t \tag{4.27}$$

where z_h^t is the output of perceptron at node z_h, and for hidden layer update rule is calculated as:

$$\Delta w_{hj} = \eta \sum_t (r^t - y^t) v_h z_h^t (1 - z_h^t) x_j^t \tag{4.28}$$

In this thesis, Google's Tensorflow library is used to implement MLP to classify the reconstructed signal portion as pathological or healthy. In the second method, reconstructed signal is divided into small windows to feed MLP. However, since in the first method, the signal has already been divided, there is no need to divide the signal into windows. Throughout training of the network, many different structures with three hidden layers to eight hidden layers have been tested. The number of perceptrons in

Table 4.2. Confusion Matrix of Crackle Detection Algorithm

	Predicted 0s	Predicted 1s
Actual 0s	17270	2357
Actual 1s	649	349

the input layer, which is equal to the length of feeding window, is equal to 468, and the number of the output layer is equal to 2. These two layers are always fixed in all of the tested structures. The structure that gives more accurate results than the other tested structures can be depicted in Figure 4.7.

4.2. Results

The results of the crackle detection algorithm are below the expected targets. Our expectation was that accuracy, sensitivity and specificity of the system would be about 80% each. However, after training the structure with preprocessed data, results are far from the expectation. The accuracy is found as 85.4%, sensitivity is 35% and specificity is 88%. Although these values are not good enough to justify the use of the algorithm for crackle detection, they will be used as features for the final classifier. Confusion matrix of the structure is given in Table 4.2.

5. DISEASE DETECTION

It is known that the acoustic characteristics of normal lung sound are quite different from those of pathological lung sound [6]. Apart from wheeze and crackle, the spectral content of pathological lung sounds is different from normal lung sounds. In order to differentiate pathological lung sound from normal lung sound, a new disease detection algorithm is developed.

In the following sections, techniques and methods used in this algorithm are explained.

5.1. Methodology

In order to detect the presence of any pathological condition in a respiratory sound record, CNN is fed with preprocessed data coming from Electrosalus's database containing records of healthy subjects and pathological subjects. Block diagram of the detection process is shown in Figure 5.1.

Applied steps are described as follow:

(i) Lung sound records from the Electrosalus's database are labeled as healthy or pathological.
(ii) STFT is applied to each record.
(iii) The signal is fed to CNN for training and testing.

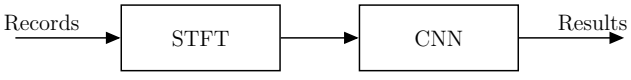


Figure 5.1. Disease detection algorithm block diagram

5.1.1. STFT

STFT is used for time-frequency representation (TFR) of a signal showing the frequency content of the signal in each short time interval by taking Fourier transform of each short time intervals of the longer signal. There are many signal analysis and detection methods based on STFT in literature [30,31]. In the thesis, CNN is employed to detect the presence of any pathological condition in a lung sound record and it requires an image like data input. Since lung sound records are time series data, STFT is applied to this lung sound record to obtain an image like data representation which is given to the CNN as input.

In order to compute STFT of a signal, $x(t)$, is first divided into smaller equal length segments and multiplied with a specific windowing function such as Hann window, Hamming window and Tukey window. Then, Fourier transform is applied to each segment. The STFT for continuous-time signals is calculated as follows:

$$\mathbf{STFT}\{x(t)\}(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t}\,dt \tag{5.1}$$

where τ is time index and ω is angular frequency, $w(\tau)$ is the window function, $X(\tau,\omega)$ is basically the Fourier Transform of $x(t)w(t-\tau)$. For a discrete-time, signal $x[n]$, it is calculated as follows:

$$\mathbf{STFT}\{x[n]\}(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \tag{5.2}$$

where $w[m]$ is the discrete window function as in the continuous-time, m and w are discrete time index and quantized.

STFT is performed in the computers using fast Fourier transform. Therefore its computation cost is insignificant. Segments overlap each other to retrieve transients between the windows. On the other hand, the main drawback of the STFT is that resolution stays the same at all locations.

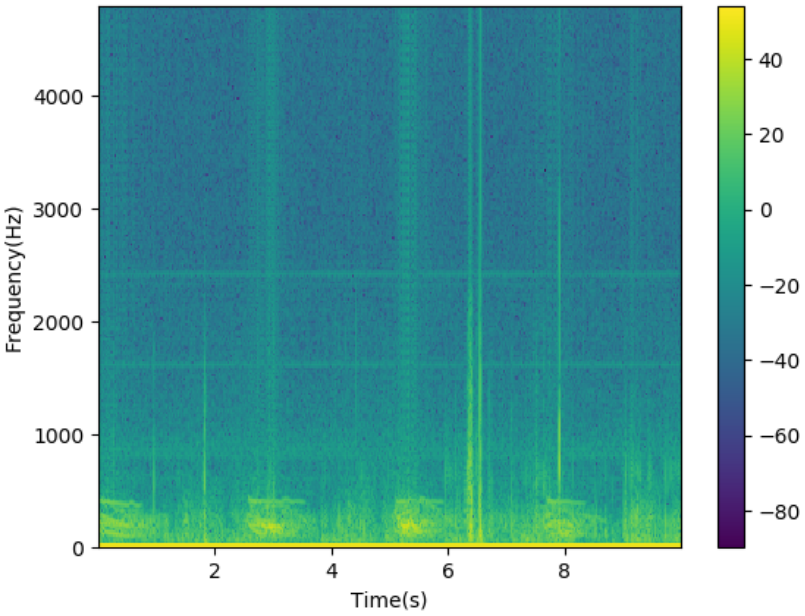


Figure 5.2. Spectrogram of a record

The spectrogram of the signal extracted from the magnitude squared of the discrete STFT shows the magnitude of frequency vs. time graph and is defined as

$$\text{spectrogram}\{x[n])\}(m,\omega) \equiv \left| \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \right|^2 \tag{5.3}$$

In the thesis, STFT is calculated using Hamming window function whose window length is 120 ms samples, the number of Fourier Transform points is 1920, and the overlapping ratio is 50%. These values are chosen for the best results obtained from Electrosalus's database after training and testing the data. Spectrogram of a record is shown in Figure 5.2.

5.1.2. CNN

CNNs are feedforward deep artificial neural networks comprised of at least one convolutional layer with pooling part to extract features from very large receptive fields and they are generally used for classification of images and object recognition in the scenery. For example, they are used for face recognition and natural language processing. Like MLP, CNN is also supervised learning technique and utilizes a back-propagation algorithm for training its network. It is mostly applied to visual data. However, sound data can also be represented visually using spectrogram and CNN can be applied to this visualization. Since this visualization has some distinctive features for detecting any pathological condition and CNN has a growing potential for evaluating visual data, in this thesis, CNN is chosen as a machine learning technique to be applied to STFT of the lung sound data for classification between a healthy subject and a pathological subject. In the Tensorflow, CNN can be represented as in Figure 5.3 [32].

CNNs require relatively minimal pre-processing [33], which is achieved with local connections and tied weights followed by some form of pooling. Therefore CNN learns the filters on its own. This makes it easy to train the network and reduces the number of parameters computed in CNN compared to MLP with the same number of hidden layers. In this thesis, Tensorflow's CNN library is used for disease detection over Electrosalus's database. Tensorflow was chosen for its ability to utilize GPU, which decreases the training time significantly and its popularity among researchers.

CNN is designed to learn directly and independently spatial hierarchies of features from data using patterns and CNN eliminates the need for manual feature extraction. A CNN architecture typically includes several stacks of three types of layers which are convolution, pooling, and fully connected layers. The first layer is convolution layer and it consists of a combination of linear and nonlinear mathematical operations, such as convolution and activation operations. The second layer is pooling layer where down-sampling operations are performed in order to to introduce a translation invariance to small shifts and distortions, and reduce the number of learnable parameters. In this layer, filter size, stride, and padding are used as variable to adjust down-sampling

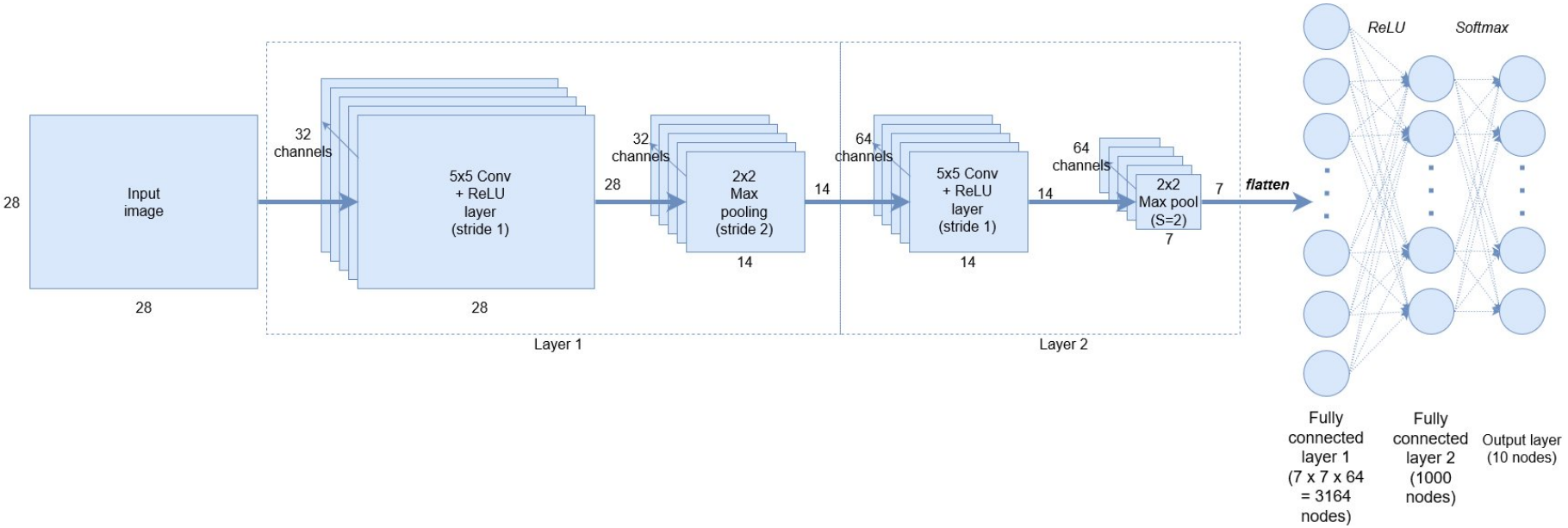


Figure 5.3. Typical Tensorflow CNN structure

operation. The third layer is a fully connected layer and generally used to map the extracted features into final classifier. Fully connected layers are generally followed by a nonlinear function, such as rectified linear unit (ReLU) defined as $f(x) = max(0, x)$...

Feature extraction is mostly performed in the first two layers. These layers is where forward propagation occurs. Kernel which is a small grid of parameters is an optimizable feature extractor and a filter applied at each image position. Backpropagation algorithm is used to minimize loss function or cost by optimizing parameters, kernels in convolution layers and weights in fully connected layers.

The CNN is implemented in python with a six hidden-layer network, shown in Figure 5.4. Constructed structure consists of 2 convolutional layers, 2 pooling layers, 2 dense layers and 1 dropout layer. In the Figure 5.4, $conv2d$ indicates two dimensional (2D) convolutional operation and $conv1$ and $conv2$ are convolutional layers obtained from $conv2d$ operation, $pool1$ and $pool2$ are pooling layers, *dense* is hidden dense layer, *dense_1* is output dense layer, *dropout* represents the dropout layer, and *enqueue_in* is randomly shuffled queued input and $softmax\_cross\_entropy$ gives the final result.

The $conv1$ layer has 32 filters and uses 8x8 convolution kernels with stride 5. Its activation function is a ReLU. The $pool1$ layer has a 5×5 kernel with stride 2. The $conv2$ layer has 64 filters and uses 10x10 convolutional kernel with stride 1. The $pool2$ layer has 2x2 kernel with stride 2. The *dense* layer has 128 nodes and uses ReLU. And the dropout rate is equal to 0.2. Finally, the output layer with 2 nodes takes inputs from the dropout layer and utilizes a softmax function for classification.

5.2. Results

The same metrics as in previous chapters, which are accuracy, sensitivity, and specificity, are used to evaluate the success of the proposed structure. Confusion matrix of the structure is given in Table 5.1. In the table, 0s indicate that evaluated record is healthy and 1s indicate that pathological condition exists in the evaluated record and numbers refer to lung sound record counts. Since the database is relatively small

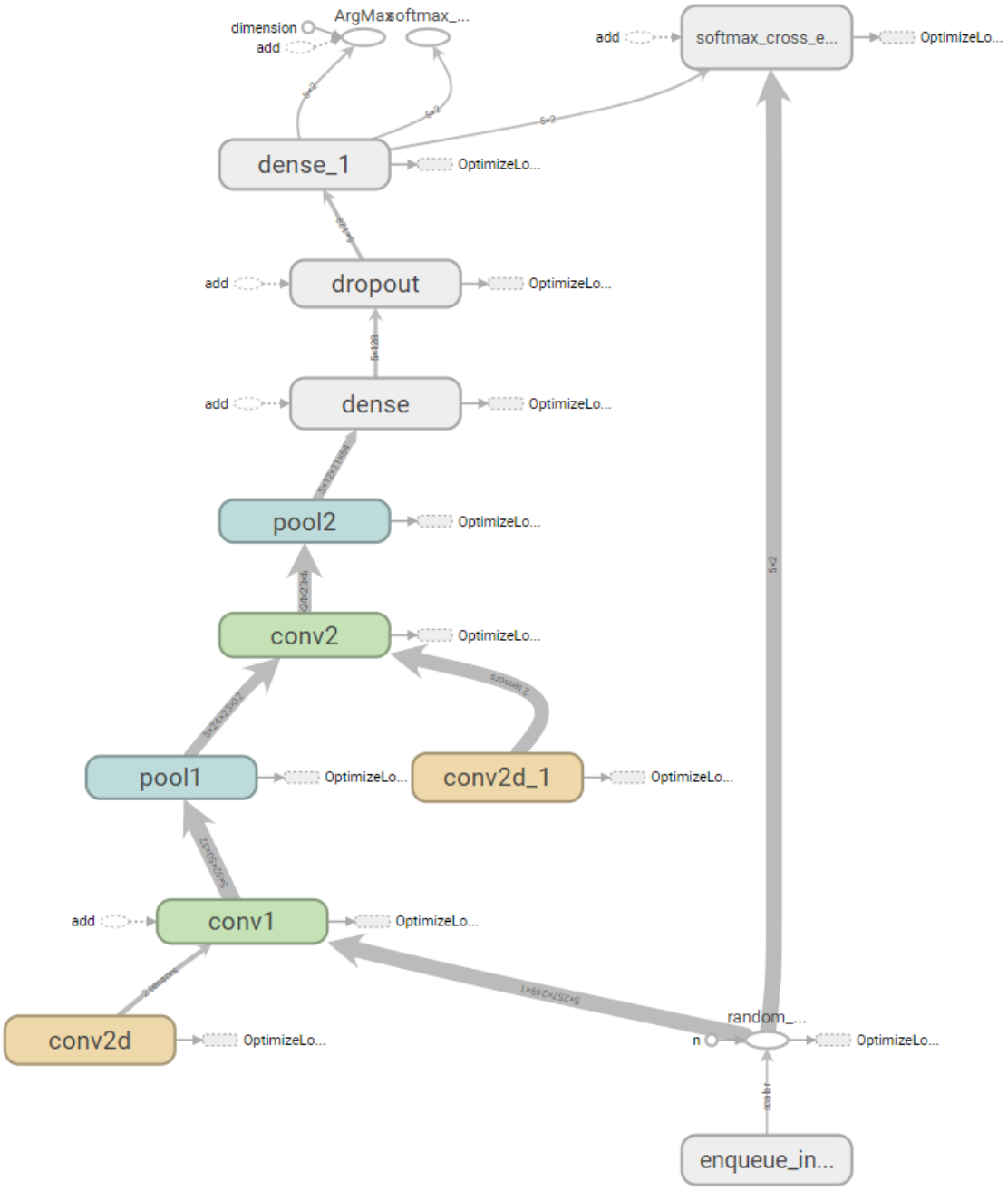


Figure 5.4. Tensorboard CNN structure

Table 5.1. Confusion Matrix of Disease Detection Algorithm

	Predicted 0s	Predicted 1s
Actual 0s	16	8
Actual 1s	4	36

to train and test the proposed network, leave-one-out method is used to increase the performance. Leave-one-out method is a time-consuming method. Even though GPU was used to train network, complete run over the database takes 12.43 hours. The accuracy of the system is calculated as 81.25%, sensitivity is 90% and specificity is 66.67%.

6. CLASSIFICATION

Determining whether a subject is healthy or pathological is very important to avoid unnecessary further examinations, which take time and sometimes cause the subject to be exposed to radiation, or to avoid taking precautions before it is too late and to help doctors diagnose the subject's diseases. Also it is a very useful approach in mass screening of population like school children , army soldiers etc. For this purpose, output of the three algorithms which are wheeze detection, crackle detection and disease detection, is combined to produce more accurate results.

In the following sections, techniques and methods used in this algorithm are explained.

6.1. Methodology

Block diagram for the classification is illustrated in Figure 6.1 and all applied steps are described as follows:

(i) Each record is resampled and reshaped.
(ii) AR model parameters are calculated for each record.
(iii) Percentile frequencies are computed for each record.
(iv) Wheeze detection algorithm is applied to each record.
(v) Crackle detection algorithm is applied to each record.
(vi) Disease detection algorithm is applied to each record.
(vii) All features from above steps are fed into an SVM for final classification.

A variety of preprocessing steps are needed to meet the sampling rate and duration requirements of the classification for a record. First, if a record has not been sampled with a rate of 9600 sps, then sample-rate conversion would be applied to the record to obtain a signal with a sample rate of 9600 sps. Besides, the record must be 10 second long. If its duration is longer then only the first 10 seconds part is taken to

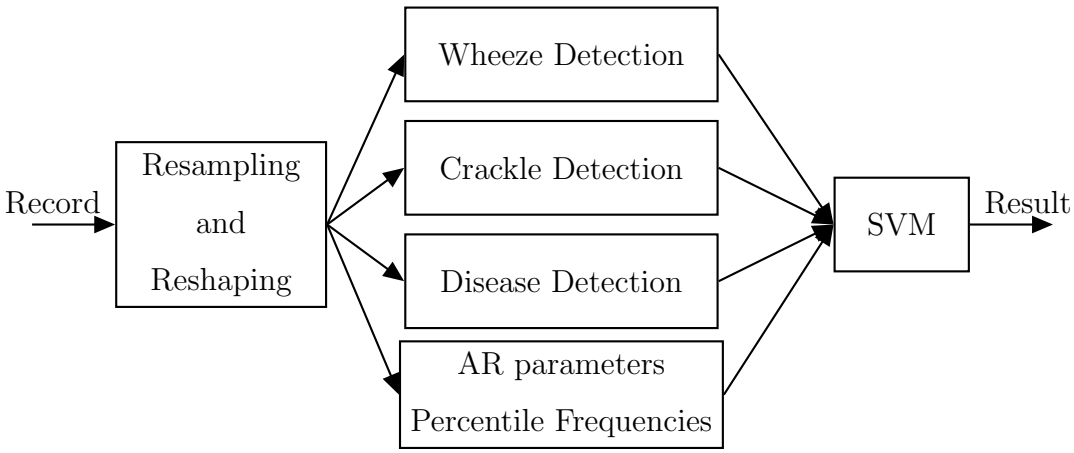


Figure 6.1. Block diagram for classification

form a 10 second long signal. If it is shorter then it is not used in the algorithm since the disease detection algorithm is trained for 10 second long signals. All computations are made using python since it provides many libraries needed for calculations and it is a fast, simple, open source platform to develop programs.

Sixth order AR model parameters are calculated using Yule-Walker equations in python. It is expected that AR parameters of normal lung sounds are different from those of pathological sounds. Detecting this difference may help to classify a respiratory sound record. Window size for AR parameter calculation is chosen as 200 ms. In this step, 6 different coefficients are produced and they are used as features for the final classification in the SVM.

Percentile frequencies are also expected to differ from normal lung sounds to pathological lung sounds. They are computed for each record using Welch method where window length and number of Fourier Transform points are 1920, overlapping ratio is 25% and Hamming window is used. $f_{25}, f_{50}, f_{75}, f_{90}$ percentile frequencies are calculated and later used as features in the SVM classifier.

Presence of a wheeze is an indication of many respiratory disorders. Therefore, it is important to detect whether any wheeze exists in a record. For this purpose, wheeze

detection algorithm is applied to each record. Since wheeze detection algorithm is used only for small windows, first, a record is divided into small windows and the algorithm is applied. Predicted labels are summed to obtain the total number of wheeze present in the processed record. This is another feature for the final classifier.

Since crackles correspond directly to some pulmonary infections and disorders, it is expected that determining their number in a record helps to improve the performance of the final classifier [16]. However, due to the very low sensitivity of the algorithm, predicted number of crackles available in the processed record is not as important feature as the others.

It is assumed that any respiratory disorder causes a change in the characteristics of the lung sound. Therefore, the disease detection algorithm is applied to records. Although it has already labeled a record as pathological or healthy, the final classifier uses this label as an additional feature to increase both accuracy and sensitivity of the overall system.

After all features are calculated, they are fed into an SVM. SVM is chosen for its ability to predict nonlinear relations among the features and its easier implementation. Leave-one-out method is used for training and testing of the SVM and SVM parameters are optimized for the best performance. Electrosalus's database, which is also used in disease detection algorithm, is used for evaluation of the proposed structure.

6.2. Results

Electrosalus's database which contains records labeled as healthy or pathological was used in the experiments to observe any improvement on the results of the disease detection algorithm. In order to evaluate the success of the final algorithm, accuracy, sensitivity and specificity are employed as in previous chapters. Confusion matrix of the structure is given in Table 6.1. In the table, 0s indicate that evaluated record is healthy and 1s indicate that pathological condition exists in the evaluated record and numbers refer to lung sound record counts. The accuracy is found as 82.8%,

Table 6.1. Confusion Matrix of General Disease Detection Algorithm

	Predicted 0s	Predicted 1s
Actual 0s	17	7
Actual 1s	4	36

sensitivity is 90% and specificity is 70.8%. Accuracy, sensitivity, and specificity of the final classifier are increased by 2%, 0%, and 6.2% respectively over the disease detection algorithm's performance metrics. From these results, it is concluded that there is a small improvement on the evaluation metrics and this improvement may be greater as the number of respiratory sounds used in the classification experiments is increased.

7. CONCLUSION

In this thesis, characteristics of respiratory sounds and differences between pathological sounds and normal pulmonary sounds are investigated. The importance of non-subjective diagnosis techniques for pulmonary diseases is stated. Then proposed machine learning based electronic lung disease detection system is explained.

The main purpose of this thesis is to develop a new noninvasive method to detect whether a pathological condition exists in a respiratory sound record which is recorded on the posterior chest wall of a subject. For this objective, a system is designed and implemented for analyzing recorded lung sound data. This system contains several different techniques such as digital signal processing methods, statistical analysis, and various machine learning algorithms. Constructed system takes a record from a user and then processes the record and finally gives its result that labels the record as healthy or pathological.

In this thesis, the lung disease detection system consisting of different algorithms each of which is specialized in different subject is introduced. These algorithms are wheeze detection algorithm, crackle detection algorithm and disease detection algorithm. They are explained in the related chapters in detail. Moreover, data sets which were used in training and validation of these algorithms are stated. Characteristics of these data sets and acquisition systems are explained in detail.

In chapter 3, wheeze detection algorithm and its steps are discussed. It is applied to the lung sound records to determine number of wheezes in the record. Kurtosis, Renyi Entropy, f_{50}/f_{90} ratio, and Zero Crossing Count are calculated for each record and they together construct feature vector of the algorithm. The algorithm is trained and tested on wheeze data set and its results are given in chapter 3. The accuracy of the system is found as 90%, sensitivity is 84% and specificity is 90%.

Crackle detection algorithm is used to find the number of crackles in the processed lung sound records. Two different approaches, whose steps are defined in chapter 4, is used to implement the algorithm. Several digital signal processing techniques and MLP are employed in the algorithm. This algorithm is trained and tested over crackle data set. This algorithm gives estimated number of crackles in a record. Validation results are also given in chapter 4 and the accuracy is calculated as 85.4%, sensitivity is 35% and specificity is 88%. Since the sensitivity of the algorithm is very low, it is not feasible to use it as a disease indicator. In order to improve its performance, thresholding method may be improved and more data is needed.

Physicians use a stethoscope to find out whether any adventitious sound exists in respiratory sound of a patient. Based on their observations, they will decide what to do in the next step. In addition to disease labels, by using the proposed system, numbers of wheezes and crackles which are detected in the wheeze detection part and crackle detection part are also provided to these physicians to help diagnosis of disease if a pathological condition exists.

Apart from adventitious sounds, the spectral content of pathological lung sounds is different from normal lung sounds. In order to observe this phenomenon , a new disease detection algorithm is developed and is presented in chapter 5. This algorithm calculates STFT of a record and then feeds it to the CNN, which was trained over pathological data set. Finally, it labels the record as healthy or pathological. Performance of the algorithm is given in the related chapter. The accuracy of the system is measured as 81.25%, sensitivity is 90% and specificity is 66.67%. In the pulmonary sound database, healthy records constitute less than 40%. Increasing the healthy sound records may increase the specificity of the algorithm.

Finally, the overall classification algorithm is presented. It is used to combine outputs of the previous three algorithms along with AR parameters and percentile frequencies of the records to produce the final result. It uses the SVM as a final classifier and labels a record as healthy or pathological. Test results of the algorithm are given in chapter 6. Its performance metrics are slightly better then the disease

detection algorithm. Its accuracy is found as 82.8%, sensitivity is 90% and specificity is 70.8%. In order to improve the performance metrics of this final classifier, performance improvements in the other algorithms, especially in the crackle detection algorithm are needed. Moreover the data set needs to be expanded.

To sum up, a new electronic lung disease detection algorithm is designed and implemented and its success is evaluated. This system can be used for determining the presence of any respiratory disorder in a subject and eliminating unnecessary further examinations. For future studies, database size may be increased, and a new method may be employed for crackle detection algorithm to improve the success of the designed system.

REFERENCES

1. H. Pasterkamp , S. S. Kraman and G.R. Wodicka, "Advances Beyond the Stethoscope", *American Journal of Respiratory and Critical Care Medicine*, Vol. 156, pp. 974–987, September 1997.

2. L. Vanuccini, J. E. Earis, P. Helisto, B. M. G. Cheetham, M. Rossi, A. R. A. Sovijarvi and J. Vanderschoot, "Capturing and Preprocessing of Respiratory Sounds", *Europian Respiratory Review*, Vol. 10, No. 77, pp. 616–620, 2000.

3. Arati Gurung, Carolyn G Scrafford, James M Tielsch, Orin S Levine, and William Checkley, "Computerized Lung Sound Analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis", *Respiratory Medicine*, Vol. 105, pp. 1396–1403, September 2011.

4. Sandra Reichert, Raymond Gass, Christian Brandt, and Emmanuel Andrès, "Analysis of Respiratory Sounds: State of the Art", *Clin Med Circ Respirat Pulm Med.*, pp. 45–58, 2008.

5. A. R. A. Sovijarvi, L. P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, J. E. Earis, "Characteristics of Breath Sounds and Adventitious Respiratory Sounds", *Europian Respiratory Review*, Vol. 10, No. 77, pp. 591–596, 2000.

6. A. R. A. Sovijarvi, F. Dalmasso, J. Vanderschoot, L. P. Malmberg, G. Righini and S. A. T. Stoneman, "Definition of Terms for Respiratory Sounds", *Europian Respiratory Review*, Vol. 10, No. 77, pp. 597–610, 2000.

7. L. P. Malmberg, A. R. Sovijarvi, E. Paajanen, P. Piirila, T. Haahtela, and T. Katilla, "Change in frequency spectrum of breath sounds during histamine challenge test in adult asthmatics and healthy control subjects", *Chest*, Vol. 105, No. 1, pp. 122–131, January 1994.

8. J. E. Earis and B. M. G. Cheetham, "Feature Perspective for Respiratory Sounds", *Europian Respiratory Review*, Vol. 10, No. 77, pp. 641–646, 2000.

9. Mete Yeginer and Yasemin P. Kahya, "Feature extraction for pulmonary crackle representation via wavelet networks", *Computers in Biology and Medicine*, Vol. 39, pp. 713–721, August 2009.

10. R. L. Murphy, S. K. Holford, and W. C. Knowler, "Visual lung-sound characterization by time-expanded wave-form analysis", *N Eng J Med*, Vol. 296, No. 17, pp. 968–971, April 1994.

11. Sandra Reichert, Raymond Gass, Christian Brandt, and Emmanuel Andrès, "Analysis of Respiratory Sounds: State of the Art", *Clinical Medical Insights: Circulatory, Respiratory and Pulmonary Medicine*, pp. 45–58, 2008.

12. Sezer Ulukaya, Ipek Sen, and Yasemin P. Kahya, "Feature Extraction Using Time-Frequency Analysis for Monophonic-Polyphonic Wheeze Discrimination", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2015)*, August 2015.

13. Ipek Sen and Yasemin P. Kahya, "A multi-channel device for respiratory sound data acqusition and transient detection", *Engineering in Medicine and Biology Society (EMBS)*, pp. 6658–6661, 2005, 27th Annual International Conference of the IEEE.

14. Taplidou SA, Hadjileontiadis LJ, Kitsas IK, Panoulas KI, Penzel T, Gross V, and Panas SM, "On Applying Continuous Wavelet Transform in wheeze analysis", *Conf Proc IEEE Eng Med Biol Soc.*, Vol. 5, pp. 3832–5, 2004.

15. Sergul Aydore, Ipek Sen, Yasemin P. Kahya and M. Kivanc Mihcak , "Classification of Respiratory Signals by Linear Analysis", *31st Annual International Conference of the IEEE EMBS*, September 2-6, 2009.

16. P. Piirilä, A.R.A. Sovijärvi, "Crackles: recording, analysis and clinical significance", *ERS Journals Ltd*, 1995.

17. Sandra Reichert, Raymond Gass, Amir Hajjam, Christian Brandt, Emmanuel Nguyen, Karine Baldassari, and Emmanuel Andres, "The ASAP project: A first step to an auscultation's school creation", *Respiratory Medicine CME*, pp. 7–14, 2009.

18. Murphy, R., Del Bono, E. and Davidson, F., "Validation of an Automatic Crackle (Rale) Counter", *American Review of Respiratory Disease*, Vol. 140(4), pp. 1017–1020, 1989.

19. Holford SK., "Discontinuous adventitious lung sounds: measurement, classification and modeling.", *Doctoral thesis. Cambridge; Massachusetts Institute of Technology*, 1982.

20. Bennett, S., Bruton, A., Barney, A., Havelock, T. and Bennett, M., "The Relationship Between Crackle Characteristics and Airway Morphology in COPD", *Respiratory Care*, Vol. 60(3), pp. 412–421, 2014.

21. "American Thoracic Society. Updated nomenclature for membership reaction. Reports of the ATS-ACCP ad hoc committee", *Am Thorac Soc News*, Vol. 3, pp. 5–6, 1977.

22. Bulent Sankur, E. Cagatay Guler, and Yasemin P. Kahya, "Multiresolution Biological Transient Extraction Applied to Respiratory Crackles", *Comput. Biol. Med*, Vol. 26, No. 1, pp. 25–39, 1996.

23. Bulent Sankur, Yasemin P. Kahya, E. Cagatay Guler, and Tanju Engin, "Comparison of AR based classifiers for respiratory sounds classification", *Comput. Biol. Med*, Vol. 24, No. 1, pp. 67–76, 1994.

24. Yasemin P. Kahya, Mete Yeginer and Bora Bilgic, "Classifying Respiratory Sounds with Different Feature Sets", *Proceedings of the 28th IEEE EMBS Annual International Conference*, Aug 30-Sept 3, 2006.

25. C. S. Burrus, R. A. Gopinanth and H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, 1998.

26. Daniel T. L. Lee and Akio Yamamoto, "Wavelet Analysis: Theory and Applications", *Hewlett-Packard Journal*, 1994.

27. I. Daubechies, "Orthonormal bases of compactly supported wavelets", *Comm. Pure Appl. Math.*, Vol. XLI, pp. 909–996, 1988.

28. Ethem Alpaydin, *Introduction to Machine Learning*, The MIT Press, Cambridge, MA, USA, 3rd edn., 2014.

29. K. Hornik and M. Stinchcombe, and H. White, "Multilayer Feedforward Networks Are Universal Approximators", *Neural Networks*, Vol. 2, pp. 359–366, 1989.

30. Lin, B.S., H.D. Wu, F.C. Chong and S.J. Chen, "Wheeze recognition based on 2D bilateral filtering of spectrogram", *Biomed. Eng. Applic. Basis Commun.*, Vol. 18, pp. 29–38, 2006.

31. S.A. Taplidou and L.J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds", *Comput. Biol. Med.*, Vol. 37, pp. 1073–1083, 2007.

32. "Convolutional Neural Networks Tutorial in PyTorch", `http://adventuresinmachinelearning.com/wp-content/uploads/2017/04/CNN-example-block-diagram.jpg`, accessed on June 16, 2018.

33. Y. LeCun, F. Huang, and L.Bottou, "Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting", *Proc. CVPR*, 2004.