ADDIS ABABA SCIENCE AND TECHNOLOGY UNIVERSITY

Robust Cough Analysis System for Diagnosis of Tuberculosis Using Artificial Neural Network

BY

Amsalu Fentie Jember

A Thesis Submitted as a Partial Fulfillment to the Requirements for the Award of the Degree of Master of Science in Electrical and Computer Engineering

(Computer Engineering)

to

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL AND MECHANICAL ENGINEERING

Sep 2021

Declaration

I hereby declare that this thesis entitled "**Robust Cough Analysis System for Diagnosis of Tuberculosis Using Artificial Neural Network**" was prepared by me, with the guidance of my advisors. The work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted, in whole or in part, for any other degree or professional qualification.

Author:	Signature,	Date:
Amsalu Fentie		06/09/2021
Witnessed by:		
Name of student advisor:	Signature,	Date:
Dr. Taye Girma		06/09/2021
Name of student co-advisor:	Signature,	Date:
Yehualashet Megersa		06/09/2021

Approval

This is to certify that the thesis prepared by Mr. Amsalu Fentie entitled "**Robust Cough Analysis System for Diagnosis of Tuberculosis Using Artificial Neural Network**" and submitted as partial fulfillment for the award of the Degree of Master of Science in Electrical and Computer Engineering (Computer Engineering) complies with the regulations of the university and meets the accepted standards concerning the originality, content, and quality.

Signed by Examining Board:

Advisor: Dr. Taye Girma
Signature, Date: 06/09/2021

External Examiner: Beakal Gizachew (Ph.D)
Signature, Date: 10-Sep-2021

Internal Examiner: Solomon Zemene
Signature, Date: 8/9/21

Chairperson: Yonas Tesfaye
Signature, Date: 8/9/21

DGC Chairperson: Fisha A.
Signature, Date: 08/10/21

College Dean/Associate Dean for GP: Muluneh Mekonnen Tulu (PhD) Associate Dean for College Electrical and Mechanical Engineering
Signature, Date: 11/10/21

College of Electrical & Mechanical Engineering
Science & Technology University

Abstract

This research proposed a robust and easily applied method for tuberculosis (TB) screening system based on the analysis of patients' cough sounds. There are various existing diagnostic tests for TB, but they are expensive and require highly skilled physicians and laboratory facilities. Therefore, there is a need for a low-cost, quick-to-diagnose, and easily accessible solution for diagnosing TB in developing countries using a patient's cough sound. The coughing sound of patients with TB have distinct mathematical features or information that can indicate a disease. The use of patients' cough sounds to diagnose pulmonary diseases is an active research field with promising results; however, a robust system for diagnosing tuberculosis using cough sounds is currently unavailable commercially.

For this research, a dataset of 6476 cough and non-cough sound events was collected from patients with various respiratory diseases from Bahir Dar Felege Hiwot compressive specialized hospital using three different recorders. An automatic cough detection and classification system were implemented using an artificial neural network (ANN) and a support vector machine. The algorithms used Mel frequency cepstral coefficient (MFCC) features to detect cough sound from the recording, and then classify it as TB or non-TB. The MFCCs are machine-based methods for detecting and classifying sounds by mimic human hearing perception. Audio signal processing was done to extract the robust MFCC features, which were achieved by pre-processing and feature engineering efforts. The ANN outperforms the SVM in cough detection, with a 98.2% accuracy and an F1-score of 98.1%, and in TB/non-TB classification, with a 92.3 % accuracy and an F1-score of 87.7%.

The result shows the potential of the proposed cough sound analysis framework for the diagnosis of TB. This study contributes to the development of a robust TB diagnosis system that addresses fundamental gaps in the cough sound analysis area and can be transformed into a cost-effective alternative to the existing diagnosis.

Key Words: **Mel frequency cepstral coefficient, Neural Network, Tuberculosis, Robust**

Acknowledgment

I'd want to convey my deepest gratitude to my advisor, Dr. Taye Girma, for his endless support during my studies.

My warmest gratitude goes out to my co-advisor Mr. Yehualashet Megersa. For this thesis, his guidance and ideas were extremely helpful. I appreciate everything I've learned from him, as well as the critical comments he's provided, as well as his encouragement from the beginning of this research to the end.

For their assistance in gathering cough data, I would like to express my gratitude to the entire personnel at Bahir Dar Felege Hiwot compressive specialized hospital.

I'd also like to express my gratitude to AASTU's Electrical and Computer Engineering staff for creating such a good atmosphere throughout my studies.

Finally, I want to thank my parents and friends for their unwavering love, commitment, and encouragement.

Contents

List of Abbreviations

ANN	Artificial Neural Network
CCD	Cough Classification Dataset
CDD	Cough Detection Dataset
CD	Compact Disc
CT	Computed Tomography
DCT	Discrete Cosine Transform
DLN	Deep Learning Networks
DNN	Deep Neural Network
EX-TB	Extra Tuberculosis
FFT	Fast Fourier Transform
LogE	Log Energy
MFCC	Mel Frequency Cepstral Coefficient
MLP	Multilayer Perceptron
MTB	Mycobacterium Tuberculosis
NGS	non-Gaussianity Score
NN	Neural Network
PCA	Principal Component Analysis
PTB	Pulmonary Tuberculosis
SD	Standard Deviation
STE	Short Term Energy
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TB	Tuberculosis
WHO	World Health Organization

List of Figures

List of Table

Chapter 1
Introduction

1.1. Overview and Motivation

Tuberculosis (TB) is a bacterial infection caused by Mycobacterium tuberculosis (MTB) which mainly affects the lungs. It is one of the top 10 killers worldwide, with 30 countries including Ethiopia accounted for 90% of the incidence [1]. The symptoms of TB include coughing that lasts three or more weeks, fatigue, fever, night sweats, and loss of appetite. Provided the correct medication that allows the patient to take antimicrobial drugs for 6 months, TB is a curable and treatable disease [2]. There are various existing tests for TB such as chest radiography, tuberculin skin test, and gene-Xperts, but they are expensive or need highly skilled physicians and laboratory facilities [1], [3]. Therefore, there is a need for an inexpensive, quick diagnosis process and easily accessible solution for TB diagnosis in developing countries.

The coughing sound of a patient with pulmonary TB contains information indicative of the infection [4]. Features like the power spectrum of cough sound and mathematical features are used to train intelligent algorithms to enable the automatic detection of TB [4]. This hypothesis is central to the proposed research, which will robust cough analysis system for diagnosis of TB from cough sound. Cough analysis has evolved to be automated and robust and it is clear that the modeling of the human auditory system is the direction of computer cough detection and classification. The motivation in doing this research was to develop a fully automatic and robust method for diagnosing TB from a patient's cough sound using a computer, which could be easily applied in resource-poor countries, in rural areas, refugee camps, and displaced accommodations.

1.2. Background

This section describes five subtopics connected with the study. The first is the human respiratory system specifically the lung, which is the source of the sound signal including cough. Secondly, tuberculosis, which is a pulmonary disease caused by lung-attacking bacteria. Third, the causes of coughing, and diagnosis of cough. Fourth, the human auditory system (modeled in this study), which is used to sense the acoustic environment so that sound can be heard, and the perceived system and the range of human hearing. Finally, hypotheses for the automatic cough analysis system for the diagnosis of pulmonary disease are mentioned.

1.2.1. The Human Respiratory System

The human Respiratory System is a system that processes oxygen inhalation and carbon dioxide exhalation to fulfill energy requirements and consists of all the organs involved in respiration such as, nasal cavity, oral cavity, pharynx, epiglottis, larynx, trachea, and lungs. Its main purpose is to supply oxygen to the body and dispose of carbon dioxide, but it's also used to warming, filtering, and humidifying the inhaled air [5]. During inhalation, the oxygen first reaches the mouth or nose and travels into the larynx and the trachea, which then branches into two bronchi. Then each bronchus bifurcates to form smaller branches and form several pathways within the lung and ends with alveoli. In the alveoli, gas exchanges take place. After the gas exchange, exhalation starts and the air containing carbon dioxide starts the return journey through the nose and mouth back to the outside [6]. Filtering, warming, and humidifying the air is the main function of the upper tract (organs outside the chest cavity), and the lower tract (organs almost entirely inside the chest cavity) is for the exchange of gases [7]. Different respiratory diseases are present and may affect any part of the respiratory system. Their differentiation lies in the cause of infection or area they have struck. Bacteria, viruses, fungi, or toxins may cause respiratory diseases [8].

Depending on the region impacted, respiratory system infections can be classified into upper tract infection (UTI) and lower tract infection (LTI). UTI, which can affect the nose, sinus, and throat consists of diseases, including influenza, the common cold, tonsillitis, and, sinusitis [9]. LTI consists of diseases that can affect the bronchi and lungs, such as tuberculosis, bronchitis, and pneumonia.

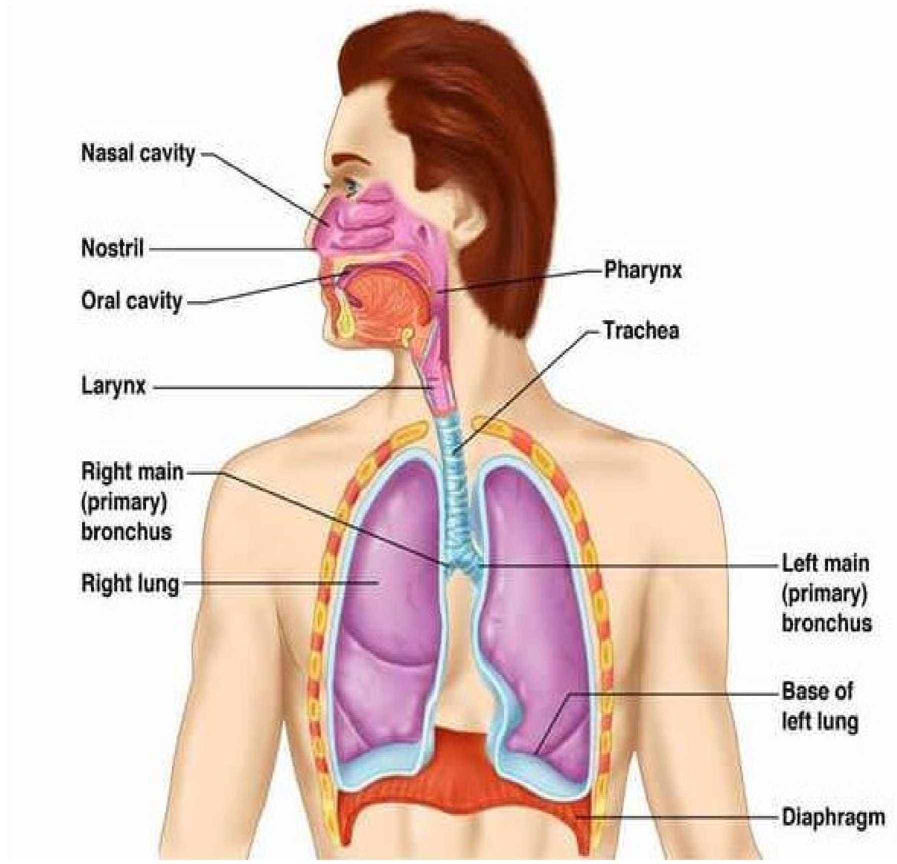


Figure 1.2-1 Human respiratory system, taken from [10]

The symptoms of respiratory system infections may differ depending on the type of germs affecting the body and include coughing, chest pain, and fatigue [9].

The respiratory system consists of several organs, as shown in Figure 1.2-1, the lungs, however, are the main respiratory organs where gas exchange takes place. To stay alive and safe, each cell of our body requires oxygen but also requires the removal of carbon dioxide [11]. The heart is responsible for pumping blood and supplying oxygen to the organs of the body. During expiration, the carbon dioxide-carrying blood cells return to the lungs and exhale the carbon dioxide [12]. Different lung diseases are present, some of them are:

- Asthma: - is the most common chronic lung condition, and attacks happen when the airways tighten and narrow, slowing down airflow. The lungs also become swollen and inflamed.

- Bronchitis: - this chest infection happens in the main airways, the bronchi. It may be due to a viral or bacterial infection.
- Pneumonia: -This is a chest infection deep in the bronchioles and alveoli. Pus and mucus can build up, and the lungs may swell. This makes it difficult to breathe.
- Tuberculosis: This is a bacterial infection spread through air droplets from coughs and sneezes.

There are many lungs functional tests, such as arterial blood gas tests, blood tests, chest X-rays, exhaled nitric oxide tests, lung diffusion capacity, pulse oximetry, spirometry, sputum (spit), or mucus sample [11].

1.2.2. Tuberculosis

Tuberculosis is an infectious disease caused by Mycobacterium tuberculosis, and typically affects the lungs but can affect other parts of the body as well. The disease is transmitted via droplet infection while people infected with TB expel the bacilli's when coughing, talking, and sneezing. Tuberculosis is one of the world's top 10 killer diseases, and an estimated 10.0 million people with TB have fallen ill. In 2019, 1.2 million TB deaths were registered among HIV-negative individuals and an additional 208 000 deaths among HIV-positive individuals [1]. About one-third of the world's population, are estimated to have latent TB and are thus at the risk of developing active TB. World health organization (WHO) lists the 30 highest TB burden countries, which accounted for 90% of the world's cases. Ethiopia is one of the highest TB burden countries [1]. There are various existing tests for TB, with their pros and cons, some of them are listed in table 1.2-1.

Table 1.2-1. Summary of traditional diagnosis techniques for TB.

Test Name	Short Descriptions	pros	cons
Smear Microscopy	Sputum samples are viewed under a microscope to check for MTB.	Inexpensive [3]	low accuracies [3].
Sputum Culturing	This process involves growing the TB bacteria on solid media from a sputum sample.	more sensitive	need a higher infrastructure and longer diagnosis process [1].

Chest Radiography	CT-scan and X-ray can be applied to the chest to view TB manifestations in the lungs. An X-ray will show areas of the lung that are clogged up or scarred.	effective and fast	need expensive equipment and experienced doctors [3].
Tuberculin Skin Test	Tests the hypersensitivity of a patient to a derived form of MTB by injecting a small amount of fluid called tuberculin into the patient arm.	inexpensive	interpretation of the result is difficult; the effectiveness of the test relies on the level of medical expertise [13].
GeneXpert	new diagnosis and recommended by WHO.	comprehensive and effective	expensive [13]

These diagnosis techniques are expensive or need highly skilled physicians and laboratory facilities. The world needs an inexpensive, quick diagnosis process, and available rapid point-of-care testing for diagnosing TB to end the TB epidemic.

1.2.3. Causes of Coughing and Diagnosis of Cough.

Coughing is one of the signs and symptoms of TB and other lung diseases. Coughing is the normal biological protective mechanism, which is an expulsive reflection of the body in the existence of a chemical or physical substance unfamiliar to the organ [14]. Coughing can occur when a cough receptor is triggered by external substances such as bacteria, viruses, smoke, or dust. Cough receptors can be found in several regions of the respiratory system and are linked to the cough center in the medulla oblongata. If cough receptors are triggered, the cough center sends signals to the respiratory system to contract to remove the foreign substance [15].

Coughs have various features and can provide a significant indicator of respiratory changes that can be related to clinical diagnosis [16]. As an indicator of respiratory changes, subjective measures of the cough events frequency have been used in the past, however, the self-report measures may be inaccurate [17]. Since it depends heavily on the patient's ability to explain the cough accurately and on the doctor's understanding and hearing ability.

1.2.4. The Human Auditory System and Range of Hearing

The human auditory system is compacted with two auditory sensor organs, the ears, attached to the brain stem from either side of the head with a network of high-speed nerve cells. The brain stem redirects audio input from the ears to the auditory cortex, a part of the brain that specializes in audio processing. The ears are organs that provide two primary functions, hearing and balance, which rely on hair cells called specialized receptors. The ear seen in Figure 1.2-2 can be divided into, the outer ear, the middle ear, and the inner ear [18].

The outer ear contains an ear canal lined with hair and wax-secreting glands and protects the middle and inner ear, channels sound to the middle ear, and aids in sound localization. The middle ear consists of an eardrum and three small bones that hold it in place. The primary middle ear's purpose is to pass sound from the eardrum to the inner part of the ear and to balance the impedance between the medium by which the sound waves propagate (mostly air) and the fluid inside the cochlea. The inner ear regulates the sense of balance of the body and includes the hearing organ. The cochlea is located in the inner ear, a snail-shaped cavern filled with hair cells transmitted via the auditory nerve that translate sound into neural signals. The basilar membrane located in the inner ear has the main feature, in the beginning, it is thin and stiff and at the end, it is wide and sloppy When rolled out, the hairs closest to the middle ear connection are short and rigid, and the farther away from the middle ear, the longer and softer they become [19]. This allows different hairs to have different frequencies of resonance, allowing us to discriminate between different pitches. At a certain level, a particular wave frequency can interact perfectly with the fibers, causing them to vibrate rapidly. The position of these hairs represents a log-like distribution with hairs that resonate at lower frequencies more sparsely spaced and hairs resonating at a higher frequency more tightly spaced. This log-like spacing allows our hearing spectrum at lower frequencies to be more.

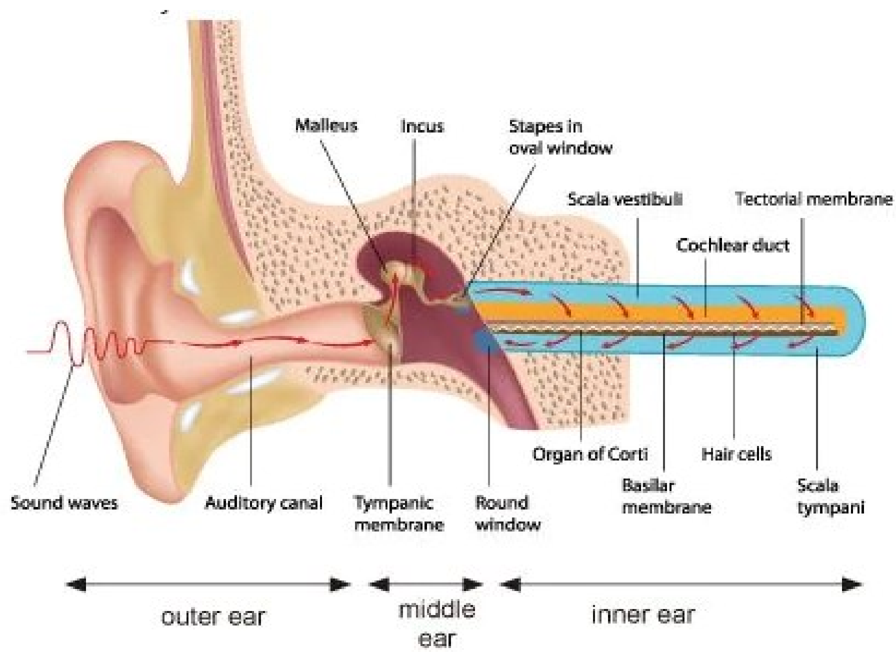


Figure 1.2-2. Anatomy of the human ear, taken from [20]

The cochlea function can be considered to be identical to the Fourier transform, transforming raw sound vibrational waves into neural signals in the frequency domain [21].

A healthy person's natural human hearing range is typically between 20Hz and 20000Hz frequencies [22]. Sounds above 20kHz are called ultrasounds and below 20Hz are called infrasound. Compared to humans, many animals have a larger hearing range and pick up infrasonic and ultrasonic signals. The perceived sound signal, however, is influenced not only by the pitch (frequency) but also by the loudness (amplitude) [23]. Decibels (dB) are the unit used to measure sound loudness. The required loudness for perceived hearing at 20 kHz exceeds the pain threshold, leading to the end of the human audible range. Note that the hearing threshold is at the lowest point in the 3kHz to 5kHz range, which means our ears are most sensitive to sounds within this range.

To understand complex sounds, we must be able to differentiate between the different pitches that make up the sound. As sound waves force the hairs to vibrate inside the cochlea, the hairs which resonate with the vibrational frequency are activated, but some nearby hairs are also stimulated, forming the crucial band. Within the same critical band, two sounds cannot be distinguished in terms of pitch [24]. Low-frequency audio sounds have narrower critical bands, making them more suited to identifying pitch at lower frequencies than higher-frequency sounds.

1.2.5. Automatic Cough Analysis System

Cough analysis has evolved to be automated and robust and it is clear that the modeling of the human auditory system is the direction of computer cough detection and classification. Cough offers significant information on the health of the airways and is very useful in assessing the disease [25] [26]. Analysis of the cough sound using digital signal processing is to calculate the spectral features of cough sound events, which are then used by intelligent algorithms to diagnose various pulmonary diseases. Detection and classification of cough sounds have been an area of research since 1989 to the diagnosis of different pulmonary diseases [27]. Cough analysis consists of two main stages, cough event detection and cough event classification to a specific disease. There are several studies on cough detection and classification, discussed in detail in chapter 2 of this paper. This research will focus on methods of both detection and classification of cough events for the diagnosis of TB. The coughing sound of a patient with pulmonary TB contains distinct information indicative of the infection [28]. The power spectrum of cough sound and mathematical features are used to train intelligent algorithms to enable the automatic detection of TB [4]. This hypothesis is central to the proposed research, which will automatically detect TB from cough sounds. The algorithms and techniques used for cough detection and classifications are discussed in detail in chapter 4 of this paper. The process of cough detection and classification is a relatively well-researched field, however, to the best of my knowledge no robust system for diagnosis of TB using cough sound is commercially available at this time.

1.3. Statement of the Problem

The traditional TB diagnostic techniques are time-consuming, or expensive, need experienced experts to interpret results. But highly trained physicians and state the art TB diagnosis equipment are scarce in resource-poor countries in general and specifically, in rural areas. Therefore, there is a need for an innovative point-of-care test that is quick, low cost, user-friendly, and highly accurate so that it aids the clinician to effectively diagnosis pulmonary TB. Despite efforts and promising results in the use of cough sound for diagnosis of various pulmonary diseases, robustness to various recording environments and different types of recorders, as well as fully automating the process by integrating cough sound detection and classification, remains a challenging task.

The previous study for TB diagnosis used data from TB patients and healthy people, but data from patients with diseases similar to TB were not included. This TB-positive and healthy individuals' cough classification yielded less robust. The classification between coughs caused by TB and coughs caused by other lung diseases has yet to be investigated.

The second limitation of the previous study for tuberculosis diagnosis was that cough sounds were recorded in a specially designed facility under controlled environments. There was no background noise, and the silences had a consistent energy level, but this differs from realistic conditions. This thesis proposes a robust signal preprocessing method for removing background noise and versatile energy level silences from data recorded in noisy environments (clinical settings in real-world environments). This method is advantageous because it is analogous to real-world conditions.

In the previous study, humans manually detected and extracted cough events among other non-cough sounds by listening to recordings. The procedure is time-consuming, difficult to implement in practice, and requires extra effort. In this research, the proposed method automatically extracts cough events from recordings and categorizes them as TB or non-TB cough. It is simple and effective to implement into practice.

In general, the sight of this study will be to implement an algorithm that is robust to various pulmonary diseases, recording environments, and different recording devices.

1.3.1. Research Questions

The primary research goal of this study is to address the following questions.

RQ1. How to remove silences from sound recordings recorded in real-world noisy environments, and how to compensate for waveform amplitude variations?

RQ2. Is there a mathematical feature that distinguishes a patient's cough from other sounds such as laugh, speech, or throat clearing?

RQ3. Can a computer mimic and enhance human hearing mechanisms to distinguish between TB patients' coughs and coughs of other lung diseases?

1.4. Objective

The general objective of this thesis research is to design and implement a robust automatic cough analysis system for the diagnosis of TB.

1.4.1.1. Specific Objective

Specific objectives are:

- To collect and construct a robust cough sound dataset for cough detection and classification models.
- To investigate currently available methods by using the constructed dataset and investigate robust feature engineering on signal pre-processing and feature extraction.
- To extract features from cough sounds for cough detection and features from cough sounds that can discriminate TB-induced cough from other types of coughs.
- To investigate learning algorithm classifier models, and compare their performance.

1.5. Scope and Importance of the Research

Many pulmonary diseases have the unique signature of cough sounds, in this research only the robust classification of coughs as either TB or non-TB were performed through spectral cough audio analysis.

In developing countries like Ethiopia, TB is one of the most common and deadliest respiratory diseases. The traditional techniques of TB diagnosis need expensive diagnosis equipment with highly trained physicians. But, the scarcity of physicians, longer diagnosis process, expensive equipment, and the high cost of diagnosis in rural areas, refugee camps, displaced accommodations, and countries at war prevent an early diagnosis. To effectively end the epidemic WHO sets three strategic pillars, one is intensified research and innovation.

So, this study contributes by providing a research-based low-cost alternative point of care solution for the early diagnosis of TB.

1.6. Methodology

The methodology followed in this research was:

- Related Works were reviewed to investigate the problem and potential solutions to the cough sound analysis.
- Medical data, that was cough and other sound recordings were collected from Bahirdar Felege Hiwot compressive specialized hospital.
- The dataset was constructed from the collected recordings using the audio editing software tool audacity.
- The signal was pre-processed to remove background noise and silence from the recordings, then normalized and segmented each sound event.
- Robust features (Mel Frequency Cepstral Coefficients (MFCCs)) were extracted from cough and non-cough sounds.
- Cough detection and classification learning models have been implemented.
- Finally, the results were discussed and evaluated.

1.7. Contribution of the thesis

This research contributes to the advancement and novel methods of automatic cough analysis methods for TB diagnosis using cough analysis. The contributions are:

- The development of a robust cough dataset in this study will be useful for future research in this area. The dataset used in this study was robust in terms of dataset size, the number of patients, diversity of diseases (twelve different pulmonary diseases), and also diversity of recording devices (three different recording devices to record sounds).
- This study adds a robust technique for signal preprocessing and feature extraction for tuberculosis diagnosis using patients' cough sounds.

- Another novel contribution of this study is the integration of an automated cough detection system with a TB/non-TB cough classification system. This is the first study in this field to differentiate TB cough from other similar lung diseases cough with high classification accuracy, and the proposed method can help clinicians in resource-limited areas.

1.8. Outline of the Thesis

The thesis is organized into six chapters as follows:

- Chapter 2, describe the literature review part of the thesis about automatic cough detection and classification.
- Chapter 3, explains the data acquisition systems, and dataset preparations
- Chapter 4, the applied methods and approaches used for automatic cough detection and classification.
- Chapter 5, describes the results and discussion of the proposed robust cough detection and a classification framework.
- Chapter 6, concludes the work and presents recommendations for future improvements.

Chapter 2
Literature Review

2.1. Introduction

Researches related to cough sound analysis could be cough detection, cough counting, and cough classification. It uses spectral features and is usually used for the analysis of cough frequency, to a specific disease. Cough detection also is used in spirometry [29]. Several methods were proposed in recent times for the automatic detection, counting, and classification of cough as described in section 2.2.

2.2. Specific Researches on Cough Detection and Classifications

A cough detection algorithm for continuous cough counting systems using an event detection algorithm by thresholding a smoothed energy measure has been designed in [30], [31]. A cough monitoring system for patient recovery from pulmonary TB has been designed in [31]. The system was implemented in three different classifiers multilayer perceptron (MLP) neural network, sequential minimal optimization, and SVM using MFCC features. They reported 88.2 % accuracy and 81% sensitivity for MLP and 86.4 % accuracy and 81% sensitivity for sequential minimal optimization.

Cough sound analysis has been designed to distinguish between a cough and non-cough sounds from patients that have pulmonary diseases has been designed in [32], [33]. The recording was performed at a 44 kHz sampling rate and 16 bits per sample resolution in [33]. Features such as kurtosis, skewness, the slope of the power spectral density, Mel - frequency Cepstral coefficients (MFCC) are used to develop models. The system was implemented using decision trees and ANN. They reported 28% sensitivity and 99% specificity using the decision tree and 82% sensitivity and 96% specificity using the ANN classification method.

Leicester cough monitor system has been developed using hidden Markov models for counting cough in [34], [35]. The collected sound signals were from 15 patients with different diseases such as asthma, eosinophilic bronchitis, and gastro-oesophageal reflux

[35]. The algorithm of cough detection was based on the technique used effectively in the recognition of speech, which is a keyword spotting approach to detect segments of a cough in the same manner as phonemes are detected in speech recognition. This method achieved 86% sensitivity and 99% specificity for detecting cough sounds.

Automatic cough events detection was constructed using spectral features from acoustic signals using a logistic regression model to classify cough from non-cough signals, in [36]. The performance of the algorithm was evaluated on 980 coughs and more than 1000 non-cough sounds events, from 43 patients. The algorithm achieved specificity, sensitivity, and F1 scores of 98.14%, 90.31%, and 88.70% respectively.

In [37], a cough detector using a wearable microphone was developed based on neural networks. The model is trained with audio recordings collected from 9 pulmonary disease patients. A frame of 200ms was split into four windows of 50ms, and 42 features (13 MFCCs, 13 deltas, 13 delta-delta, and 3 log energy) were computed from each window. A total of 168 features were used by the DNN model to effectively discriminate coughing from background noise and achieved a specificity of 93.7% and a sensitivity of 97.6%.

Cough sound signals were pre-processed to suppress the noise and classified using SVM into various respiratory disorders in [38], [39]. The fourth-order Butterworth high pass filter was used to reduce the effect of noise from the cough sound in [39]. Shannon entropy, MFCCs, and ZCR features were extracted after the suppression of noise from cough signals. The SVM has been trained with a 374x46 feature size and tested with a 94x46 feature size and provided 98.9% accuracy.

Local Hu moments as a feature set was proposed in [40], for automatic cough event detection. They used Hu moments to collect the same information from cough events, despite the level of contamination due to background noise. This algorithm shared a feature set with MFCC, which is widely used in audio signal processing. The relationship between frequency bands among windows using block processing is taken into account by Hu moments, but MFCC does not. They designed this feature to enhance the robustness against MFCC noise, which unexpectedly resulted in MFCC becoming more robust for their metrics. The use of spectral features and two classifiers (SVM and k-NN) for the detection of cough in noisy environments was presented and the evaluation was conducted using 16-hour recordings from pulmonary patients. Using local Hu moments as a feature set and a k-NN classifier, the best efficiency they recorded was 88.51% sensitivity and 99.77% specificity, but MFCC had the best performance.

For the detection of cough, principal components analysis (PCA) and deep learning networks (DLN) based on TensorFlow were used in [41]. Feature extraction was done using PCA, model training using DLN, and graph model computation was used by TensorFlow. The model was trained with a dataset consisting of 810 events (303 coughs, and 507 other non-cough sounds) recorded from eight volunteers. PCA+DLN achieved an accuracy of 0.989, while DLN achieved an accuracy of 0.983, showing that the PCA+DLN model performed better than the DLN model.

The author in [42], proposed a machine hearing system for audio-based cough segmentation based on a high-level representation of band-specific audio features. The record sound signals from 13 adult patients in waveform audio file format, at 44.1 kHz sample rate, with 16 bits resolution, and manually annotated. To compute the mean and standard deviation of short-term descriptors in 300ms lengthy frames, five frequency bands have been defined and cough detection was performed using an SVM trained with data from different noisy situations. Using twenty-nine short-term features, they tested the system and obtained 92.71 % sensitivity and 88.58 % specificity.

In [43], a dry and wet cough sound classifier has been designed with 178 cough events from 46 subjects. Features like bispectrum score, non-Gaussianity score (NGS), formants frequencies, log energy (LogE), kurtosis, and MFCC were used for classification. The system was implemented using logistic regression models and resulted in 72% specificities and 79% sensitivity.

In [44], cough sound analysis has been designed for pneumonia and asthma classification using ANN for 18 patients with a total of 674 coughs (412 were from nine patients with pneumonia and 262 from nine patients with asthma). In this system, 22 features were computed (13 MFCCs,_first five formant frequencies, zero-crossing rate, NGS, and

Shannon entropy). They used an ANN with 1 input layer with linear activation function, 2 hidden layers with sigmoid activation function, and an output layer. They reported 88.9% of sensitivity and 100% specificity.

In [45], a cough analysis system has been developed to diagnose pertussis by analyzing cough signals using logistic regression. Features like MFCC, crest factor, maximum frequency, spectral roll-off, spectral kurtosis, spectral slope, band power, spectral flatness, the spectral standard deviation are extracted from 38 audio recordings and used to train and test the cough sound, detection model. The algorithm provided 92% accuracy.

Automatic detection of TB has been designed to differentiate TB positive from healthy controls using cough sounds analysis, in [4]. The dataset was built from 17 TB-infected individuals and 21 healthy individuals. A total of 746 cough events were extracted and then used for logistic regression classifier training. Features such as log spectral energies and MFCC are used to develop models and reported Sensitivity of 95% and Specificity of 72%. Limitations of this work are data from patients with diseases similar to TB did not include, the only used classifier was logistic regression and other classifiers would not investigate, feature selection could not be used to lower the computational cost and the integration of the automatic cough annotation system and the cough classifiers could not be implemented.

In, [46], a general framework for the analysis of cough sound includes automatic cough segmentation, extraction of features, and classification for different pulmonary diseases. For analysis, kurtosis, variance, and zero-crossing irregularity features were extracted. The dataset was constructed using voluntary cough data from 54 patients and 33 healthy individuals. The model was trained using the dataset, and the accuracy was 81% for asthma vs. COPD, 74% for healthy control vs. unhealthy, and 80% for obstructive vs. non-obstructive as calculated by the AUC of a ROC curve.

2.3. Summary of Literature Review

In digital signal processing, cough detection is a technique of separating cough events from other sounds, which is useful for obtaining data on the frequency and strength of cough and can provide valuable insight into the treatment of patients and the seriousness of the disease. Cough classification is aimed at diagnosing particular diseases or disorders by analyzing cough sounds. Several studies were considered for each field (cough detection and cough classification), and their results and findings were reported. Some researches were limited to only cough detection or cough classification, they were not fully automated by integrating the two methods. The robustness of various recording environments whether the recordings were performed in a clinical setting in real-world conditions or the noise-controlled environment were one issue of previous studies. Diversity of similar pulmonary diseases in the datasets for classification and attention for TB were also issued (investigated or not explored) in previous studies. Table 2.3-1 Summarize some similar studies, with the tick symbol (✓) indicating that the study meets the criteria and the cross symbol (X) indicating the study's limitations.

Table 2.3-1. Summary of the related works.

Autor	**Automatic Cough detection**	**Cough classification**	**Diversity of diseases (more than 3)**	**Real cough sound in noisy environments**	**For TB diagnosis**
Brian et al. [31]	✓	X	X	✓	✓
Martinek et al [33]	✓	X	X	✓	X
Birring et al. [35]	✓	X	✓	X	X
Pramono et al [36]	✓	X	✓	X	X
Kadambi et al. [37]	✓	X	✓	✓	X
Monge et al. [40]	✓	X	X	✓	X
Khomsay et al [41]	✓	X	X	X	X
Alvarez et al. [42]	✓	X	X	✓	X
Swarnkar et al [43]	X	✓	✓	X	X
Amrulloh et al [44]	X	✓	X	X	X
Pramono et al [45]	✓	✓	✓	X	X
Botha et al. [4]	X	✓	X	X	✓
This work	✓	✓	✓	✓	✓

According to the table, there are efforts in the use of cough sound for the diagnosis of pulmonary diseases. However, using cough classification for TB diagnosis has yet to experience attention. The robustness of various recording devices, as well as the ability to discriminate TB cough from other lung diseases cough, were not explored. It requires focusing on fully automating and integrating the two methods of cough sound detection and cough classification. By building robust cough datasets, feature engineering efforts, and investigating and selecting the best classifier models used in this cough analysis field, this study would cover the gap in the robustness of the preceding works.

Chapter 3

Data Collection and Preparation of Datasets

This chapter describes the data collection, the materials used for data collection, and the preparation of datasets for this research. The data were collected by using three different recording devices. All Cough events and other non-cough events were manually segmented from the audio recordings. To the best of my knowledge, standard cough sound datasets are still not publicly available. Then the prepared datasets were for both cough detection and cough classification models.

3.1. Data Collection

The recordings of sound data (both cough and non-cough sound) were collected from January 2020 to March 2020 from Bahir Dar Felege Hiwot compressive specialized hospital. The collected data is only sound recordings (cough and non-cough events) from patients who have a cough. Table 3.1-1 shows several patients with coughs due to different respiratory diseases, that are included in this data collection. The recording data were collected from patients after their cases are identified by the medical experts (shown in Appendix A). The case of the patients was identified using clinical diagnoses such as sputum culture, complete blood count (CBC), GeneXpert, erythrocyte sedimentation rate test (ESR test), chest x-ray, computed tomography scan (CT scan), and Bronchoscopy.

Table 3.1-1. The number of patients who participated in this data collection with their cases.

	Case	Number of patients
1	TB	15
2	Pneumonia	38
3	B. Asthma	4
4	Bronchiectasis	6
5	Obstructive airway disease, Pleural effusion, Rt lung mass, URTI, Cardiomegaly, Lung Cancer, Allergic rhinitis, and Interstitial lung disease (ILD)	1 each

The cough data were obtained from seventy-one (71) volunteers suffering from cough due to respiratory disease, from the total recordings 15 records are from patients with TB. These cough sounds were labeled as TB cough, pneumonia cough, etc. on the grounds of their clinical findings. For the training and testing processes of the learning models, these labels were used. From each patient, more than five cough events were recorded within 6-25 minutes. All cough data and multiple non-cough events were collected from the adult ward. All cough data were recorded within the room shown in Figure 3.1-1, at Felege Hiwot Compressive Specialized Hospital.

Figure 3.1-1. Recording setup at the hospital

All the subjects (patients) who satisfied the inclusion criteria involved in the cough recordings were given informed consent (shown in Appendix B). Inclusion criteria are patients with known respiratory tract infections having cough, and willing to fill out the written consent form. But patients having droplet precautions, who are in critical conditions and unwilling to fill written consent forms were excluded.

3.2. Materials Used for Data collection.

Sound recordings were collected using three different recorders, a Philips VoiceTracer Digital Voice Handheld Recorder, an HM 1000 microphone, and an Infinix Hot 8 smartphone shown below in Figure 3.2-1(a), (b), and (c) respectively.

Figure 3.2-1. (a) Handheld Recorder, (b) HM microphone, (c) Infinix smartphone.

A Philips VoiceTracer Digital Voice Handheld Recorder: is a small device recording audio of any kind and more advanced types of built-in microphones. HM, 1000 microphone is used to record sounds, it provides a 10m cable. The Infinix Hot 8 smartphone was used for sound recording in addition to handheld and HM1000 recorders.

The reason why these three recorder systems are used is to build a robust system for various recording devices. There had been variations in the recording instruments, noise levels, and sampling frequencies, associated with each recording, as the recordings were collected using three different recording devices. Recordings were performed with a resolution of 16 bits per sample, a sampling rate of 44.1kHz, which is CD standard [47]. The microphone was angled straight to the patient mouth and held 30-50 cm away from the mouth and the handheld recorder and the phone were set on the table when recordings were performed. There was some unwanted speech noise originated from outside the room, at the time of recordings.

3.3. Preparation of Datasets

The sound database consisted of cough and non-cough events collected from 71 patients. The dataset contains a rich variety of different pulmonary disease coughs such as TB coughs, pneumonia coughs, asthma coughs, etc. The non-cough events contain many sounds easily confused with coughing such as throat clearing, sneezing, laughing, and other sounds. To construct the dataset for which the automatic cough detector and classifier were trained, manual sound event extraction was performed from the recordings.

Each recording was loaded into Audacity software and, listening to the audio, and detected the cough events and extracted each cough event with 16 bits per sample and a sampling rate of 44.1kHz. Figure 3.3-1 shows a raw recording of a patient recorded using an HM 1000 microphone.

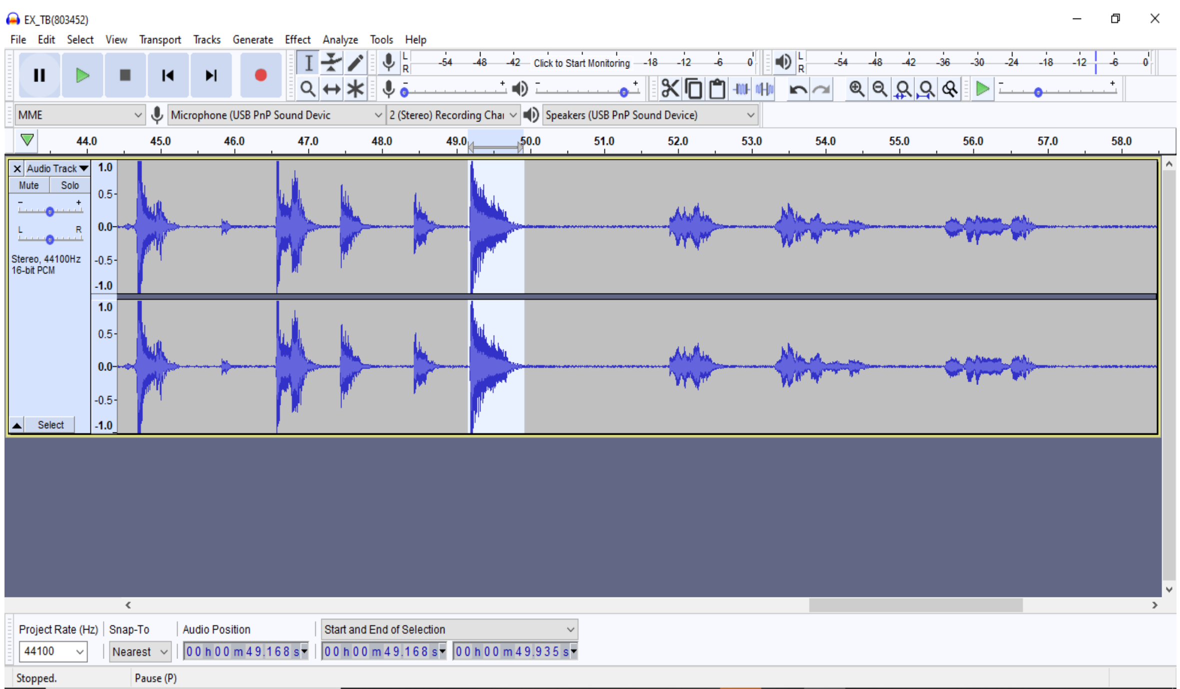


Figure 3.3-1 The waveform of a raw recording of a patient.

Cough events were manually marked and segmented in each recording by listening to the events carefully and concurrently looking at the waveform shown on the screen on Audacity. The selected segments in Figure 3.3-1 indicate a coughing event that can be extracted as a single cough event. Figure 3.3-2 shows a single cough event (only the selected segments in Figure 3.3-1) waveform after being listened to extract as a single (.wav) file format.

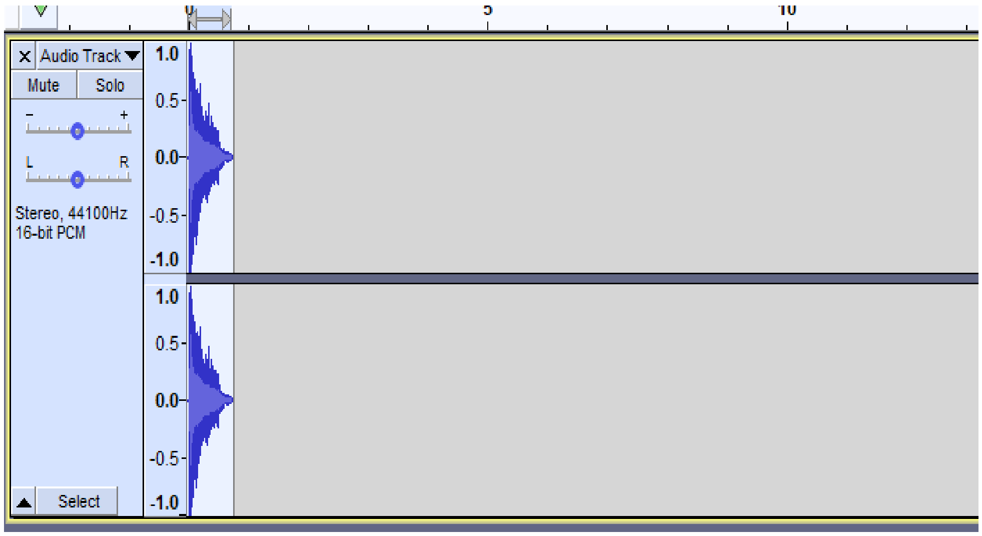


Figure 3.3-2. A cough event sample.

All sound events in this dataset were extracted in this way, and have a length of one or less than one second. A total of 6476 sound events are represented in this cough detection dataset (CDD), among them 3238 events were labeled as cough sounds and 3238 were non-cough events. This CDD was used for training and testing for the cough detection algorithm. A total of 3238 sound events are represented in the cough classification dataset (CCD), among them 1080 cough events were labeled as TB cough sounds and 2158 were non-TB cough events. The CCD was used for training and testing for the cough classification algorithm. Represented within these cough events are robust data, which is a multitude of different pulmonary diseases recorded by a microphone, handheld recorder, and phone. One aim of this research is to develop a robust system for various recording devices and different cases of pulmonary disease.

Multiple non-cough events were extracted from the collected recordings. But some non-cough sound files which could not record during data collection time on the hospital were also collected from the public universal-soundbank. Table 3.3-1 shows the composition of cough events with three recorders (microphone, portable recorder, and phone) and different pulmonary diseases.

Table 3.3-1. Composition of the cough data.

Cases	No. of Patient	Cough events were recorded using			Total
		Microphone	Hand-held recorder	Phone	
Pneumonia	38	444	436	444	1346
TB	15	374	366	374	1080
B. Asthma	4	63	68	68	199
Bronchiectasis	6	59	60	61	180
Obstructive airway disease	1	0	13	13	26
Plural effusion	1	20	20	20	60
Rt lung mass	1	10	10	10	30
URTI	1	19	19	19	57
ILD	1	15	15	15	45
Cardiomegaly	1	23	20	22	65
Lung Cancer	1	40	40	40	120
Allergic rhinitis	1	15	0	15	30
Total cough sound events					3238
Total non-cough sound events					3238
Total sound events in the dataset					**6476**

Represented within these non-cough events are a multitude of other audio sources such as speech, ceiling fan, footsteps, sounds from the outside environment (ambulance), sounds of closing the doors, sounds typical for walking, motor vehicles, laughter, sounds as the mobile device moved about, recording office sound effects, free tools recording effects and the phone ringing, and other types of background sounds. The silence is removed at the preprocessing stage, and it is not counted as either a cough or a non-cough event. All cough sound events and other non-cough sounds in these datasets are audible.

All data collects were approved by Felege Hiwot Compressive Specialized Hospital Out-Patient Department. The ethical clearance for this dataset is shown in Appendix C.

Chapter 4
Cough Detection and Classification Methods

This chapter discusses the proposed method for robust cough detection and classification. This thesis aims to develop a robust cough analysis method for different types of recorders in a moderately noisy environment and different lung diseases. It focuses on robust feature extraction for differentiating cough sounds from non-cough sounds within recordings, and then classifying coughs as TB cough or non-TB cough. This was accomplished through pre-processing and feature engineering efforts. The remaining of this chapter is organized as follows: The method for cough detection is explained in section 4.1 and the cough classification of the proposed algorithms is presented in section 4.2.

This automatic cough detection and classification method have two major steps. The first is cough detection, which involves separating cough events from non-cough sounds in a recording by removing undesirable signals via pre-processing and then extracting features for classifiers. The second stage is cough classification, which extracts discriminative features for classifiers to classify the detected cough events in the first step as TB cough or non-TB cough. A general overview of the proposed system is shown in Figure 4-1.

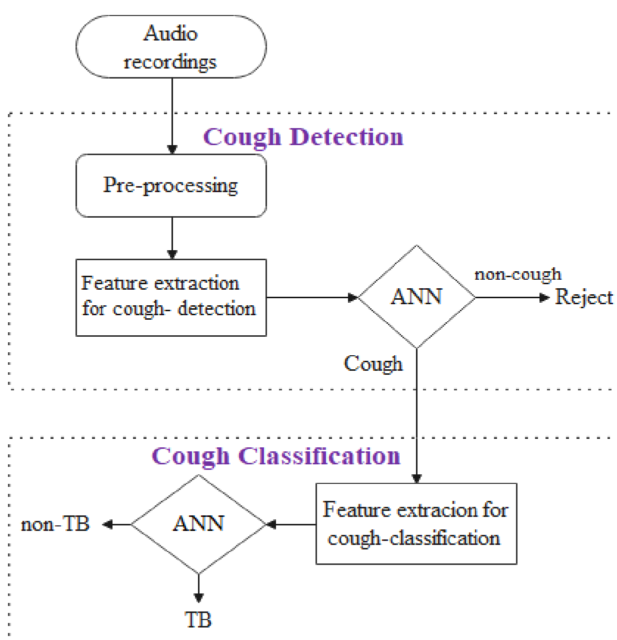


Figure 4-1. Automatic cough detection and classification system overview

4.1. Cough Detection Method

Cough detection is a technique of separating cough events from other sounds. Manual cough detection by listing each recording is very time-consuming. The sound events in CDD were used to train the learning models, which were then used to detect cough automatically, from the newly recorded data. The cough segment detected by the model can then be used as the expected input for subsequent cough classifiers. Before performing the feature extraction, the sound signals were pre-processed, then detection was performed using those extracted features. The cough detector learning algorithms' efficiency depends on both how well the features are extracted and how well the sound signals are pre-processed. The general workflow for automatic cough detection is displayed in Figure 4.1-1.

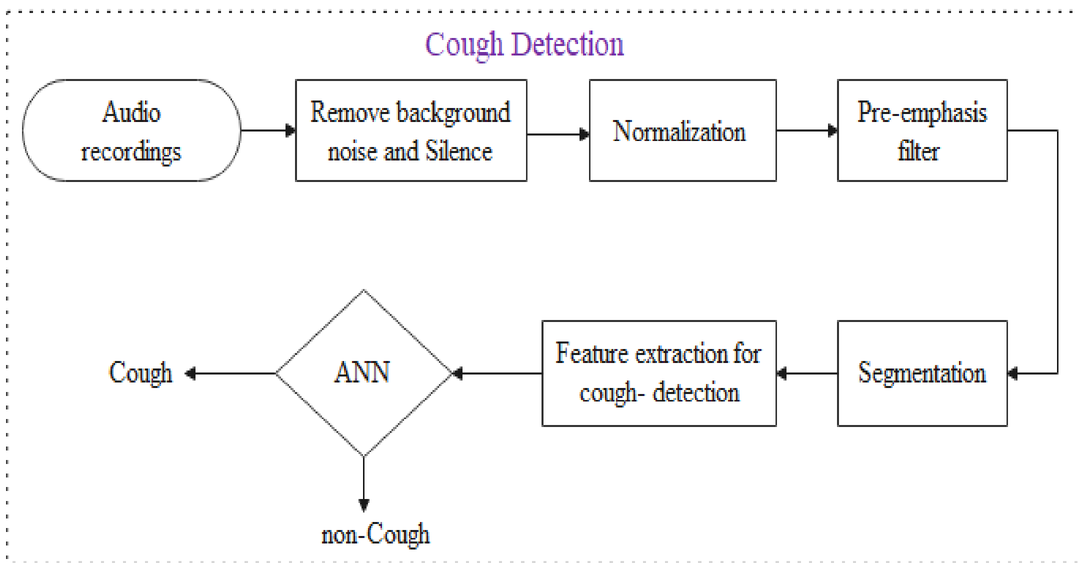


Figure 4.1-1. The general workflow for automatic cough detection.

The first step aims at removing background noise and silence within the recordings. Normalization and filtering were performed on the audio recording and then segmenting each recording into short events and features were extracted for each sound event. Finally, with those extracted features, the learning models training, and testing were performed. The outcome of each step is highly dependent on the previous steps. To enhance the system's robustness capacity to identify cough events, each processing step was investigated and improved.

4.1.1. Background Noise and Silence Removal

At clinics, when coughing sound was recorded, there were moderate levels of background noise obtained from the process of data collection, and unwanted silence was recorded. Equation 4.1 below models the discrete-time audio recordings with their components, which are used to demonstrate the changes after pre-processing.

$$x[n] = x_a[n] + x_{bn}[n] + x_s[n] \tag{4.1}$$

where $x[n]$ is the audio recording, $x_a[n]$ are the audio components (cough sound and non-cough sound), $x_{bn}[n]$ represents the background, and $x_s[n]$ is the silence.

To suppress the background noise $x_{bn}[n]$ from $x[n]$, a Butterworth bandpass filter was used to pass frequencies within a range. The Butterworth filter is a flat band filter at most, and the passband or stopband does not have a ripple. It has a wide region of transition from passband to stopband, and the response of frequency, group delay, the impulse response is much better and more practical than other filters [48]. This Butterworth filter was implemented to reduce the low-frequency noise such as noise coming from the vibration of microphone stands and high-frequency noise mostly Gaussian noises. It has a low and high cut-off frequency of $Lf_c = 20Hz$ and of $Hf_c = 20kHz$ respectively, which is the lower and upper boundaries of human hearing. Figure 4.1-2 shows the effect of Butterworth filters on the raw recording audio signal to suppress the background noise component.

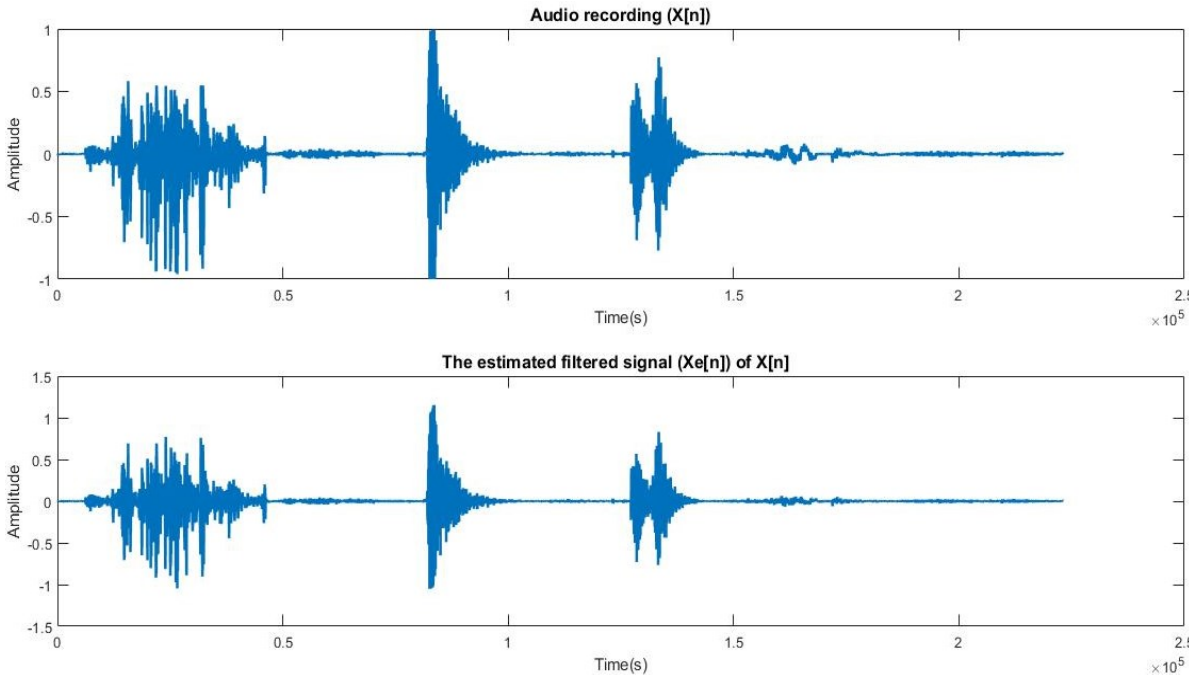


Figure 4.1-2. Audio recording before and after a Butterworth bandpass filter.

After eliminating the background noise using the Butterworth bandpass, the estimated audio recording $x_e[n]$ of the audio recording $x[n]$ is given below in equation 4.2.

$$x_e[n] = x_{ea}[n] + x_{es}[n] \tag{4.2}$$

where $x_{ea}[n]$ is the estimates of is the audio components (cough sound and non-cough sound), and $x_{es}[n]$ is the estimate of silence.

The next step is removing silence $x_{es}[n]$ from the estimated audio recording $x_e[n]$. The removal of silence focuses on the detection of frames that do not contain sound events relative to cough events or other non-cough events. Different methods are used to identify and remove the silence. The audio recordings are conducted at clinics and have moderate levels of noise, so this issue must be resolved by the silence removal process. The standard deviation was used to solve this problem since it is the effective method of silence removal for performing under moderate noisy conditions [49].

4.1.1.1. Standard Deviation

Standard deviation (σ) is a measure used to quantify a signal's variance or dispersion [49]. It is the square root of variance by evaluating the variation of each data point (amplitude value) relative to the mean. The standard deviation of audio recording is calculated as follows:

1. The first step is to split the audio recording into several short frames.
2. The mean (μ_i) of a frame (equation *4.3*) is determined by summing all the data points and dividing them by the total number (N) of points in a frame.

   $$\mu_i = \frac{\sum_{n=1}^{N} x_{en}}{N} \tag{4.3}$$

   Where, x_{en} the data points of estimate frames of audio recording and N is the number of data points in a frame.
3. The variance ($\sigma_i{}^2$) for each data point of a frame i is calculated using equation *4.4*, by subtracting the mean (μ_i) from each data point. Then, each of the resulting values is squared and the results are summed up and divided by the number of points minus one.

   $$\sigma_i{}^2 = \frac{\sum_{k=1}^{W} |x_{ei}\,(k) - \mu_i|^2}{\mathrm{W} - 1} \tag{4.4}$$

   Where, $x_{ei}\,(k)$ are the estimated frames of audio recording, W is the window size, i is the current frame, and k is the current sample.
4. By square rooting the variance resulting in number 3, the standard deviation is obtained.

There is a higher variance within a recording if the data points are further from the mean. For each audio recordings, the audio event and silence were determined using a threshold (T) value as shown in equation *4.5*.

$$T = \mu(min_\sigma) + \sigma(min_\sigma) \tag{4.5}$$

A higher standard deviation refers to audio bursts and a lower standard deviation would be correlated with silence. The raw audio, standard deviation, and measured threshold for a given audio recording are presented in Figure 4.1-3.

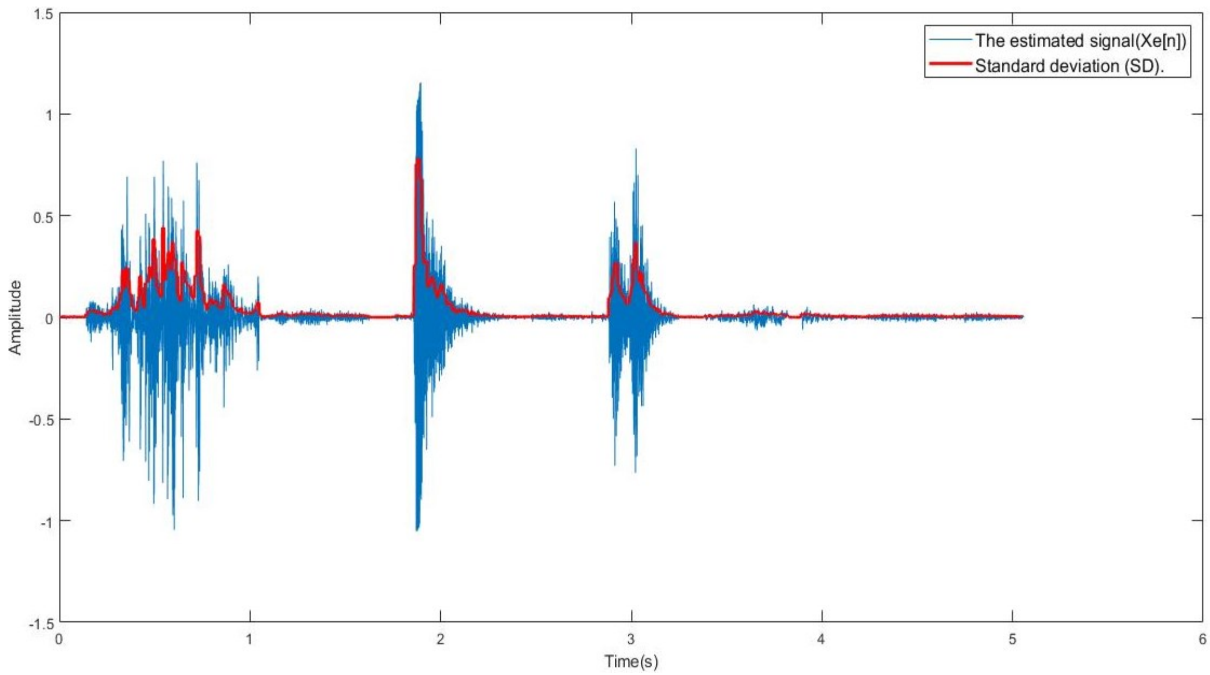


Figure 4.1-3. Estimated audio signal, with SD of the frames.

The SD of sound events is higher than the SD of silence, as shown in Figure 4.1-3. The sound events are identified as time regions where the SD reaches a specified threshold value of equation 4.5.

4.1.2. Amplitude Normalization

Audio recordings have variations in waveform amplitudes due to patients sitting at various distances from the recorders, different recording devices, or the naturally different sound loudness of the patients. Audio normalization can compensate for those differences by boosting the sound to a target level by altering the overall audio recordings by the same amount, without clipping and distorting the peak. For each sound recording event, the normalization was performed as follow:

1. Split the sound event into non-overlapping frames.

2. Compute the energy of each frame. It can be calculated by:

$$E(i) = \frac{1}{\mathrm{N}} \sum_{n=1}^{N} |x_i\,(n)|^2 \qquad (4.6)$$

Where N is the length of the frame, i is the current frame, and n is the current sample.

3. Find maximum standard deviation (σ_{max}) of all event energy frames and the energy standard deviation (σ_{energy}) of each frame.
4. Calculate the ratio $r = \frac{\sigma_{energy}}{\sigma_{max}}$ for all samples and multiply all samples with the mean.
5. The waveform has been scaled such that the energy of the events among sound events is normalized.

4.1.3. Pre-emphasis and Segmentation

A pre-emphasis is used to amplify the magnitude of higher-frequencies components of the sound events, to enhance the Signal to Noise Ratio (SNR). Pre-emphasis reduces the adverse effects of events such as recording device distortion in subsequent parts of the environment and flattens the spectrum. At this stage, equation 4.7 was used on sound events to enhance SNR with a $\varepsilon = 0.96$ which is the cut-off standard for the pre-emphasis [50].

$$x_i[n] = x_i[n] - \varepsilon x_i[n-1] \qquad (4.7)$$

This process (pre-emphasis) increases the energy of sound signals at a higher frequency and gives more information. After preprocessing, the next step was to segment the filtered sound events into a 100ms-size non-overlapping block. The energy of each segment was computed after segmentation, and a segment with a pick value was selected. To reduce the computational complexity of processing all segments of a sound event, it is possible to represent a sound event with a pick-value segment. Then, for further processing of the feature extraction, this pick-value segment of a sound event was used.

4.1.4. Feature Extraction

Feature extraction is about extracting information from the segmented sound event for reducing the dimensionality for the detection and classification model. It is a technique of feature engineering to derive relevant features from sound events. In sound signal analysis, feature engineering is a central task that is the process of converting raw sound events into features that better represent the underlying problem to enhance the accuracy of machine learning or deep learning model on unseen data. The innovative aspect of feature engineering is to find ways to develop the model by extracting different unique features used to discriminate cough events from other non-cough events and further classification of coughs. It helps the algorithms to understand data and decide patterns that can enhance the learning algorithms' efficiency. Much of the effectiveness of machine learning is the success of the feature engineering process that a learning model can comprehend [51]. Figure 4.1-4, is a feature engineering block diagram that converts inputs into features that can be understood by the learning algorithm.

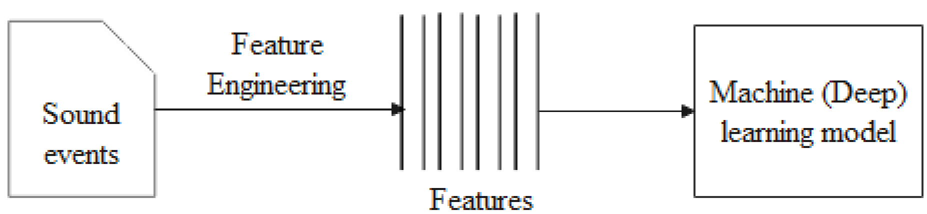


Figure 4.1-4. Block diagram of feature engineering

The aim of feature extraction is not only dimensionality reduction, but also extracting unique features present in the cough events, as well as to reduce the risk of overfitting, speeding up training, and reducing the complexity of computations.

There are a variety of sound features in the literature aimed at the identification and classification of cough signals. Several techniques are used for this process, two of the most are Linear Predictive Cepstral Coefficients (LPCC) and MFCC. The LPCC models the vocal tract correspond to the articulatory system of humans and the sound signal is modeled as a linear combination of its previous and current input [52]. The MFCC uses a spectral decomposition to break down sound signals into a non-linear distribution that mimics a biologically-inspired human auditory system response to sound [53]. Both of these (LPC and MFCC) approaches reduce the complexity of sound signals. In this paper, MFCC was used, because it is the prominent and more robust feature in cough signal processing [40] [53] and it gives consistent and robust results to noise because it is based on human perception of hearing [54].

4.1.4.1. Mel Frequency Cepstral Coefficients

MFCCs are typical features used in cough signal analysis, which represent the short-term spectral of audio based on the human hearing mechanism [55]. In nature, cough sounds are complex signals, indicating that the respiratory tract (system) carries essential information and provides the tract with cough sound from its substructure [39]. MFCC produces a sound signal representation, varying from other cepstral features (e.g., LPCC) is it use the Mel-scale in the frequency bands. In MFCC the bands are arranged logarithmically, which mimic the human hearing mechanism. The efficacy of MFCCs is their ability to efficiently represent the significant part of the vocal tract (shape of the vocal tract, which originates cough or non-cough sound) [4]. Figure 4.1-5 explains the concrete steps of the MFCC.

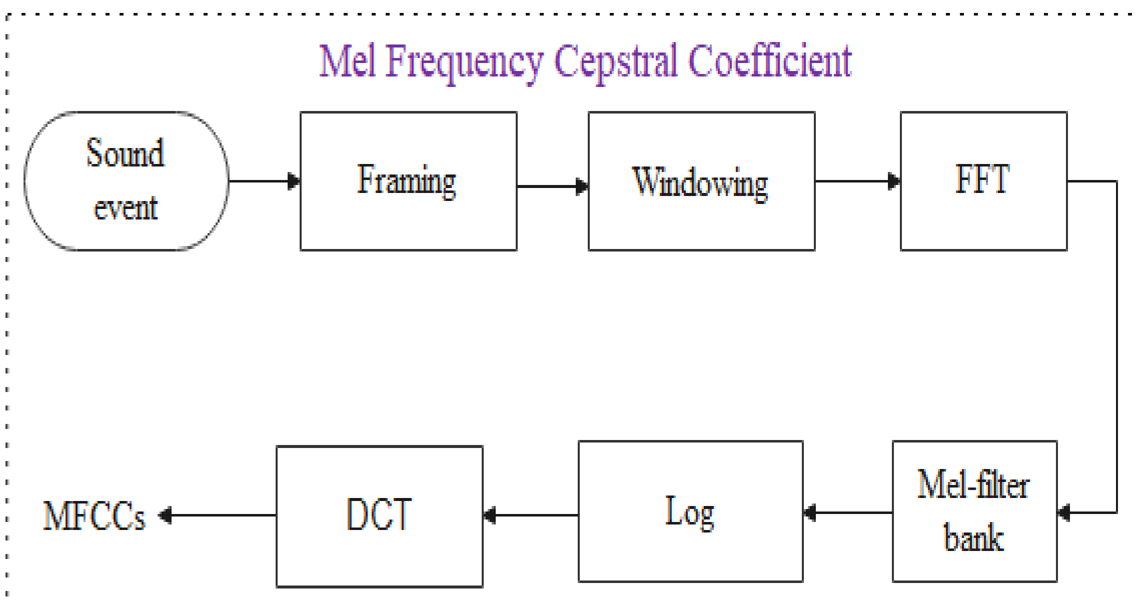


Figure 4.1-5 MFCCs features calculation flow diagram

A. Framing

In MFCC, the selected sound event is split into frames of short length. A spectral-domain analysis (e.g., Fourier transform) yields good results for stationary signals, but not for non-stationary signals. Cough sound signals are volatile and have non-stationary characteristics, so computing a single Fourier transformation to the entire cough event or record is meaningless. But for the frame of short length, the properties of the cough signal are considered stationary, and thus spectral analysis can be applied to it [56]. For this reason, the spectral domain analysis is computed for successive frames of the sound signal. The frame duration is usually between 20-40ms. Each frame was overlaid on the front frame to smooth the transition by accounts frames at window edges and provide equal weight between frames.

B. Windowing

The next step in the MFCC is to eliminate discontinuities or spectral leakage at the edges of the frames by windowing all frames. There are several window functions, but the hamming window is used to conduct windowing, which is better to eliminate spectral leakage [57]. Equation 4.8 is a Hamming window that is used at both edges of a frame to reduce spectral leakage.

$$w(m) = 0.54 - 0.46 \cos\left(\frac{2\pi m}{N_m - 1}\right) \quad 0 \leq m \leq N_m - 1 \tag{4.8}$$

Where, $w(m)$ is the window function, N_m is the number of samples within each frame. Overlapping Hamming windows were used to eliminate the degraded events at the boundaries, which improves how well the MFCC can describe different sounds. Before the fast Fourier transform, the framed signals $\mathrm{x}(m)$ were multiplied with the hamming function. The output signal after windowing is presented in equation 4.9.

$$\mathrm{y}(m) = \mathrm{x}(m)w(m) \quad 0 \leq m \leq N_m - 1 \tag{4.9}$$

Where, $y(m)$ is the output windowed signal, $x(m)$ is the input framed signal, $w(m)$ is the Hamming window shown in equation 4.9, and N_m is the number of samples within each frame.

C. Fast Fourier Transform

The FFT is an algorithm that converts time-domain frames into the spectrum (frequency-domain). FFT is an efficient method of discrete Fourier transform (DFT) since it can compute the N-element vector with $O(N \log N)$ the operation, but for the same vector, the DFT requires $O(N^2)$ operation. The result of FFT is a spectrum or periodogram. FFT is widely used for sound analysis. The cochlea function can be considered to be identical to the Fourier transform, transforming raw sound vibrational waves into neural signals in the frequency domain. The frequency distribution of a sound should be assessed since the human ear exercises this skill in hearing [21].

D. Mel-Filter Banks

In the FFT spectrum, the frequency range is wide and has a linear spectrum (follow a linear scale). But the frequency perception of the human ear follows a non-linear scale, which is linear up to 1000 Hz and logarithmic above. MFCCs use a non-linear frequency scale since it is the perception of human hearing [58]. The filter Bank of melody scale called Mel-Scale, which describes the frequency perception of the human ear was used as a passband filter at this stage. The tone of the sound signal with a real frequency is measured in Hz, and the perceptual scale of pitches is measured on Mel-Scale.

As mentioned above, the Mel-Scale is linear up to 1000 Hz and logarithmic above, so the following formula approximates each real linear frequency scale to the Mel-Scale.

$$f_m = 2595 \log_{10}(1 + {}^{f}/_{700}) \qquad (4.10)$$

Where f is the real frequency in Hz, and f_m is the perceptual Mel-Scale frequency in a melody (Mel).

In MFCC processing, Mel-frequency warping is realized by triangular bandpass filter banks in the frequency domain. The spectrum of signals (cough and non-cough sound segments) is passed through the Mel-filter banks, which spaced non-uniformly with Mel scale, then normally obtain the perceptual frequency, that can properly simulate auditory processing [59].

E. Log

After acquiring the Mel-spectrum, the next step is calculating the logarithm (log) of the squared magnitude or power spectrum of the output. The reason for this is that the dynamic range of amplitudes can be compressed by a logarithm. Since the response of the human ear to the sound signal level is logarithmic, it is less sensitive to small-amplitude differences. This makes the estimated frequency less sensitive to small amplitude changes due to the patient mouth being closer or far to the recorders.

F. Discrete Cosine Transform (DCT)

In this step, the DCT was carried out on the log Mel-spectrum to convert it back to the time domain, and the result is called MFCC. Cepstrum is the inverse Fourier transform of the spectrum. The DCT decorrelates the filter bank coefficients and generates a compressed representation of the log filter banks. Then, the DCT of the log Mel-scale of the power spectrum can be estimated by equation 4.11.

$$MfCC_w = \sum_{k=1}^{K} L_{Ms} \cos\left[w(k - \frac{1}{2})\frac{\pi}{K}\right], \quad w = 1,2,\ldots\ldots W \tag{4.11}$$

Where, W is the required number of MFCC coefficients, L_{Ms} is log Mel-scale of the power spectrum, and K is the number of filters.

Typically, for cough sound analysis the first 13 coefficients are retained for each window and the rest are discarded. The set of 13 MFCC is called acoustic vectors. Therefore, each input sound signal is transformed into a sequence of MFCCs.

4.1.5. Cough Detection Learning Algorithms

The cough detection phase was carried out after obtaining the feature vector, to discriminate cough sound events from no cough sound events. For cough detection and classification two popular and suitable algorithms in fields (ANN [32], [35], [38], [41] and SVM [31], [36], [37], [39]) were trained and tested to select the one with the highest numerical performance. The F1-score and accuracy were used as a metric to choose optimal models and compare results during optimization.

4.1.5.1. Artificial Neural Network

The ANN operated using an algorithm to interpret nonlinear data which is independent of sequential patterns. The ANN consists of neurons, organized into input, hidden, and output layers between the input and the results, which act like biological neurons, and are learned through a technique of backpropagation [60]. The strength typically offered by ANN is its capability of extracting hidden linear and classified data in complex and high-dimensional data like cough datasets based on a supervised learning technique using non-linear decision boundaries [61]. Concerning activated functions, the input of ANN was transferred into the output, and the result of each neuron was multiplied by weights and added with biased values from the neurons of the preceding level.

The learning models have hyper-parameters, that can influence the model's performance number of hidden layers and neurons used in each layer, for example, would be considered as hyperparameters while creating an ANN system. Overfitting can be avoided by optimizing model hyper parameters on a development set (selected from the training set). The hyper-parameters of a model were optimized using a brute force method called grid search to achieve the best results. Grid search is a method for training and validating the model for all possible hyper-parameter combinations. The procedure begins by separating the dataset into K equal sections, with the validation set being chosen from one of the K folds and the remaining K - 1 folds set as the training set. This was performed until all K folds have been tested, and the final evaluation metrics were calculated by averaging all K iterations.

The ANN structure consisted of a feed-forward network and all of the layers used the sigmoid transfer function to transform activation levels to output, and the approaches for this study were written in the MATLAB script.

4.1.5.2. Support Vector Machines

The SVM is a supervised machine learning algorithm for detecting and analyzing relationships. It operates by analyzing data sets using a set of parameters that are used to solve classification and regression problems. SVM determines the desired hyperplane for maximizing the distance between any two classes [62].

The hyperplane is a decision boundary that aid in the classification of data into various classes based on its attributes. The number of features of the data determines the dimension of the hyperplane. When the data has two features, the hyperplane is a line; when the data has three features, the hyperplane becomes a plane (two-dimensional); when the number of features exceeds three, it becomes difficult to picture [62]. The position of the hyperplane is influenced by data points called support vectors that are closer to the hyperplane. SVM aims to find a plane with the greatest distance between support vectors referred to as the maximum margin distance, to both classes. Grid search was used to fine-tune model parameters, and a radial-type kernel was selected to project the data, allowing for more complex class separation.

4.2. Cough Classification

The second phase of this study was classifying the cough events (coughs detected in section 4.1) as TB disease and non-TB diseases. The method used to classify cough events as TB and non-TB diseases is extremely similar to the technique of cough detection described in this study in Section 4.1. There are two differences between the two methods.

The first difference is the dataset used for training and testing the classifier model. The dataset used for the classifier model is the cough classifier dataset (CCD). The CCD contains only cough events collected from71 patients with various lung diseases. There are

3238 cough events in CCD, of which 1080 cough events were obtained from patients with TB and 2158 cough events were obtained from 12 other different pulmonary diseases.

The second differences are at the stage of pre-processing and extraction of features. The input for the classifier (cough classification artificial neural network (CC-ANN) system is the pre-processed cough events obtained from the cough detection system, so no further pre-processing is required. The classifier model begins with the extraction of features from cough events, from those previously obtained in cough detection artificial neural network (CD-ANN). Both models (CD-ANN and CC-ANN) use MFCCs features, but the extraction process has changed slightly. Figure 4.2-1 demonstrates the general workflow for the automatic cough classification system.

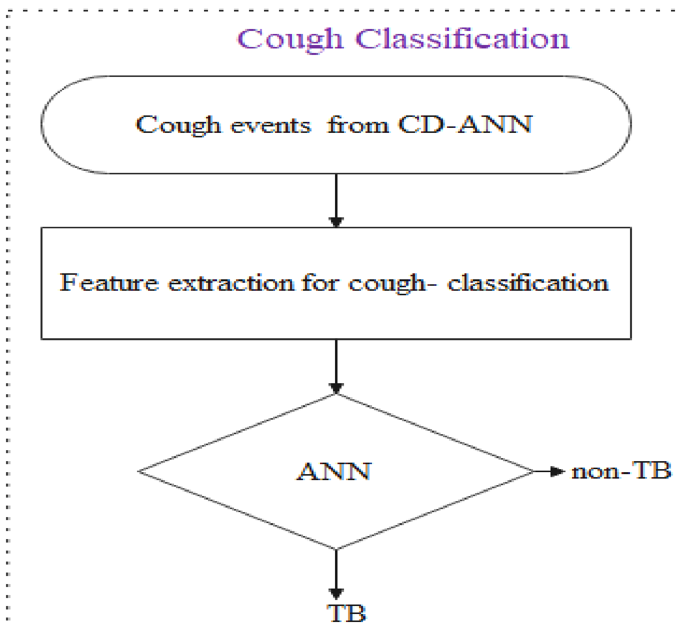


Figure 4.2-1. The general workflow for the cough classification (TB/non-TB).

The MFCC features were extracted from the cough events obtained from the CD-ANN model. The process of extracting features is almost similar to the technique previously used in section 4.1.4. The differences are, in CC-ANN the frame length is 20ms instead of 40ms. The differences are relatively small, but they combine to make the classification neural network complex, which boosts the accuracy.

Chapter 5
Implementation Result and Discussion

This chapter presented the results and discussion of this thesis work stage by stage, starting from the prepared input dataset to the final result. First, the robustness of the prepared dataset used in this study was compared to that of some related works. The results of various stages of pre-processing were then illustrated and discussed in section 5.2, along with their effect on the sound signal. Finally, in sections 5.3, to 5.6 the features used in cough detection and classification algorithms, as well as a comparison of the two models using various evaluation metrics, were discussed.

5.1. The Dataset Preparation Result and Discussion

The dataset (CDD and CCD) was used to train and test the models in this study. Using the technique presented in chapter 3 of this paper, 6476 sound events were labeled in the CDD and 3238 cough events were labeled in the CCD. The data summarized in Table 3.3-1 was obtained from multiple pulmonary patients in a clinical setting using three different recording devices in moderately noisy environments. This system is robust to differences in recording equipment, noise levels, and sampling frequencies associated with each recording because the sounds were recorded from patients with various lung diseases. The dataset was validated since it was manually segmented using a human scorer. Table 5.1-1 compares the datasets used in this research to those used in some similar studies listed in Chapter 2.

Table 5.1-1. Comparison of the dataset used in this thesis to the related works.

Autor	**# Sound events in the dataset**	**Number of patients**	**Variety of diseases**
Botha et al. [4]	746	17	2
Brian et al. [31]	620	62	2
Martinek et al. [33]	1706	41	2
Birring et al. [35]	Not specified	15	7
Pramono et al. [36]	1980	43	7

Kadambi et al. [37]	5670	9	4
Monge et al. [40]	78	13	3
Khomsay et al. [41]	810	8	2
Alvarez et al. [42]	1648	13	3
Swarnkar et al. [43]	178	46	4
Amrulloh et al. [44]	674	18	2
Pramono et al. [45]	1850	38	4
This work	**6476**	**71**	**12**

Since one of the goals of this study was to create a robust dataset and investigating learning models using it. The dataset used in this study is robust and comparable to the studies described in Chapter 2 as shown in table 5.1-1 in terms of dataset size, number of patients, diversity of diseases, and also diversity of recording devices.

5.2. Result and Discussion on Pre-Processing Phases

The sound signals were pre-processed to eliminate background noise before the features were extracted, and then detection and classification models were developed using the extracted features. The performance of the classifier models was determined by how well the sound signals were pre-processed as well as how well the features were extracted.

The recording sound signal was passed through a bandpass Butterworth filter, as described in Section 4.1.1, with low and high cut-off frequencies of 20Hz and 20kHz, respectively, to minimize the background noise. As illustrated in Figure 4.1-2, the filter suppressed the low and high-frequency noise components of the signal and lies within a fixed boundary. The silence was removed from the recording using the SD method after the noise from audio recording signals was suppressed. Figure 5-2-1 shows the audio recording before and after the silence was removed.

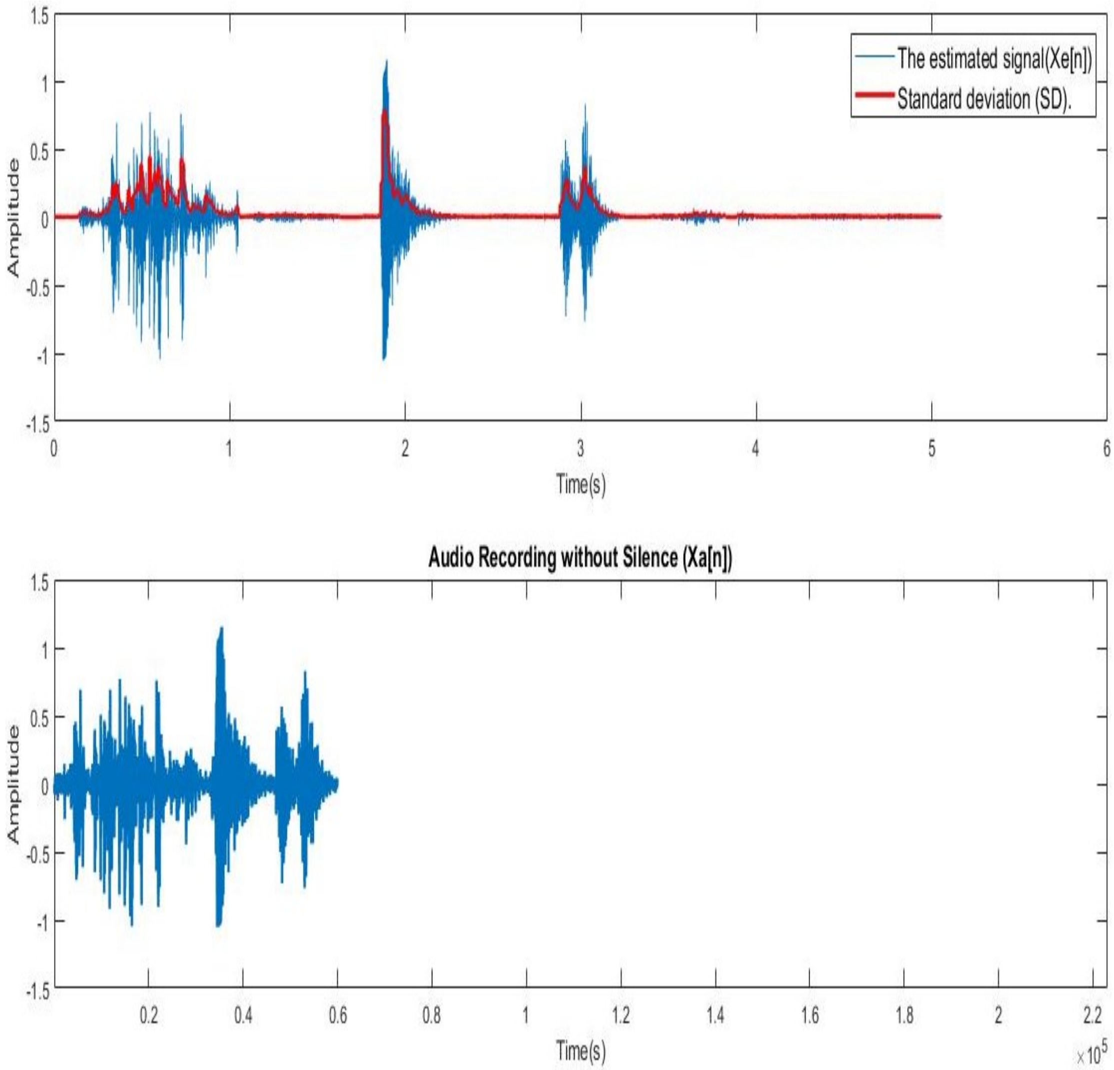


Figure 5.2-1. Waveforms, before silence removal, and after silence removal.

There was a higher variance within a recording if the data points were further from the mean. For each audio recordings, the audio event and silence were determined using a threshold (T) value using equation 4.5. A higher standard deviation refers to an audio event and a lower standard deviation would be correlated with silence. The audio recordings were conducted at clinics and have moderate levels of noise, so this issue was solved by the standard deviation silence removal method and it is an effective method for using in real-time conditions.

Waveform amplitudes in audio recordings varied due to patients sitting at different distances from the recorders, different recording devices, or the patients' naturally variable sound loudness. Audio normalization adjusted for these variations by raising the sound to a target level by adjusting the entire audio recordings using the technique stated above in section 4.1.2. The waveform before and after normalization is illustrated in Figure 5 2 2.

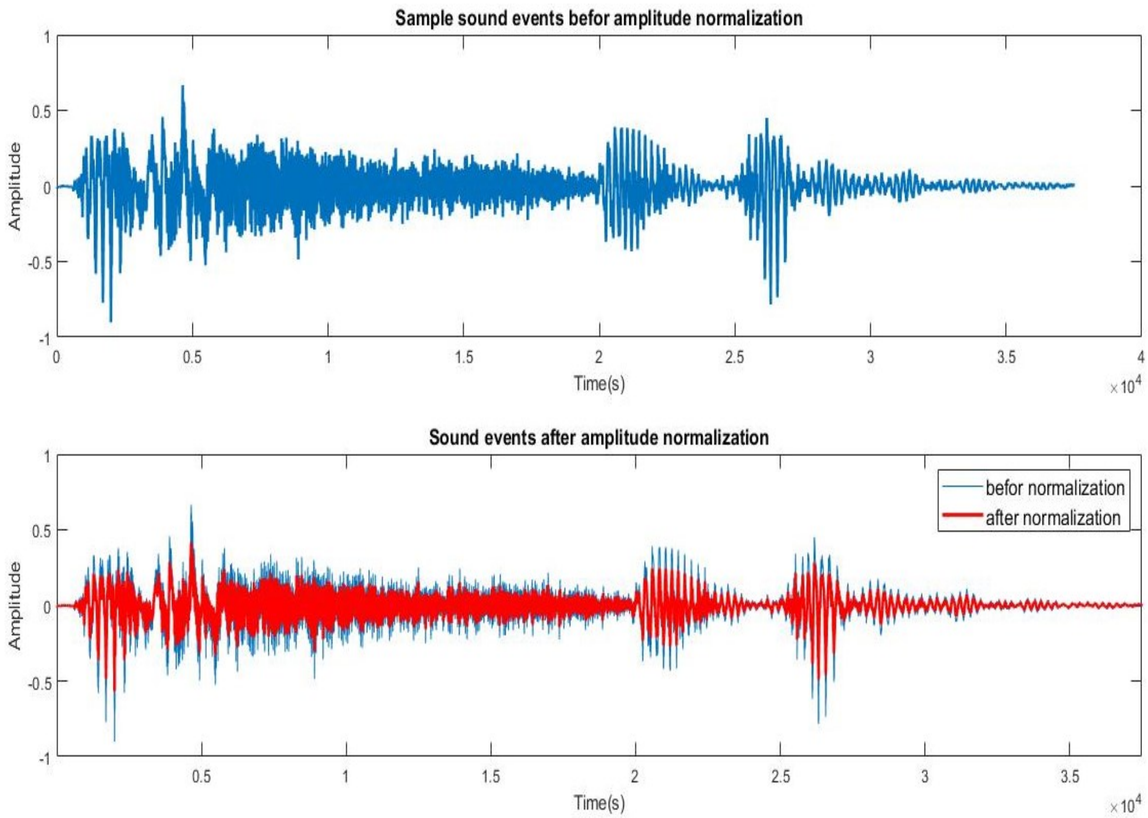


Figure 5.2-2. Sample waveform and its normalization result.

The amplitude of the sound events has been normalized by different scaling factors based on its waveform variation. After normalization, a pre-emphasis was used to enhance the magnitude of the sound events' higher-frequency components. The filtered sound events were segmented into a 100ms non-overlapping block. After segmentation, the energy of each segment was calculated, and a segment with a pick value was chosen as shown in Figure 5.2-3.

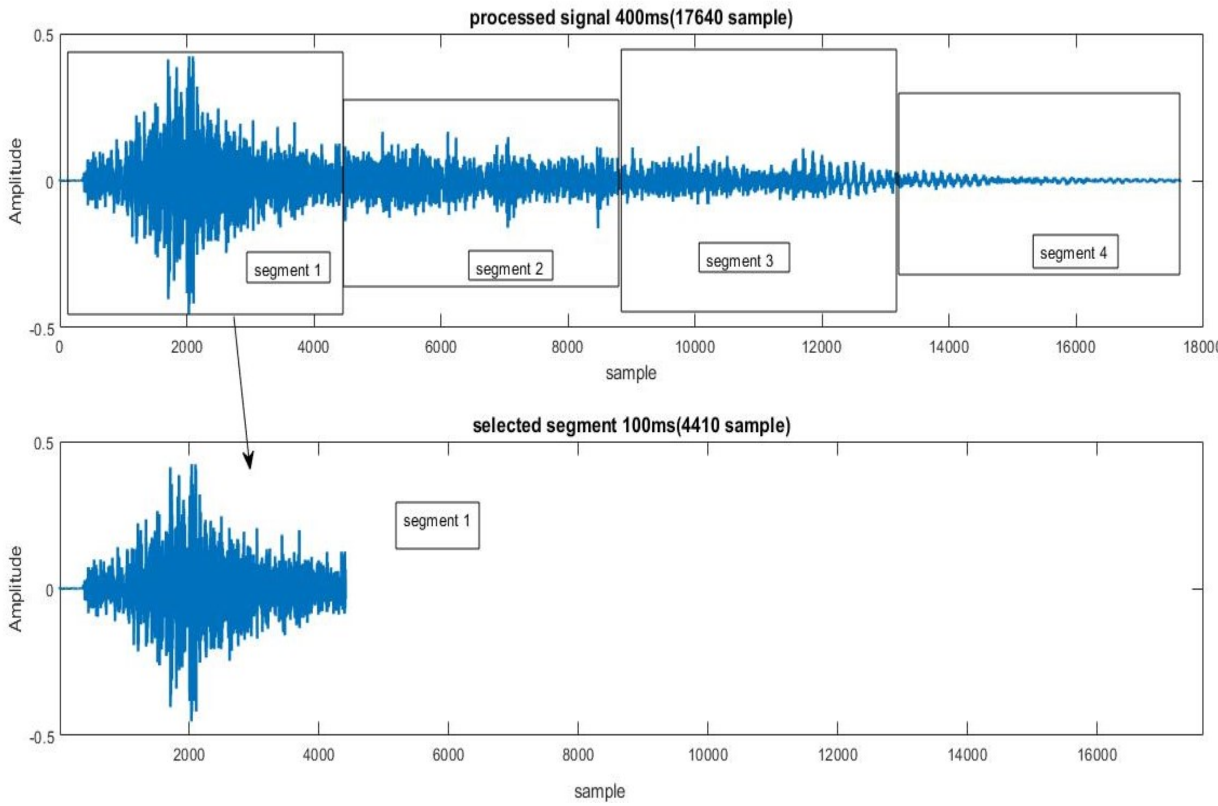


Figure 5.2-3. Sample cough event and segmentation result.

This segmentation process reduces the computational complexity of processing all segments of a sound event, it is possible to represent a sound event with a pick-value segment and feature extraction was performed on this segment.

5.3. Results of Feature Extraction process

For cough detection, a segment of each input sound event was split into frames of short length (40ms duration with 50% overlap), considering it as stationary signal and thus spectral analysis was applied to it. The 100ms segment in Figure 5.2-3 was split into 4 overlapped frames. By taking into account discontinuities at window edges, the frame was overlaid to smooth the transition.

The spectral leakages at the edges of the frames were eliminated by windowing all frames using hamming window. A framed signal, as shown in Figure 5.3-1 (b), was multiplied with the Hamming window shown in Figure 5.3-1(c) before the fast Fourier transform.

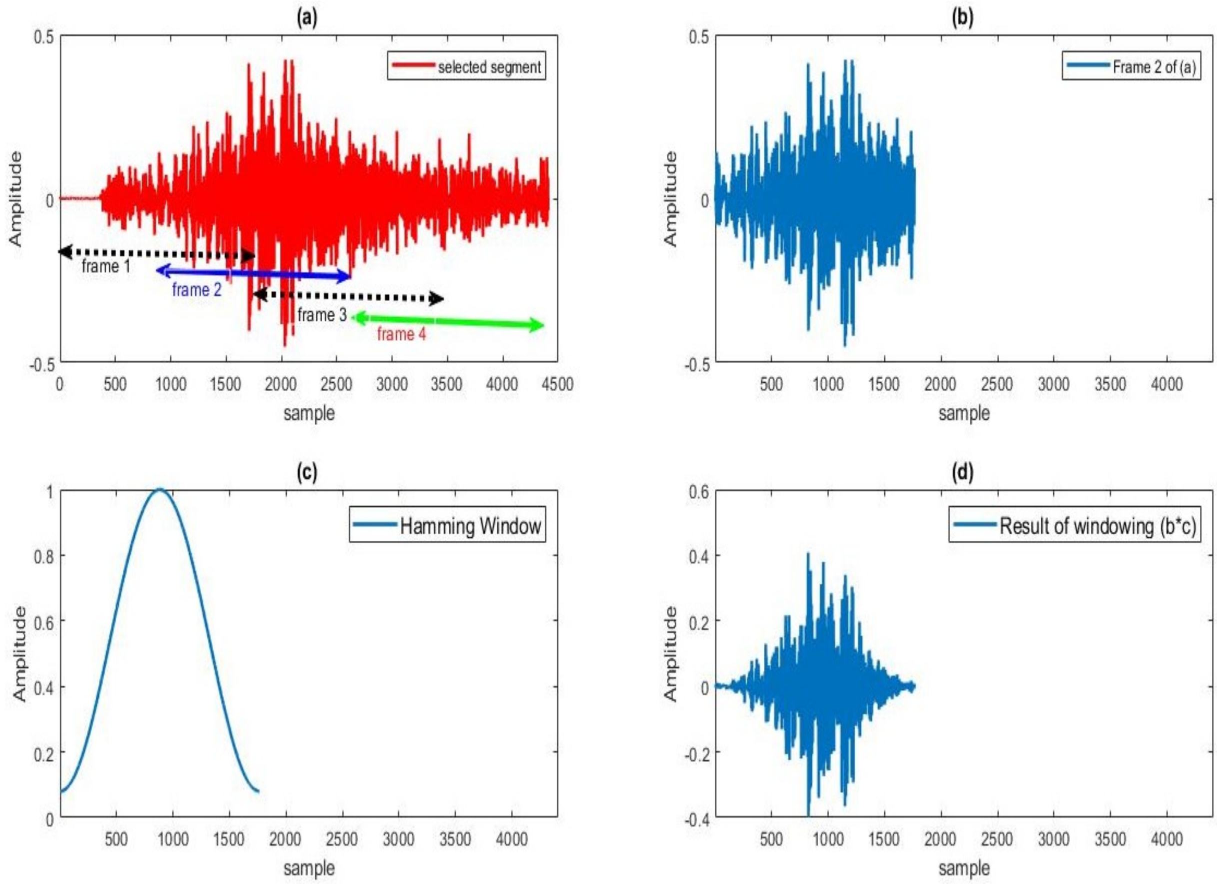


Figure 5.3-1. (a) Overlapped frames, (b) frame 2, (c) window, (d) windowed frame.

Figure 5.3-1 (d) shows a framed signal after windowing, which reduces the effects of FFT leakage. FFT was calculated after windowing, and the time-domain frames were transformed to the spectrum (frequency-domain). The Fourier transform function is similar to that of the cochlea in the human ear. The function of the cochlea is to transform raw sound vibrational waves into frequency domain neural impulses.

The frequency range in the Fourier spectrum was large and followed a linear scale, whereas the human ear's frequency perception follows a non-linear scale. The Mel-Scale filter Bank, which describes the frequency perception of the human ear was used as a passband filter at this stage. An example of Mel-scale filter banks is shown in Figure 5.3-2 (this is purely for demonstration purposes; the spacing and number of filters are not similar to those used in this study due to the difficulty of visualizing and plotting them).

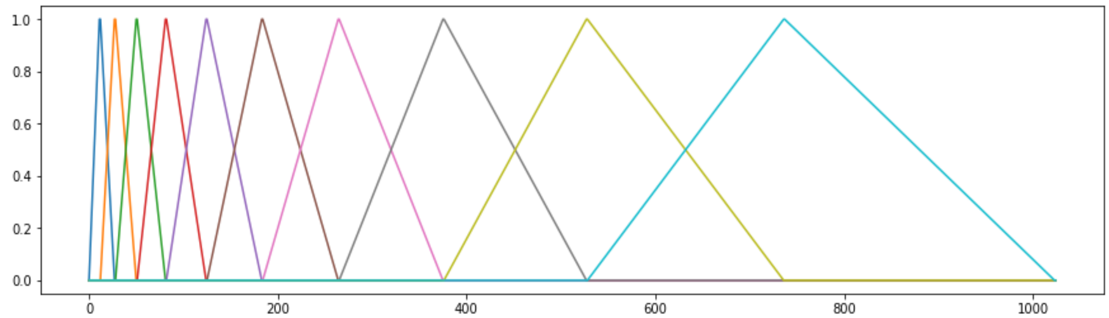


Figure 5.3-2. Mel-scale triangular filter banks.

The FFT spectrums of frames were passed through Mel-filter banks, which were spaced non-uniformly with Mel scale (equation 4.10), to obtain the perceptual frequency, which can accurately mimic auditory processing.

The human ear is less sensitive to small-amplitude changes of sound, but the amplitude of the FFT result had a dynamic range. By calculating the logarithm of the spectrum, the dynamic range was compressed, making the estimated spectrum less responsive to small amplitude changes. Finally, the DCT was carried out on the log Mel-spectrum to convert it to cepstrum. The DCT decorrelates the filter bank coefficients and generates a compressed representation of the log filter banks. The set of 13 cepstrum coefficients (MFCCs) for each window were held for cough sound analysis. Table 5.3-1 displays the MFCCs of the cough event described in Figure 5.3-1 (a).

Table 5.3-1. MFCCs of a sample cough event.

Frames	MFCCs (MFCC1- MFCC13)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-8.14	-0.46	1.09	-0.33	-0.15	-0.51	-0.37	-0.03	-0.31	-0.65	-0.05	0.14	-0.05
2	-6.44	-1.34	0.61	-0.72	0.36	-0.51	-0.55	-0.19	-0.06	-0.56	-0.25	0.11	0.25
3	-7.40	-1.64	0.58	-0.62	0.44	0.20	-0.49	-0.22	-0.01	-0.42	-0.46	0.08	0.24
4	-7.88	-1.13	0.54	-0.14	0.23	0.16	-0.77	-0.56	0.06	-0.21	-0.41	0.06	-0.12

These MFCCs are used as features for the classifier models. A vector containing 4X13=52 MFCC features was formed from the 100ms segment of every sound event. This feature vector was then supplied to a classifier for the cough detection system.

The second phase of this study was classifying the detected cough event as TB disease and non-TB disease cough. The method used to classify cough events as TB and non-TB diseases used the CCD dataset and it was extremely similar to the technique of cough detection. The MFCC features were extracted from the cough events obtained from the cough detection system. For cough classification, a segment of each input sound event was split into 20ms frames with 50% overlap. The 100ms segment was split into nine overlapped frames because this parameter used above for cough detection has changed slightly. The modification enhances the effectiveness of classifier models for differentiating Tb coughs from other coughs, but it comes at the cost of increased computational complexity compared to the cough detection system. The 100ms segment of every (TB/non-TB) cough event was used to create a vector with 9X13=117 MFCC features. This feature vector was then fed into a cough classification system classifier.

5.4. Hyper-parameters Optimization of the Models

The hyper-parameters of ANN were optimized using a brute force method called grid search to achieve the best results. The number of hidden layers and the appropriate learning algorithms were selected by comparing their performance. The other hyper-parameters were adjusted or tuned using a variety of parameter combinations, and the validation method was k-fold cross-validation. The first main hyper-parameter for ANN was the number of hidden layers. The scoring method used during hidden layer optimization was root means square error. Figure 5-4.1 depicts the optimization surface for the number of hidden layers with their root means square error. The number of hidden layers was selected 28 since it had the minimum root means square error value as shown in Figure 5.4-1.

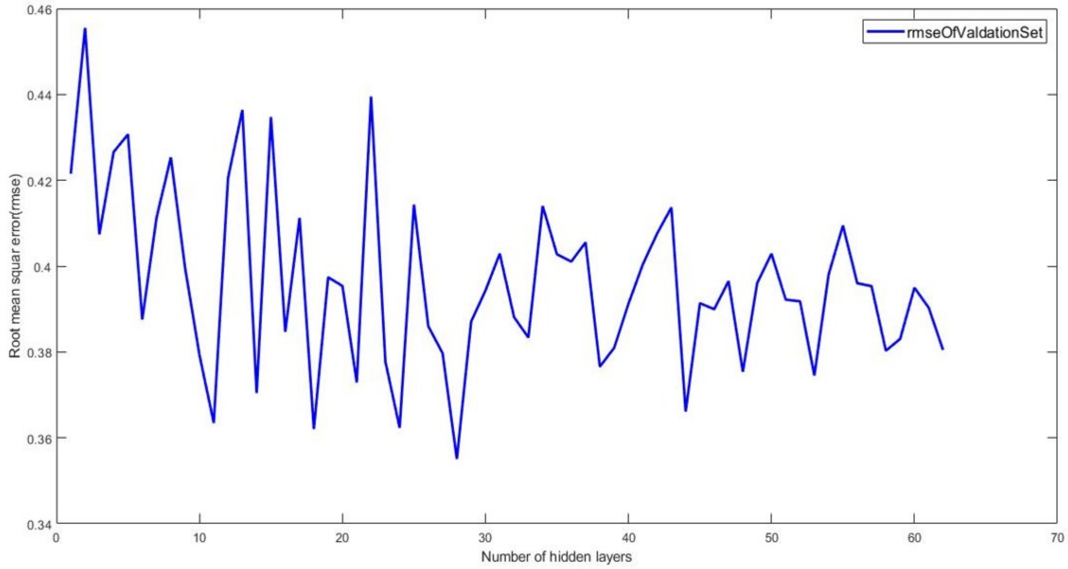
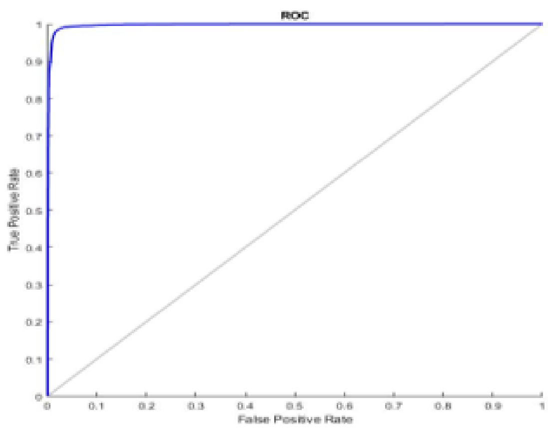
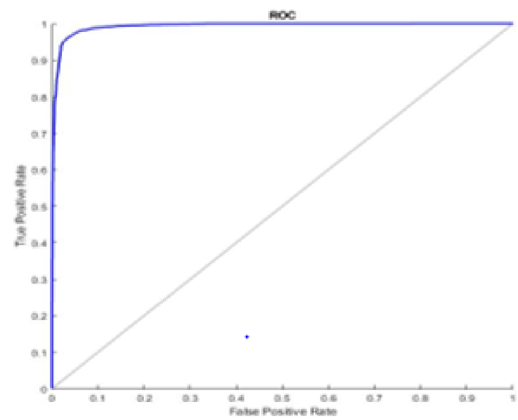


Figure 5.4-1. The rmse value with the number of hidden layers.

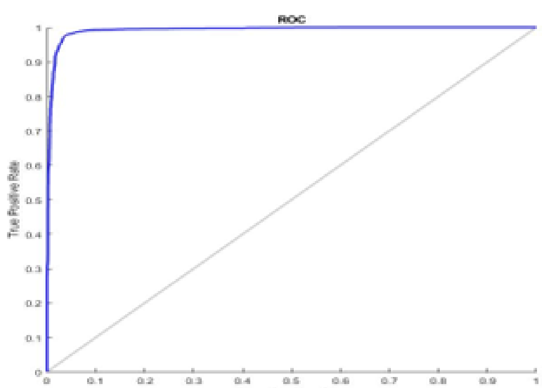
The second main hyper-parameter for ANN was the learning algorithm. Figure 5-4-2 shows the receiver operating characteristic (ROC) curve of different ANN learning algorithms.



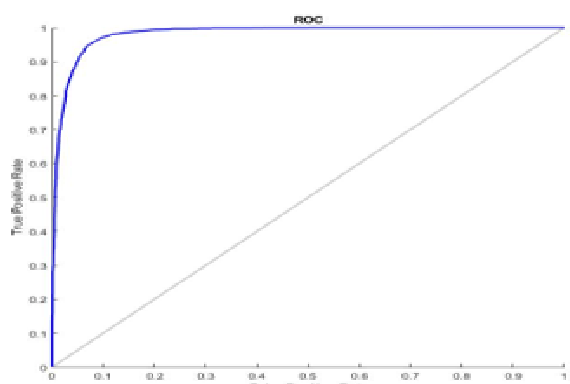
ROC of Levenberg-Marquardt's algorithm



ROC of One Step Secant algorithm



ROC of Resilient Backpropagation algorithm



ROC of Gradient Descent algorithm

Figure 5.4-2. ROC of ANN learning algorithms for cough detection

Table 5.4-1 depicts the ANN learning algorithms with their accuracy for both cough detection and classification systems.

Table 5.4-1. ANN learning algorithms with their accuracy.

Systems	Levenberg-Marquardt (LM)	Gradient Descent	One Step Secant	Resilient Backpropagation
Cough Detection	**98.2%**	93.7%	96.2%	96.9%
Cough Classification	**92.3%**	80.9%	88.0%	92.1%

Levenberg-Marquardt's learning algorithm was selected for both cough detection and classification with high accuracy of 98.2% and 92.3% respectively. Because of its ability to solve difficult nonlinear problems, the Levenberg-Marquardt algorithm outperformed other algorithms. It is a hybrid of Gauss-Newton and the gradient descent method; when the parameters are near to their optimal value, it acts more like the Gauss-Newton, and when they are far from their objective function, it works more like a gradient-descent.

The other hyperparameter is epochs, which represent one complete pass of the data (training data) through the learning algorithm. Figure 5-3-4 depicts the optimization surface for the performance of the model with their number of epochs for the cough detection model.

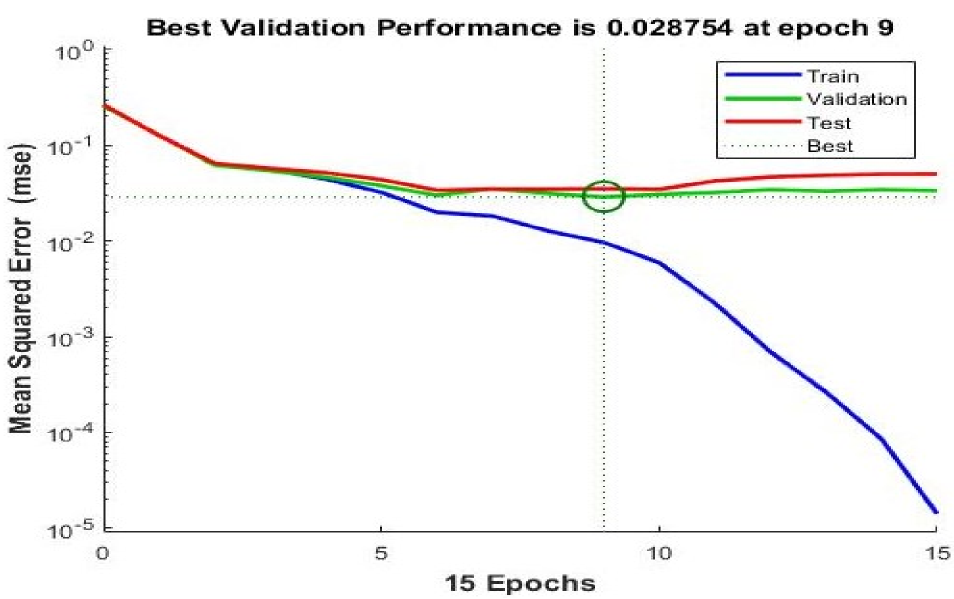


Figure 5.4-3. Performance with number of epochs for cough detection.

The best validation performance for cough detection was 0.028 at epoch 9 and the network was trained for 15 epochs. Similarly, figure 5-3-4 depicts the optimization surface for the performance of the model with their number of epochs for the cough classification model.

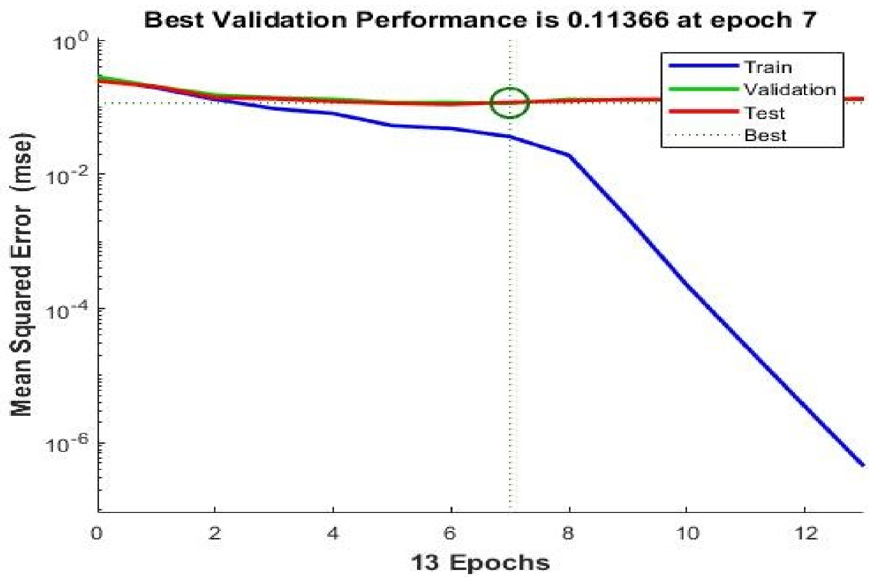


Figure 5.4-4. Performance with the number of epochs for cough classification.

The best validation performance for cough classification was 0.11367 at epoch 7 and the network was trained for 13 epochs.

Grid search was also used to optimize the hyper-parameters of SVM. Obtaining the desired hyperplane (kernel) for optimizing the distance between any two classes is the key parameter for SVM. Table 5.4-2 lists the SVM kernel functions for cough detection and classification systems, along with their accuracy.

Table 5.4-2. SVM kernel functions with their accuracy for both cough detection and classification system.

Cough Systems	Linear	Radial Basis Function (RBF)	Polynomial power 2	Polynomial power 3
Detection	86.9%	**97.1%**	91.7%	91.8%
Classification	66.5%	**89.4%**	75.4%	75.4%

The position of the hyperplane is influenced by data points called support vectors that are closer to the hyperplane. RBF kernel outperforms other kernels as shown in Table 5.4-2 due to its function space flexibility in projecting high-dimensional non-linear data. RBF kernel was selected for both cough detection and classification with high performance to project the data.

5.5. Performance Comparison of the Selected ANN and SVM Models

The accuracy and F1-score validation metrics were used to compare the performance of the two models (ANN and SVM with their selected optimal parameters). The accuracy (Equation 5.1) is calculated by dividing the total number of correct classifications by the total number of classifications.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

Where TP is True Positives, the number of positive samples in which the system classified as positive, TN is True Negatives, the number of negative samples in which the system is classified as negative, FP is False Positives, the number of positive samples in which the system classified as negative, and FN is False Negatives, the number of negative samples in which the system classified as positive. Accuracy is a simple metric, it is inadequate in certain situations, such as with unbalanced datasets, to evaluate a model's efficiency solely based on accuracy. As a result, F1-score was used to evaluate the models. The F1-score is a metric for determining the effectiveness of a classifier. It is the harmonic mean of precision and recall, in such a way that the lowest value is emphasized, the formula is:

$$F1 - score = \frac{2 * (precision \ * recall)}{precision \ + recall} \tag{5.2}$$

Where, precision (positive predictive value) is calculated by dividing the number of TP results by the total number of positive results, which includes those that were incorrectly classified and obtained using Equation 5.3, and recall (also called sensitivity) is the number of TP results divided by the number of all samples that classified as positive, which can be calculated using Equation 5.4.

$$precision = \frac{TP}{TP + FP} \quad (5.3)$$

$$recall = \frac{TP}{TP + FN} \quad (5.4)$$

The F-1 score is a good indicator for classification problems on unbalanced datasets, and it's the metric to use when you're looking for a good balance of recall and precision

Based on the ANN and SVM, the full dataset used in the current study has 6476 and 3238 sound events for detection and classification respectively, from the datasets 70% were used for training, 15% were used for validation, and 15% were used for the test. The confusion matrix for cough detection ANN is displayed in Figure 5.5-1.

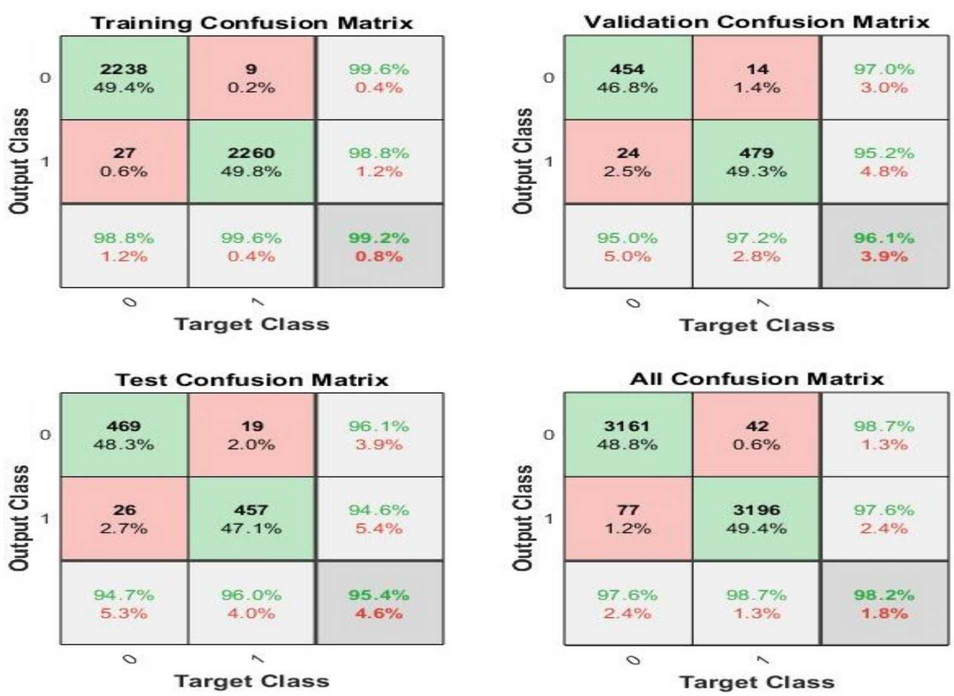


Figure 5.5-1. The confusion matrix for cough detection ANN

The accuracy and F1-score of the models were evaluated on overall sets. Table 5.3-1 shows the performance of cough detection and cough classification systems on overall sets for ANN and SVM classifiers.

Table 5.5-1. Overall accuracy and F1-score of the models.

Models	Cough Detection		Cough Classification	
	Accuracy	F1-score	Accuracy	F1-score
ANN	**98.2%**	**98.1%**	**92.3%**	**87.7%**
SVM	97.1%	96.6%	89.4%	82.2%

In both cough detection and cough classification, the two models, ANN and SVM, achieved acceptable accuracy and F1-score values. This indicates that MFCCs, which are based on perceptual models of the human auditory system, can uniquely represent sound signals using feature engineering effort. Due to its capacity to adjust the size of a network, the ANN outperforms SVM models in solving high dimensionality nonlinear problems.

5.6. Comparison of the proposed method to a previous study

Previously, the most similar study, automatic detection of tuberculosis was designed to distinguish TB positive from healthy controls using cough sounds analysis (*Botha et al 2018)* [4]. The research gaps of the previous work done by *Botha et al 2018*, were one of the impetus for us for conducting this research. Both studies were conducted to diagnose TB using the cough sound of patients. Table 5.6-1 compares the methods used in this research to a previous study (*Botha et al 2018).*

Table 5.6-1. Comparison of this study to the most closely related one.

Comparison parameters	Most similar research (*Botha et al.2018)*	This work
Datasets	• Only data from TB patients and healthy individuals voluntary cough were collected, but data from patients with diseases similar to TB were not included. • It severely limits the classification robustness of various diseases. • The dataset contained 746 cough events collected from 17 tuberculosis patients and 21 healthy individuals.	• Collected cough data was the multitude of twelve different pulmonary diseases such as TB, pneumonia, B. Asthma, etc. • It is more robust in terms of disease classification. • The dataset contained 6476 sound events collected from 71 patients suffering from twelve different pulmonary diseases.

Recording Environment	• Cough sounds were recorded in a specially designed facility under controlled environments. • There was no background noise, and the silences have a fixed level of energy. • Differ from realistic conditions.	• Cough sounds were recorded in clinical settings in real-world environments. • There were background noises, and the silences had different levels of energy. • Similar to real-world conditions.
Preprocessing and feature engineering efforts.	• No need for preprocessing effort to remove background noise and silences. • Without any preprocessing, features were extracted directly from cough sounds. • It is difficult to apply in practical applications (real-world scenarios).	• Before feature extraction, background noise was removed using preprocessing efforts. • The silences in the recordings were removed based on their energy standard deviation with the dynamic energy threshold value. • Designed for a real-world scenario.
Fully automated	• Only cough classification. • Humans manually detected and extract cough events among other non-cough sounds such as laughing, speech, throat clearing, etc. • Time-consuming, difficult to apply in practice, and requires additional effort.	• Fully automated, have both cough detection and cough classification. • The system extracts cough events from the recordings and categorizes them as TB or non-TB cough. • Simple and efficient to put into practice.
Segmentation and feature	• The MFCCs features were extracted from the entire cough event, with no segmentation. • Complex in terms of computation.	• MFCCs features extraction were performed on the selected segment after segmentation • Effective in terms of computation
Experimental Result	• 82% accuracy for classification	• 92.3% accuracy for classification

We note that the methods we used in this study outperform previous work on all of the comparison parameters listed in table 5.6-1. The reason for this is that the goal of this research is to fill gaps in previous work. Even though the experiments were conducted under diverse circumstances (different datasets and feature engineering efforts), a comparison of the accuracy of the two studies in table 5.6-1 shows that the proposed work achieves better performance. This indicated that the robust pre-processing technique and feature engineering efforts used in this study are essential for developing cough diagnosis systems in real-world scenarios. We can consider the proposed methods in this study to be an advancement of previous work (*Botha et al.2018*).

Chapter 6
Conclusion and Recommendation

6.1. Conclusion

The use of cough data obtained from patients with various pulmonary diseases and recorded using three different recorders in a clinical setting in a moderately noisy environment to develop a robust system for TB diagnosis was investigated in this study. The dataset utilized in this study was compared to data from other studies, and the data in this study is robust in terms of dataset size, the number of patients, disease diversity, and recording device diversity. Audio signal processing was done to extract the robust MFCC features, which were achieved by pre-processing and feature engineering efforts. An ANN and SVM were used to create an automated cough detection and classification system, and the ANN with the best performance was selected. The ANN algorithms for cough detection achieved an accuracy of 98.2% and F1-score of 98.1%, and for TB/non-TB classification, an accuracy of 92.3% and an F1-score of 87.7%. This indicates that ANN is a better classifier for high-dimensional and non-linear complex data like cough sound classification. As a result, we can conclude that the system can differentiate between TB patients' coughs and coughs caused by other lung diseases that aren't audible by a human listener. This shows that using cough sound MFCCs features as a diagnostic tool for TB testing can be a viable solution.

This research contributes to the advancement of automatic cough analysis methods for TB diagnosis. The dataset prepared in this study will be helpful for future research in this area. In the cough analysis areas, this study contributed by developing a robust automatic system for cough detection from continuous sound recordings, as well as a robust cough sound classification method for TB diagnosis. The development of a robust diagnostic system was aided by the use of various sound recorders to obtain the cough signal from patients with various pulmonary diseases in a moderately noisy environment. Signal preprocessing was performed well to compensate for the differences in data collected from various recorders and to remove noise from the recordings. Using feature engineering of sound signals, unique features of TB patients' coughs were extracted to differentiate them from non-TB pulmonary patients. The integration of an automated cough detection system with a cough classification system was another achievement of this research. All of these efforts were made to develop a robust system and overcome the challenges of TB diagnosis in resource-constrained environments.

Finally, this study investigates that cough sound analysis can be used to diagnose tuberculosis. However, before putting the proposed approach into practice, it is important to note that the promising outcomes of this research should be followed up with further research.

6.2. Future Development

To advance this research, the dataset must first be analyzed and given attention. Even if the dataset used in this research was robust and comparable to those used in previous similar studies, the development of the best reliable database remains a priority. A comprehensive dataset could be built by increasing the dataset's size, including coughs from patients with other lung diseases such as Coronavirus (COVID-19), including data from pediatric patients, and collecting sound recordings from various settings.

In some clinics, the cough recording space is too small, resulting in echo, an echo cancellation algorithm from recordings would be useful to improve the cough processing algorithm. More research should be conducted in the future to resolve these issues, and the method should be developed into a smartphone application.

References

[1] World Health Organization, "Global Tuberculosis Report," ISBN 978-92-4-1565646, Geneva, 2020.

[2] R. Horsburgh, C. E. Barry, and C. Lange, "Treatment of Tuberculosis," *The New England Journal of Medicine,* no. 373:2149-60, 2015.

[3] S. Singer-Leshinsky, MSEd and RPAC, "Pulmonary tuberculosis: Improving diagnosis and management," *American Academy of Physician Assistants,* vol. 29, no. 2, 2016.

[4] G. Botha, G. Theron, R. Warren, M. Klopper, K. Dheda, P. Helden and T. Niesler, "Detection of tuberculosis by automatic cough sound analysis," *Physiological Measurement, IOP,* vol. 39, 2018.

[5] I. Peate, "Anatomy and physiology, 10. The respiratory system," *British Journal of Healthcare Assistants,* vol. 12, no. 4, pp. 178-181, 2018.

[6] J. Tu, K. Inthavong and G. Ahmadi, "The Human Respiratory System," in *Computational Fluid and Particle Dynamics in the Human Respiratory System*, SN - 978-94-007-4487-5, 2013, pp. 19-44.

[7] A. Patwa and A. Shah, "Anatomy and physiology of respiratory system relevant to anesthesia," *Indian Journal of Anaesthesia,* vol. 59, no. 9, 2015.

[8] K. Kon and M. Rai, The Microbiology of Respiratory System Infections, Academic Press, 2016.

[9] S. Cinaroglu, "Prevalence of upper respiratory tract infections and associated factors among children in Turkey," *Journal for Specialists in Pediatric Nursing-Wiley,* 2019.

[10] Fun Science, "Human Respiratory System," [Online]. Available: https://funscience.in/human-respiratory-system/. [Accessed 18 9 2020].

[11] H. Aung, A. Sivakumar, K. Gholami, P. Venkateswaran, B. Grain and S. Md, "An Overview of the Anatomy and Physiology of the Lung," in *Nanotechnology-Based Targeted Drug Delivery Systems for Lung Cancer*, Elsevier, 2019, pp. 1-20.

[12] L. Sherwood, "The Respiratory System," in *Human Physiology From Cells to Systems, Seventh Edition*, Cengage Learning, 2014, pp. 461-510.

[13] C. Catharina, M. Boehme, P. Nabeta, D. Hillemann, N. Mark P, S. Shenai, F. Krapp, J. Allen, R. Tahirli and R. Blakemore, "Rapid molecular detection of tuberculosis and rifampin resistance," *New England Journal of Medicine, N ENGL J MED,* vol. 363, no. 11, 2010.

[14] S. Raju, R. Revathi, and M. Mahalingam, "Exploration of Cough Recognition Technologies Grounded on Sensors and Artificial Intelligence," on *Internet of Medical Things for Smart Healthcare*, Springer, 2020, pp. 193-215.

[15] J. Widdicombe, "Neurophysiology of the cough reflex," *European Respiratory Journal,* vol. 8, pp. 1193-1202, 1995.

[16] Korpa´s, J. Sadlonova and M. Vrabec, "Analysis of the cough sound: An overview," *Pulmonary Pharmacology,* vol. 9, pp. 5-6, 1996.

[17] S. Birring and A. Spinou, "How best to measure cough clinically," in *Current opinion in pharmacology*, vol. 22, Elsevier, 2015, pp. 37-40.

[18] E. Eric, "Form and function of the mammalian inner ear," *Journal of Anatomy,* vol. 228, pp. 324-337, 2016.

[19] P. Ziegler, P. Wahl, and P. Eberhard, "Vibration of the Basilar Membrane and Fluid Pressure Distribution in the Human Cochlea," in *GAMM, Proceedings in Applied Mathematics and Mechanics (PAMM)*, DOI 10.1002/pamm.10084, 2017.

[20] T. Corp, "YAMAHA," 2016. [Online]. Available: https://uk.yamaha.com/en/products/contents/proaudio/docs/audio_quality/04_audio_quality.html.

[21] F. Hu and X. Cao, "An Auditory Feature Extraction Method for Robust Speaker Recognition," in *IEEE 14th International Conference on Communication Technology*, doi: 10.1109/ICCT.6511354., 2012.

[22] A. Huet, C. Batrel, Y. Tang, G. Desmadryl, J. Wang, J.-L. Puel and J. Bourien, "Sound coding in the auditory nerve of gerbils," *Hearing Research: Elsevier,* vol. 338, pp. 32-39, 2016.

[23] D. Riordan, P. Doody, and J. Walsh, "Modelling of the human perception of sound using ANNs," in *25th IET Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014)*, IEEE, 2014.

[24] H. Fastl and E. Schorer, "Critical bandwidth at low frequencies reconsidered," in *Auditory Frequency Selectivity*, Boston, Springer, 1986, pp. 311-318.

[25] J. Smith, H. Ashurst, S. Jack, A. Woodcock, and J. Earis, "The description of cough sounds by healthcare professionals," *BioMed Centra,* vol. 2, no. 1, January 2006.

[26] J. Smith and A. Woodcock, "New developments in the objective assessment of cough," *Lung,* vol. 186, no. 1, pp. 48-54, 2007.

[27] P. Piirila and A. Sovijarvi, "Differences in Acoustic and Dynamic Characteristics of Spontaneous Cough in Pulmonary Diseases," *CHEST: ScienceDirect,* vol. 96, no. 1, pp. 46-53, 1989.

[28] D. Miranda, A. Diacon and T. Niesler, "A Comparative Study of Features for Acoustic Cough Detection Using Deep Architectures," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, doi: 10.1109/EMBC..8856412., 2019.

[29] M. Solinski, M. Lepek and L. Koltowski, "Automatic cough detection based on airflow signals for portable spirometry system," *Informatics in Medicine Unlocked, Elsevier,* 2020.

[30] S. Barry, A. Dane, A. Morice, and A. Walmsley, "The automatic recognition and counting of cough," *BioMed Central,* vol. 2, no. 8, 2006.

[31] T. Brian, G. Comina, S. Larson, M. Bravard, J. López, and H. Robert, "Cough Detection Algorithm for Monitoring Patient Recovery from Pulmonary Tuberculosis," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 2011.

[32] J.Liu, M.You, Z.Wang, G.Li, X.Xu and Z.Qiu, "Cough event classification by pre-trained deep neural network," *BMC Medical Informatics and Decision Making,* vol. 15, no. 2, 2015.

[33] J. Martinek, P. Klco, M. Vrabec, T. Zatko, M. Tatar, and M. Javorka, "Cough sound analysis," *ACTA MEDICA MARTINIANA,* vol. 1, 2013.

[34] A.Raj and S.Birring, "Clinical assessment of chronic cough severity.," *Pulmonary Pharmacology and Therapeutics,* vol. 20, p. 34–337, 2007.

[35] S. Birring, T. Fleming, S. Matos, and A. Raj, "The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough," *European Respiratory Journal,* vol. 31, no. 5, 2008.

[36] R. Pramono, A. Imtiaz and E. Rodriguez-Villegas, "Automatic Cough Detection in Acoustic Signal using Spectral Features," *IEEE,* vol. 19, no. 978-1-5386-1311-5, 2019.

[37] P. Kadambi, A. Mohanty, H. Ren, J. Smith, K. McGuinnes, K. Holt, A. Furtwaengler, R. Slepetys, Z. Yang, J.-s. Seo, J. Chae, Y. Cao, and V. Berisha, "TOWARDS A WEARABLE COUGH DETECTOR BASED ON NEURAL NETWORKS," *IEEE,* no. 978-1-5386-4658-8/18, 2018.

[38] A.Shankar, V.Bhateja, A.Srivastava and A.Taquee, "Continuous Wavelets for Pre-Processing and Analysis of Cough Signals," in *Proc. of Third International Conference on Smart Computing and informatics (SCI)*, pp. 1-8, 2018.

[39] V. Bhateja, A. Taquee, and K. Sharma, "Pre-Processing and Classification of Cough Sound in Noisy Environment using SVM," *4th International Conference on Information Systems and Computer Networks (ISCON), IEEE,* no. 978-1-7281-3651-6/19, Nov 21-22, 2019.

[40] J. Monge-Álvarez, C. Hoyos-Barceló, P. Lesso and P. Casaseca-de-la-Higuera, "Robust Detection of Audio-Cough Events using local Hu moments," *IEEE Journal of Biomedical and Health Informatics,* 2018.

[41] S. Khomsay, R. Vanijjirattikhan, and J. Suwatthikul, "Cough detection using PCA and Deep Learning," *IEEE,* no. 978-1-7281-0893-3/19, 2019.

[42] J. Alvarez, C. Hoyos-Barcelo, L. M. San-Jose-Revuelta, and P. Casaseca-de-la-Higuera, "A Machine Hearing System for Robust Cough Detection Based on a High-Level Representation of Band-Specific Audio Features," *IEEE Transactions on Biomedical Engineering,* 2018.

[43] V. Swarnkar, U. Abeyratne, Y. Amrulloh, and A. Chang, "Automated algorithm for wet/dry cough sound classification," *IEEE Engineering in Medicine and Biology Society,* no. 34th Annual International Conference of the IEEE EMBS, 2012.

[44] Y. Amrulloh, U. Abeyratne, V. Swarnkar, and R. Triasih, "Cough Sound Analysis for Pneumonia and Asthma Classification in Pediatric Population," *IEEE, "6th International Conference on Intelligent Systems, Modelling and Simulation",* 2015.

[45] R. Pramono, S. Imtiaz and V. Rodriguez, "A cough-based algorithm for automatic diagnosis of pertussis," *PLoS ONE,* vol. 11, no. 9, 2016.

[46] C. Infante, D. Chamberlain, R. Fletcher, Y. Thorat, and R. Kodgule, "Use of Cough Sounds for Diagnosis and Screening of Pulmonary Disease," *IEEE,* no. 978-1-5090-6046-7/17, 2017.

[47] J. REISS, "A Meta-Analysis of High-Resolution Audio Perceptual Evaluation," *Audio Engineering Society,* vol. 64, no. 6, 2016.

[48] E. Rao, P.Muralidhar and S.Raghuramakrishna, "Audio Equalizer with Fractional Order Butterworth Filter," *International Journal of Engineering and Management Research,* vol. 5, no. 5, pp. 266-272, 2015.

[49] M. Cohen-McFarlane, R. Goubran, and F. Knoefel, "Comparison of Silence Removal Methods for the Identification of Audio Cough Events," *IEEE,* no. 978-1-5386-1311-5/19, 2019.

[50] V. Swarnkar, U. Abeyratne, Y. Amrulloh, C. Hukins, R. Triasih and A. Setyati, "Neural network-based algorithm for automatic identification of cough sounds," *Annu Int Conf IEEE Eng Med Biol Soc,* 2013.

[51] F. Nargesian, H. Samulowitz, U. Khurana, E. Khalil, and D. Turaga, "Learning Feature Engineering for Classification," in *International Joint Conference on Artificial Intelligence*, University of Toronto, 2017.

[52] S. Jain and B. Kishore, "Comparative study of voiceprint Based acoustic features: MFCC and LPCC," *International Journal of Advanced Engineering, Management and Science (IJAEMS),* vol. 3, no. 4, 2017.

[53] P. SAI, N. RAO, N. KUMAR, P. BRAHMAIAH, and D. AJAY, "Cough classification tool for early detection and recovery monitoring of tuberculosis and asthma," in *4th International Conference on 'Computing, Communication and Sensor Network, CCSN2015*, 2015.

[54] P. Prithvi and K. Kumar, "Comparative Analysis of MFCC, LFCC, RASTA -PLP," *International Journal of Scientific Engineering and Research (IJSER),* vol. 4, no. 5, 2016.

[55] S. Mermelstein and P. Davis, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE,* Vols. VOL. ASSP-28, NO. 4, AUGUST 1980.

[56] J. Knocikova, J. Korpas, M. Vrabec, and M. Javorka, "Wavelet Analysis of Voluntary Cough Sound in Patients with Respiratory Diseases," *Journal of Physiology Pharmacology,* vol. 59, pp. 331-340, 2008.

[57] A. C. Bhatnagar, L. Sharma, and R. Kumar, "ANALYSIS OF HAMMING WINDOW USING ADVANCE PEAK WINDOWING METHOD," *International Journal of Scientific Research Engineering &Technology (IJSRET),* vol. 1, no. 4, July 2012.

[58] Y. Shi, H. Liu, Y. Wang, M. Cai, and W. Xu, "Theory and Application of Audio-Based Assessment of Cough," *Hindawi,* no. 9845321/18, 2018.

[59] L. Muda, M. Begam and I. Elamvazuthi, "Voice recognition algorithms using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques," *Journal of Computing,* vol. 2, no. 3, 2010.

[60] C. Kaushik and A. Sahi, "artificial neural network-based model for orphan GPCRs. Neural. Comput.," *HOGPred,* Vols. 10.1007/s00521-016-2502-2506, p. 985–992, 2018.

[61] Smitha, S. Shetty, S. Hegde, and T. Dodderi, "Classification of Healthy and Pathological voices using MFCC and ANN," in *Second International Conference on Advances in Electronics, Computer, and Communications (ICAECC)*, IEEE, 2018.

[62] C. CORTES and V. VAPNIK, "Support-Vector Networks," *Machine Learning, Kluwer Academic Publishers,* vol. 20, no. 3, pp. 273-297, 1995.

Appendix A: Statements of Diagnosis Medical Report of patients from Physician

FELEGE HIWOT COMPRESSIVE SPECIALIZED HOSPITAL

STATEMENT OF DIAGNOSIS MEDICAL REPORT

PART A: Patient's Detail

Patient's Name: [redacted] Age 28 Sex: M

Diagnosis Date: 10/03/2020 Card No.: 80 8549

Address: Mota

PART B: Authorization

I hereby authorize my attending physician Dr. Yenew to furnish and disclose all facts concerning my medical condition that are within his or her knowledge and to allow inspection, and provide copies of any medical records concerning my medical condition that are under his or her control. This authorization shall be valid only for this particular research and I agree that a photocopy of this authorization shall be as valid as an original. I understand that if I do not sign this authorization my medical information cannot be disclosed. Also, I understand that I can revoke this authorization in writing. I also understand that the research project will keep confidential all of the information which is provided pursuant to this authorization, and that the information will be used solely for research purpose only.

Patient's\Guardian's Signature

10/03/2020

Date

PART C: Statement of Diagnosis Medical Report

DIAGNOSIS MEDICAL REPORT

Performed Tests:

, CBC, G-expert, X-ray

Findings:

PTB

PART D: Certification of Medical Condition

ANRS Bureau of Health

Felege Hiwot Referral Hospital

Appendix B: Informed Consent Form

መረጃ የመሰብሰብም ሆነ የማስራጨት ፈቃድ (Infromed Concent From)

የትዉልድ ዘመን/እድሜ:__________________________

ጾታ:__________________________

ታካሚ ከሆኑ ህክምና የጀመሩበት ጊዜ:__________________________

የሚኖሩበት ክልል/ከተማ:__________________________

ካርድ ቁጥር:__________________________

ሆስፒታል:__________________________

ምርምሩ ርዕስ : አቶማቲክ ካፍ አናሊስስ ሲስተም ፎር ዳያግኖሲስ ኦፍ ቱበርክሎሲስ (ዲዛይን እና ኢምፕልመንቴሽን)

ተመራማሪዉ ስምና አድራሻ : አምሳሉ ፈንቴ（+251945563670）

ምርምሩ ዋና ግብ: የዚህ ምርምር ዋና አላማ ዝቅተኛ ወጪ የሚጠይቅ የቱበርክሎሲ ህመም መለያ በማሺን ለርኒንግ ላይ ተመሰረተ ሶፍትዌር ማበልጸገ ነዉ። የበሽታዉን ምንነት ለመገንዘብ እና ማሽኑ እንዲረዳዉ ለማድረግ የህሙማንን ሳል ማይክራፎንን ወይም በድምጽ መቅጃን በመጠቀም ከቱበርክሎሲ ፣ ከሌሎች የመተንፈሻ አካል ህመምተኞች እና ከጤኛ ዎች የሳል ድምጽ መረጃዎችን መሰብሰብ ያስፈልጋል።

ረጃ ለመሰብሰብ የሚፈጀው ጊዜ : ከ20 እስከ 30 ደቂቃ

ረጃ በሚሰበሰበበት ጊዜ ከእኔ ምን እንደሚጠበቅብኝ ገለጻ ተደርጎልኛል።ከ10 እስከ 20 የሚደርሱ ቮለንተሪ እና ንቮለንተሪ የሳል ድምጽ መረጃዎችን ለመስጠት ተስማምቻለሁ፤ መረጃዉን የምሰጠዉ ሙሉ በሙሉ በገዛ ፈቃዴ ሲሆን ማኘኝዉም ጊዜ ማቋረጥ እችላለሁ። ይህ ምርምር ለሳይንሳዊ ጥናትና ህትመት ሲዉል የእኔን ማንነት በማይገልጽ መልኩ ዉ።የጥናቱን ዉጤት ለማወቅ ተመራማሪዉን በአድራሻዉ መጠየቅ እችላለሁ።

ኛዉምንም አይነት ተጨማሪ መረጃ ለማግኘት ወይም ቅሬታ ለማቅረብ የጎንደር ዩኒቨርቲ ሆስፒታል የምርምር ይሮክተሬት ወይም የአዲስ አበባ ሳይንስ ና ቴክኖሎጂ ዩኒቨርቲ የኤለክትሪካልና ኮምፒዉተር ምህንድስና ት/ክፍልን ማናገ ቻላል።

ላይ የተገለጸዉን አንብቤ ተረድቸዋለሁ። ለጥያቄየም ተገቢዉን ምላሽ አግኝቻለሁ። ይህንንም በፊርማየ አረጋግጣለሁ።

መረጃ ሰጪው ስምና ፊርማ:____

ተመራማሪዉ ስምና ፊርማ: አምሳሉ ፈንቴ____________________________

Appendix C: Ethical Clearance for Dataset

Ethical Clearance for Dataset

DATE: 12/22/2020

TO: Amsalu Fentie
Addis Ababa Science and Technology University (AASTU)

Yehualashet Megersa
Addis Ababa Science and Technology University (AASTU)

FROM: Dr Getahun Hailu Mengesha Out Patient Department Service
Phone: +251946128786
Bahir Dar, Ethiopia

RE: PERMISSION FOR USE OF DATA (Cough Sounds)

On behalf of Dr Getahun Out Patient Department Service:

- I give permission to the students to analyze the data (cough sounds) collected with the approval of our institution during their MSc and PhD dissertation.
- The students are given a right to use the data for article (journal and proceeding), thesis and dissertation publication.
- The students must fully and correctly reference Felege Hiwot Compressive Specialized Hospital as the source of the data.
- The students must maintain Felege Hiwot Compressive Specialized Hospital as a copyright holder of the data unless another written official letter is given in addition to this permission letter.
- The students must remove name, age and all attributes related to patient before using the data to keep privacy and confidentiality of the patient and if the students are failed to do this, they will be accountable for any damage caused to the patients.

Name Getahun Hailu Signature ________ አቶ ጌታሁን ኃይሉ Getahune Hailu ነርስ /BSC/

On behalf of Addis Ababa Science and Technology University:

- The University assure that the students use the data (cough sounds) for academic purposes stated above.

Name________________________________Signature __________________________

Appendix D: Certificate of Appreciation awarded to this research


CERTIFICATE OF APPRECIATION

AWARDED TO

Amsalu Fentie

IN RECOGNITION TO YOUR CONTRIBUTION

AS Oral Presenter

IN THE 1st AFRICAN GRADUATE STUDENTS CONFERENCE

(AGSC-BDU-2021)

DURING MAY 28-29, 2021

Dr. ESHEY KEBEDE
ACADEMIC VICE PRESIDANT

Dr. FIREW TEGEGNE
PRESIDENT

Appendix E: Publication

1. *Amsalu et al.,* "Automatic detection of cough using artificial neural network", Under review at International Journal of Computer Trends and Technology (IJCTT), Article id: *CTT21SEP101*.