University of South Dakota

USD RED

Dissertations and Theses

Theses, Dissertations, and Student Projects

12-2021

Respiratory Sound Analysis for the Evidence of Lung Health

Priyanka Sreerama

Follow this and additional works at: https://red.library.usd.edu/diss-thesis

Part of the Computer and Systems Architecture Commons

RESPIRATORY SOUND ANALYSIS FOR THE EVIDENCE OF LUNG HEALTH

By
Priyanka Sreerama
B.Tech., Anna University, 2007

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science

Department of Computer Science

Computer Science Program
In the Graduate School
The University of South Dakota
December 2021

Copyright by
PRIYANKA SREERAMA
2021
All Rights Reserved

The members of the Committee appointed to examine
the Thesis of Priyanka Sreerama
find it satisfactory and recommend that it be accepted.

DocuSigned by:
kc santosh
BBEF270EA578492...

Chairperson

DocuSigned by:
Dr. Douglas R Goodman, PhD
3C0E646BD1404F3...

DocuSigned by:
Dr. Arun Singh, PhD
C9532D809A79467...

ABSTRACT

Significant changes have been made on audio-based technologies over years in several different fields along with healthcare industry. Lung sound analysis is a potential source of noninvasive, quantitative information along with additional objective on the status of the pulmonary system. Recognition of abnormal respiratory sounds with a stethoscope known as auscultation is important in diagnosing respiratory diseases and providing first aid. At times, possibility of inaccurate interpretation of respiratory sounds happens because of clinician's lack of considerable expertise or sometimes trainees such as interns and residents misidentify respiratory sounds. We have built a tool to distinguish healthy respiratory sound from non-healthy ones that come from respiratory infection carrying patients. The audio clips were characterized using Linear Predictive Cepstral Coefficient (LPCC)-based features and the highest possible accuracy of 99.22% was obtained with a Multi-Layer Perceptron (MLP)-based classifier on the publicly available ICBHI17 respiratory sounds dataset [1] of size 6800+ clips. The system also outperformed established works in literature and other machine learning techniques. In future we will try to use larger dataset with other acoustic techniques along with deep learning-based approaches and try to identify the nature and severity of infection using respiratory sounds.

Thesis/Dissertation Advisor: ______________________
KC Santosh, PhD
Chair, Department of Computer Science

ACKNOWLEDGEMENTS

I would like to express my special gratitude, appreciation, and thanks to my thesis advisor Dr. KC Santosh, PhD (Computer Science, Chairperson), for his constructive instructions, immense support, guidance, and encouragement throughout my thesis term. Without his thoughtful encouragement and careful supervision, this thesis would never have taken shape. I would also like to thank my thesis committee members, Dr. Douglas R Goodman, PhD (Computer Science, Thesis Committee Member) and Dr. Arun Singh, PhD (Basic Biomedical Sciences, Thesis Committee Member) for their contributions to the direction and richness of this work and also for accepting the invitation to become the committee members. I am also thankful for Himadri Mukherjee for his guidance throughout my research and thesis work. I am also grateful to various other researchers and scientists who are working in the field of Computer vision and helping the community with their rich work and ideas. Finally, I extend my deepest thanks to my family who supported, encouraged, and motivated me every step of the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Our research work is published in Journal of Medical Systems. For further reference you can check out the research paper by clicking on the below link

https://doi.org/10.1007/s10916-020-01681-9

CHAPTER 1

INTRODUCTION

1.1 Introduction

Respiratory diseases are leading causes of death and disability in the world. The poorest regions of the world had the greatest disease burden. Ageing and risk factors including smoking, environmental pollution, and body weight also play a key role, say the researchers. Chronic respiratory diseases pose a major public health problem and about 65 million people suffer from chronic obstructive pulmonary disease and with an estimated 3.91 million deaths in 2017 which accounts for 7% of all deaths worldwide and its third leading cause of death. Between 1990 and 2017, the number of deaths due to chronic respiratory diseases increased by 18%, from 3.32 million in 1990 to 3.91 million in 2017. About 334 million people suffer from asthma, the most common chronic disease of childhood affecting 14% of all children globally.

Respiratory diseases like Pneumonia kills millions of people annually and is a leading cause of death among children under 5 years old. Over 10 million people develop tuberculosis (TB) and 1.4 million die from it each year, making it the most common lethal infectious disease. Lung cancer kills 1.6 million people each year and is the deadliest cancer. Globally, 4 million people die prematurely from chronic respiratory disease. Respiratory diseases make up five of the 30 most common causes of death: COPD is third; lower respiratory tract infection is fourth; tracheal, bronchial and lung cancer is sixth; TB is twelfth; and asthma is twenty-eighth [1]. Altogether, more than 1 billion people suffer from either acute or chronic respiratory conditions. The stark reality is that each year, 4 million people die prematurely from chronic respiratory disease [2]. Infants and young children are particularly susceptible. A total of 9 million children under 5 years old die annually, and pneumonia is the world's leading killer of these children [1].

People often take breathing and our respiratory health for granted, but the lung is a vital organ that is vulnerable to airborne infection and injury. Respiratory system diseases affect people's social, economic and health life significantly. Social deprivation was the most important factor affecting rates of death and disability, with the highest rates seen in the poorest regions of the world. Lower mortality was seen in more affluent countries, reflecting better access to health services and improved treatments.

So, treatment of lung diseases, which are the most common cause of death in the world, is of great importance in the medical field. For these reasons, a lot of research are going on for early diagnosis and intervention in respiratory diseases. In order to accurately identify health problem regarding this information requires experience and time, but according to the World Health Organization (WHO) statistics [3], 45% of the WHO Member States report to have less than 1 physician per 1000 population, the WHO ratio recommendation. Considering these statistics into account, to study individually and diagnose every patient by a health specialist who are already overbooked, mistakes can happen. This is why finding new ways to help doctors to save time is a priority. Hence, automatic and reliable tools can help in diagnosing more people and it can also help specialists to make less mistakes due to the work overload.

1.2 Motivation

As rapid growth of respiratory diseases is witnessed around the world, medical research field has gained interest in integrating potential audio signal analysis-based technique. From the past few decades, computer science constantly improving the ability to analyze media data automatically and with the help of diagnosis tools we are able to process image and/or audio information. Hence, Computer science could help nursing staff or doctors for diagnosis by proposing faster and reliable tools and by giving customizable tools for medical monitoring to the patient. Like in other application domains, audio signal analysis tools can potentially help in analyzing respiratory sounds to detect problems

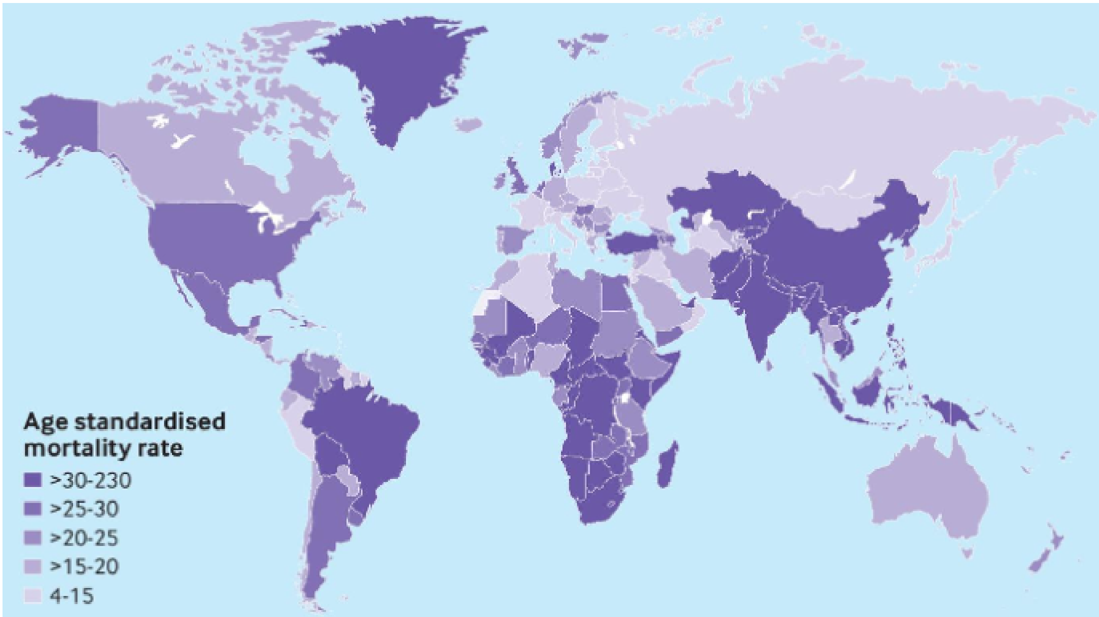

Figure 1.1: Global age standardized mortality rate per 100000 people of chronic obstructive pulmonary disease for both sexes in 195 countries and territories in 2017
(Source: https://www.bmj.com/content/368/bmj.m234)

in the respiratory tract. Audio analysis aids in timely diagnosis of respiratory ailments more effortlessly in the early stages of a respiratory dysfunction. Apart from respiratory check-ups, every cardiac assessment also includes an audio auscultation in which one the medical specialist listens to sounds from the patient body with different tools like stethoscope or sonography. This shows how important sound analysis is for heart and lungs disease detection.

Respiratory sounds may be acquired by the easy and non-invasive auscultation procedure. Auscultation is an effective technique in which physicians evaluate and diagnose the disease after using a stethoscope for lung disease. This method is inexpensive and easy as it does not require internal intervention into the human body. However, traditional stethoscopes may be exposed to external noise sounds and cannot filter the audio frequencies of the body in auscultation and cannot create permanent recordings in monitoring of the disease course. Also, there is a possibility of untrained physicians incorrectly recognizing abnormalities, which can be due to not calibrating the instrument and/or due to noisy environment, is very high using this method.

As lung and heart diseases remains the leading cause of death globally, there are many studies about lung and heart sound identification. Since then, there are lots of improvements, for processing records taken in noisy environments. Furthermore, new kinds of methods drastically improve the domain, as machine learning and deep learning. These approaches contribute a lot to computer vision, or audio analysis. This gives more relevant information from respiratory sounds extracted and contribute to reducing the time for diagnosis, consequently increasing treatment efficiency. Thus, an automated algorithm developed to recognize abnormalities in respiratory sounds may be of great relevance to clinical diagnosis. Also researchers are looking in to combining speech and signal processing tools techniques with image analysis-based tools techniques [4, 5, 6] can also help doctors predict or guess about the presence of respiratory diseases based on verbal communication before they even start with the X-ray screening or other procedures.

Machine learning has proven to be an effective technique in recent years and machine learning algorithms have been successfully used in a large number of applications. The development of computerized lung sound analysis has attracted many researchers in recent years, which has led to the implementation of machine learning algorithms for the diagnosis of lung sound. In our research we have used machine learning techniques in computer-based lung sound analysis. A brief description of the types of lung sounds and their characteristics is provided. We examined specific lung sounds/disorders, the number of subjects, the signal processing and classification methods and the outcome of the analyses of lung sounds using machine learning methods that have been performed by previous researchers. Before diagnosing disease based on their types, it is important to first ensure that whether a person has any lung infection. True positive case can then be pushed for further processing, such as diagnosis.

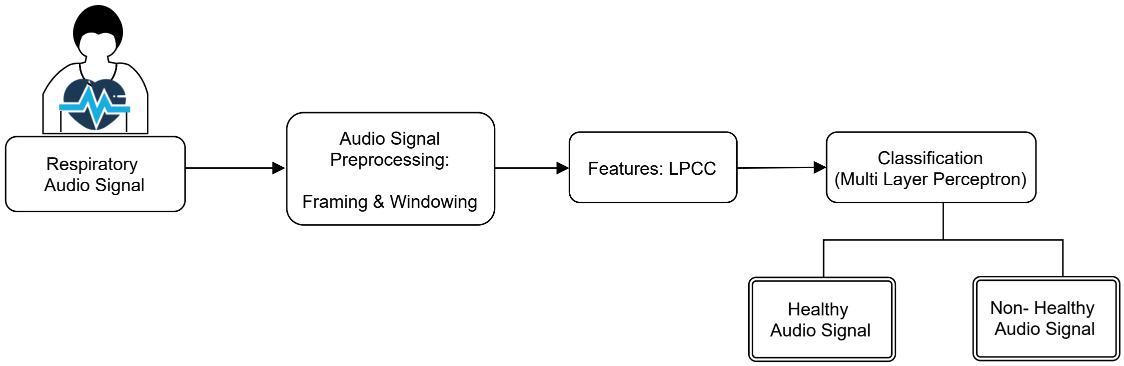

Figure 1.2: Block diagram of the proposed work.

In this research, we developed an automated tool to distinguish healthy respiratory sound from and non-healthy ones that come from respiratory infection carrying patients, where LPCC-based features are employed. Using over 6800 clips, we obtained the highest accuracy of 99.22%. A brief description on the previous works is also included and in conclusion, the review provides recommendations for further improvements.

1.3 Methodology

Respiratory conditions are diagnosed through spirometry and lung auscultation. Spirometry is measuring the volume of air mobilized in respiration. Even though, this method is one of the most commonly available lung function tests and well validated for the diagnosis and monitoring of upper and lower airway abnormalities [1], it is limited to patient's cooperation and therefore, is error prone. Moreover, traditional spirometers are normally used only in clinical settings due to their high cost and required calibration [2] along with challenges in patient guiding. Auscultation is other technique which involves listening to the internal human body sounds with the aid of a stethoscope and typically performed on the anterior and posterior chest. From past few years, it has been an effective tool to understand lung disorders and possible abnormalities. However, this process is limited to physicians as they are well trained. For various reasons like faulty instrument or noisy environment, false positives can happen. Therefore, it opens a door to develop computerized lung sound analysis tools/techniques, where automation is the integral part.

1.4 Contribution Outline

Sounds heard over the chest wall are useful tools for diagnosing pulmonary diseases. Modern lung sound analysis, which began in the last four decades, is focused on digital sound processing and graphic representation of the signals [7]. As Computerized lung sound analysis and diagnosis is the main goal of the researchers in this field, several different approaches are being continuously evaluated by researchers to help medical professionals. However, lung sound analysis continues to attract researchers because past researchers focused on identifying lung sounds and very few researchers concentrated on developing lung disorder diagnostic tools. Therefore, this research area appears incomplete and has thus attracted many researchers in recent years. Thus, an objective and reliable diagnostic tool for the detection of pulmonary diseases is aimed.

Previous researchers used three notable databases namely, Marburg European project CORSA [8], Respiratory Sounds (MARS) [9] and R.A.L.E. repository [10]. However, R.A.L.E. repository used to be commercially available database. The Marburg Respiratory Sounds (MARS) database was compiled using Lung sound CDs which are commercially available for training doctors and nurses to understand lung sounds [9]. The European project CORSA was developed with an intension of standardizing the recording process of respiratory sounds [8]. However, In 2017, the largest publicly available respiratory sound database was compiled and encouraged the development of algorithms that can identify common abnormal breath sounds (wheezes and crackles) from clinical and non-clinical settings.

Machine learning algorithms are currently used in many applications which possess artificial intelligence that learns from past experiences and allow the tools to function

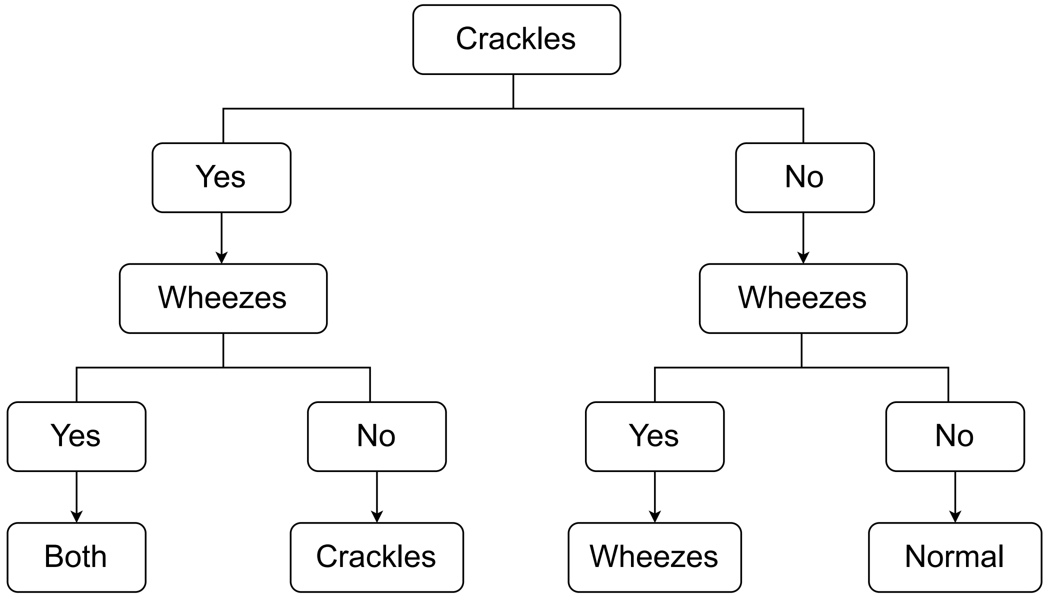

Figure 1.3: Decision tree for anomaly detection

more accurately [11, 12]. In addition, the previous research on computer-based lung sound analysis using machine learning algorithms, such as artificial neural networks (ANNs), the hidden Markov model (HMM), k-nearest neighbor (k-NN) algorithm, Gaussian mixture model (GMM), genetic algorithms (GAs).

Initially the ANN and k-NN algorithms are the machine learning techniques that are mostly used. The use of support vector machines (SVMs) was found to be very limited in the literature. The most commonly used machine learning methods used for lung sound analysis are ANN and k-NN. The classification accuracy reported by Kandaswamy et al., was 100% for training and 94.02% for testing using ANN in classification of normal, wheeze, crackle, squawk, stridor, and rhonchus respiratory sounds [13]. This shows the effectiveness of ANN in classifying the lung sounds. The ANN has the ability to adapt well with complex non-linear data and classify it accurately and effectively [14]. The k-NN classifier is another machine learning technique which has attracted researchers to use it in lung sound classification. The advantage of using k-NN is its simplicity and robustness [15]. The work of Alsmadi and Kahya has reported a classification accuracy of

96% in real time using k-NN classifier [16]. Their developed system can recognize normal and abnormal lung sounds and they trained the model with a large dataset comprising of 42 subjects. In spite of its advantages, the ANN and k-NN have few disadvantages too. The disadvantage of using ANN and k-NN in classification would be the computational burden caused for training the model and also it is required to have a very large dataset to train the model to effectively recognize the lung sounds accurately [14, 15]. In spite of its disadvantage, ANN and k-NN serves as the most commonly used machine learning algorithms in lung sound analysis due to its ability to achieve better classification accuracy and detected the lung sounds accurately compared to other methods.

Machine learning algorithms allow the computer to make decisions based on its previous experiences [17, 18]. In the past decade, machine learning has been used in many research areas and its diversity has attracted the use of these algorithms for different applications. In the past few years, researchers have used machine learning algorithms in computer-based lung sound analysis. However, the use of machine learning techniques in computer-based lung sound analysis is still preliminary. The work of Guler, who used genetic algorithm-based artificial neural networks for the classification of lung sounds [19], shows the importance of using hybrid machine learning algorithms in computer-based lung sound analysis. Their resulting classification accuracy using GA-based ANN algorithms was reported to be 83–93%, which shows the significant improvement that can be achieved through the use of hybrid machine learning algorithms. The use of hybrid machine learning algorithms in lung sound analysis is very limited. However, the exploration of hybrid machine learning algorithms might help researchers improve the classification accuracy. It was observed from the literature that ANN yields good results in almost all the previous works and hence combining other methods with ANN would most probably yield better classification accuracy. The ability of ANN to discriminate both linear and non-linear data accurately gives it an advantage over other methods [20, 21]. Alsmadi and Kahya developed a real-time classification system with a classification accuracy of 96% [16], which is satisfactory. Their system provides sufficient evidence that demonstrates the high possibility of the development of real-time computer-based lung sound analysis systems. The advantages of using a computer-based lung sound analysis algorithm include that this method is non-invasive, less time consuming and more accurate than other methods. In spite of its advantages, the computer-based lung sound analysis has not yet been developed to a level that can be used in a clinical setting. The development and commercialization of real-time computer based-lung sound analysis systems is a major area for future research approaches.

Though there has been development of disparate systems for lung sound analysis, but the number of misclassifications has not been very low. Moreover, non-healthy cases are composed of several conditions. Distinguishing healthy conditions from non-healthy conditions is very challenging when the non- healthy cases consist of multiple problems. Shallow learning based systems are preferred over deep learning-based systems where computational resource is an issue. The Shallow learning also need to be robust enough to be able to effectively model healthy and problematic cases considering different problematic cases. The main contribution of this work is to suggest a new approach in audio classification. In some cases, here for lung pathologies, machine learning for audio classification based on sound content is not the best solution, or at least not alone. In this study, a machine learning approach is presented and outperforms the previous state of the art. Using this classification model and extrapolating the results to take a decision on the patient level leads to better results.

Secondly, prior to deeper analysis of problematic cases, it is essential to distinguish healthy and non-healthy cases. A hierarchical approach can aid to reduce the workload of doctors considering the shortage of medical facilities in resource constrained areas. After ensuring that whether a person has any lung infection or not, the true positive case can then be pushed for further processing.

In this research, we developed an automated tool, where LPCC-based features are employed. LPCC-based features were chosen due to its ability of modelling different type of audio signals [22, 23]. Using over 6800 clips, we obtained a highest accuracy of 99.22%. The block diagram of the proposed methodology is presented in Figure 1.2.

1.5 Organization Of Thesis

Next, we will review the terminology and an explanation of the physiological origin of respiratory sounds used by medical practitioners, which are also studied by many engineers in the electronic respiratory sound analysis field. These include the two main categories of i.) normal and ii.) adventitious respiratory sounds. Respiratory sounds are difficult to analyze and distinguish because they are non-stationary and non-linear signals. Several techniques were implemented to recognize lung disorders and possible abnormalities. Automated analysis was made possible with the use of electronic stethoscope.

The audio clips were characterized using Linear Predictive Cepstral Coefficient (LPCC)-based features and the highest possible accuracy of 99.22% was obtained with a Multi-Layer Perceptron (MLP)-based classifier on the publicly available ICBHI17 respiratory sounds dataset [24] of size 6800+ clips.

The rest of the thesis is structured as follows:

Chapter 2: In this chapter, we present a little background about the topic of thesis and we also briefly discuss some relevant work and discussed about Shallow learning and deep learning that were important for this work.

Chapter 3: Description of the dataset that was used in this work to develop the classification methods along with describing the signal processing methodology. Then we present the experimental methodology for comparing results of different methods. We also discuss the challenges and our proposed solutions concerning the application of our method and the search for the best classification method.

Chapter 4: In this chapter, we present the results of our proposed methods by comparing with the other methods. We then interpret the results, comparing each method and showing the weaknesses and strengths of the methods.

Chapter 5: We finish by summarizing the work, the challenges we faced, our solutions. Also, we present the results we obtained along with a brief proposal for the future work.

CHAPTER 2

RELATED WORKS

2.1 Background

As the respiratory diseases are increasing worldwide, it is extremely important of timely diagnosis of the issue. Prevention and early detection are essential steps in managing respiratory disease. Auscultation is an essential part of clinical examination as it is an inexpensive, noninvasive, safe, easy-to-perform, and one of the oldest diagnostic techniques used by the physician to diagnose various pulmonary diseases. The drawbacks of this procedure are that doctors require experience and ear acuity to provide a more accurate diagnosis to the patient. It is especially hard since some sounds are harder to detect because of the limitations of the human ear. Automatic lung health screening using respiratory sounds meant to help physician by successfully detecting and classify the adventitious sounds in the lung sound with the help of digital signal and using a combination of signal processing techniques with shallow learning technique, deep artificial neural networks.

2.2 Related works

In what follows, we categorized previous works into, Aykanat et al. [25] presented a convolutional network as well as mel frequency cepstral coefficient, support vector machine-based approach for lung sound classification. The two feature extraction methods are mel frequency cepstral coefficient (MFCC) feature extraction and spectrogram generation using short-time Fourier transform (STFT). They used MFCC features combined with SVM which is a generally accepted practice for audio classification. In sound processing, the mel frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a non-linear mel scale of frequency. MFCCs are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip. MFCC features are also used in[26] where clips are first preprocessed in the form of framing and windowing followed by extraction of MFCC features. Also to handle the uneven and large dimensionality problems in the subsequent paragraphs, the second level MFCC-2 feature values are computed. A spectrogram is a visual representation of the spectrum of frequencies in a sound or other signal as they vary with time or some other variable. They are used extensively in the fields of music, sonar, radar, and speech processing and seismology. Since MFCC features are widely used in audio detection systems, the experiments they ran using the MFCC features which enabled to find a base value for accuracy, precision, recall, sensitivity, and specificity. Spectrogram images are also used in audio detection. However, they were never tested in respiratory audio with CNNs. MFCC datasets were built using SciPy library. They used support vector machines to process these datasets. The spectrogram dataset was built using a combination of open-source graph generation library Pylab and various open-source image processing libraries. The original spectrograms generated were 800 × 600 RGBA, and since it's too large for computer's memory in experiment they changed the algorithm to generate them 28 × 28 grayscale to fit them into the memory for CNN to process. They used a dataset of 17930 sounds from 1630 subjects and experimented with four different scenarios which involved both the proposed approaches. They reported an accuracy of 86% using both SVM and CNN for healthy- pathological classification. Finally, they concluded that spectrogram image classification with CNN algorithm works as well as the SVM algorithm, and given the large amount of data, CNN and SVM machine learning algorithms can accurately classify and pre-diagnose respiratory audio.

Pramono et al. [27] evaluated disparate features for classifying normal respiratory sounds and wheezes. This study evaluated the discriminatory ability of different types of feature used in previous related studies, with the dataset consisted of 38 recordings from disparate sources. It had 425 events out of which 223 were wheezes and the rest were normal. They demonstrated that certain individual fea- tures (MFCC, tonality index) are much more accurate in detection of wheezes. However, their computation requirements are higher than those of simpler time-domain features. In addition, it has also been shown that while the use of multiple features does increase the classification accuracy in some cases, the gain in performance becomes very limited after a certain number of features. They concluded by mentioning, while the classifier used in this work is very simple, the use of other more complex classifiers such as support vector machines, artificial neural networks, etc. may help to increase the classification performance at the added cost of computational complexity. Thus, it is important to take all the competing requirements into account when selecting a feature for wheeze detection in different applications. They experimented with different features and the results are presented in [27].

Acharya et al. [28] presented a deep learning-based approach for lung sound classification. Deep learning has gained a lot of attention in recent years due to its unparalleled success in a variety of applications including clinical diagnostics and biomedical engineering. A significant advantage of these deep learning paradigms is that there is no need to manually craft features from the data since the network learns useful features and abstract representations from the data through training. As the dataset is relatively small for training a deep learning model, they used several data augmentation techniques to increase the size of the dataset. Aside from increasing the dataset size, these data augmentation methods also help the network learn useful data representations in-spite of different recording conditions, different equipment's, patient age and gender, inter-patient variability of breathing rate etc. For feature extraction they have used Mel-frequency spectrogram with a window size of 60 ms with 50% overlap. Each breathing cycle is then converted to a 2D image where rows correspond to frequencies in Mel scale and columns correspond to time (window) and each value represent log amplitude value of the signal corresponding to that frequency and time window. They proposed a hybrid

CNN-RNN model that consists of three stages: the first stage is a deep CNN model that extracts abstract feature representations from the input data, the second stage consists of a bidirectional long, short term memory layer (Bi-LSTM) that learns temporal relations and finally in the third stage they have fully connected to softmax layers that convert the output of previous layers to class prediction. While these type of hybrid CNN-RNN architectures have been more commonly used in sound event detection due to sporadic nature of wheeze and crackle as well as their temporal and frequency variance, similar hybrid architectures may prove useful for lung sound classification. Since deep learning models require much larger amount of data for training, they faced an issue. To address these shortcomings of existing methods, they proposed a patient specific model tuning strategy that can take advantage of deep learning techniques even with small amount of patient data available. In this proposed model, the deep network is first trained on a large database to learn domain specific feature representations. Then a smaller part of the network is re-trained on the small amount of patient specific data available. This enabled them to transfer the learned domain specific knowledge of the deep network to patient specific models and thus produce consistent patient specific class predictions with high accuracy. In their proposed model they trained the 3-stage network on the training samples. Then, for a new patient, only the last stage is re-trained with patient specific breathing cycles while the learned CNN-RNN stage weights are frozen in their pre-trained values. They reported their hybrid CNN-RNN model produced a score of 66.31% scores on 80–20 split for four-class respiratory cycle classification. Then they proposed a patient screening and model tuning strategy to identify unhealthy patients and then built patient specific models through patient specific re-training which provided significantly more reliable results for the original train-test split achieving a score of 71.81% for leave-one-out cross-validation on the ICBHI17 dataset.

Dokur [29] first used a rectangular window formed from one cycle of respiratory sound (RS) windowed time samples are then normalized. In order to extract the features, the normalized RS signal is partitioned into 64 samples of long segments. The power spectrum of each segment is computed, and synchronized summation of power spectra components is performed. Feature vectors are formed by the averaged power spectrum components, yielding 32-dimensional vectors. In this study, classification performances of multi-layer perceptron (MLP), grow and learn (GAL) network and a novel incremental supervised neural network (ISNN) are comparatively examined thirty-six patients for the classification of nine different RS classes: Bronchial sounds, bronchovesicular sounds, vesicular sounds, crackles sound, wheezes sound, stridor sounds, grunting sounds, squawk sounds, and sounds of friction rub. They have performed analysis of respiratory sounds in three stages: Normalization process, feature extraction process, and the classification of the respiratory sounds by artificial neural networks (ANNs). In the first stage, a rectangular window is formed so that one cycle of RS is contained in this window. The window comprises of 8,192 samples. Then, the windowed time samples are normalized so that the power of the respiratory signals in the window is set to 1. In the second stage, feature vectors are formed by using the normalized data in the window. Finally, in the last stage, classification of the RSs is realized by using artificial neural networks and have reported an accuracy of 92% in this study using multi-layer perceptron.

Shivakumar [30] classified respiratory sounds with a CNN-based technique and experiments were performed with two kind of sounds namely crackles and wheezes. After pre-processing the audio files they developed a Neural network in which they modified an existing CNNs to create the base model for dataset. Later they used an Adam optimizer with learning rate 0.009 and batch size of 64. For the first model, author used both wheezes and crackles simultaneously for 10 epochs and then split the dataset and ran the model on wheezes and crackles separately again for 10 epochs. When used both a 90-10 and 80- 20 train-test split – the results for both were the same. Author also demonstrated that splitting the sounds up into different models is very beneficial. Two models proposed in this study produced test accuracies of 50% and 100% respectively.

Faustino [31] presented a CNN-based technique for detection of wheeze and crackle on the ICBHI 2017 dataset. The study involved extraction of MFCC and power spectral density values from the audio clips. These were fed to a CNN for classification. They found that utilizing a Mel Spectrogram for lung sound classification utilizing a Convolutional Neural Network architecture is more beneficial than utilizing MFCC features. However, these results were not better than the results obtained in the other study that also utilizes the same dataset but uses a RNN architecture with MFCC features. Based on these findings, they infer that utilizing a Recurrent Neural Network architecture combined with the use of MFCCs is a better approach than utilizing a convolutional based approach, for the classification of lung sounds. The MFCC method utilizes the discrete cosine transform to compress and decorrelate the signal features which explains why it works better when combined with a RNN instead of a CNN. A CNN architecture takes advantage of local patterns in data; therefore, it makes inefficient use of the MFCCs. An RNN is built using a FNN as the interior network, which has access to all input features without the utilization of shared parameters, combined with the temporal context of the data, making it a much better architecture for interpreting MFCC input. Finally, using a fivefold cross validation technique, 43% test accuracy was reported.

Ma et al. [32] presented a system that has incorporated the non-local block in the ResNet architecture to distinguish respiratory sounds. They proposed a LungRBN model, which uses short-time Fourier transform (STFT) and wavelet feature extraction methods together with a product of two ResNet models through a fully connected layer to achieve the best state-of-the-art accuracy. However, less attention has been paid to finding ways to automatically augment existing data to achieve a significant breakthrough in detection accuracy. To overcome this challenge, they proposed an improved adventitious Lung Sound Classification, LungRN+NL, incorporate a non-local layer in ResNet neural network with a mixup data augmentation method. Considering the key discrimination among different categories, we choose short-time Fourier trans- form (STFT), a time-frequency analysis method, to extract fea- tures from lung sounds. Experiments were performed on the ICBHI 2017 dataset and an accuracy of 52.26% was reported.

Emmanouilidou et al. [33] proposed a robust approach to identify respiratory sounds in the presence of noise. The proposed framework addressed the need for improved lung sound quality by using noise-suppression techniques suitable for auscultation applications. They developed noise-suppression scheme which eliminates ambient sounds, heart sounds, sensor artifacts and crying contamination and tackled various noise- sources including ambient noise, signal artifacts, patient- intrinsic maskers. The improved high-quality signal is then mapped onto a rich spectro temporal feature space before being classified using a trained support-vector machine classifier. Individual signal frame decisions are then combined using an evaluation scheme, providing an overall patient-level decision for unseen patient records. They composed a dataset with the aid of over 1K volunteers and reported an accuracy of 86.7

Sen et al. [34] experimented with distinction of respiratory sounds from healthy and non-healthy subjects. This study explored a useful methodology for the classification of the three-class structure (healthy- obstructive-restrictive) by using 14-channel pulmonary sounds data are modeled using a second order 250-point VAR model, and the estimated model parameters are fed to SVM (of discriminative type) and GMM (of generative type) classifiers designed in various classifier configurations. The adventitious sound components (e.g., crackles and wheezes), which are indicators of pathological conditions, are informative about the disease by their timing within the respiration cycle as well as their other (spectral, temporal, and spatial) characteristics. To make use of their distinctive information, the six subphases of the flow cycle are considered separately, until being suitably combined at the decision level. The linear kernel function is adopted for the SVM classifier since it yields satisfactory results with low computational complexity. They concluded that hierarchical approach to be adopted for diagnostic classification of pulmonary conditions, i.e., first, a discrimination between healthy versus pathological conditions, second, a discrimination between obstructive versus restrictive types under the pathological condition. Although the GMM classifier has been shown to be more successful compared to the SVM classifier, the probabilistic variants of the SVM classifier are still suggested for future studies, depending on the performances obtained in the augmented feature space. The methodology of this study is proposed as a promising diagnostic framework to consider for clinical purposes. They collected data from 20 healthy and non-healthy subjects which were fed to gaussian mixture model and support vector machine-based classifiers. Among them, the gaussian mixture model-based classifier produced an accuracy of 85

Demir et al. [35] used a CNN-based approach for lung sound classification from the ICBHI 2017 dataset. They proposed a new pretrained Convolutional Neural Network (CNN) model such as VGG16 and AlexNet is proposed for the extraction of deep features. However, sound characteristics are not fully represented since these CNN models have not been trained on sound datasets. Hence, the proposed CNN model was trained with spectrogram images based on lung sounds. In addition, the parallel-pooling structure was employed in order to boost classification performance in the proposed CNN architecture. In the CNN architecture, an average-pooling layer and a max-pooling layer are connected in parallel in order to boost classification performance. The deep features are utilized as the input of the Linear Discriminant Analysis (LDA) classifier using the Random Subspace Ensembles (RSE) method. They reported a highest accuracy of 83.2% for the healthy class and an overall accuracy of 71.15%.

Chen et al. [36] used a S-transform-based approach coupled with deep residual networks for separating respiratory sounds. First, the raw respiratory sound is processed by the proposed OST. Then, the spectrogram of OST is rescaled for the Resnet. After the feature learning and classification are fulfilled by the ResNet, the classes of respiratory sounds are recognized. In order to evaluate the effectiveness of the proposed OST and ResNet for the triple-classification of respiratory sounds, the three rescaled feature maps of STFT, ST and OST are applied to the ResNet-50 with different batch sizes and iterations. The proposed OST highlights the features of wheeze, crackle, and respiratory sounds, and the deep residual learning generates discriminative features for better recognition. The experimental results show that the proposed OST and ResNet is excellent for the multi-classification of respiratory sounds like crackle, wheeze and normal sounds and reported an accuracy of 98.79

Kok et al.[37] used several features including MFCC, DWT and time domain metrics for distinguishing healthy and non-healthy cases. A number of features were investigated, and Wilcoxon Rank Sum statistical test was used to determine the significance of the extracted features. The significant features were then passed to a feature selection algorithm based on mutual information, to determine the combination of features that provided minimal redundancy and maximum relevance. The instances were classified random under sampling and boosting method. They reported accuracy specificity and sensitivity values of 87.1%, 93.6% and 86.8%.

Chambers et al. [38] presented a system in patient level to identify healthy/ non healthy situation by proposing a method divided in two parts. The first part is about the classification of the respiratory cycles depending on the adventitious sounds and the second part is about extrapolating the classification results to consider the patient classification. These parts are respectively named the micro-level part and the macro-level part. The micro-level part is to classify individually every respiratory cycle depending if adventitious sounds are detected or not. For that, all records are taken one by one, and for each record, features are extracted on the signal window containing every cycles. The classification of each cycle is computed with a boosted decisional tree, which gives, according to the features, the probability to belong to every class. The macro-level part is to suggest a "diagnostic" taking into account the totality of the predictions previously computed. As a doctor not only listens one time the lung of his patient, but several times at different area of the body, they computed the different kind of cycle ratio, predicted for all the cycles of one patient. Depending on what kind of cycles appears the most, a decision is taken. With all these several spectral, rythm, SFX and tonal features coupled with decision tree-based classification they reported an accuracy of 85%.

Altan et al.[39] presented a deep learning-based approach for detection of chronic obstructive pulmonary disease. Their study focused on analyzing multichannel lung sounds using statistical features of frequency modulations that are extracted using the Hilbert-Huang transform. Deep-learning algorithm was used in the classification stage of the proposed model to separate the patients with COPD and healthy subjects. The methodology involved the use of Hilbert-Huang transform on multichannel respiratory sounds and an accuracy of 93.67% was re- ported in segregation of healthy and non-healthy patients.

Rao et al.[40] acoustic techniques for pulmonary analysis. They talked about the acoustic aspects of different lung diseases. A discussion is also provided regarding the physic of human thorax and techniques of measuring respiratory sounds. The authors have also discussed in detail about different signal processing techniques which are required to analyze these sounds along with disparate classifiers.

Cohen and Landsberg [41] classified 7 different type of breath sounds using linear predictive coefficient-based technique. The classification is performed in two levels, with the first level based on linear prediction coefficients and the second level on energy envelope features. Each type of breath sound is represented by its mean feature vector and by its covariance matrix. These are acquired by training set classified by a physician. The distance measure is defined and used to compare unknown breath sounds. The unknown signal is hypothesized to belong to that type which distance is minimal. So, in their research, rather than trying to automatically diagnose lung diseases they quantitatively characterized and automatically classify breath sounds by providing physician with a diagnostic assist device. They performed experiments with 105 instances out of which 100 were classified correctly.

Table 2.1: Overview of previous work on Shallow Learning

Author	Method	Dataset (Size)	Performance (ACC|AUC |SEN |SPEC)
Pramono et al.[27]	Compared performance of different features	Multiple repositories (38)	MFCC |–|0.8919 |83.86%| 81.19%
Dokur[29]	MLP, ISNN, GAL	Individual patient data and RALE (180)	ISNN |98%|– |– |– GAL |92%|– |– |– MLP |92%|– |– |–
Emmanouilidou et al.[33]	Biomimetic approach with SVM classifier	PERCH Study (250 hours)	SVM |86.67%|– |86.82%| 86.55%
Sen et al[34]	SVM and GMM classifiers	Individual patient data (40 subjects)	GMM |85%|– |90% |90%
Kok et al. [37]	RUSBoost Algorithm	ICBHI'17 dataset(920)	RUSBoost Algorithm |87.1%|– |86.8% |93.6%
Chambers et al [38]	Combined multiple features like spectral, rythm, SFX and tonal features coupled with decision tree-based	ICBHI'17 dataset(920)	Macro level |85%|– |– |–
Rao et al. [40]	Review on different Acoustic techniques	Multiple sources	SVM |90.77%|– |– |– KNN |93 - 95%|– |– |–
Cohen and Landsberg [41]	Linear prediction coefficients and energy envelope features	Individual patient data (105 instances)	LPC |95.2%|– |– |–

Table 2.2: Overview of previous work on Deep Learning

<table>
<tr><th>Author</th><th>Method</th><th>Dataset (Size)</th><th>Performance (ACC|AUC |SEN |SPEC)</th></tr>
<tr><td>Aykanat et al.[25]</td><td>CNN and SVM algorithms</td><td>Electronically recorded (17,930)</td><td>CNN |86%|– |86% |86%
SVM |86%|– |87% |82%</td></tr>
<tr><td>Acharya et al.[28]</td><td>Deep CNN -RNN model</td><td>ICBHI’17 dataset(920)</td><td>Deep CNN-RNN |96%| – |48.63% |84.14%</td></tr>
<tr><td>Shivakumar[30]</td><td>CNNs</td><td>ICBHI’17 dataset(920)</td><td>1st Model |50%|– |– |–
2nd Model |100%|– |– |–</td></tr>
<tr><td>Faustino[31]</td><td>CNNs</td><td>ICBHI’17 dataset(920)</td><td>CNN |43%|– |51% |36%</td></tr>
<tr><td>Ma et al.[32]</td><td>LungRN+NL model</td><td>ICBHI’17 dataset(920)</td><td>LungRN+NL |–|– | 41.32% |63.2%</td></tr>
<tr><td>Demir et al.[35]</td><td>CNN model</td><td>ICBHI’17 dataset(920)</td><td>CNN |71.15%|– |– |–</td></tr>
<tr><td>Chen et al.[36]</td><td>Optimized S-transform (OST) and deep residual networks (ResNets)</td><td>ICBHI’17 dataset(920)</td><td>ResNet with OST |98.79%|– |96.27% |100%</td></tr>
<tr><td>Altan et al.[39]</td><td>Deep Learning model with the Hilbert- Huang transform</td><td>NA</td><td>Deep learning model |93.67%|– |91% |96.33%</td></tr>
</table>

2.2.1 Shallow Learning

In the last decades many machine learning (ML) approaches have been introduced to analyze respiratory cycle sounds including crackles, coughs, wheezes [42, 43, 44, 45, 46, 47]. In many researches, conventional ML models solely rely on shallow learning as deep learning may not be suitable in all the experiments. Thus, merely deep learning based models may not be robust to external/internal noises in lung sounds and may not generalize their performance across different software's and measuring devices. Furthermore, highly complex preprocessing steps are required to make use of designed features [45, 46, 47].

Shallow learning is a type of machine learning where we learn from data described by pre-defined features. Shallow learning refer to properties derived using various algorithms using the information present in the image itself. The Shallow learning were commonly used with "traditional" machine learning approaches for object recognition and computer vision like Support Vector Machines, for instance. However, "newer" approaches like convolutional neural networks typically do not have to be supplied with such shallow learning, as they are able to "learn" the features from the image data. In this research we developed an automated tool to classify lung sounds using our shallow learning feature, where LPCC-based features are employed. As our dataset is audio files and the clips are of different lengths, the clips were first framed into short sections and then windowed as part of preprocessing. Next, standard LPCC features were extracted from the clips. In order to tackle the problem of uneven dimensionality, we have done grading and standard deviation. Later it is classified using an MLP(multi-layer perceptron) classifier.

2.2.2 Deep Learning

The first research that modelled Artificial Neural Networks (ANN) was from Warren McCulloch and Walter Pitts in 1943, with their paper A Logical Calculus of Ideas Immanent in Nervous Activity [48]. Though there were some research into ANNs through the 1950s and 1960s, limited computing power prevented experimentation with large ANNs. Nearly 15 years later with the invention of the Backpropagation algorithm, research into ANNs became popular again. However, ANNs gave away to simpler classifiers such as SVMs, which outperformed ANNs in both accuracy and training time. The way that ANNs worked was using simple Perceptron Units in one or two hidden layers and using weighted connections to an input and an output layer (Figure 2.1). Running networks with more hidden layers was usually infeasible, again due to limited computational power.

In the 21st century, research into ANNs have again become popular, but in the form of Deep (Neural) Networks and Deep Learning. Deep Learning is a technique of hierarchical machine learning using multiple layers of non-linear processing. One of the successful approaches to Deep Learning have been with Deep Networks, which have become a re-branding or buzzword for Artificial Neural Networks. Deep Neural Networks are basically ANNs with multiple hidden layers, which presents the opportunity of creating more complex models of non-linear structures, but also increases the time and space complexity of training models in the same way ANNs were limited by in earlier research. The reason optimization problems in Deep Neural Networks have a high time complexity is due to its iterative nature in training.

In our research, we are using shallow learning feature as deep learning features are automatically extracted and may not give the feature we are looking for in our research. The ability to process large numbers of features makes deep learning very powerful when dealing with unstructured data. Occasionally, deep learning algorithms can be overkill for less complex problems because they require access to a vast amount of data to be effective. If the data is too simple or incomplete, it is very easy for a deep learning model to become overfitted and fail to generalize well to new data. As a result, deep learning models are not as effective as other techniques.

2.3 Discussion

Deep learning methods are becoming increasingly popular because of their impressive classification performance. However, it is known that they typically require a large training sample to achieve that accuracy and features are automatically extracted and it may not generate the features we exactly need to have. Meanwhile, Shallow learning have been implemented for decades and still serve as a powerful tool when combined with machine learning classifiers as they are expert based.

CHAPTER 3

DATASET

3.1 Foreword

In this chapter, we discuss on ICBHI dataset used, annotations and challenges.

3.2 Dataset

The lung sounds that are heard over the chest wall are caused by the airflow in the lungs during the inspiration and expiration phases. These sounds are non-stationary and non-linear signals, which make it difficult for physicians to recognize any abnormalities [13]. The types and characteristics of lung sounds are listed in Fig. 3.1 [49, 50, 51, 52, 53, 54, 55, 56]. Abnormal breath sounds include the absence or reduced intensity of sounds where they should be heard or, by contrast, the presence of sounds where there should be none, as well as the presence of adventitious sounds. As opposed to those classified as "normal", abnormal sounds are those which may indicate a lung problem, such as inflammation or an obstruction. Each lung disorder is associated with one or more lung sounds [13]. The disorders that are associated with each sound are also detailed in Fig. 3.1. The dominant frequency of heart sounds is typically below 150Hz, whereas the dominant frequency of lung sounds ranges between 150 and 2000Hz. This difference in the frequencies makes it easier to filter the heart sounds from the lung sounds. The durations of the different types of lung sounds are also mentioned in Fig. 3.1.

The ICBHI (International Conference on Biomedical and Health Informatics) dataset [24] was originally compiled to support the scientific challenge on respiratory data analysis organized in conjunction with the 2017 Int. Conf. on Biomedical Health Informatics (ICBHI). The current version of this database is made freely available for research which

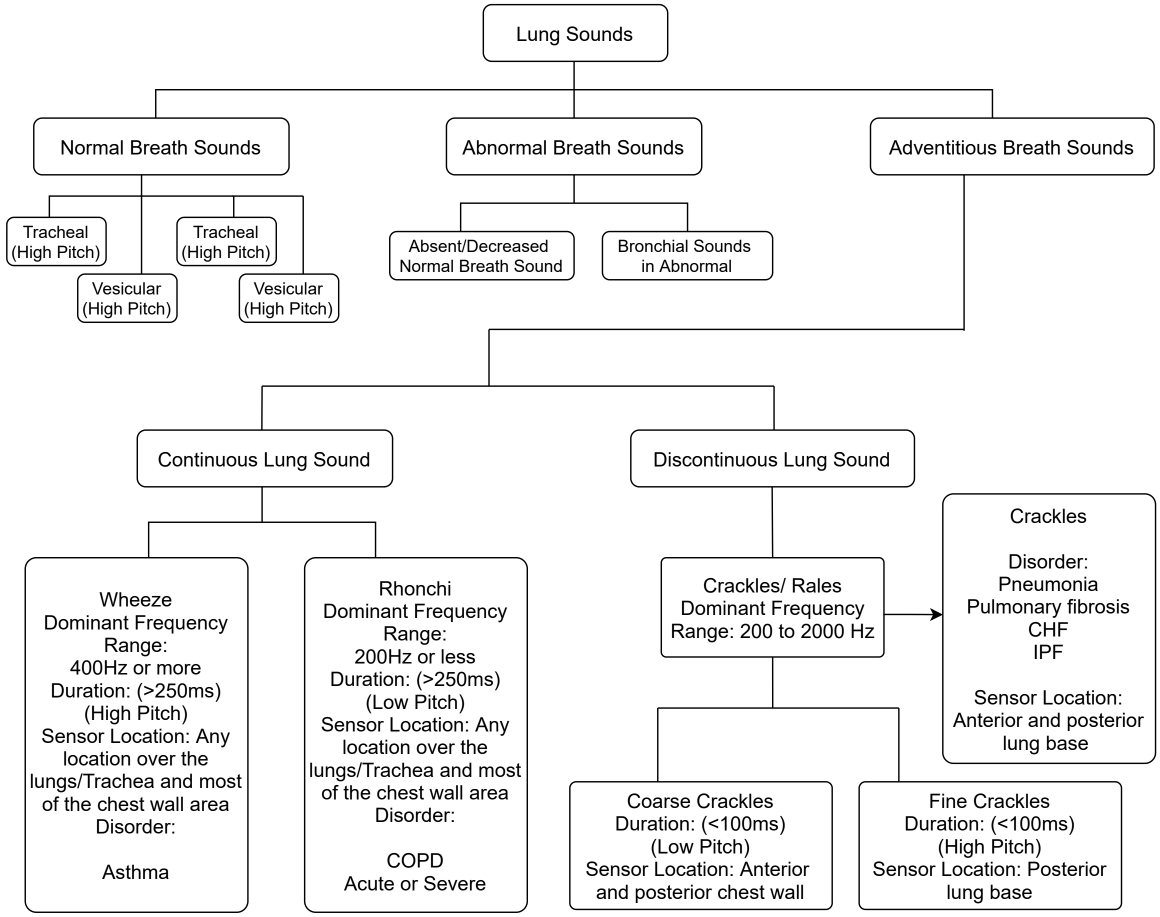

Figure 3.1: characteristic of lung sound

contains both the public and the private dataset of the ICBHI challenge. The Respiratory Sound Database contains audio samples that were collected independently by two research teams in two different countries (Greece and Portugal) over several years. The data collection required several years, and the final dataset consists of 920 labeled audio tracks from 126 distinct participants. It is currently the largest annotated, publicly available dataset.

The two independent research groups are

(1) Respiratory Research and Rehabilitation Laboratory (Lab3R), School of Health Sciences, University of Aveiro, Aveiro, Portugal and

(2) Papanikolaou General Hospital and the General Hospital of Imathia, Aristotle Univer-

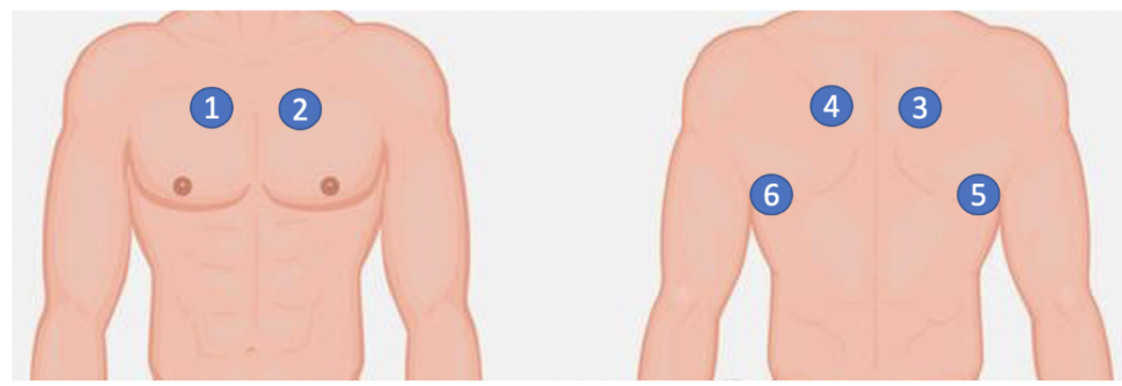

Figure 3.2: Locations from which respiratory sounds were collected: right anterior (1), left anterior (2), right posterior (3), left posterior (4), right lateral (5) and left lateral (6).

sity of Thessaloniki and the University of Coimbra, Thessaloniki, Greece.

These audio signals were recorded using one of the following stethoscope systems:

(1) Electronic Stethoscope 3200, 3M Littmann,

(2) Classic II SE Stethoscope, 3M Littmann

(3) C417 L Professional Lavalier Microphone, AKG HARMAN, and

(4) Meditron Master Elite Electronic Stethoscope, Welch Allyn. The sounds were collected from six different positions (left/right anterior, posterior and lateral) as illustrated in Figure 3.2.

The audios were collected in both clinical and non-clinical settings from adult participants of disparate ages. Participants encompassed patients with lower and upper respiratory tract infections, pneumonia, bronchiolitis, COPD, asthma, bronchiectasis, and cystic fibrosis.

3.3 Annotations

The ICBHI sound data were provided with two types of annotation: i) for each respiratory cycle, whether or not crackles and/or wheezes are present, and ii) for every patient, whether or not a specific pathology from a set of predetermined categories is present.

The ICBHI database consists of 920 annotated audio samples from 126 subjects and so it is used as a benchmark in the field. Each respiratory cycle in the dataset is annotated with 4 classes. The annotations basically cover 2 broad groups-normal and problematic. The problematic section is further divided into wheeze and crackle with some cycles having both issues. Among 6898 cycles totaling to 5.5 hours, 3642 cycles are healthy while the remaining 3256 are problematic. Out of these problematic cycles, 1864 cycles have crackles while 886 have wheezes. There are 506 cycles which have both wheezes and crackles. Overall, there were 3642 healthy breath cycles and 3256 problematic breath cycles.

A single-channel respiratory sound is composed of a certain number of cycles, which in turn include four main components, two pauses, and two distinctive patterns. Discarding fine-grain variations, mostly due to the conversion of air vibrations to electrical signal, a respiratory cycle is conventionally described as follows: it starts from the inspiratory phase, which is characterized by a lower amplitude and a regular pattern, then it follows with an expiratory phase, which shows one or multiple peaks, a decreasing amplitude pattern, and is usually characterized by a higher average energy. As previously mentioned, the respiratory cycles were annotated by domain experts to state the presence of crackles, wheezes, a combination of them, or no adventitious respiratory sounds. More in detail, the annotation style format includes the beginning of the respiratory cycle(s), as well as the end of the respiratory cycle(s), the presence or absence of crackles, and the presence or absence of wheezes. The recordings were collected using heterogeneous equipment, with duration ranging from 10 s to 90 s. The average duration of a respiratory cycle is 2.7 s, with a standard deviation of about 1.17 s; the median duration is about 2.54 s, whereas the duration ranges from 0.2 s to above 16 s. Moreover, wheezes are characterized by an average duration of about 600 ms, with a relatively high variance, and a minimum and maximum duration value ranging between 26 ms and 19 s; conversely, crackles are characterized by an average duration of about 50 ms, smaller variance, and a minimum and maximum duration values of 3 ms and 4.88 s, respectively.

Table 3.1: Cycle Breakdown Of ICBHI 2017 Challenge Dataset

Number of Cycles	Total
With crackles	1864
With wheezes	886
With crackles + wheezes	506
Normal cycles	3642
Total number of cycles	6898

It is important to note that the detection range for crackles and wheezes lies within 100 to 2500 Hz, therefore any other sounds that are outside this range, such as noise, can be safely discarded or filtered without significant loss of quality of the adventitious sounds.

3.4 Challenges

While recording, the participants were seated. The acquisition of respiratory sounds was performed on adult and elderly patients. Many patients had COPD with comorbidities (e.g., heart failure, diabetes, hypertension). Further, there was also presence of noise like the rubbing sound of the stethoscope with the patient's dress, background talking etc. Such varieties in the data made it very challenging to identify problems in the respiratory sounds. One of the most challenging aspects of the audio clips was the presence of heartbeat sound along with the breath sounds. No preprocessing was performed to remove the heartbeat sounds. Pictorial representations of 200 audio clips from the healthy and non-healthy class are shown in Figure 3.3.

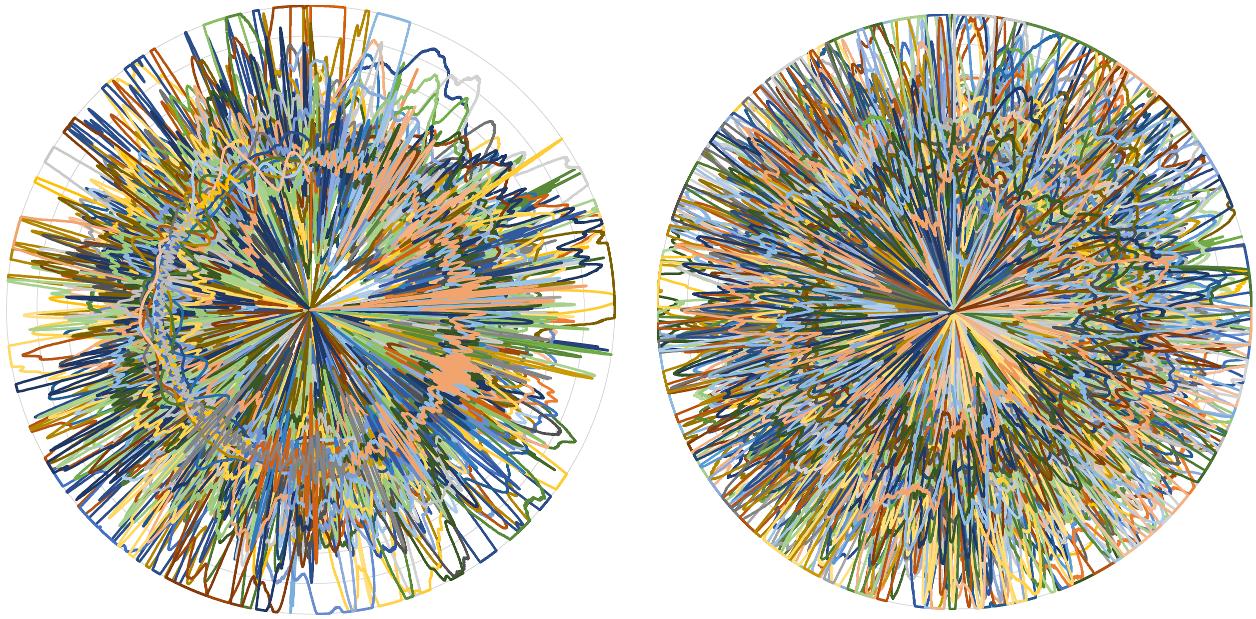

Figure 3.3: 200 audio clips (original): healthy class (left) and non-healthy class (right).

3.5 Summary

We have discussed about the dataset used, annotations and the challenges faced while using the annotated audio files in this chapter. In the next chapter we will examine about the Methodology, how we have preprocessed and extracted features from our data.

CHAPTER 4

SHALLOW LEARNING

4.1 Foreword

In this chapter, we describe about the framing and windowing technique used for pre-processing. Then linear predictive analysis is done along with grading and standard deviation measurement for feature extraction. Next by using Multi layer perceptron we classified healthy vs non healthy respiratory sounds.

4.2 Data Preprocessing

Typically, to evaluate robustness of algorithms, health professionals detect adventitious respiratory sounds by annotating sounds with the help of Respiratory Sound Annotation Software (SAS). As audio clip contains high deviations across its entire length, its analysis is not trivial. Therefore, each audio clip is broken down into smaller segments called frames to facilitate analysis. In our research, we divided the clips into frames consisting of 256 sample points with a 100-point overlap in between them. The parameters were chosen based on [22]. The same 200 audio clips (as in Figure 3.3) are shown in Figure 4.1 after framing. The number of Sz sized overlapping frames Of with O overlapping points for a signal having S points is presented below:

$$O_f = \left\lceil S - S_z O + 1 \right\rceil. \tag{4.1}$$

After framing the signal into shorter segments, it was observed that in various instances the starting and ending points were not aligned in a frame. These discontinuities/ jitters lead to smearing of power across the frequency spectrum. This posed a problem in the form of spectral leakage during frequency domain analysis which produced additional frequency components. To tackle this, the frames were subjected to a window function. Hamming window was chosen for this purpose due to its efficacy as demonstrated in [22]. Post framing, jitters might be observed in them which interfere with the Fourier Transformation of the same in the form of spectral leakage. In order to minimize such problems, the frames are usually multiplied with a windowing function which approaches 0 towards its ends and reaches its peak in the middle. Amidst various such windowing functions, Hamming Window function is one of the popularly used windowing functions. The same frames (Figure 4.1) are presented in Figure 4.2 after windowing. The hamming window is mathematically illustrated below:

$$A(z) = 0.54 - 0.46 \cos\left(\frac{2\pi z}{S_z - 1}\right), \tag{4.2}$$

where $A(z)$ is the hamming window function and z is a point within a frame.

4.3 Feature extraction

After frame extraction, we performed Linear Predictive Coefficient(LPC) analysis [23] on each of them. A present sample is represented in terms of previous samples. The previous P samples are used to present the rth sample in a signal s() as presented below:

$$s(r) \approx \; p_1 s(r-1) + p_2 s(r-2) + p_3 s(r-3) +, \ldots, + p_P s(r-P), \tag{4.3}$$

where p_1, p_2,..., p_P are the LPCs or predictors. The error of this prediction $E(r)$ bounded by the actual and predicted samples: ($s(r)$ and $\hat{s}$(r)) can be explained as

$$E(r) = s(r) - \hat{s}(r) = s(r) - \sum_{k=1}^{P} p_k s(r-k). \tag{4.4}$$

The error of sum of squared differences (as shown below) is minimized to generate the unique predictors for a x sized frame, which can be expressed as

$$E_r = \sum_x \Big[s_r(x) - \sum_{k=1}^{P} p_k s_r(x-k) \Big]^2. \tag{4.5}$$

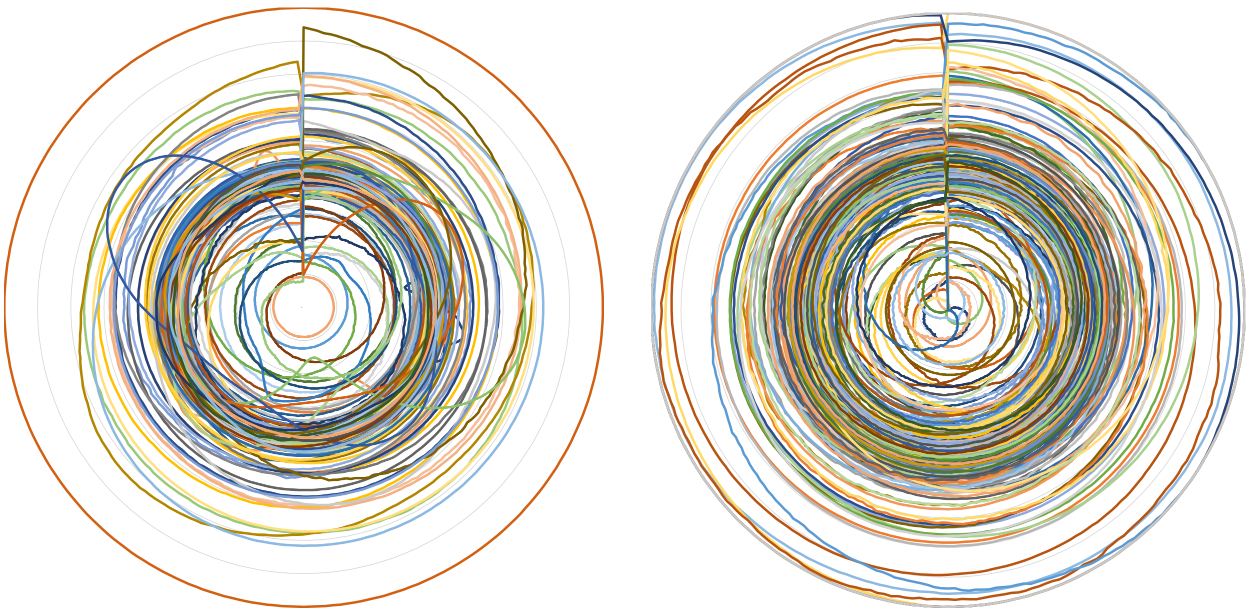

Figure 4.1: The same 200 audio clips (as in Fig. 3.3) after framing: healthy class (left) and non-healthy class (right).

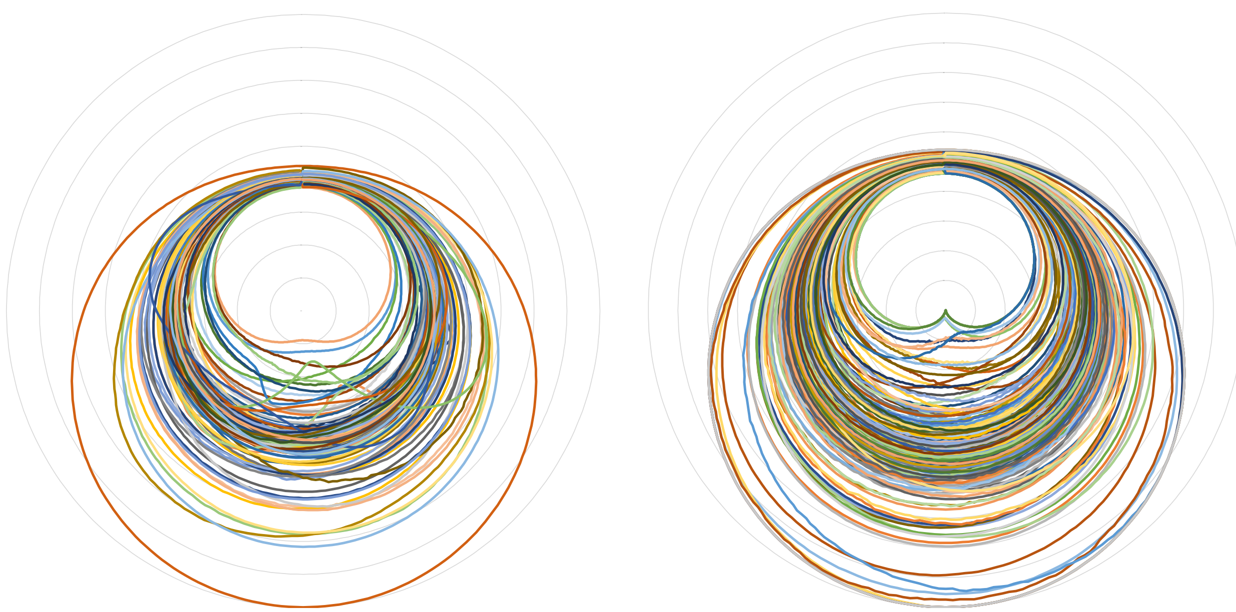

Figure 4.2: Representation of the same 200 audio clips (as in Fig. 3.3) after windowing: healthy class (left) and non-healthy class (right).

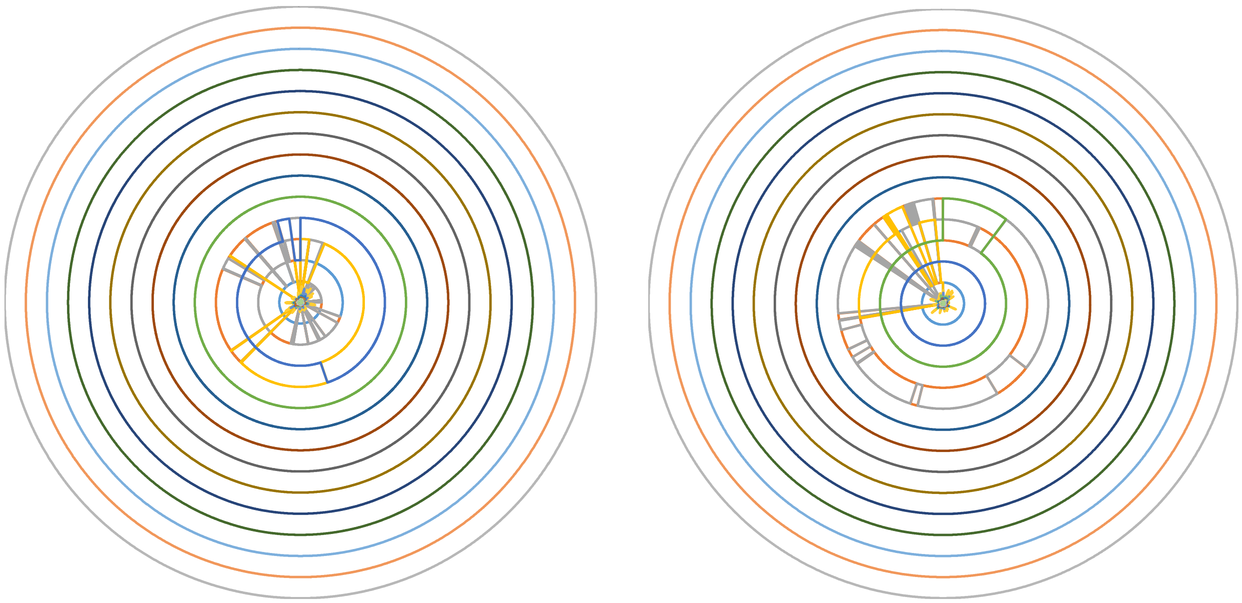

Figure 4.3: Representation of 30 dimensional features for the audio clips: healthy class (left); non-healthy class (right).

Thereafter, a recursive technique is used to compute the Cepstral coefficients (C), which is expressed as

$$C_0 = \log_e P$$

$$C_r = p_r + \sum_{q=1}^{r-1} qrC_q p_{r-q}, for\ 1 < r \leq P\ and$$

$$C_r = \sum_{q=r-P}^{r-1} qrC_q p_{r-q}, for\ r > P \tag{4.6}$$

Since clips in the dataset were of unequal lengths and number of frames obtained varied. When features were extracted in frame level, it produced different dimensions. To handle this, we performed two operations: a) grading and b) standard deviation measurement.

1. Firstly, the sum of LPCC coefficients in each of the frequency ranges (bands) across all the frames was computed. Based on the sum of these energy values, bands were graded in an ascending order. This sequence of band numbers was used as features that helped in identifying dominance of different bands for the clips from various categories.

2. Secondly, standard deviation was computed for every band. These two metrics were

stacked to form the feature, which is independent of the clip length. 10, 20, 30, 40 and 50 dimensional features were extracted for the 2 classes. The trend of the 30-dimensional feature values (best result) for the 2 classes is shown in Fig. 4.3

4.4 Classification

4.4.1 Multi-layer perceptron(MLP)

Multilayer perceptron's (MLPs) otherwise called as Feedforward neural networks (FNNs) are the archetypes of deep learning models. These networks were inspired by neuroscience and how we believe neurons work in the brain.

The purpose of these networks is to approximate some function f by mapping an input domain to an output domain, which can be applied to solving complex problems such as prediction or classification from high dimensional data to a set of labels.

These networks consist of multiple layers, where the first layer is the input layer and the last is the output layer. The intermediate layers in the network are called the hidden layers and their number can vary. The use of multiple layers is what originated the term "Deep Learning", with each additional layer creating an additional level of abstraction or representation.

Each layer is comprised of a number of neurons that represent activation values, and it determines the width of that layer. Each neuron has a number of input weights that connect to each of the neurons of the previous layer, with the exception of the neurons in the input layer.

The activation values of the input layer are propagated forward in the direction of the output layer with no feedback connections where the outputs of the neurons are fed to previously activated neurons, hence the designation of "feedforward".

The network is associated with a directed acyclic weighted graph describing how the

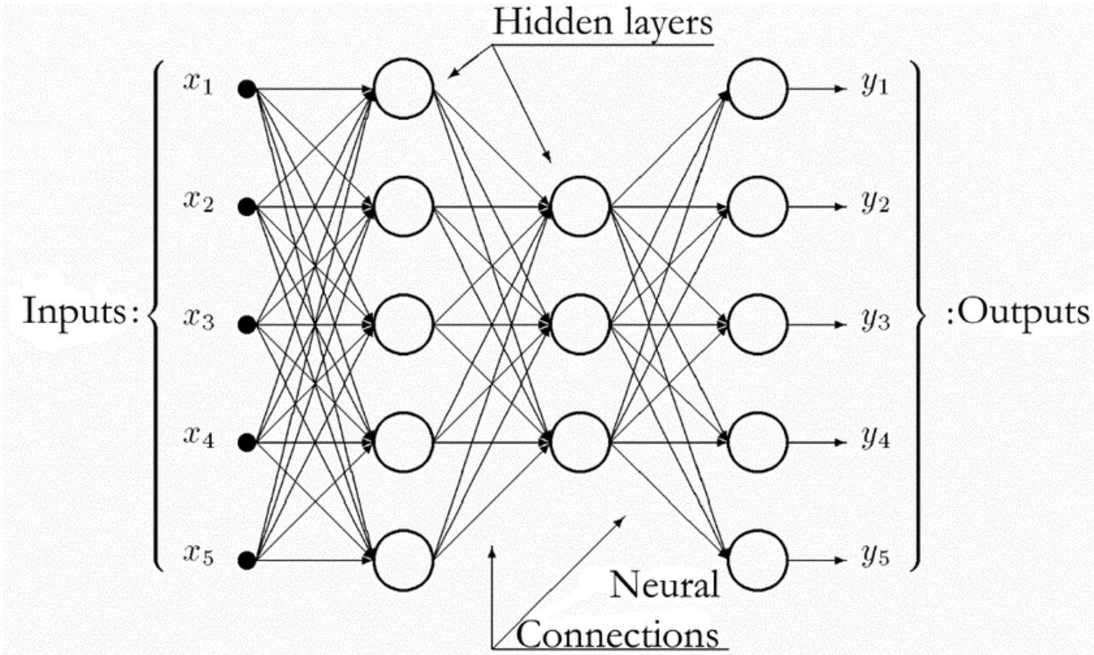

Figure 4.4: Structure of a feed-forward ANN with two hidden layers

functions are composed together. The network's parameters consist of the weights and biases between layers. The output activation values of a layer are represented as a vector, with each entry of the vector representing the activation value of a single neuron. The size of the vector corresponds to the number of neurons in that layer.

The weights between layers are represented as a 2D matrix, with each entry of the matrix at coordinates i,j representing the weight connecting the neuron i from layer l - 1 to the neuron j in the layer l. The biases between layers are represented as a vector with the same size as the number of neurons in the next layer.

The mathematical equation for the calculation of the output of each layer of the feed-forward model is defined as:

• hl = gl(Wlhl - 1 + bl), the activation values of a layer. With Wlhl - 1 being the dot product operation between the weight matrix of the current layer and the output values of the previous layer.

• y = hL - 1 , the activation values of the final output layer of the network

4.4.2 Layers

We employed MLP classifier, feed-forward artificial neural network – for classification purpose [57]. Feedforward neural networks are made up of the input layer, output layer and hidden layer. It is a supervised learning algorithm trained on a dataset using a function f() : Zn $\rightarrow$ Zo, where n and o represent the dimensions for input and output. For a given set of features $\mathrm{P} = \mathrm{p}_1, p_2, ..., p_n$ and aim x, a non- linear function is learned for classification. The difference between MLP and logistic regression lies in the existence of one or more non-linear layers (hidden layers) between the input and the out- put layer. MLP consists of three or more layers (input layer, output layer and one or more hidden layers) of non-linear activating neurons. The number of hidden layers can be increased according to the requirement of developing a model to accomplish certain task. The initial layer is the input layer which comprises of a set of neurons $\mathrm{p}_i|p_1, p_2, ..., p_n$ denoting the features. Each neuron of the hidden layer modifies the values from the previous layer using sum of weights as: $\mathrm{w}_1 p_1 + w_2 p_2 +, ..., +w_2 p_n$.
The activation function that represents the relationship between input and output layer in of non-linear nature. It makes the model flexible in defining unpredictable relationships. The activation function can be expressed as:

$$y_i = \tanh(w_i) \quad \text{and} \quad y_i = (1 + e^{w_i})^{-1}, \tag{4.7}$$

where yi and wi denotes the outcome of the ith neuron and weighted sum of the input features. The values from the ultimate hidden layer are provided to the output layer as output values. Each layer of MLP contains several fully connected layers as each neuron in a layer is attached to all the neurons of the previous layer. The parameters of each neuron are independent of the remaining neurons of the layer ensuring possession of unique set of weights. The initial momentum and learning rate were set to 0.2 and 0.3 respectively.

4.5 Summary

We have discussed about the methodology we used to preprocess, extract features and then classify the respiratory sounds. In the next chapter we will look into evaluation where we discuss and analyze the results.

CHAPTER 5

RESULTS AND ANALYSIS

5.1 Evaluation Metric and Protocol

Not just Accuracy it is also very much important to analyze the disparate misclassifications. Hence, to evaluate our tool, the following performance metrics are used: Precision, Accuracy, Sensitivity (Recall), Specificity, and Area under ROC curve (AUC). They are computed as,

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, \tag{5.1}$$

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. Accuracy is a great measure only when datasets are symmetric where values of false positive and false negatives are almost same. Therefore, looking at other parameters to evaluate the performance of model is important.

$$\text{Precision} = \frac{T_P}{T_P + F_P}, \tag{5.2}$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$\text{Sensitivity (Recall)} = \frac{T_P}{T_P + F_N},$$

$$(5.3)$$

Sensitivity (Recall) is the ratio of correctly predicted positive observations to the all observations in actual class

$$\text{Specificity} = \frac{T_N}{T_N + F_P}, \text{ and} \tag{5.4}$$

Specificity is the metric that evaluates a model's ability to predict true negatives of each available category. These metrics apply to any categorical model. The equation for recall looks exactly the same as the equation for sensitivity and when to use either term depends on the task at hand.

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{5.5}$$

F1 score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. F1 is usually more useful than accuracy, especially in an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

where TP , TN , FP , and FN refer to true positive, true negative, false positive, and false negative, respectively.

Table 5.1: Performance of different feature dimensions using MLP.

Feature dim.	Accuracy(%)
10	93.91
20	90.01
30	**99.07**
40	89.19
50	98.78

To avoid possible bias in evaluation, 5-fold cross validation was used. Cross validation was used because it subjects each instance of the dataset to testing and training at least once. This also helps to avoid biased modeling when outliers are present.

5.2 Our Results

The performance of the different features are provided in Table 5.1. It is observed that the best result was obtained with 30 dimensional features and it's corresponding confusion matrix is provided in Table 5.2.

As compared to the default scenario, there were 4 more misclassifications in the case of the healthy cases (and 9 less misclassifications for the non-healthy cases).

Next, the momentum was varied from 0.1 to 0.5 with a step of 0.1, and results are provided in Table 5.3.

The best result was obtained for a momentum of 0.1 whose interclass confusions are provided in Table 5.4.

Finally, the momentum was varied from 0.1-0.6 with a step of 0.1 whose results are provided in Table 5.5. In our experiment, the highest performance was obtained when a learning rate of 0.5 was selected. We presented a confusion matrix for this setup in Table 5.6. It is observed that the number of misclassifications for both classes was reduced as

Table 5.2: Inter-class confusions for the 30 dimensional features (Best result) using MLP.

	Healthy	Non-healthy
Healthy	3611	31
Non-healthy	33	3223

Table 5.3: Performance for different momentum values on 30 dimensional features with learning rate of 0.3.

Momentum	Accuracy(%)
0.1	**99.14**
0.2	99.07
0.3	99.04
0.4	99.07
0.5	99.12

compared to the initial setup. The misclassified instances were analyzed, and it was found that many of them had heartbeat sounds. Along with this, other unwanted artefacts, such as talking, and movement of the probe helped in misclassifying.

It is observed that the misclassified instances were reduced by almost 15.63% as compared to the original setup using default settings. As compared to best result, after momentum tuned, a decrease of nearly 8.47% occurred for the misclassified instances. A deeper analysis of the misclassifications revealed that approximately 0.74% of the healthy

Table 5.4: Inter-class confusions for momentum value of 0.1 on 30 dimensional features.

	Healthy	Non-healthy
Healthy	3607	35
Non-healthy	24	3232

Table 5.5: Performance for different learning rates with momentum of 0.2.

Learning rate	Accuracy(%)
0.1	99.03
0.2	99.13
0.3	99.07
0.4	99.06
0.5	**99.22**
0.6	99.13

Table 5.6: Interclass confusions for learning rate of 0.5 and momentum of 0.2 on 30 dimensional features.

	Healthy	non-healthy
Healthy	3615	27
non-healthy	27	3229

Table 5.7: Performance metrics for default scenario, best results after tuning momentum value and best result after tuning learning rate.

Metrics	Default	Best momentum	**Best learning rate**
Sensitivity	0.9915	0.9904	**0.9917**
Specificity	0.9899	0.9926	**0.9926**
Precision	0.9909	0.9834	**0.9917**
False positive rate	0.0101	0.0074	**0.0074**
False negative rate	0.0085	0.0096	**0.0083**
Accuracy(%)	99.07	99.14	**99.22**
F1 score	0.9912	0.9919	**0.9917**
AUC	0.9994	0.9995	**0.9993**

Table 5.8: Performance of different classifiers on the 30 dimensional features.

Classifier	Accuracy(%)
BayesNet	98.26
Naïve Bayes	88.98
SVM	98.59
RBF Network	95.82
LibLINEAR	98.59
Simple Logistic	98.70
Decision Table	98.62
RNN	93.82
Multilayer Perceptron	**99.22**

cases were misclassified as opposed to non-healthy. In the case of non-healthy instances, approximately 0.83% of the clips were misclassified as healthy, which we call false negative. The different performance metrics were computed for the default setup, best momentum, and best learning rate (overall highest). Such results are provided in Table 5.7. The ROC curves for these scenarios are shown in Fig. 5.1.

5.3 Comparative study

The reason why we have done a comparative study is to know how much we did and the performance of several other classifiers was compared in order to establish the efficacy of MLP. For comparison, the 30-dimensional feature set (best performance) was chosen. We experimented with BayesNet, SVM, RNN, Naive Bayes, RBF network, Decision Table, LibLINEAR, and Simple logistic. The results are provided in Table 5.8. We also compared the performance of our system with reported works by Kok et al. [37] and Chambers et

Table 5.9: Comparison with reported works.

Work	Accuracy(%)
Kok et al. [**?**]	87.10
Chambers et al. [**?**]	85.00
Proposed technique	**99.22**

al. [38]. The average accuracies for both the systems along with the proposed system are provided in Table 5.9.

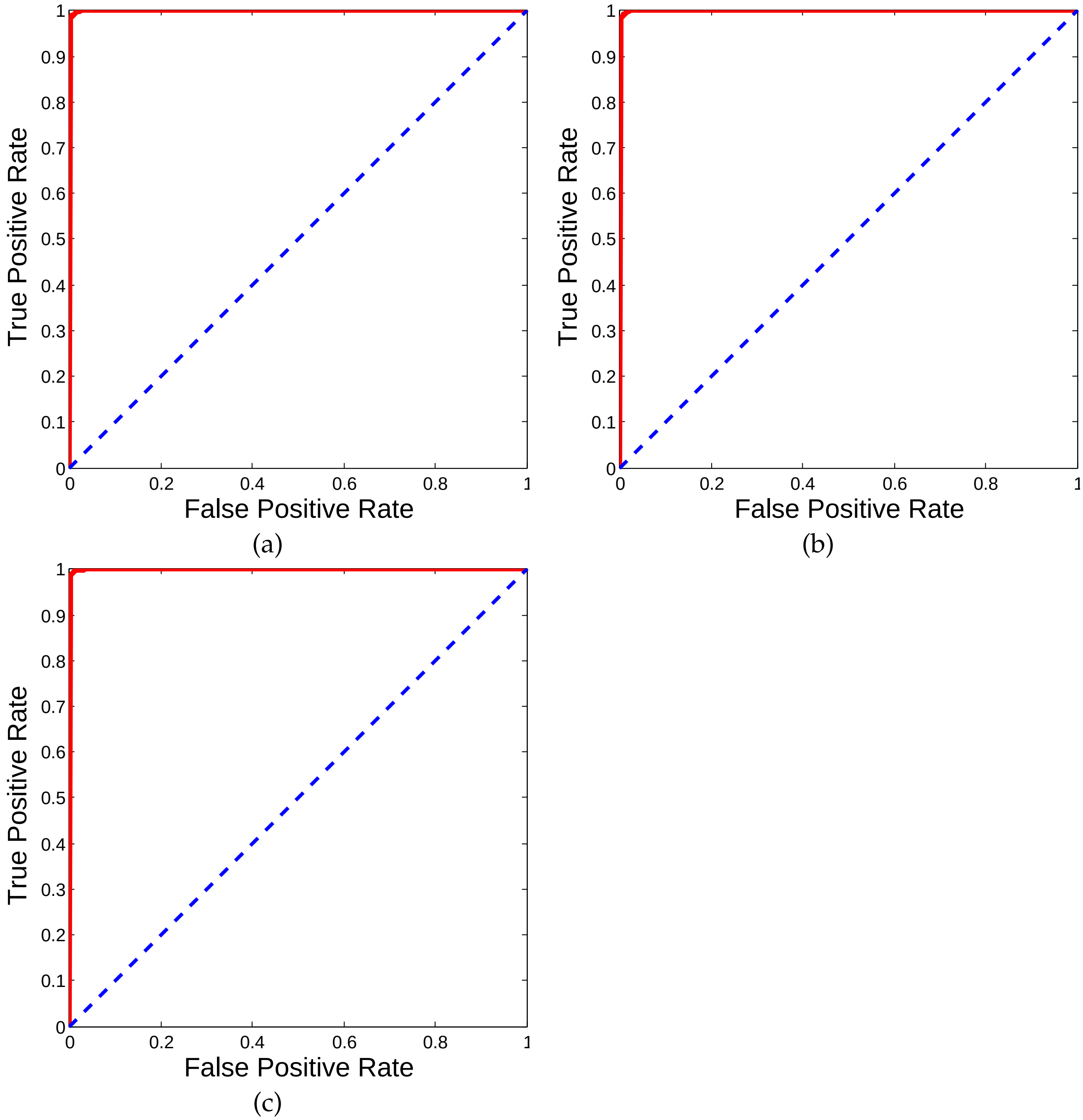

Figure 5.1: ROC curves: a) default settings, b) best momentum value (0.1), and c) best learning rate (0.5, overall highest result).

CHAPTER 6

CONCLUSION

Looking at the audio content, it is difficult to classify respiratory sounds. In our research, a system is presented for distinction of healthy and non-healthy lung sounds which is very important prior to further diagnosis of the type and severity of infection. We have performed our experiments using a publicly available dataset and evaluations indicate that the highest accuracy of 99.22% with an AUC value of 0.9993 is obtained.

Automated adventitious sounds detection or classification provides a promising solution to overcome the limitations of conventional auscultation. In future the subject area for future investigation will be:

1. To use larger dataset and test further on robustness in presence of higher percentages of noise.
2. Attempts will also be made towards isolation of breath sounds from the ambient noises and heart- beat sounds [58] for better analysis.
3. Other acoustic techniques [59] will be applied for even better modelling of the lung sounds along with deep learning-based approaches.
4. To have clinical use in pulmonary health screening and as a tool in differential diagnosis of pulmonary diseases.
5. Finally, we will be trying to identify the nature and severity of infection from the breath sounds.

BIBLIOGRAPHY

[1] Tea Lallukka, Anoushka Millear, Amanda Pain, Monica Cortinovis, and Giorgia Giussani. Gbd 2015 mortality and causes of death collaborators. global, regional, and national life expectancy, all-cause mortality, and cause-specifi c mortality for 249 causes of death, 1980-2015: a systematic analysis for the global burden of disease study 2015 (vol 388, pg 1459, 2016). *Lancet*, 389(10064):E1–E1, 2017.

[2] Global status report on noncommunicable diseases 2014. geneva, world health organization, 2014. *Available from: http://www.who.int/nmh/publications/ncd-status-report-2014/en/*.

[3] World health organization. density of physicians. *http://www.who.int/gho/health workforce/physicians density/en/, 2017. [Online; accessed 18-May-2018]*.

[4] KC Santosh. Speech processing in healthcare: Can we integrate? In *Intelligent Speech Signal Processing*, pages 1–4. Elsevier, 2019.

[5] Himadri Mukherjee, Subhankar Ghosh, Shibaprasad Sen, Obaidullah Sk Md, KC Santosh, Santanu Phadikar, and Kaushik Roy. Deep learning for spoken language identification: Can we visualize speech signal patterns? *Neural Computing and Applications*, 31(12):8483–8501, 2019.

[6] Himadri Mukherjee, Sk Md Obaidullah, KC Santosh, Santanu Phadikar, and Kaushik Roy. Line spectral frequency-based features and extreme learning machine for voice activity detection from audio signal. *International Journal of Speech Technology*, 21(4):753–760, 2018.

[7] Victor A McKusick, John T Jenkins, and George N Webb. The acoustic basis of the chest examination; studies by means of sound spectrography. *American review of tuberculosis*, 72(1):12–34, 1955.

[8] ARA Sovijarvi, J Vanderschoot, and JE Earis. Standardization of computerized respiratory sound analysis. *European Respiratory Review*, 10(77):585–585, 2000.

[9] Volker Gross, LJ Hadjileontiadis, Thomas Penzel, Ulrich Koehler, and C Vogelmeier. Multimedia database" marburg respiratory sounds (mars)",". In *Proc. 25th Annual Int Engineering in Medicine and Biology Society Conf. of the IEEE*, volume 1, pages 456–457, 2003.

[10] Rale: A computer-assisted instructional package. *Respir Care 1990;35:1006.*

[11] William H Wolberg, W Nick Street, and Olvi L Mangasarian. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer letters*, 77(2-3):163–171, 1994.

[12] Sotiris B Kotsiantis, I Zaharakis, P Pintelas, et al. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007.

[13] Arumugam Kandaswamy, C Sathish Kumar, Rm Pl Ramanathan, S Jayaraman, and N Malmurugan. Neural classification of lung sounds using wavelet coefficients. *Computers in biology and medicine*, 34(6):523–537, 2004.

[14] Jack V Tu. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11):1225–1231, 1996.

[15] Marcin Raniszewski. The edited nearest neighbor rule based on the reduced reference set and the consistency criterion. *Biocybernetics and Biomedical Engineering*, 30(1):31–40, 2010.

[16] Sameer Alsmadi and Yasemin P Kahya. Design of a dsp-based instrument for real-

time classification of pulmonary sounds. *Computers in biology and medicine*, 38(1):53–61, 2008.

[17] Geert Meyfroidt, Fabian Güiza, Jan Ramon, and Maurice Bruynooghe. Machine learning techniques to examine large patient databases. *Best Practice & Research Clinical Anaesthesiology*, 23(1):127–143, 2009.

[18] Shijun Wang and Ronald M Summers. Machine learning and radiology. *Medical image analysis*, 16(5):933–951, 2012.

[19] İnan Güler, Hüseyin Polat, and Uçman Ergün. Combining neural network and genetic algorithm for prediction of lung sounds. *Journal of Medical Systems*, 29(3):217–231, 2005.

[20] Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, and Wei-Yang Lin. Intrusion detection by machine learning: A review. *expert systems with applications*, 36(10):11994–12000, 2009.

[21] Jyh-Shing Roger Jang, Chuen-Tsai Sun, E Mizutani, and Y-C Ho. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence. *PROCEEDINGS-IEEE*, 86(3):600–603, 1998.

[22] Himadri Mukherjee, Sk Md Obaidullah, Santanu Phadikar, and Kaushik Roy. Misna-a musical instrument segregation system from noisy audio with lpcc-s features and extreme learning. *Multimedia Tools and Applications*, 77(21):27997–28022, 2018.

[23] Himadri Mukherjee, Ankita Dhar, Sk Md Obaidullah, KC Santosh, Santanu Phadikar, and Kaushik Roy. Linear predictive coefficients-based feature to identify top-seven spoken languages. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(06):2058006, 2020.

[24] Bruno M Rocha, Dimitris Filos, Luís Mendes, Gorkem Serbes, Sezer Ulukaya, Yasemin P Kahya, Nikša Jakovljevic, Tatjana L Turukalo, Ioannis M Vogiatzis, Eleni Perantoni, et al. An open access database for the evaluation of respiratory sound classification algorithms. *Physiological measurement*, 40(3):035001, 2019.

[25] Murat Aykanat, Özkan Kılıç, Bahar Kurt, and Sevgi Saryal. Classification of lung sounds using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, 2017(1):1–9, 2017.

[26] Himadri Mukherjee, Sk Md Obaidullah, KC Santosh, Santanu Phadikar, and Kaushik Roy. A lazy learning-based language identification from speech using mfcc-2 features. *International Journal of Machine Learning and Cybernetics*, 11(1):1–14, 2020.

[27] Renard Xaviero Adhi Pramono, Syed Anas Imtiaz, and Esther Rodriguez-Villegas. Evaluation of features for classification of wheezes and normal respiratory sounds. *PloS one*, 14(3):e0213659, 2019.

[28] Jyotibdha Acharya and Arindam Basu. Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. *IEEE transactions on biomedical circuits and systems*, 14(3):535–544, 2020.

[29] Zümray Dokur. Respiratory sound classification by using an incremental supervised neural network. *Pattern Analysis and Applications*, 12(4):309–319, 2009.

[30] Vinita Shivakumar. Classification of respiratory sounds.

[31] Pedro Sousa Faustino. Crackle and wheeze detection in lung sound signals using convolutional neural networks. 2019.

[32] Yi Ma, Xinzi Xu, and Yongfu Li. Lungrn+ nl: An improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation. In *Interspeech*, pages 2902–2906, 2020.

[33] Dimitra Emmanouilidou, Eric D McCollum, Daniel E Park, and Mounya Elhilali. Computerized lung sound screening for pediatric auscultation in noisy field environments. *IEEE Transactions on Biomedical Engineering*, 65(7):1564–1574, 2017.

[34] Ipek Sen, Murat Saraclar, and Yasemin P Kahya. A comparison of svm and gmm-based classifier configurations for diagnostic classification of pulmonary sounds. *IEEE Transactions on Biomedical Engineering*, 62(7):1768–1776, 2015.

[35] Fatih Demir, Aras Masood Ismael, and Abdulkadir Sengur. Classification of lung sounds with cnn model using parallel pooling structure. *IEEE Access*, 8:105376–105383, 2020.

[36] Hai Chen, Xiaochen Yuan, Zhiyuan Pei, Mianjie Li, and Jianqing Li. Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access*, 7:32845–32852, 2019.

[37] Xuen Hoong Kok, Syed Anas Imtiaz, and Esther Rodriguez-Villegas. A novel method for automatic identification of respiratory disease from acoustic recordings. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2589–2592. IEEE, 2019.

[38] Gaëtan Chambres, Pierre Hanna, and Myriam Desainte-Catherine. Automatic detection of patient with respiratory diseases using lung sound analysis. In *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6. IEEE, 2018.

[39] Gokhan Altan, Yakup Kutlu, and Novruz Allahverdi. Deep learning on computerized analysis of chronic obstructive pulmonary disease. *IEEE journal of biomedical and health informatics*, 24(5):1344–1350, 2019.

[40] Adam Rao, Emily Huynh, Thomas J Royston, Aaron Kornblith, and Shuvo Roy.

Acoustic methods for pulmonary diagnosis. *IEEE reviews in biomedical engineering*, 12:221–239, 2018.

[41] Arnon Cohen and Dorota Landsberg. Analysis and automatic classification of breath sounds. *IEEE Transactions on Biomedical Engineering*, (9):585–590, 1984.

[42] M Bahoura and C Pelletier. Respiratory sounds classification using cepstral analysis and gaussian mixture models. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 9–12. IEEE, 2004.

[43] P Mayorga, C Druzgalski, RL Morelos, OH Gonzalez, and J Vidales. Acoustics based assessment of respiratory diseases using gmm classification. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 6312–6316. IEEE, 2010.

[44] Rajkumar Palaniappan, Kenneth Sundaraj, and Sebastian Sundaraj. A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. *BMC bioinformatics*, 15(1):1–8, 2014.

[45] MARIO Milicevic, IGOR Mazic, and MIRJANA Bonkovic. Classification accuracy comparison of asthmatic wheezing sounds recorded under ideal and real-world conditions. In *15th International Conference on Artificial Intelligence, Knowledge Engineering and Databases (AIKED 2016), Venice*, 2016.

[46] BM Rocha, L Mendes, I Chouvarda, P Carvalho, and RP Paiva. Detection of cough and adventitious respiratory sounds in audio recordings by internal sound analysis. In *International Conference on Biomedical and Health Informatics*, pages 51–55. Springer, 2017.

[47] Gorkem Serbes, Sezer Ulukaya, and Yasemin P Kahya. An automated lung sound

preprocessing and classification system based onspectral analysis methods. In *International Conference on Biomedical and Health Informatics*, pages 45–49. Springer, 2017.

[48] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[49] J Earis. Lung sounds. *Thorax*, 47(9):671, 1992.

[50] Hans Pasterkamp, Steve S Kraman, and George R Wodicka. Respiratory sounds: advances beyond the stethoscope. *American journal of respiratory and critical care medicine*, 156(3):974–987, 1997.

[51] SK Chowdhury and AK Majumder. Frequency analysis of adventitious lung sounds. *Journal of biomedical engineering*, 4(4):305–312, 1982.

[52] Zbigniew Korona and Mieczyslaw M Kokar. Lung sound recognition using model-theory-based feature selection and fusion. *Applied Signal Processing*, 5(3):152, 1998.

[53] Howelldcj.in:geoffreyjl,stevends,editors. signs of respiratory disease: lung sounds.

[54] RB Urquhart, J McGhee, JES Macleod, SW Banham, and F Moran. The diagnostic value of pulmonary sounds: a preliminary study by computer-aided analysis. *Computers in biology and medicine*, 11(3):129–139, 1981.

[55] Mcgee s. auscultation of the lungs. philadelphia: W.b. saunders;. 2012: 251–66 [chapter 28].

[56] Steven m. auscultation of the lungs. saint louis: W.b. saunders;. 2007: 326–45 [chapter 27].

[57] Sankar K Pal and Sushmita Mitra. Multilayer perceptron, fuzzy sets, classifiaction. 1992.

[58] Kun-Hsi Tsai, Wei-Chien Wang, Chui-Hsuan Cheng, Chan-Yen Tsai, Jou-Kou Wang, Tzu-Hao Lin, Shih-Hau Fang, Li-Chin Chen, and Yu Tsao. Blind monaural source separation on heart and lung sounds based on periodic-coded deep autoencoder. *IEEE Journal of Biomedical and Health Informatics*, 24(11):3203–3214, 2020.

[59] Mukherjee Himadri, Hanan Salam, and KC Santosh. Lung health analysis: Adventitious respiratory sound classification using filterbank energies. *IEEE Journal of Biomedical and Health Informatics*, 2021.