A Comprehensive Respiratory Evaluation: Incorporating Lung Sound and Disease Classification Along with Spirometry Assessment

by

Labiba Shayetreen
23241136
Tasfia Mehnaz Tazin
19101136
Ummea Marzan
19101651
Syeda Raisa Afsar
23141097
Afra Anani
19341006

Department of Computer Science and Engineering
School of Data Science
Brac University
September 2023

Declaration

It is hereby declared that

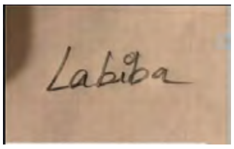
1. The thesis submitted is our own original work while completing the degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

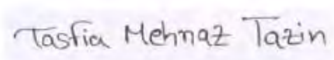
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



Labiba Shayetreen

23241136



Tasfia Mehnaz Tazin

19101136

Ummea Marzan

19101651

Raisa	Afra Anani
Syeda Raisa Afsar	Afra Anani
23141097	19341006

Approval

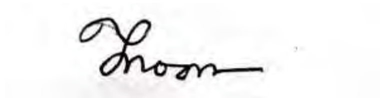
The thesis titled " A Comprehensive Respiratory Evaluation: Incorporating Lung Sound and Disease Classification Along with Spirometry Assessment" submitted by

1. Labiba Shayetreen (23241136)
2. Tasfia Mehnaz Tazin (19101136)
3. Ummea Marzan (19101651)
4. Syeda Raisa Afsar(23141097)
5. Afra Anani (19341006)

Of Summer, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on 18th September, 2023.

Examining Committee:

Supervisor:
(Member)



Jannatun Noor Mukta

Assistant Professor
Department of Computer Science Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Associate Professor
Department Of Computer Science And Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Respiratory disease, also known as pulmonary disease or lung diseases mainly affects the airways and hinders important functions of the lungs (NCI Dictionary of Cancer Terms). Some widely known respiratory diseases include asthma, pneumonia, Bronchiectasis, Bronchiolitis, chronic obstructive pulmonary disease (COPD), pulmonary fibrosis, upper respiratory tract infection (URTI), lower respiratory tract infection (LRTI), and lung cancer. Lung sounds are acoustic signals generated during breathing, commonly referred to as breath noises or respiratory sounds. They can offer insightful information about the condition of a patient's lungs. Wheezing, crackles, or other abnormal lung noises can be a sign of underlying respiratory problems. On the other hand, procedures like Spirometry analyzes the volume and flow of air as a person breathes in and out to determine lung function. Spirometry may not always give a complete picture of a patient's respiratory condition. This is where including lung sound analysis can be really helpful. Spirometry and lung sounds are both crucial instruments for evaluating respiratory health, but they have different roles and yield different kinds of data. While lung sounds provide qualitative details about the noises made when breathing, spirometry concentrates on quantitative measurements of lung function. In this paper, we explore ways in which we can make lung sound results more accurate and classifiable by using respiratory sound readings and by processing the data using machine learning and deep learning. We will be able to classify lung sound data into multiple categories. We will also be classifying spirometry data. In this research, we rigorously compare several machine learning and deep learning models to ascertain how well they classify lung sound and spirometry data. Gated Recurrent Unit (GRU), Support Vector Machine (SVM), Decision Tree, Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN) with different feature extractions, Stacked Autoencoder with SVM, and Attention and Vision Transformer are just a few of the models being examined. Through this assessment, we hope to find the best appropriate model(s) for improving the precision and usefulness of respiratory health evaluations, advancing the level of diagnostic capacities in the field of respiratory medicine.

Keywords: Machine Learning, Respiratory Disease, Classify, Multi-class, Lung Sound database, Convolutional Neural Network (CNN), Feature extraction.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ANN Artificial Neural Network

ATT Attention based model

CNN Convolutional Neural Network

$COPD$ Chronic Obstructive Pulmonary Disease

$CRNN$ Convolutional Recurrent Neural Network

$FEV1$ Forced Expiratory Volume

FOT Forced Oscillation Technique

FVC Forced Vital Capacity

GRU Gated Recurrent Unit

$LRTI$ Lower Respiratory Tract Infection

$LSTM$ Long Short-Term Memory

PFT Pulmonary Function Tests

QLD Quantitative Lung Data

RNN Recurrent Neural Network

SDA Stacked Denoising Autoencoder

SVM Support Vector Machine

$URTI$ Upper Respiratory Tract Infection

$VTLP$ Vocal Tract Length Perturbation

Chapter 1

Introduction

Respiratory diseases are mostly lung and pulmonary diseases. The lung is a very important organ in the human body as it is the main organ that is responsible for supplying oxygen to the cells of the human body. If the lung is not healthy, it can hamper the functions of the body. The most common lung diseases are COPD, asthma, lung cancer, pulmonary fibrosis, upper respiratory tract infection (URTI), and lower respiratory tract infection (LRTI) such as pneumonia, acute bronchitis, Bronchiectasis, Bronchiolitis, etc. and these diseases can be caused due to reasons like, air pollution, dust, occupational chemicals, frequent lung infection from childhood as well as smoking as stated by the National Cancer Institute (NIH)[8].

Numerous lung conditions are included under COPD. But the leading conditions of COPD are emphysema and chronic bronchitis and the disease is characterized by restrictive airflow and airway inflammation. Emphysema is caused by damaged alveoli or air sacs in the lungs and poor elasticity of the lung walls which ultimately affects the amount of air we can in and out[7].

In the year 2019, almost 3.2 million people died due to COPD, becoming the third leading cause of death throughout the world. Among those, 90% of the people were below the age of 70[69]. These lung dysfunctions can not be cured but can be controlled with proper treatment and a healthy lifestyle. In our research, we aim to classify these respiratory diseases better to help the patients as well as the medical caregivers.

Rocha et al. [34] claim that a lung's condition can be indicated by its sound. A person's breathing sound is interconnected with the airflow in that environment, the condition of the lung tissues and the positions of secretion within the lung first hand. A wheezing sound is emitted from a person when their airway is restricted or oxygen can not fully flow through the air pipes which indicates unhealthy lungs or respiratory disease like COPD or asthma[34].

Lung sounds play a vital role in distinguishing between a healthy and unhealthy lung. Healthy lungs sound a lot different than unhealthy lungs because they have different acoustic properties, pitch and depending on the anatomic condition of the area where the readings are being taken. Healthy lung sounds are known as bronchial, bronchovesicular and vesicular sounds.[2] [17]

Unhealthy sounds can be heard along with the raw sounds that the lung normally makes. Wheezes, rhonchi and crackles are the most commonly known abnormal lung sounds. The first feature that helps in classifying an unhealthy lung sound is whether it's constant or irregular. For instance, wheeze and rhonchi are ongoing sounds but crackles are more of discrete acoustical sounds where their interruption period is observed to be less than 25 milliseconds [55]. Crackles are produced when narrow airways crack open on inspiration sounding[10] like dropping a marble on the floor but wheezes create a musical sound due to air flowing through restricted air pipes such as bronchioles. Wheezes are expiratory sounds or both expiratory and inspiratory but not inspiratory alone[1].

However, as stated by Meslier et al.[2], these sounds are somewhat nonspecific without a precise clinical context. In alignment with that, this paper aspires to classify lung diseases by analyzing lung sounds which will lead the medical specialists to come to a valuable decision sooner.

Spirometry is the most prevalent form of PFT or Pulmonary Function Test[65]. This test assesses lung function by measuring the amount of air inhaled and exhaled, as well as the ease and speed of the expulsion of air from the lungs. Spirometry may be conducted if a patient is experiencing wheezing, shortness of breath, or a cough (American Lung Association, 2023). Additionally, spirometry may be conducted prior to surgical procedures to evaluate the patient's lung function. For individuals being treated for chronic lung diseases like COPD, asthma, or pulmonary fibrosis, spirometry helps track the progression of the disease. This test can be performed either in the clinic or in a lab specializing in PFTs.

In our work, we propose a machine learning and deep learning approach to process lung sounds to classify lung disease and sound and additionally include spirometry to gain more understanding about lung diseases.

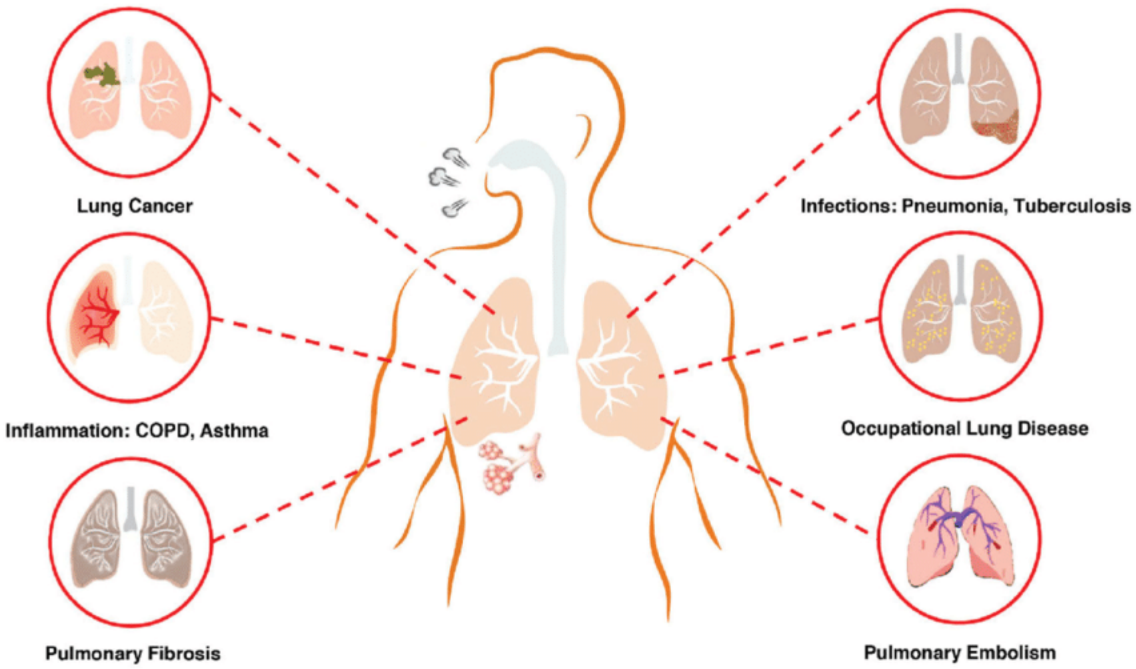


Figure 1.1: Types of respiratory diseases[43].

1.1 Problem Statement

According to the World Health Organization(WHO), COPD and Asthma are two majorly known respiratory diseases that are chronic [70]. Both of these diseases affect the air pipes through which the air is flown. COPD has the symptom of breathlessness additionally chronic cough which can get worse by the time. Advanced stages of COPD, it gets difficult for the lungs to pump oxygen. As a result of which the heart has to pump more blood through the lungs [28].

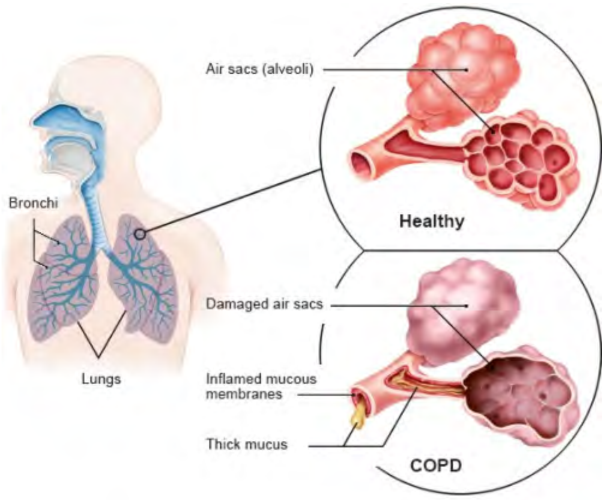


Figure 1.2: COPD affected lungs[28].

In addition to that, Asthma is a disease that causes breathlessness or shortness of breath and a whistling sound while breathing called wheezing as a result of the restricted airways[66].

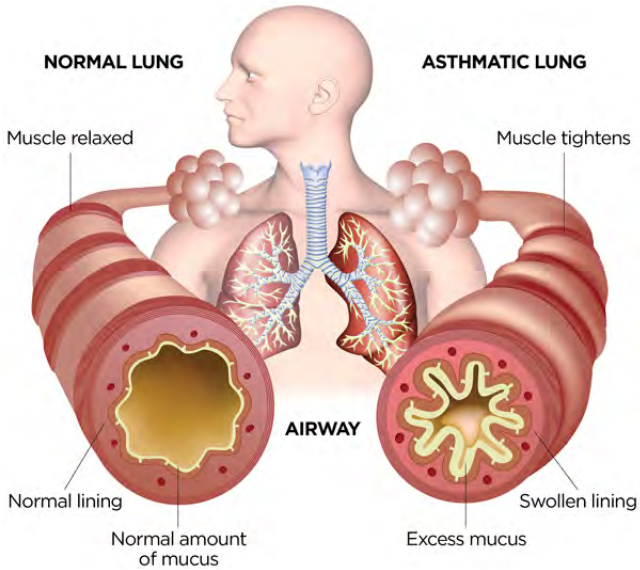


Figure 1.3: Normal vs Asthma affected lungs[64] .

Lung cancer also has symptoms like shortness of breath, wheezing as well and coughing that get worse by the time [62]. As stated by the National Health Service(NHS), Bronchiectasis and Bronchiolitis both have symptoms of coughing but Bronchiectasis comes with sputum and the latter has an additional symptom of cold and sneezing mostly occurring in infants or or children below 2 years[67] [68]. The World Health Organization(WHO) stated in 2017 that 650 thousand deaths occur each year due to various respiratory diseases. [74].

It can be interpreted from the symptoms that almost all respiratory diseases have similar symptoms, which can often get very tough to classify. When a doctor deals with multiple patients per day, or if they are tired, stressed or sleep deprived, they might miss some crucial symptoms, or the patient might not be paying attention to their symptoms entirely and fail to explain to the doctor, which can result in deterioration of the health of the patient or even cause fatality. To solve that problem, we hypothesize creating a multiclass classification of these respiratory diseases and sound using respiratory sound databases along with spirometry.

Sometimes, in extreme health conditions, it gets difficult for patients to go to a specific place for a particular health examination. The method this paper proposes is to classify lung sound and disease that does not require the patient to go to particular health examination centers for the lung examination.As the test readings are taken through the electronic stethoscope[46], which is a portable medical tool, the readings can be taken easily from where the patient is. Lung sound readings are taken through electronic stethoscopes which is also a noninvasive way of health checkup that does not require breaking the skin or entering any medical tools inside of the body.

This paper also examines the ways spirometry can provide in-depth understanding of different lung illnesses. The incorporation of lung sound analysis in addition to spirometry has possible outcomes of early detection and intervention of respiratory diseases which will open opportunities for the patient to seek proper healthcare at an early stage. The main challenge is to create a reliable and accurate classification system that can divide respiratory data into many categories and provide a complex assessment of lung health. Furthermore, the model has the potential to improve diagnostic accuracy, facilitate individualized treatment planning, and support disease surveillance. Healthcare providers can enhance patient care and make knowledgeable decisions about respiratory diseases by leveraging the combined strength of both of these techniques.

1.2 Research Objective

The main objective of our research is to do a classification of lung sounds in order to classify respiratory diseases as well as respiratory sounds with the help of a respiratory sound database and spirometry. The results that we acquire from the data training process would be evaluated in order to classify respiratory diseases.

1. Developing a classification model that can classify between different respiratory diseases and respiratory sounds using various dataset contained several types of audio recordings. The goal is to create and identify the most effective model for this classification task.

2.Investigate several feature extraction methods for respiratory sound data to find the most important aspects that support precise classification.

3. Not only lung sound, but also building a classification model to classify respiratory diseases with spirometry measurements.

4. To build a model that will be able to analyze lung sound and its features including wheezing and crackling or both.

5. Comparing the performance of different machine learning algorithm and find out which algorithm will be able to rapidly classify lung diseases based upon lung sounds as well as spirometry.

6. Help the medical professionals to classify respiratory diseases more accurately.

7. Ensure the patients regarding error-free classification.

8. Investigate how the proposed classification model will affect improving respiratory disease diagnosis and treatment planning.

9. By highlighting the interpretability of our model's predictions, our goal is to build a connection between the healthcare industry's requirements for comprehensible insights and the complexity of machine learning.

1.3 Research Contribution

Respiratory Disease Classification:

1. Utilized machine learning techniques, including Decision Trees, Convolutional Neural Networks (CNNs), Support Vector Machines (SVM) and LSTM/GRU models, to classify respiratory diseases based on lung sound analysis.

2. Achieved varying levels of accuracy, with CNNs demonstrating the highest accuracy of 95%.

3. Evaluated model performance using metrics such as accuracy,precision, recall, and F1-score, highlighting CNN-Linked Features as the most accurate and balanced model.

Lung Sound Classification:

1. Employed models like SVM with MFCC, CNN, Attention-based Models (Deit base+Att+CNN), and Stacked Denoising Autoencoders (SDA) with CNN to classify lung sounds into four classes.

2. SVM with MFCC and CNN exhibited competitive accuracy rates of approximately 74% and balanced precision, recall, and F1 scores of 74%.

3. Deit base+Att+CNN and SDA models showed lower accuracy and less balanced precision and recall.

Spirometry and Lung Function Assessment:

1. Calculated FEV1, FVC, FEV1/FVC, FEF25-75 from raw spirometry data
2. Made multiple calculations using GLI-12 equations using demographic data
3. Discussed the importance of utilizing LLN and Z-scores

These findings underscore the potential of machine learning and medical data integration in enhancing respiratory disease detection and lung health assessment, ultimately leading to more accurate diagnoses and personalized treatment strategies in the field of respiratory medicine.

1.4 Thesis Organization

Chapter 1: In Chapter 1, we laid the foundation for our research on respiratory diseases. We introduced the significance of healthy lungs, discussed common respiratory conditions like COPD and asthma, and emphasized the role of sound as a diagnostic tool, particularly in identifying wheezes and crackles. Additionally, we explored the importance of spirometry as a diagnostic test. We highlighted the challenge of distinguishing between respiratory diseases with similar symptoms and set clear research objectives, aiming to develop accurate classification models for respiratory diseases and sounds, explore feature extraction methods, and provide valuable insights for medical professionals. This chapter provided essential context for our in-depth study into respiratory disease classification and its potential impact on healthcare.

Chapter 2: In Chapter 2, we delve into the technical aspects of our research, exploring various algorithms for the classification of lung diseases and lung sounds. These include Support Vector Machine (SVM) for supervised learning, Decision Trees for structured classification, Convolutional Neural Networks (CNNs) for image-based analysis, Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) for sequential data modeling, Vision Transformers (ViT) for visual and audio data, and a Stacked Autoencoder with CNN hybrid model , for feature extraction and classification. These algorithms collectively form the core of our research approach, facilitating accurate classification and valuable insights.

Chapter 3: In Chapter 3, we conduct a comprehensive review of related research on the classification of respiratory diseases using both respiratory sound databases and spirometry data. We examine various studies that employ machine learning models, to classify respiratory conditions such as asthma, pneumonia, bronchiectasis, bronchiolitis, pulmonary fibrosis, upper respiratory tract infections (URTI), lower respiratory tract infections (LRTI), and chronic obstructive pulmonary disease (COPD) as well as sound like wheezes,ceackle,both,normal. The literature review highlights the potential of combining lung sound analysis and spirometry results for more accurate disease diagnosis and monitoring, emphasizing the importance of machine and deep learning techniques in enhancing patient care and early detection.

Chapter 4: In Chapter 4, we outlined the methodology employed in our research. We depicted a workflow for classifying lung sounds and diseases, utilizing machine learning models. The dataset used was a respiratory sound database containing recordings from individuals with respiratory conditions, including asthma, pneumonia, bronchitis, and COPD. Additionally, we discussed a raw spirometry dataset and the importance of accurate spirometry calculations. Data preprocessing steps were performed, including addressing missing data and class imbalance. Feature extraction techniques, such as STFT, chroma, spectral contrast, Mel spectrogram, Mel spectrogram with VTLP, and Tonnetz, were applied to prepare the audio data for analysis. These methods enable us to extract relevant information from the respiratory sound recordings for disease classification and evaluation.

Chapter 5: In Chapter 5, we conducted an experimental evaluation of various machine learning and deep learning models for the classification of respiratory diseases and lung sound patterns. We employed models such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Decision Trees, LSTM/GRU, Attention-based Models (Deit base+Att+CNN), and Stacked Denoising Autoencoders (SDA). The models were trained and tested using different feature extraction techniques, including MFCCs, Mel spectrograms, and linked features. We analyzed the performance of these models based on accuracy, precision, recall, and F1-score, providing insights into their strengths and weaknesses. Additionally, we discussed the application of Z-scores and the lower limit of normal (LLN) in spirometry data analysis, highlighting their importance in accurately diagnosing lung conditions. Overall, this chapter presents a comprehensive evaluation of our classification models and their potential implications in the field of respiratory health.

Chapter 6: In our study, we delved into the realm of respiratory health assessment, leveraging lung sounds and spirometry data to classify various respiratory diseases. Employing cutting-edge machine learning techniques such as CNNs and SVMs, we explored the potential of merging technology and medicine for more accurate disease detection and personalized treatment in respiratory medicine. Our rigorous experimental evaluation of diverse classification models highlighted the significance of metrics like accuracy, precision, recall, and F1-score. Moving forward, enhancing model reliability and robustness through larger and more diverse datasets, conducting real-world validation studies, and exploring integration with electronic health record systems are pivotal steps to realize the full potential of this research in transforming respiratory healthcare.

Chapter 2

Background

2.1 Model Description:

For the purpose of our research, various algorithms have been applied. In this part, the following are explained: For Lung Diseases classification we used:

1.Support Vector Machine
2.Decision Tree
3.CNN
4.LSTM/GRU/CNN

For Lung Sound classification we used:

1.Support Vector Machine
2.CNN
3.Attention and Vision Transformer
4.Stacked Denoising Autoencoder with CNN

2.1.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression problems. However, we see it used in classification more.In this algorithm, each of the data points are plotted as a point in n-dimensional space. (n is the number of features you have). Each of the value is depicted by a particular co-ordinate.

2.1.2 Decision Tree

Decision tree is widely used for classification.There are two factors we need to think about while applying decision trees which are Nodes and Rules (tests).We have to construct a tree where each node reflect a test on an attribute.The fundamental idea

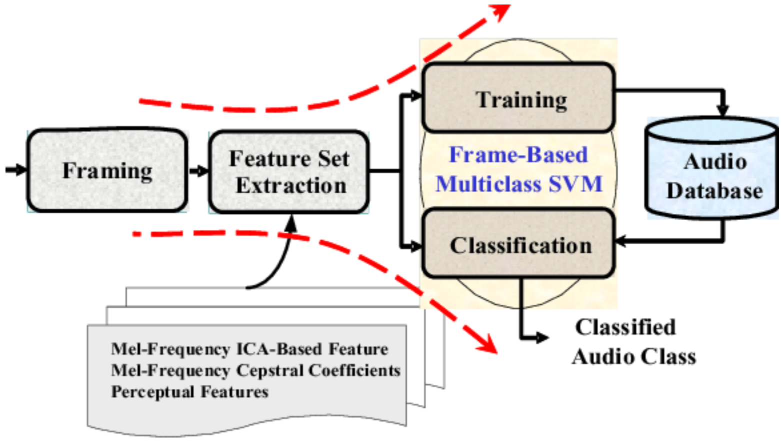


Figure 2.1: An arbitrary audio classification system.[6]

of this algorithm is to draw a flowchart diagram which includes a root node on top. All other (non-leaf) nodes depict a test until you reach a leaf node (final result). [14]

It's a traditional instance-based learning algorithm that emphasizes classification rules represented as decision trees derived from a collection of disorderly and irregular instances. This method follows a top-down iterative approach, where it examines attributes at internal nodes of the decision tree, evaluates the downward branches using various attributes of the node, and derives conclusions from the leaf nodes within the decision tree. Each path from the root to a leaf node corresponds to a conjunctive condition, and the entire tree represents a set of disjunctive conditional expressions. You can think of the decision tree as a Boolean function, where the input is the object or all the situation's properties, and the output is a "yes" or "no" decision.

In the decision tree, each of the tree nodes correspond to an attribute test, each leaf node corresponds to a Boolean value(such as "0" or "1"), ("Yes" or "No") and each of the branches correspond to a specific outcome or decision path. [14]

2.1.3 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a subclass of deep learning models that have shown outstanding performance in handling image processing, computer vision, and other grid-like data analysis tasks. The purpose of CNNs is to automatically and adaptively learn spatial feature hierarchies from the input data.They consist of neurons with biases and weights that are learnable. Each neuron processes a few inputs, conducts a dot product, and may optionally do a non-linearity as a follow-up. [71]

The primary elements of a typical CNN are as follows:

Convolutional Layers(CONV): The foundation of CNNs are convolutional (CONV) layers. Convolutional layers build feature maps that depict the existence of certain features in the input by applying a number of filters to the input data. Moving the filters, often referred to as kernels, over the input image (or the output from a previous layer), computing the dot product between the weights of the filter and the input, and then creating an output matrix known as a feature map or convolved feature are the steps that enable this process.

Activation Functions: An activation function is performed following each convolution operation to add non-linearity to the model, enabling the network to learn more complicated information. The activation procedures: (ReLU) $f(x) = tanh(x)$ (ii) the saturating hyperbolic tangent $f(x) = tanh(x), f(x) = |tanh(x)|$, and (iii) the sigmoid $function f(x) = (1 + ex)1$ [48]

Pooling Layers: These layers help to lessen the computational complexity of the model, increase model invariance to small translations, and manage overfitting by reducing the spatial dimensions (i.e., width and height) of the input. Max pooling and average pooling are the two most used types of pooling.

Fully Connected (FC) Layers: These layers are referred regarded as "fully connected" (FC) layers because they link every neuron in one layer to every neuron in the following layer. They are often applied after a number of convolutional and pooling layers near the network's finish. The fully connected layers' primary function is to carry out high-level reasoning and reach the final classification conclusion using the features discovered by the convolutional layers.

SoftMax Function: The final layer of a CNN frequently employs the SoftMax function, which generates a vector that depicts the probability distributions of a number of possible outcomes. It is mostly applied to multi-class classification issues. [48]

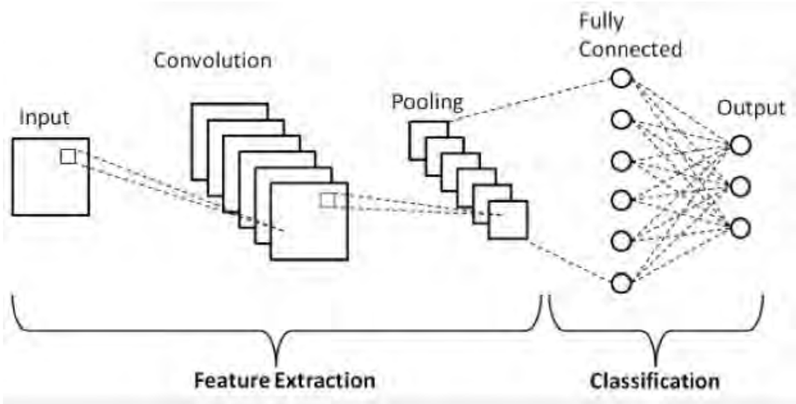


Figure 2.2: Basic CNN architecture.[33]

2.1.4 LSTM

Long Short-Term Memory, or LSTM, is a developed type of recurrent neural network (RNN), which specializes in the successful modeling of sequential data. Through memory cells and gating mechanisms, it overcomes the difficulty of collecting both short- and long-term relationships in sequences, minimizing the vanishing gradient problem in deep learning. In recognition of its versatility and competence in handling complicated sequential patterns, LSTMs are widely used across many different disciplines, including time series analysis and natural language processing.[36].

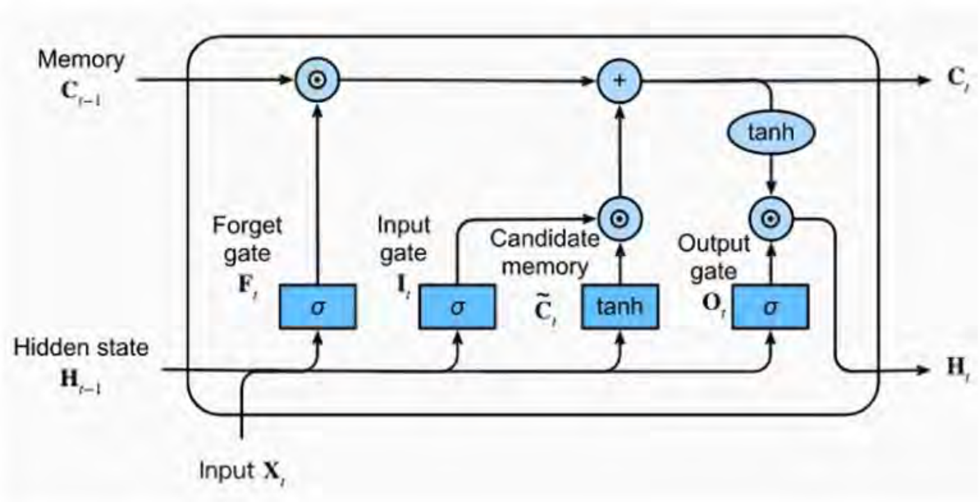


Figure 2.3: Architecture of LSTM cell.[51]

GRU:GRU, first introduced by Cho et al. in 2014, has some similarities with LSTM however stands out due to its computational performance and ease of installation. A common GRU design consists of two crucial gates: the reset gate (typically denoted by "r") and the update gate (often denoted by "z"). The reset gate's function is similar to the function of the LSTM's forget gate as it influences how much data from previous time steps is retained. Furthermore, the update gate determines the extent to which updated data impacts the present status. This concise explanation emphasizes GRU's key characteristics and utility, particularly in situations that require the fast processing of sequential input.. [15].

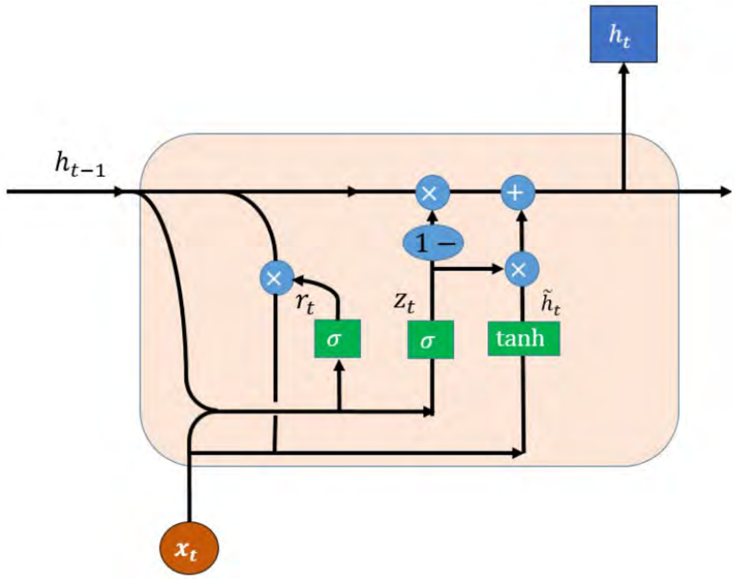


Figure 2.4: Architecture of a GRU cell.[29]

2.1.5 Attention and Vision Transformer

As per Li et al., a vision transformer(ViT) is a type of deep learning model that extends the success of transformer models in natural language processing. These transformer models are designed to work with visual data like videos and images. however, this model can also work with audio or sound data with certain preprocessing being done[59]. This approach is executed by breaking down the image into smaller patches and organizing them as a sequence of data. Multi-head self-attention focuses on finding a relation between the patches and it looks into different regions of the patches at the same time to find a relation. [60]

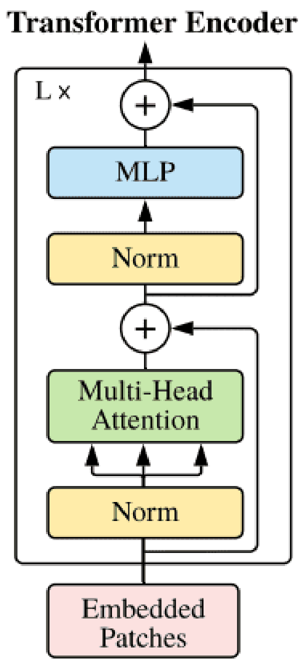


Figure 2.5: Workflow of a vision transformer (ViT)[45]

2.1.6 Stacked Autoencoder with CNN

Autoencoder is an artificial neural network consisting of three layers (Wang et al., 2016). It is widely used to reduce the dimensionality of data. It performs encoding or decoding on the input data according to its type and tries to generate results as accurately as possible. Not only can it reduce the dimensionality and noise of the data, but also it can detect repetitive data, as inferred by Wang et al. (2016)[16]. Stacked Autoencoder with CNN is a hybrid model that is suitable for extracting relevant features from audio data and and classify accordingly. The stacked autoencoder plays an essential role in capturing features from data and CNN is efficient in classifiying into various sectors[72][40].

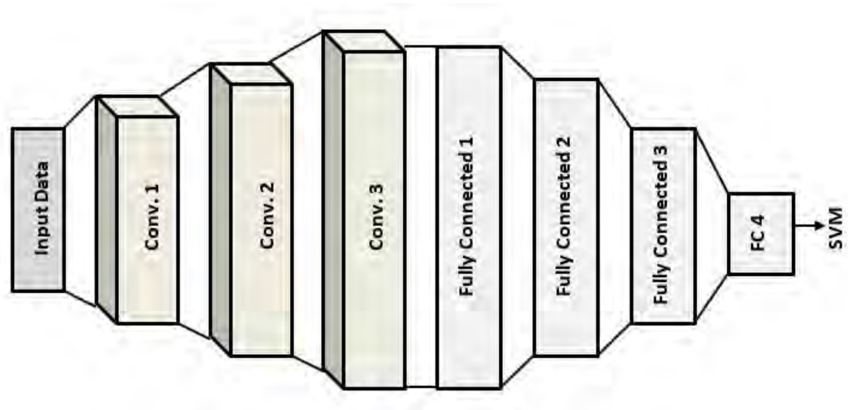


Figure 2.6: Architecture of a stacked autoencoder.[35]

Chapter 3

Related Work

Each year, a large number of people get affected by lung disease, leading to a lifetime of morbidity or sometimes even death. Lung diseases like COPD are one of the three main reasons for mortality worldwide. Therefore, accurate and timely classification of lung diseases is crucial. Doctors treat the disease immediately after classifying it, so accurate classification helps plan and execute the treatment. Spirometry has become a recognized approach for classifying different lung diseases more accurately.

The primary purpose of this literature review is to examine the classification of respiratory diseases using the respiratory sound database and spirometry. We explore relevant literature on our topic and assess the datasets using machine learning models such as Convolutional Neural Network (CNN), Support Vector Machine (SVM), and Decision Tree. We evaluate the accuracy of the results concerning spirometry. Furthermore, this paper aims to classify respiratory diseases such as asthma, pneumonia, bronchiectasis, bronchiolitis, pulmonary fibrosis, upper respiratory tract infection (URTI), lower respiratory tract infection (LRTI), and chronic obstructive pulmonary disease (COPD).

Singh et al. [44] conducted a study on COPD and asthma patients and a control group. They perform pulmonary function tests (PFT) using a store-bought spirometry device and record the lung and heart sounds from four regions using an electronic Littman stethoscope. The results of asthma patients demonstrate more significant reversibility than COPD patients. However, there is some overlap in the reversibility values, suggesting the requirement for supplementary diagnostic approaches. The study concludes that denoised heart and lung sound signals and advanced techniques assist in precise disease diagnosis. Future efforts involve collecting more data and creating a multimodal system to enhance diagnosis. Computerized techniques provide faster and more accurate outcomes, leading to enhanced patient care[44].

Rudraraju et al. [41] state that cough sounds contain valuable information about the respiratory system and related diseases. They establish a relationship between cough patterns and various respiratory conditions. Furthermore, they discovered a strong correlation between cough sound characteristics and airflow parameters measured by Spirometry. They develop a machine-learning model and validate it using K-fold cross-validation with reliable data to enhance the accuracy of pattern predictions.

Ma et al. [52], in the year 2022, stated that auscultating lung sounds are an affordable, straightforward and noninvasive approach to identifying respiratory ailments.

However, the proficiency of individual medical experts can lead to differences in diagnostic outcomes. They developed a deep learning model to categorize lung sounds to address this issue, ensuring medical experts have a more uniform point of reference for precise diagnoses. Utilizing a lung sound dataset from children above one month and under 18 years, they introduced a model incorporating refined data preprocessing techniques with a DenseNet169 Convolutional Neural Network (CNN) model.

In their research paper, the authors utilized pre-trained ResNet models as the foundational architectures for the classification of abnormal lung sounds and respiratory diseases [49]. They transferred knowledge from these pre-trained models using a variety of methods, including standard fine-tuning, stochastic normalization, and a combination of co-tuning and stochastic normalization. Furthermore, they tackled the issue of class imbalances in the ICBHI dataset and their multi-channel lung sound dataset by employing data augmentation techniques in both the time and time-frequency domains. Additionally, they introduced spectrum correction to address variations in recording device properties within the ICBHI dataset. The systems they proposed consistently outperformed all existing state-of-the-art approaches for the classification of abnormal lung sounds and respiratory diseases in both datasets, as demonstrated in their study by Nguyen and Pernkopf in 2021.

The primary method for screening and diagnosing lung diseases involves listening to respiratory sounds through auscultation. Integrating automated analysis with digital stethoscopes holds significant potential for enabling remote screening of life-threatening lung conditions, as phrased by Gairola et al. [38]. Deep neural networks (DNNs) have emerged as an optimistic choice for handling such challenges. However, DNNs require a significant amount of data, and even the most enormous available respiratory dataset, ICBHI, contains only 6898 instances of breathing cycles, which remains insufficient for training a robust DNN model. The authors propose a straightforward CNN-based model accompanied by a set of innovative techniques. These techniques include fine-tuning specific to the recording device, augmentation through concatenation, removal of blank regions, and padding. These innovations allow us to leverage the limited size of the dataset effectively. The authors conducted thorough evaluations using the ICBHI dataset, achieving a 2.2% improvement over the current state-of-the-art results for 4-class classification [38].

In their study, Sharan et al. [24] investigate cough sounds and their relationship with pulmonary function tests using spirometry. The subjects undergo spirometry tests, and their cough sounds are recorded. The researchers employ linear and support vector regression (SVR) models to estimate spirometry readings based on cough features and demographic information. The results indicate that including demographic features enhances the accuracy of the predictions. The study also explores the correlation between cough sounds and spirometry readings, revealing a high to moderate positive correlation. The researchers conclude that cough sounds contain sufficient information to estimate spirometry parameters. They further propose a method that utilizes a smartphone as a recording device and computing platform, making it applicable in ambulatory clinical settings outside a pulmonary function laboratory. However, the study does not directly compare the diagnostic outcomes based on cough-based estimations with those derived from laboratory spirometry measurements.

Using spirometry data, Kammoun et al.[21] compare two methods for diagnosing and grading obstructive ventilatory defects (OVD). One method follows the American Thoracic and European Respiratory Societies (ATS/ERS) guidelines, defining OVD based on the lower limit of normal (LLN). In contrast, the other method uses Z-scores recommended by the Global Lung Initiative (GLI)(Kammoun et al., 2018). They assessed 1000 participants, analyzing FEV1/FVC ratios, FEV1% predicted (ATS/ERS), and FEV1 Z-scores (GLI) to diagnose and classify OVD severity. The study found that the two methods yielded different OVD frequencies and severity classifications, indicating they are not interchangeable. The GLI method's universal approach may require ethnic-specific adjustments to ensure accurate OVD diagnosis and severity assessment.

In a recent paper, Das et al. [19] discuss the application of artificial intelligence (AI) in diagnosing obstructive lung diseases. They assert that AI can automate Pulmonary Function Tests and potentially replace human physicians by replicating their cognitive abilities in interpreting the data. AI demonstrates the ability to swiftly identify patterns within data. Machine learning and artificial intelligence can be employed to analyze data from lung tests, such as spirometry and lung sounds, to obtain a diagnosis. Despite being in the early stages of development, these methods exhibit impressive results.

Oud and Maarsingh [4] develop a model to recognize airway obstruction by analyzing respiratory sounds and spirometry results of asthma patients through a computerized method. They employ a supervised neural network as a function approximation technique to establish a relationship between the spectral parameters of lung sounds and obstruction parameters. They conclude that to enhance the accuracy of assessing airway obstruction, it is recommended to explore various parameters or incorporate multiple impedance parameters.

Leuppi et al. [5] aims to evaluate the accuracy of physicians' estimation of airway obstruction through lung auscultation and compare it with spirometry measurements using the FEV1/FVC ratio. The study includes a total of 233 patients in the emergency room. After history-taking, physicians perform auscultation followed by spirometry. The results indicate that physicians' auscultation-based estimation demonstrates a weak but significant correlation with the degree of airway obstruction measured by FEV1/FVC. Normal lung auscultation is identified as an independent predictor for the absence of airway obstruction. The study concludes that although physicians can reasonably rule out airway obstruction through auscultation in emergency room settings, spirometry should still be conducted to ensure an accurate diagnosis.

Mineshita et al. [13] conducted a study involving 27 male patients with consistent Chronic Obstructive Pulmonary Disease and a smoking history. The study includes spirometry-based pulmonary function tests. They obtain lung sound recordings using the VRIxp System and calculate quantitative lung data (QLD). This study demonstrates that COPD patients exhibit altered lung sound distribution compared to healthy smokers, with correlations observed between the lower QLD/upper QLD ratio, spirometric measurements, and emphysematous lesions. The impact of emphysematous lesions on lung sound distribution varies among individuals.

Vaz Fragoso and colleagues [9] introduce an innovative approach to assess the severity of Chronic Obstructive Pulmonary Disease (COPD) in elderly individuals using spirometry data. Known as the LMS method, this approach addresses the limitations of current spirometric criteria recommended by organizations like GOLD and ATS/ERS. Unlike existing methods, the LMS approach takes into account age-related variations in pulmonary function, including variations and asymmetry in reference data. It establishes COPD severity thresholds based on percentile distributions of Z-scores for Forced Expiratory Volume in 1 second (FEV1) derived from LMS analysis. The study highlights that these Z-score thresholds, based on the median as a more suitable measure of central tendency, are associated with clinically significant health outcomes, such as mortality and respiratory symptoms. This offers a more evidence-based approach to staging COPD in older populations. While this innovative method has the potential to reduce misclassifications of COPD severity, additional research is needed to validate its applicability across diverse populations and using more up-to-date data [9].

A research paper by Bae et al. [56] stated that lung sound holds significant information in the early detection of lung diseases of high risk. The authors in this research evaluated the potential use of pre-trained models mainly created for large-scale data of pictorial or audio datasets to classify lung sounds. However, the drawback is the unavailability of large datasets. Additionally, they introduced a simple and straightforward augmentation technique called Patch-Mix, which involves randomly mixing patches from different samples, in combination with the Audio Spectrogram Transformer (AST). Furthermore, their proposal is to innovate the Patch-Mix Contrastive Learning approach to distinguish mixed representations in the latent space. The model by Bae et al. surpassed the previous results by 4.08%, outperforming by gaining state-of-the-art performance on the ICBHI dataset [56].

After reviewing the paper by Viswanath et al. [25], we observe that respiratory sound diseases can be categorized using a combination of gated CRNN (Convolutional Recurrent Neural Network) models and CNN (Convolutional Neural Network) models. This fusion enhances the accuracy of disease categorization by allowing the model to capture spatial and temporal information in respiratory sound data. Gated CRNN can also be applied to spirometry, aiding in the diagnosis and follow-up of respiratory illnesses such as asthma, chronic obstructive pulmonary disease (COPD), and restrictive lung diseases. CRNN models leverage the strengths of CNNs, RNNs, and gating processes to provide precise classification, monitoring, and early detection of respiratory diseases, thus improving patient care and outcomes.

Furthermore, in the same paper, the researchers utilize two approaches. The first approach involves traditional machine learning models, while the second approach employs convolutional neural networks (CNNs) trained on Mel-spectrogram features, as mentioned earlier. Two neural network models are utilized: a nine-layer VGG-style CNN and a three-layer Gated-Convolutional Recurrent Neural Network (Gated-CRNN). The results demonstrate that both strategies successfully detect invalid smartphone spirometry attempts with high precision and recall, with the Gated-CRNN model performing the best. This paper highlights the superior performance of CNN models and suggests that future research can explore combining other approaches with CNNs, along with spirometry, to achieve improved outcomes in classification.

We selected these papers to discover existing information demonstrating the potential benefits of using both lung sounds and spirometry results for evaluating and diagnosing lung diseases. These papers employ diverse methods for data processing and evaluation, yet all reveal a correlation between spirometry and lung sound data. The findings in these papers indicate that utilizing machine learning and evaluating both types of data can result in early detection and improved diagnosis of respiratory diseases.

Chapter 4

Methodology

The figure 4.1 shows the workflow that was followed in order to get the desired results. Lung sound and lung diseases have been classified using one Respiratory sound database implementing different machine learning models. 5 models were trained for lung disease classification and 4 models were trained for lung disease classification with data pre-processing and feature extraction.

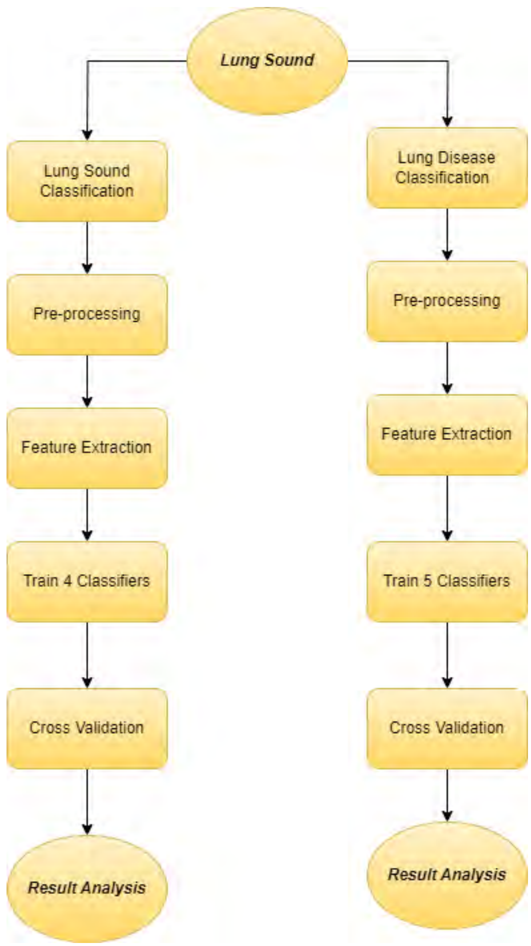


Figure 4.1: The flowchart of the proposed classification model.

DATASET	TOTAL
Crackle Cycles	1864
Wheezes Cycles	886
Combination of Crackles & Wheezes	506
Normal Cycles	3642
Total	6898

Table 4.1: Lung sound dataset description.

4.1 Dataset Description

For our research, we used a respiratory sound database (Rocha et al., 2018) from public repository kaggle created by two research teams from Portugal and Greece [34]. This Dataset contains respiratory sound recordings obtained from patients who have history with respiratory conditions, like Asthma, Pneumonia, Bronchitis, and Chronic Obstructive Pulmonary Disease (COPD), etc. The dataset has audio files accompanied by annotations.The annotation offers insights regarding the respiratory pathology we get in each recording. There are 920 .wav audio files and 920 annotations .txt files which helped us with evaluating respiratory disease. These 920 recordings have a length of 10 seconds to 90 seconds.The 920 annotated audio samples from 126 patients, totaling 5.5 hours of recordings with 6898 breathing cycles, of which 1864 have crackles, 886 have wheezes, and 506 have both.

The dataset contains data of healthy lungs as well as affected ones which helps us to differentiate between the healthy and unhealthy as well as the diseases. People from all age groups, from children to adults and elderly, all have participated in giving respiratory sounds samples to create this database.

The sound that is emitted when a person breathes has a direct connection with their lung tissue changes and the secretion's positions along with the movement of air inside the lung.This dataset contains two types of sounds, wheezing and crackling. A wheezing sound indicates that the person has airway restrictions, either their airways can be shrinked or blocked by mucus [75]. A crackling sound is a harsh and squeaking sound that comes out when the patient is breathing. This can be due to blocked airways as well. While the wheezing can be a result of Bronchitis, Pneumonia or lung cancer, crackling can be due to bronchitis, pneumonia, COPD as well[73]. The table 4.1 shows the lung sound dataset description.

Raw Spirometry Dataset: Inspiratory and Expiratory:

The dataset by Falvo et al. [57] provided us with 1,055,236 observations. A total of 129 subjects provided their expiratory and inspiratory flow volume loop data. This dataset contains six specific variables in six columns in the dataset, and each row of data corresponds uniquely to a specific subject with ID, visit, and trial, and includes a pair of measurements for flow and volume in a time series format, as outlined in Table 4.2. While Table 4.3 offers an overview of the entire dataset, when examining the data at the individual subject level, it's noted that, on average, each subject underwent 8.2 trials, with a standard deviation of 1.8 trials per subject. These trials encompass data from both the first and second test sessions.

Additionally, alongside this dataset, there is another dataset containing demographic information. This demographic dataset has one entry per subject, totaling 129 rows of data. The information provided in Table 4.3 summarizes the demographic measurements for each subject.[57].

Variable	Definition	Range
ID	Uniquely assigned numbers	101–237
Visit	Number associated with study session	1–2
Trial	Sequentially numbered for each subject and visit	1–16
Time	Time stamp of flow and volume measurement (ms)	0–10,750
Volume	Volume measurement corresponding to flow volume loop (L)	0.405–6.409
Flow	Flow measurement corresponding to flow volume loop (L/s)	8.88690–14.24400

Table 4.2: The 6 variables are shown.

Variable	Range
ID	101–237
Weight (kg)	43.00–143.00
Height (cm)	151.0–193.0
Sex	39 Male, 90 Female
Age (yr)	18–39 years
Ethnicity	19 Hispanic or Latino, 108 Not Hispanic or Latino, 2 Missing Data
Race (can be more than 1)	37 Asian, 20 Black or African American, 58 White, 7 >1 Race, 7 Missing Data

Table 4.3: Demographic information.

4.2 Spirometry Calculations

In our study, we evaluated crucial lung function metrics derived from spirometry data. Forced Vital Capacity (FVC) [3] represents the total volume of air that can be forcefully exhaled during a robust exhale, giving insight into overall lung capacity. We also measured Forced Expiratory Volume in One Second (FEV1), reflecting the amount of air exhaled during the first second of a deep and forceful breath out, helping assess the speed and efficiency of exhalation. The FEV1/FVC Ratio, presented as a percentage, signifies how much of the total lung capacity (FVC) is expelled in the initial second (FEV1), aiding in diagnosing various respiratory conditions. Furthermore, we computed Forced Expiratory Flow 25–75% (FEF25–75%), which characterizes the average airflow during the middle phase of a strong exhalation, offering valuable insights into lung function dynamics.

Accurate interpretation of pulmonary function tests, particularly spirometry, is vital in respiratory medicine, considering factors like gender, age, height, and race/ethnicity's impact on lung function [27]. Using reference values from similar racial backgrounds is recommended for precise assessments. In 2012, the Global Lung Function Initiative (GLI-2012) introduced spirometry equations covering all ages and various ethnicities, including North East and South East Asians, even offering equations for mixed ethnicities. Although these equations have generally performed well, their specific applicability for assessing spirometry in Asian Americans remains unexplored. Given the growing Asian American population in the U.S., evaluating these reference values becomes crucial for accurate lung function assessments in this demographic.

Our dataset contained vital demographic variables, including age, height, sex, and race/ethnicity. To evaluate the dataset, we employed the GLI-2012 guidelines [11], involving spline functions. These equations enabled the computation of crucial metrics for lung function assessment.

1. Predicted Value Equation (M):

$$\text{Predicted Value (M)} = M \tag{4.1}$$

2. Lower Limit of Normal (LLN, 5th percentile) Equation:

$$\text{LLN 5th percentile} = \exp\left(\frac{\ln(1 - 1.644 \cdot L \cdot S)}{L} + \ln(M)\right) \tag{4.2}$$

3. Z-score Equation (for L not equal to 0):

$$\text{Z-score} = \frac{\frac{M}{L} - 1}{L \cdot S} \tag{4.3}$$

4. % Predicted Equation:

$$\%\text{Predicted} = \left(\frac{\text{measured}}{M}\right) \times 100 \tag{4.4}$$

These equations utilize parameters L, M, and S, which depend on sex, age, height, and ethnic group. L characterizes skewness, S measures the coefficient of variation, and M represents the predicted value of lung function metrics like FEV1, FVC, or FEV1/FVC.

For our dataset spanning ages 3 to 95 years, we incorporated spline functions into the equations:

L (A parameter calculated based on age):

$$L = q0 + q1 \cdot \ln(\text{Age}) + \text{Lspline} \tag{4.5}$$

M (A parameter calculated based on age, height, race, and sex):

$$M = \exp(a0 + a1 \cdot \ln(\text{Height}) + a2 \cdot \ln(\text{Age}) + a3 \cdot \text{black} + a4 \cdot \text{NEA} + a5 \cdot \text{SEA} + \tag{4.6}$$

$$\text{Mspline}) \tag{4.7}$$

S (A parameter calculated based on age, race, and sex):

$$S = \exp(p0 + p1 \cdot \ln(\text{Age}) + p2 \cdot \text{black} + p3 \cdot \text{NEA} + p4 \cdot \text{SEA} + \text{Sspline}) \tag{4.8}$$

These spline-based equations allowed us to accurately assess and interpret lung function metrics within the context of age, height, and ethnicity, providing a robust framework for comprehensive pulmonary health evaluation.

4.3 Data Preprocessing:

Preprocessing the data is an important procedure to go through before starting to work with the algorithms. In order to preprocess a dataset, raw data is needed. Raw data is known to be some numerical set of values based on different scenarios which have not been trained or tuned before [54]. Preprocessing the data means converting the raw data into a dataset that is clean. Preprocessing method detects if there is any missing or redundant data, detects data noise as well as checks for other instability in the raw dataset. After getting the clean data, it can be executed into the preferred algorithms[22].
We applied various preprocessing techniques to prepare the audio data and extract relevant features (MFCCs) for analysis. As a preprocessing step, we removed instances related to rare diseases to address class imbalance or focus on more prevalent classes. This filtering process helps align the data with the analysis or modeling objectives, improving model performance and generalization by reducing the impact of underrepresented classes. Data preprocessing involves transforming raw data into a suitable format for analysis or modeling, handling data quality issues, addressing outliers or missing values, and ensuring proper data preparation. In addition to removing specific labels, dividing the data into train and test sets, one-hot encoding the labels, and reshaping the data for modeling, we performed other preprocessing operations. These actions prepare the data for machine learning model training and evaluation.
Some of the steps we took include checking for missing data, dropping rows with unimputable values, removing rows with multiple missing values, and imputing missing "BMI" values using the means of related rows.We also checked for data imbalance so we could get accurate representation of the diseases.

4.4 Feature Extractions

4.4.1 MFCC

MFCC stands for Mel-Frequency Cepstral Coefficients which is a popular feature extraction approach in speech and audio processing. MFCCs can be utilized to describe the spectrum properties of sound in a form that is useful in machine learning applications that include recognition of voices and audio recordings analysis[47].MFCCs, to put it simply, are a collection of coefficients that record the contours of a sound signal's power spectrum. They are created by first applying a method like the Discrete Fourier Transform (DFT) to turn the raw audio signal into a frequency domain, then utilizing the mel-scale to simulate how the human ear perceives sound frequency. The mel-scaled spectrum is used to determine the cepstral coefficients[47].

Since they highlight aspects of the audio signal that are crucial for human speech perception while ignoring less critical details, MFCCs are particularly helpful. They are therefore useful for tasks like speech-to-text conversion, speaker recognition, and emotion detection[47].

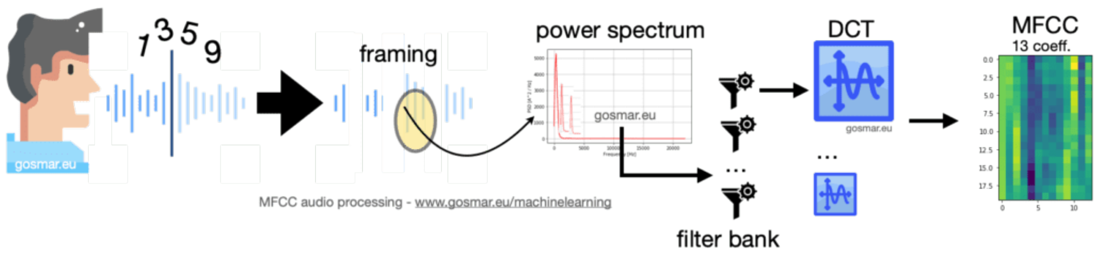


Figure 4.2: Audio recognition by MFCC. [39]

4.4.2 STFT

The short-time Fourier transform (STFT) is a series of Fourier transforms of a windowed signal. When a signal's frequency components change over time, STFT offers time-localized frequency information, whereas a typical Fourier transform offers frequency information that is averaged across the whole signal time period. STFT is used extensively in several sectors, including the processing of audio and images. Its significance comes from its capability to evaluate time-varying data, which enables scientists and engineers to better comprehend how these signals behave[42].

In order to conduct the Short-time Fourier Transform (STFT), the signal is divided into overlapping segments, with the segment size and overlap being established by the application and information properties. In order to enhance frequency resolution, each segment is given a windowing function similar to the pounding frame. Time-frequency spectra are produced by applying the Fourier Transform separately to each segment after segmentation and windowing. Segments that overlap one other prevent jarring transitions and lessen edge effects. To balance temporal and frequency resolution, researchers can modify the STFT by changing segment size and overlap. The STFT is a flexible tool for time-varying signal classification that is often used in industries including audio processing, picture analysis, and biomedical analysis of signals[63].

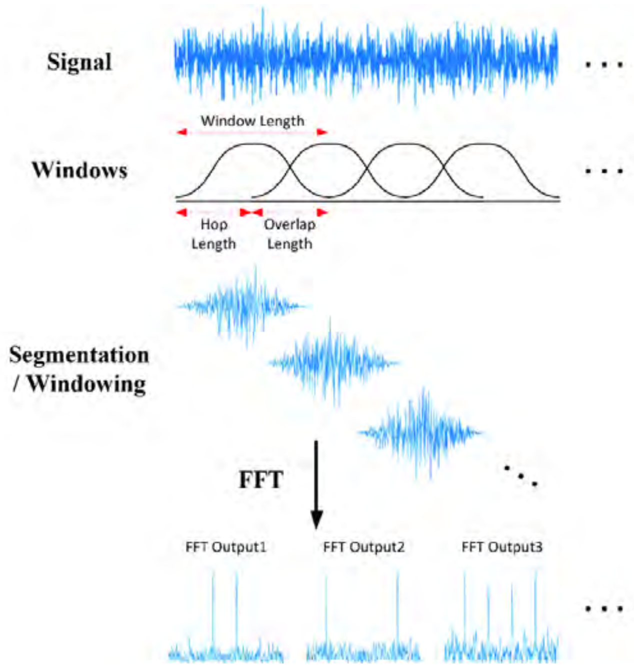


Figure 4.3: A short time fourier transform (STFT) workflow. [63]

4.4.3 Chroma:

We can understand the distribution of musical pitch classes within an audio stream thanks to the chroma feature. This trait emphasizes the importance of locating and measuring the predominance of particular musical notes throughout the audio recording. The chroma feature, which consists of 12 values and corresponds to 12 different pitch classes, was created by applying the Short-Time Fourier Transform (STFT), and it successfully captures the underlying harmonic content and tonal nuances inherent in the audio. The overall pitch profile of the audio is represented succinctly by averaging these values along axis 1. Chroma characteristics are a useful tool for many analyses of music, such as classifying musical genres, determining musical keys, and identifying chords in a composition [42].

4.4.4 Spectral Contrast:

By measuring amplitude changes between a signal's high points and low points across several frequency bands, the metric of spectral contrast provides insightful information on the dynamic characteristics of the spectral composition of an audio signal. The capacity to identify changes in the timbre and volume of the audio material is quite helpful. It evaluates the amplitude contrast within each frequency band of the audio spectrum, a procedure made easier by the Short-Time Fourier Transform (STFT), which separates the spectrum into different frequency subbands. By averaging these contrast values along axis 1, we have a thorough picture of the temporal evolution of spectral features [42].

4.4.5 Mel spectrogram:

The Mel spectrogram emphasizes frequency components important for human auditory perception as a transformative representation of audio input. It discretizes the spectrum into different frequency bins by using the Mel scale's principles, with a focus on the lower and midrange frequencies in particular since they correspond to the human auditory system's enhanced sensitivity. This transformation provides a useful depiction of how various frequencies contribute to the acoustic properties of the audio by condensing the time evolution of the spectral content of the audio. We acquire a comprehensive picture of the various frequency contributions inside the audio stream after generating the Mel spectrogram and averaging along axis 1. Mel spectrograms are essential tools for applications like voice recognition, speaker identification, and music genre in the fields of speech and music processing [42].

4.4.6 Mel spectrogram with VTLP:

A Mel spectrogram with Vocal Tract Length Perturbation (VTLP) implementation procedure comprises of several crucial steps. The audio signal is first loaded and prepared, making sure it has a constant sample rate and, if necessary, applying noise reduction or resampling[12]. The next step is to build a Mel filter bank that emphasizes perceptually important frequency components using a predetermined number of Mel filter banks. Following the application of this filter bank to the audio stream, the Mel spectrogram, which displays energy across various frequency bands, is produced. By making appropriate adjustments to the Mel spectrogram, VTLP can be used, if desired, to imitate changes in vocal tract length. These adjustments can be made to accurately mimic different vocal tract traits or other speech or sound differences. As a powerful tool for recording and analyzing spectrum information inside audio data, this solution is especially useful for tasks involving speech processing and audio analysis[12].

4.4.7 Tonnetz:

Tonnetz, also known as Tonal Centroid Features, is a useful collection of audio features designed to decipher the harmonic intricacies and tonal characteristics present in audio transmissions. These characteristics offer a way to extract the harmonic and musical information included in the audio data. Tonnetz characteristics, which are computed from the chromagram, are crucial for a number of music-related tasks, such as identifying musical keys and categorizing musical genres. In the field of music and audio analysis, they act as a crucial instrument that enables machines to fully comprehend musical compositions[42].

Chapter 5

Experimental Evaluation

5.1 Respiratory Disease Classification Outcomes:

This thesis intends to classify respiratory sound into multiple categories by applying multiclass classification which is a machine learning task. The performance on a particular model is determined by its accuracy. We are giving brief information about the proposed algorithms which we used .

5.1.1 Convolutional Neural Network:

In one of the CNN models we used MFCC (Mel-frequency cepstral coefficients) to train the model for classification task shown in Figure 5.1 . Using the Keras API of TensorFlow, the CNN model is defined. Convolutional layers are followed by max-pooling layers and dropout layers for regularization in this model. The model utilizes a GlobalAveragePooling2D layer to lessen the dimensionality of the data after the convolutional layers. The probabilities for each class are subsequently provided by the model's dense output layer, which has a softmax activation function.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 39, 861, 16)       80
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 19, 430, 16)       0
_________________________________________________________________
dropout (Dropout)            (None, 19, 430, 16)       0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 18, 429, 32)       2080
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 9, 214, 32)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 9, 214, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 8, 213, 64)        8256
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 4, 106, 64)        0
_________________________________________________________________
dropout_2 (Dropout)          (None, 4, 106, 64)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 3, 105, 128)       32896
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 1, 52, 128)        0
_________________________________________________________________
dropout_3 (Dropout)          (None, 1, 52, 128)        0
_________________________________________________________________
global_average_pooling2d (Gl (None, 128)               0
_________________________________________________________________
dense (Dense)                (None, 6)                 774
=================================================================
Total params: 44,086
Trainable params: 44,086
Non-trainable params: 0
_________________________________________________________________
184/184 [==============================] - 3s 16ms/sample - loss: 22.5919 - accuracy: 0.0163
Pre-training accuracy: 1.6304%
```

Figure 5.1: CNN with MFCC model summary

This model has been trained for 250 epochs.Test data sets have given an accuracy of 88% respectively on this model. The count of each illness class in the sound files used for training and testing the model is shown in the given CNN model's bar chart in Figure 5.2 . It offers a visual representation of how the various diseases are distributed throughout the dataset.The illness classes are shown on the bar chart's x-axis, and they are "Bronchiectasis," "Bronchiolitis," "COPD," "Healthy," "Pneumonia," and "URTI." The frequency or count of each disease class is shown on the y-axis.According to this graph, the class "COPD" appears to have the highest number of cases, followed by the classes "Pneumonia," "Bronchiectasis," "URTI," "Healthy," and "Bronchiolitis."

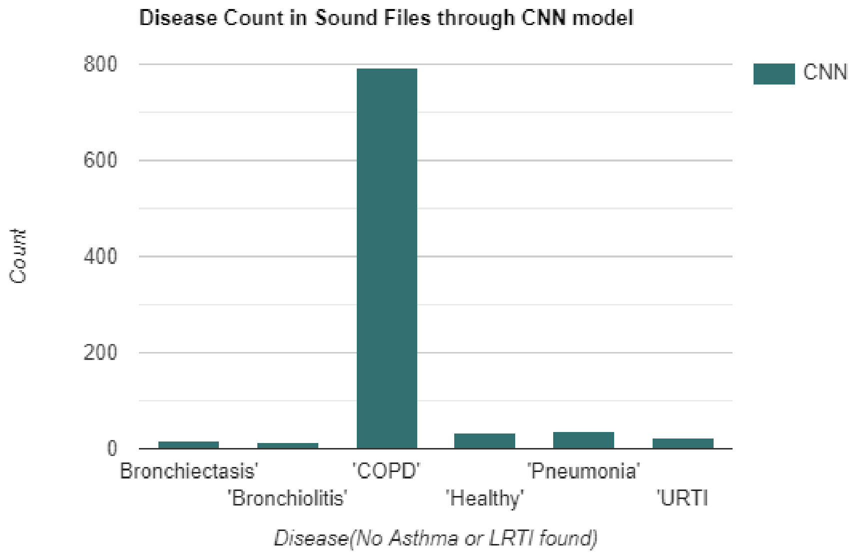


Figure 5.2: Bar chart representing the count of each illness by CNN using MFCC

We use another model of CNN with Linked features.In this model, linked features taken from audio signals are combined with CNN. It seeks to improve classification performance by introducing additional features that are learned by the CNN model. In terms of the linked features utilized in the CNN model, the code integrates different audio features extracted using librosa, including MFCCs, chroma, mel spectrogram, spectral contrast, and tonnetz. Concatenated versions of these features are used or fed as input for the CNN model. The model is able to use a variety of audio signal properties by integrating numerous characteristics to enhance classification performance.This model has been trained for 70 epochs and obtained an accuracy of 95% and a loss of 0.1928 for our dataset.CNN linked Features model Train and Validation dataset accuracy and loss comparing graphs in figure 5.3. table 5.4

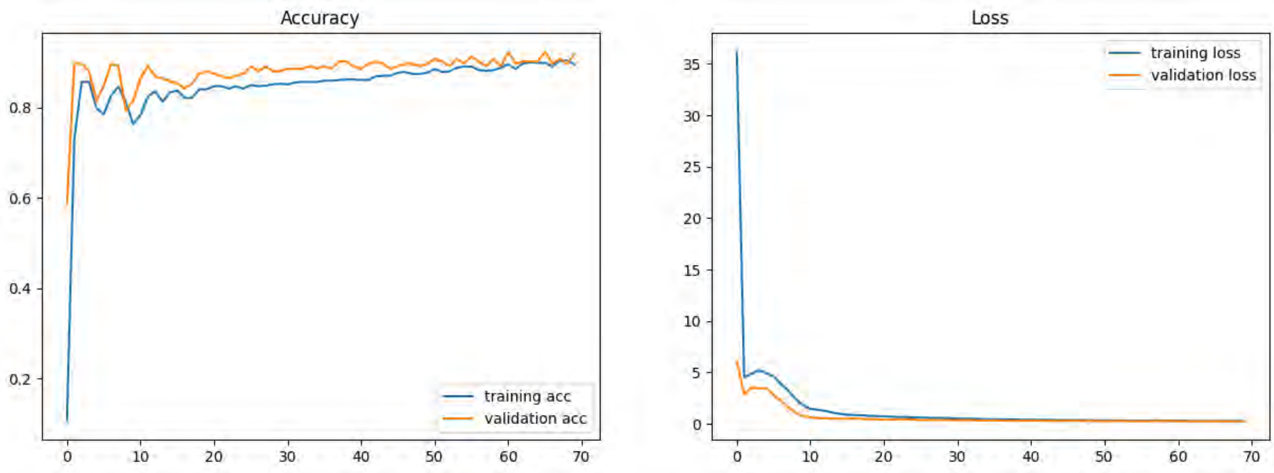


Figure 5.3: CNN linked features train vs validation accuracy and loss curve

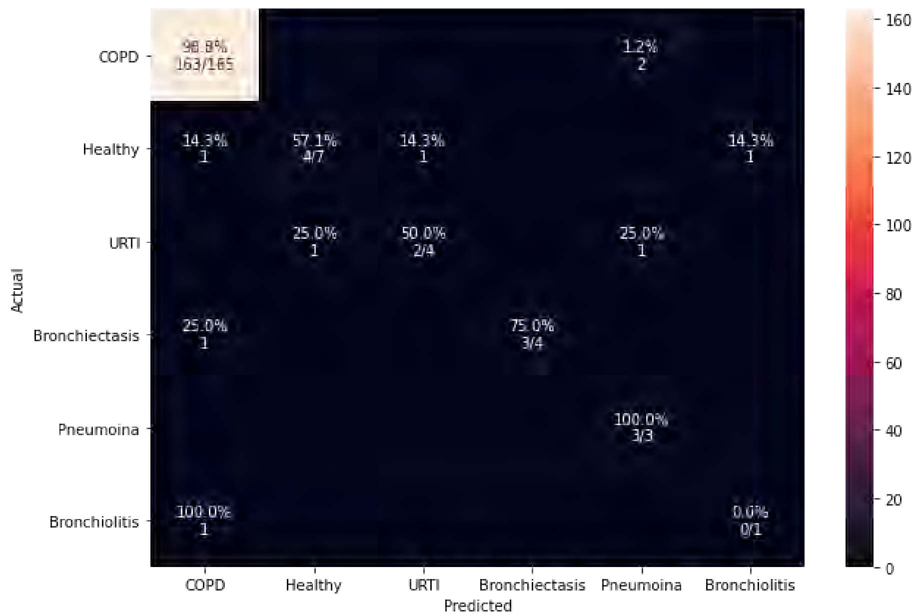


Figure 5.4: Basic CNN architecture.[33]

5.1.2 Support Vector Machine:

In feature extraction of SVM, A list of sound file paths and their associated labels are iterated through. Each audio file's features are extracted, and they are then saved in an array. To balance the dataset, the labels are altered, and the altered labels are saved in a separate array. The arrays of features, which hold the extracted features, and labels, which hold the changed labels, are actually the final output. The Figure 5.5 shows the classification of diseases found by SVM using MFCC.

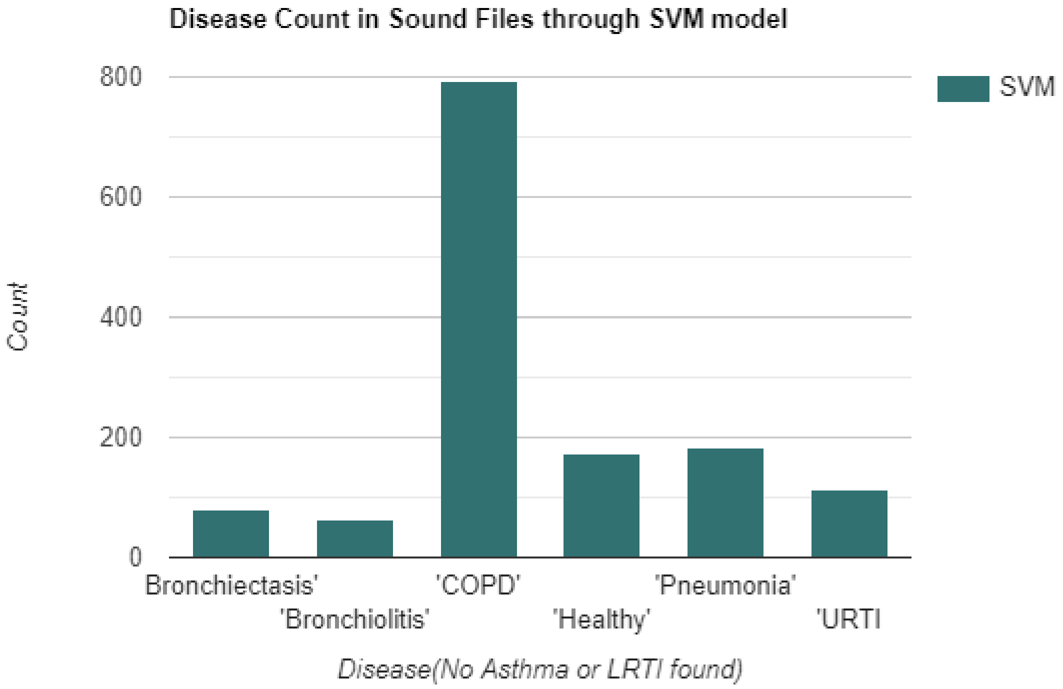


Figure 5.5: Bar chart representing the count of each illness by SVM using MFCC

5.1.3 Decision Tree:

The accuracy, confusion matrix, and classification report are a few of the parameters used to evaluate the decision tree classifier's efficiency. These parameters give insights about how well the model is functioning and how well it can classify the different classes in the dataset.

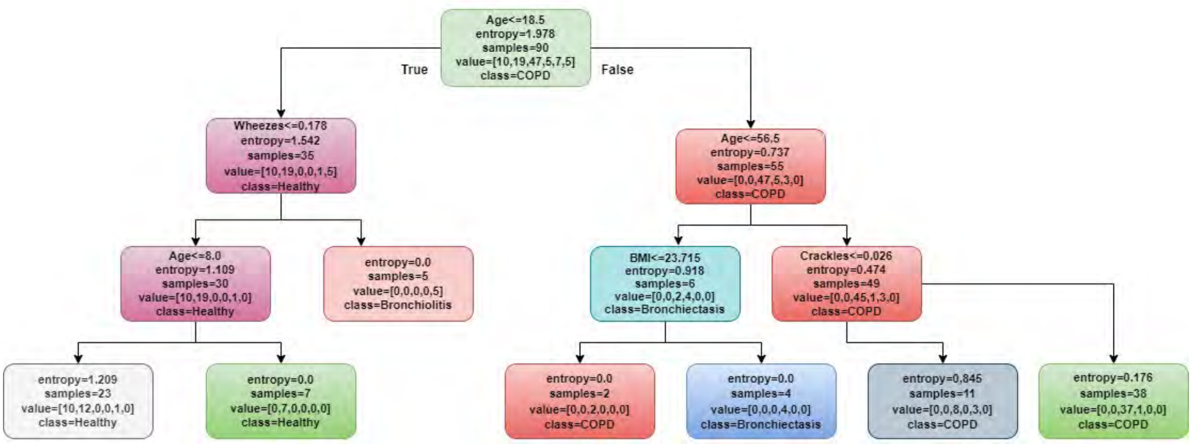


Figure 5.6: Visualization of the decision tree model

Understanding how the decision tree makes decisions and which features are most critical for classification we can use this visualization in Figure 5.6.

The graph's leaf nodes portray the final classification outcomes. Each of the leaf nodes has a distinct class label attached to it. We can find out the model's decision making by it. The splitting criteria displayed in the graph indicate the conditions that were used to divide the data at each decision node. We can discern the specific conditions employed by the model when making classification decisions by examining the splitting criteria depicted in the graph. This visualization of the graph enables us to evaluate the significance attributed to different features by the model when classifying outcomes.

5.1.4 LSTM/GRU/CNN:

LSTM and GRU play a vital role in detecting patterns, sequences or dependencies of the data. After extracting the appropriate features such as MFCC, spectrogram, and frequencies, the audio files were fed to LSTM as input. Created a sequence from the features extracted to avoid overlapping windows. The recurrent units work in noticing sequences or dependencies of the raw data which is vital to understand the audio sequences and patterns. For audio classification, it generates output as shown in Figure 5.7 .

The Figure 5.7 shows the architecture of GRU model and these sequence of layers were designed for audio data analysis.

Max Pooling: Max pooling is a type of decreased sampling process that recovers the majority of relevant characteristics from a sequence while shortening it. MaxPooling1D is commonly used after convolutional layers to minimize the dimensions of space.

Batch Normalization: Batch normalization is an approach for normalizing the responses of a layer in a neural network. The design aids in training stability and might result in quicker divergence. Normalization of batches is frequently used before the activation process takes place.
Conv1D (1D Convolutional Layer): Convolutional operations are applied to one-dimensional sequences, such as time-series data or audio spectrograms, using convolutional layers. In the input sequence, these layers pick up on spatial patterns.

GRU (Gated Recurrent Unit): They are repetitive layers that identify consecutive relationships in the data. These are often utilized to process the mastered spatial characteristics after the convolutional layers. Add Layers: These layers can be utilized to combine highlights from distinctive parts of the arrange, which can be valuable for certain errands or models.
Thick Layer: The thick layer is utilized for the ultimate transformation of highlights into the specified yield arrange.

Leaky ReLU Activation: Cracked ReLU actuation capacities can be connected after the thick layers to present non-linearity and offer assistance to relieve vanishing slope issues.

This model across 50 epochs was summarized in the training. The model's accuracy, loss, and other parameters were tracked throughout training. The training started with a reduced accuracy and a somewhat significant loss.The loss steadily dropped over the course of 50 epochs, demonstrating that the model's performance increased.The simultaneous rise in accuracy shows that the model became more accurate at identifying samples.

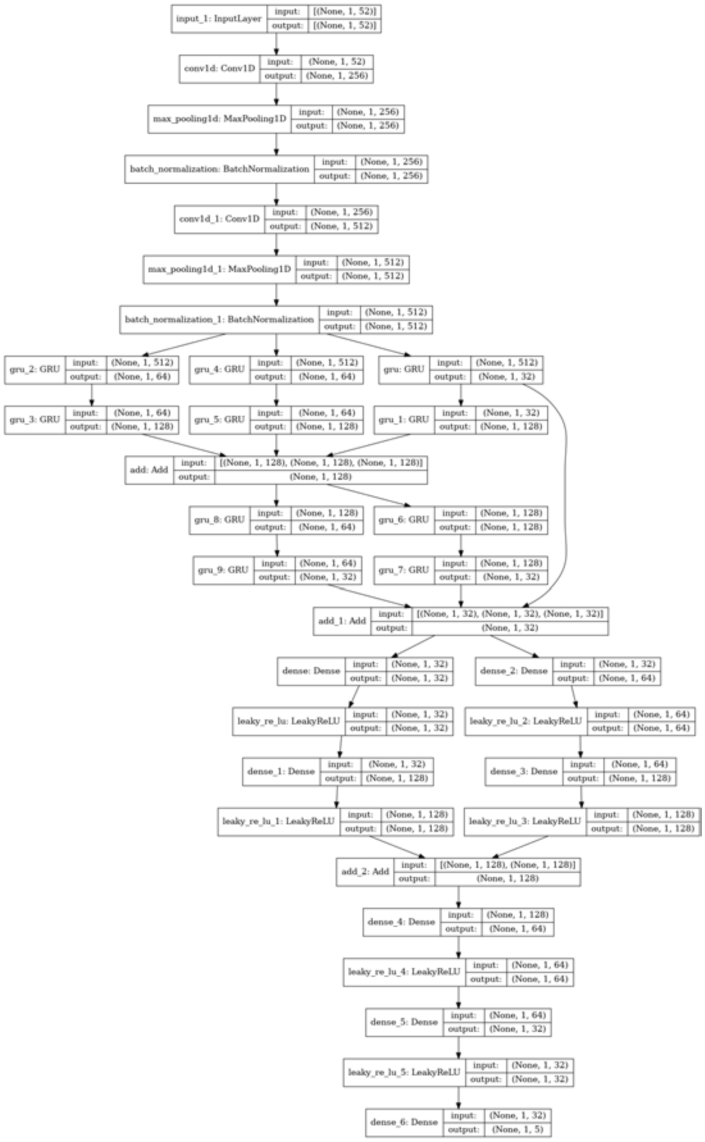


Figure 5.7: Architecture of neural network model.

Improvements in validation loss and accuracy, which evaluated the model's performance on hypothetical data, also indicated that the model was not over fit.

5.2 Lung Sound Classification Outcomes

This thesis intends to classify respiratory sound into multiple categories by applying 4-class classification which is a machine learning task.The performance on a particular model is determined by its accuracy. We are giving brief information about the proposed algorithms which we used .

5.2.1 Attention And Vision Transformer:

Even though the Attention in Vision Transformer models are majorly known for working with pictorial data, they can also detect sequences and patterns in sound data. It used a variety of audio features, such as Mel-frequency cepstral coefficients (MFCCs), Mel spectrograms, and constant-Q transform (CQT).

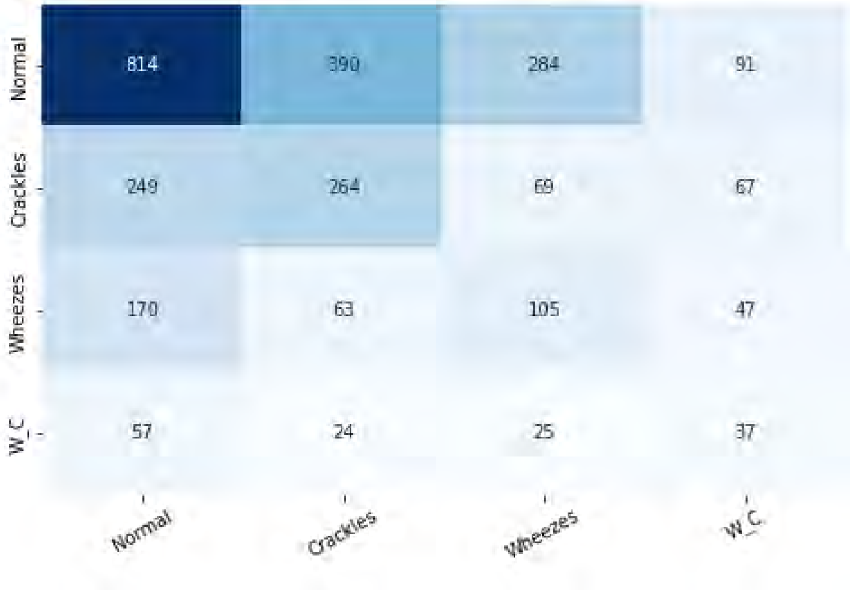


Figure 5.8: Confusion matrix of lung sound classification.

Raw audio data with main features should be fed to the model after necessary extractions. The model then creates patches of the audio data by dividing it into a certain-sized frame or segments. The Attention Mechanism here operates to find correlations or dependencies between these audio patches. It makes the classifications based on these dependencies along with class probabilities with a softmax layer. confusion matrix in seen as output in Figure 5.8 .

5.2.2 Stacked Autoencoder with CNN:

The audio data has to go through a feature extraction process to identify relevant features of the audio file and then the raw audio files are fed into the Stacked Autoencoder. These features can consist of frequencies, spectrograms or MFCCs. After that, the autoencoder's encoder works to decrease the input data's dimensionality to recognize even more features. These features are unlabelled until they are fed into CNN and the CNN classifies and gives class labels relying on the findings comprehended through training. The following graph in Figure 5.10 shows training and validation accuracy vs. loss graph.

The deep convolutional neural network (CNN) used in the model was created for picture categorization. It accepts pictures with three channels of color as well as a dimension of 250x250 pixels as input. The dimensions of space of the feature maps are gradually shrunk by the model's many convolutional layers and max-pooling layers. To capture various degrees of features, the convolutional layers include varied numbers of filters (512, 256, 128, 64) shown in Figure 5.9.

The training progress of this neural network model over 40 epochs was represented. The model's parameters were changed during training to reduce loss, and the accuracy on the training and validation datasets was tracked and displayed the learning

```
Model: "model_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 250, 250, 3)]     0
_________________________________________________________________
conv2d_24 (Conv2D)           (None, 250, 250, 512)     14336
_________________________________________________________________
max_pooling2d_12 (MaxPooling (None, 125, 125, 512)     0
_________________________________________________________________
conv2d_25 (Conv2D)           (None, 125, 125, 256)     1179904
_________________________________________________________________
max_pooling2d_13 (MaxPooling (None, 63, 63, 256)       0
_________________________________________________________________
conv2d_26 (Conv2D)           (None, 63, 63, 128)       295040
_________________________________________________________________
max_pooling2d_14 (MaxPooling (None, 32, 32, 128)       0
_________________________________________________________________
conv2d_27 (Conv2D)           (None, 32, 32, 64)        73792
_________________________________________________________________
max_pooling2d_15 (MaxPooling (None, 16, 16, 64)        0
_________________________________________________________________
conv2d_28 (Conv2D)           (None, 16, 16, 128)       73856
_________________________________________________________________
up_sampling2d_9 (UpSampling2 (None, 32, 32, 128)       0
...
Total params: 14,968,519
Trainable params: 14,968,519
Non-trainable params: 0
_________________________________________________________________
```

Figure 5.9: Stacked Autoencoder with CNN model summary

rate. As the validation dataset is often hidden during training, the aim of training is to enhance the model's performance.
The Figure shown in 5.9 consists of multiple convolutional layers, followed by max-pooling layers for feature extraction. Up-sampling layers increase spatial dimensions, and the model includes dense layers for classification. With approximately 14.97 million trainable parameters.The convolutional layers are responsible for extracting features from the input data. They have different numbers of filters, such as 512, 256, 128, and 64, to capture different levels of information.The final layers include a convolutional layer with 3 output units. The data is then flattened before passing through dense layers.There are multiple dense layers (e.g., 256 and 128 neurons) for making final predictions. A dropout layer with a 0% dropout rate is present but inactive in this model.

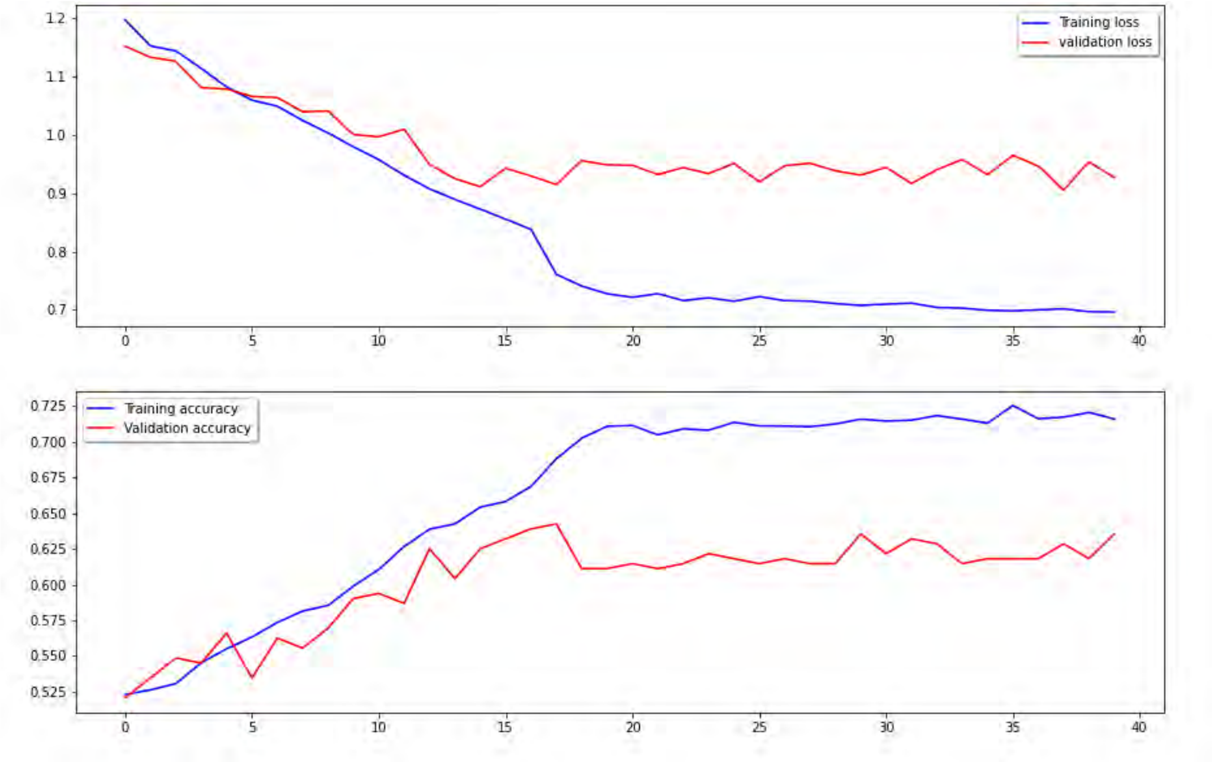


Figure 5.10: Training and validation accuracy vs. loss graph

5.2.3 Support Vector Machine(SVM):

Support Vector Machine (SVM) for Crackles and Wheezes classification operates by locating an optimum decision boundary that divides the two classes in a high-dimensional data space. It accomplishes this by choosing a hyperplane that optimizes the gap between the classes, effectively generating a distinct border between the presence and absence of crackles and wheezes. SVM can manage non-linear connections by employing kernel functions to convert the data into a higher-dimensional space where the classes become linearly separable. Figure 5.11 shows the confusion matrix .

5.2.4 Convolutional Neural Netwok (CNN):

Convolutional neural network (CNN) was applied to identify 4 classes in sound records utilizing Mel-Spectrograms as input. The sound clips are prepared in 5-second buffers and may be isolated into portions with zero cushioning to fit the buffer estimate. Amid preparing, Mel-Spectrograms are transposed and wrapped around the time-axis to permit the arrange to memorize highlights happening at distinctive times within the recording. Information increase methods, such as sound extending and Vocal Tract Length Irritation, are utilized, especially for the less common 'wheeze' and 'wheeze and crackles' classes. The labeling conspire takes after a one-hot encoding approach due to challenges experienced with a multi-label plot and a Sigmoid yield layer. The show faces challenges in precisely classifying 'wheeze' and 'wheeze and crackles,' coming about in lower review scores. Right now, the validation precision stands at roughly 73.86%. The Figure 5.12, shows the

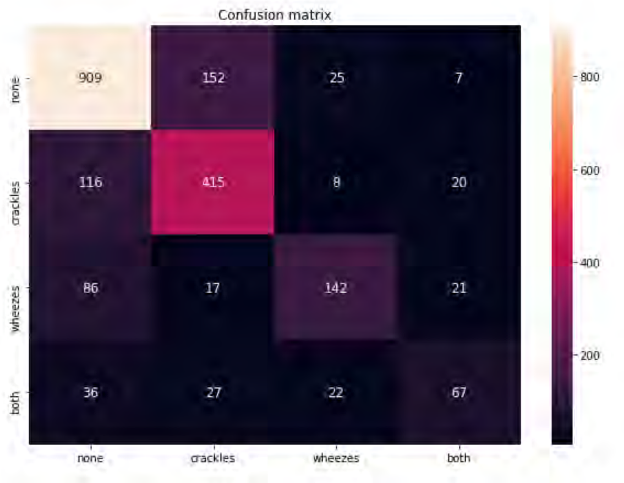


Figure 5.11: Confusion matrix of lung sound classification in SVM.

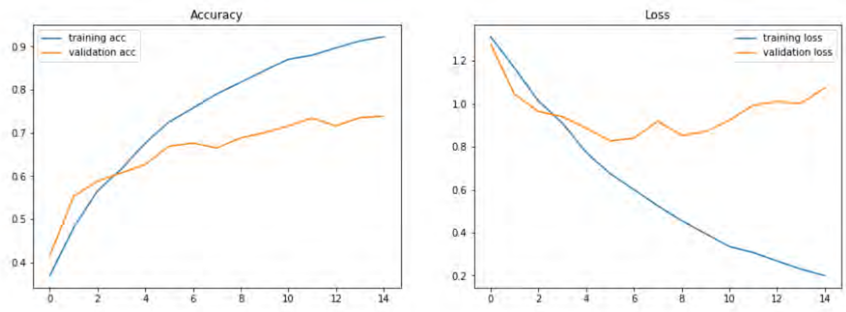


Figure 5.12: Training and validation accuracy vs. loss graph in CNN.

acquired result.

Lung Disease Result Analysis:

The accuracy of the model tells us the proportion of times it precisely predicted the whole dataset. Loss is a number that symbolizes the total of our model's errors. It evaluates the performance of our model. The loss will be high if the errors are high, indicating that the model does not perform well. Otherwise, our model performs better the lower it is [58].

We used Support Vector Machine, Decision Tree from Machine Learning, and Convolutional Neural Network with three feature Extractions (MFCC, Linked Features) from Neural Network and LSTM.The table 5.1 shows that, in Support Vector Machine, we get an accuracy of 69.96%. On the other hand, in the decision tree, we get an accuracy of 74% which implies the model will correctly classify around 74 of 100 of the samples. In CNN, with different feature extraction, we get 88% and 95% of accuracy for MFCC and Linked features respectively. CNN with linked features can classify 95 out of 100 samples which is the best out of all the trained models. The loss is also the lowest in CNN with Linked Features which is commendable[61].

In addition to accuracy, our models can be evaluated based on Precision, F1 score, and Recall. Accuracy does not give any class-specific features, like in which class

Model	Accuracy	Precision	Recall	F1-Score
SVM WITH MFCC	69.96%	64%	70%	64%
DECISION TREE	74%	56%	74%	63%
CNN WITH LINKED FEATURES	95%	96%	95%	95%
CNN WITH MFCC	88%	85%	88%	86%
LSTM/GRU/CNN WITH MFCC	88%	79%	75%	75%

Table 5.1: Identification of significant information obtained from the dataset in respiratory disease classification.

boundaries the model learned competence, but encountered confusion or faced challenges in some. Precision and recall provide more information about the models' proficiency by assessing its' performance across different classes. Depending on the needs of the problem, these metrics are applied in different ways; sometimes it makes sense to evaluate precision, and other times it makes sense to prioritize recall. Confusion matrices can be used to visualize the precision and recall metrics, which makes them easier to comprehend. A confusion matrix, which can be used to determine many different evaluation metrics (such as accuracy, precision, and recall), is an orderly representation of the predictive performance of a classifier on a dataset.

Model	Accuracy	Precision	Recall	F1-Score
SVM WITH MFCC	74.06%	74%	74%	74%
CNN	73.86%	74%	74%	74%
DeiT base+Att+CNN	44.36%	36.43%	36.68%	36.06%
SDA	67.97%	39%	42%	40%

Table 5.2: Identification of significant information obtained from the dataset in lung sound classification.

The table 5.1 signifies a more accurate identification of significant information from the dataset. By comparing the precision values, we can see that the Support Vector Machine (SVM) model has a precision of 64%, which means that it correctly identifies 64 of positive instances out of all instances predicted as positive. As opposed to the SVM model, the decision tree model achieves a precision of 56%, indicating a marginally better performance of SVM.

On the other hand, CNN-MFCC and CNN-Linked Features models' precision values, however, are noticeably higher. A precision of 85% for the CNN-MFCC model demonstrates that it is more accurate at classifying positive instances. The CNN-Linked Features model performs well in accurately identifying positive instances, achieving an even higher precision of 96%. LSTM/GRU model has higher percision value than SVM and Decision tree which is 79%.

The precision and recall test results are used to calculate the F1 score. The f1 score has a maximum value of 1, which represents an ideal precision and recall result. On the other hand, the lowest f1 score value is 0, which indicates that neither precision nor recall have any results. With a 95% f-1 score in CNN-Linked Features which is the highest ever, precision and recall are commendable. It is less efficient in SVM and Decision Tree and LSTM/GRU Models.

Lung Sound Result Analysis:

In this paper, we report the findings of a thorough performance assessment of multiple categorization models used in the context of a particular assignment. Support Vector Machine (SVM) with MFCC, Convolutional Neural Network (CNN), Attention-based Model (ATT) and Stacked Denoising Autoencoder (SDA) and LSTM are among the models under consideration.

From the table 5.2, the accuracy rates of SVM with MFCC and CNN, which were competitively close by having 74.06% and 73.86% respectively. The fact that these models continually maintained an equilibrium between their precision, recall, and F1-Score percentages, each assessed at 74%, is extremely significant. The models' capacity to accurately categorize instances that are beneficial (precision) while successfully capturing all actual positive examples (recall) may be seen in this balancing.

The accuracy of the DeiT base+Att+CNN model was 44.36%, while it's precision, recall, and F1-Score values were all lower and averaged about 36%. Our result indicates a significant issue in accurately locating and classifying positive cases in our model. This reduced efficiency can call for additional tweaking or a reassessment of the model's design.

The SDA model had a great accuracy of 67.97%, but its precision (39%) and recall (42%) percentages were much lower than the other three models. The F1-Score, a critical criterion for determining the model's overall efficacy, was consequently determined to be 40%. This implies that while the model may reach a decent degree of accuracy, it still has to be improved in order to reduce false positives and capture all real positives.

A comprehensive picture of each model's performance may be obtained by combining these intricate measurements. This extensive understanding offers important insights into the models' strengths and potential for additional enhancement, allowing in the selection of models in a way that is well-informed. These understandings are invaluable for analyzing the unique objectives and requirements of a real-world application.

5.3 Spirometry Calculation and Results

About the Dataset:
The dataset that we worked with by Falvo et al. provided us with raw spirometry data that contained the expiratory and inspiratory volumes of 129 individuals along with the times and flow. The data was collected over 2 visits for every individual and consisted of 1 to 16 trials for each visit. The dataset also contained demographic information such as age, sex, height, race, etc. for each subject. Additionally, it also provided us with the time-zero information that indicated when the first expiration began after the time started to be recorded.

What We Calculated:

In order to evaluate Spirometry data we need some crucial measurements such as FEV1, FVC, and FEF25-75 that were not included in the dataset. We needed to study the raw data that was available and calculate these values.

We calculated the following from the data set:
1.FEV1
2.FVC
3.FEV1/FVC
4.FEF25-75

These calculations were carried out using the Visit, Trial, Time Zero, Time, and Volume data for each individual. Later we used the Global Lung Function 2012 equations to find the values that our previous calculations could be compared with.

The equations that we used were:

$$
\begin{aligned}
Mspline = &b0 + b1\left(\frac{Age}{100}\right) + b2\left(\frac{Age}{100}\right)^2 \\
&+ b3\left(\frac{Age}{100}\right)^3 + b4\left(\frac{Age}{100}\right)^4 \\
&+ b5\left(\frac{Age}{100}\right)^5
\end{aligned} \tag{5.1}
$$

$$
\begin{aligned}
Sspline = &c0 + c1\left(\frac{Age}{100}\right) + c2\left(\frac{Age}{100}\right)^2 \\
&+ c3\left(\frac{Age}{100}\right)^3 + c4\left(\frac{Age}{100}\right)^4 \\
&+ c5\left(\frac{Age}{100}\right)^5
\end{aligned} \tag{5.2}
$$

$$
\begin{aligned}
Lspline = &d0 + d1\left(\frac{Age}{100}\right) + d2\left(\frac{Age}{100}\right)^2 \\
&+ d3\left(\frac{Age}{100}\right)^3 + d4\left(\frac{Age}{100}\right)^4 \\
&+ d5\left(\frac{Age}{100}\right)^5 + d6\left(\frac{Age}{100}\right)^6 \\
&+ d7\left(\frac{Age}{100}\right)^7
\end{aligned} \tag{5.3}
$$

$$
L = q0 + q1 * ln(Age) + Lspline \tag{5.4}
$$

$$
\begin{aligned}
M = &\exp\left(a0 + a1\ln(Height) + a2\ln(Age) + a3\text{black}\right. \\
&\left.+ a4\text{NEA} + a5\text{SEA} + Mspline\right)
\end{aligned} \tag{5.5}
$$

$$
\begin{aligned}
S = &\exp\left(p0 + p1\ln(Age) + p2\text{black}\right. \\
&\left.+ p3\text{NEA} + p4\text{SEA} + Sspline\right)
\end{aligned} \tag{5.6}
$$

$$
\text{LLN (5th percentile)} = \exp\left(\frac{\ln(1 - 1.644 \cdot L \cdot S)}{L} + \ln(M)\right) \tag{5.7}
$$

All these values were calculated using the demographic information provided in the dataset.

Calculation Procedure:

We carried out all our spirometry-related calculations in Google Colaboratory using Pandas, Numpy, and Matplotlib libraries in Python.

In order to calculate FEV1, FVC, FEV1/FVC, and FEF25-75, we first needed to merge two separate files, one containing the visit, trial number, time, and volume and another containing only time and time-zero information. The merging process was complicated as the file containing time-zero information only had trial numbers and no visiting numbers to ensure a seamless correspondence. Additionally, the number of rows in the time-zero file had more rows than the one that included the time and volume information. We were able to merge the files by adding some additional information to the two files that indicates when a trial number is being repeated.

After merging the files we did some data preprocessing that included, ensuring that there were no NAN values. We dropped all rows containing NAN values. We set proper data types for each attribute. We made sure that the attributes were in the correct units, for instance, we changed the time-zero values from seconds to milliseconds. Then we renamed some attributes to make the process easier. Finally, we dropped all the unnecessary rows.

Before carrying out our calculations, we wanted to plot some graphs to understand the data. We realized that the data from the second visit were easier to work with. Also, we saw the effects of calibrating the data according to the time-zero values. The Figure 5.13 and 5.14 is an example of a graph without time-zero calibration and after calibration respectively.

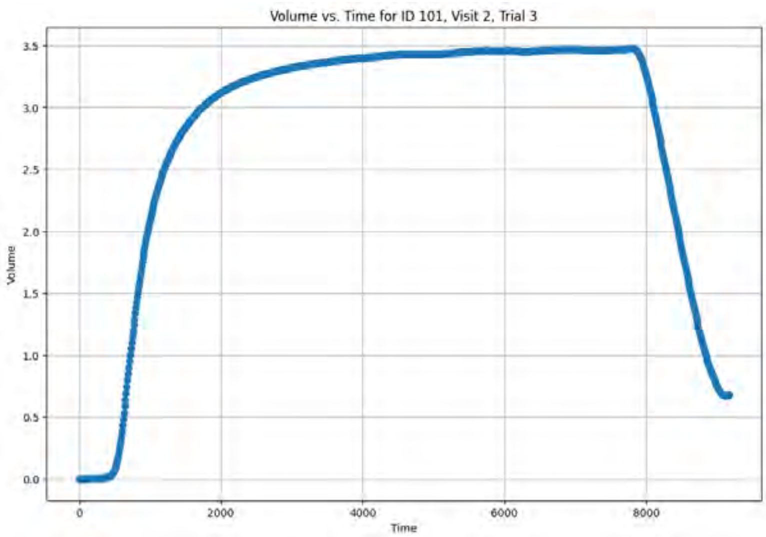


Figure 5.13: Curve without time-zero calibration

The Figure 5.15 tells how the values were calculated:
1. FVC was calculated using the maximum volume reached in each trial.
2. FEV1 was calculated using the value in the first second
3. FEV1/FVC was calculated by dividing the FEV1/FVC values
4. FEF25-75 was calculated by finding the average volume value from 0.25s to 0.75s

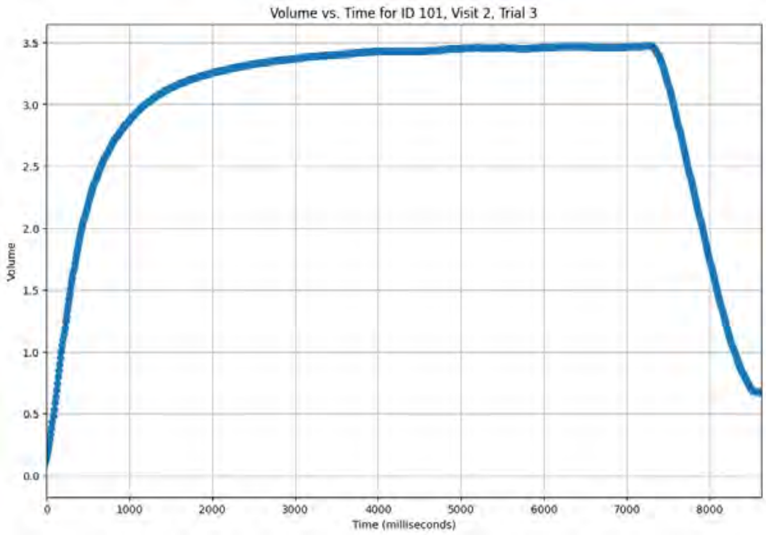


Figure 5.14: Curve ready for calculation.

After carrying out these calculations for each individual, we moved on to calculating the GLI-12 values.

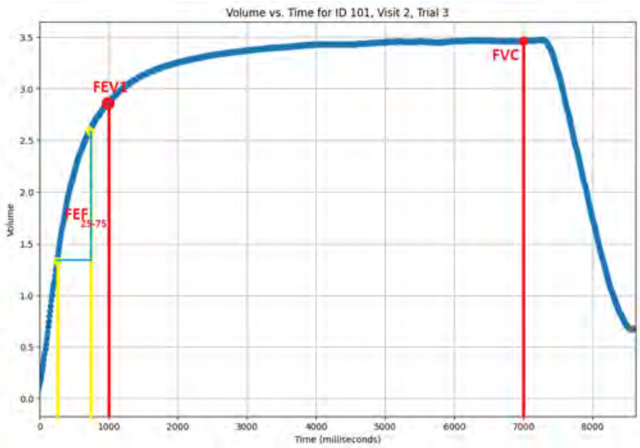


Figure 5.15: Value calculation curve.

After carrying out all these calculations we were able to calculate Z-score.

$$Z - score = ((measured/M)^L - 1)/(L * S) \tag{5.8}$$

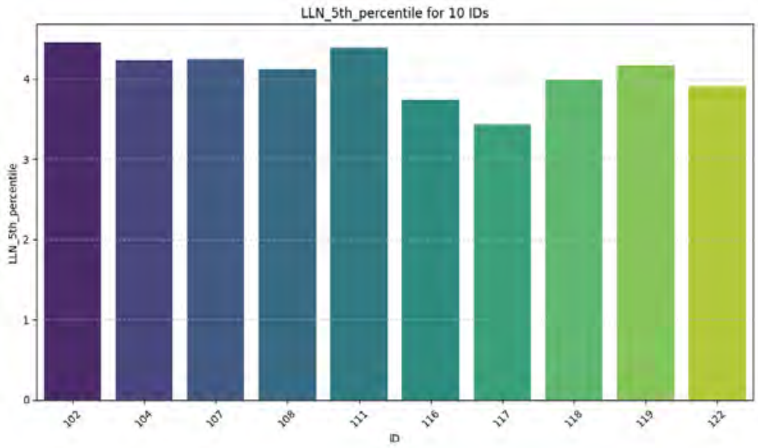


Figure 5.16: Visual demonstration for LLN for 10 ID's.

5.4 Discussion

5.4.1 Classifications of Lung Sound and Diseases

Lung sound classification was done by implementing four types of models. SVM with MFCC gave the most accuracy, which is 74.06%, and CNN showed the second most accuracy, very close to SVM with MFCC, which is 73.86%. Both of these model's precision, recall and f1-score were 74%. SDA's accuracy was less than the other two models, which was 67.97%. However, to our surprise, DeiT base + ATT + CNN, a hybrid model, gives the lowest accuracy of 44.36%. Its precision, recall and f1-score are 36% on average, which is also significantly lesser than the other models.

Additionally, while classifying lung diseases with different models, CNN-Linked features gave the highest accuracy, which is 95% with a precision of 96% Also, both recall and f1-score were 95%, making CNN-Linked features the most reliable model for lung disease classification in this paper. Both CNN-MFCC and LSTM show the second-highest accuracy at 88%. However, SVM with MFCC and Decision tree have shown 69.96% and 74% accuracy, respectively.

It can be stated that SVM with MFCC has the highest accuracy in classifying lung sounds, and CNN-Linked features have the highest accuracy in classifying lung diseases.

CLASS	PRECISION(%)	RECALL(%)	F1-SCORE(%)
NONE	79	83	81
CRACKLE	68	74	71
WHEEZES	72	53	61
BOTH	58	44	50

Table 5.3: Lung sound analysis through SVM.

CLASS	PRECISION(%)	RECALL(%)	F1-SCORE(%)
COPD	98	99	98
HEALTHY	80	57	67
URTI	67	55	57
BRONCHIECTASIS	100	75	86
PNEUMONIA	50	100	67
BRONCHIOLITIS	0	0	0

Table 5.4: Lung Disease analysis through CNN Linked Features.

The precision, recall and evaluating metrics for each classes of the best performed classification model.

Sound Classifier	Model	Study	Accuracy	Sensitivity (Recall)
	SVM	Serbes et al. [23]	49.86%	N/A
	HMM	Chambres et al. [18]	49.50%	42.32%
	SVM	Chambres et al. [18]	49.98%	48.90%
	CNN	Minami et al. [32]	52.79%	31.12%
	Bi-ResNet	Ma et al. [31]	67.44%	58.54%
	ResNet	Nguyen & Pernkopf [49]	73.69%	47.37%
	SVM with MFCC (Our Model)		**74.06%**	**74%**
Disease Classifier	VGG16 scalogram using CWT	Shuvo et al. [50]	88.58%	89%
	CNN model with scalogram using CWT	Shuvo et al. [50]	86.31%	86%
	GRU	Basu and Rana [37]	96%	96%
	RNN with MFCC	Basu and Rana [37]	95.67%	95.67%
	CNN With Linked Features (Our Model)		**96%**	**95%**

Table 5.5: Comparison between proposed model and relevant studies.

5.4.2 Spirometry

The Z-score, a valuable tool in lung function assessment, provides a clear means of gauging how an individual's lung function measurements compare to the expected norms derived from a reference population (Haynes, 2018). It accomplishes this by quantifying the standard deviation (SD) difference between measured values and their predicted counterparts, factoring in the residual standard deviation. This approach has effectively curbed false positive diagnoses often encountered when using traditional criteria like 80% predicted values or a fixed cutoff of 0.70 for defining bronchial obstruction. In the context of spirometry and distinguishing normal from abnormal lung function, the lower limit of normal (LLN) is of critical importance. When assessing lung health, Z-scores help identify values that fall below the LLN, indicating significant deviation from the expected range. For instance, in restrictive lung diseases, both FVC and FEV1 Z-scores drop below the LLN, signifying a restriction. Meanwhile, obstructive patterns in diseases like asthma are corroborated by a negative Z-score for FEV1/FVC below the LLN. The combined use of Z-scores and the LLN enhances the precision of diagnosing a wide array of lung conditions, ensuring more accurate assessments and tailored treatments[20].

The table 5.6 shows difference between NHANES III (National Health and Nutrition Examination Survey III) and GLI-12 (Global Lung Function Initiative 2012) as stated by different papers.

NHANES III (National Health and Nutrition Examination Survey III)	**GLI-12 (Global Lung Function Initiative 2012)**
1. Based on data collected from 7429 asymptomatic, lifelong non-smokers in the USA, covering Caucasians, African-Americans, and Mexican -Americans aged 8–80 years [26].	1. Values are derived from a large dataset containing 74,187 healthy non-smokers from 70 centers across the world. It covers individuals aged 3–95 years and includes diverse ethnic background covering four specific populations (Caucasian, Black, North-East Asian, South-East Asian) [26].
2. Employs piecewise polynomial regression for its reference equations, making it mathematically superior and suitable for a wide age range and diverse ethnic groups specifically for US populations (Caucasian, African-American, Mexican-American) [26] [53]	2. Uses advanced statistical methods, including piecewise polynomial regression. It is considered state-of-the-art and endorsed by major respiratory societies [26].
3. Equations have been recommended for use in the USA but are also utilized in other populations due to their reliability [26].	3. GLI-12 equations are increasingly being implemented worldwide, and they are endorsed by major respiratory societies [26].
4. Equations may still be commonly used in clinical practice despite being based on older data and may require extrapolation for patients over 70 years [26].	4. Provides the most sophisticated reference equation, corresponding well to the biological model of lung function, with a broad age and ethnic representation [26].
5. NHANES III primarily represents the US population [26] [53].	5. It is considered mathematically superior due to its advanced statistical methods, offering a more accurate representation of lung function across a wide age range [26].
6. NHANES III may result in lower disease severity categorization for some studies[30].	6. GLI 12 classifies a higher proportion of obstructed studies as moderately severe, severe, or very severe [30].
7. NHANES III was criticized for poor sampling of the elderly population, especially nonwhite individuals [30].	7. GLI 12 extended age ranges to 95 years old and included more elderly subjects [30].
8. It has a limited ability to account for variance and less comprehensive modeling of age [27].	8. Models variance and skewness and includes splines for age variation [27].

Table 5.6: Comparing GLI-12 equations with NHANES III for Spirometry.

5.4.3 Limitations

During our research, we aimed to create hybrid models, believing they could outperform traditional models. We put significant effort into crafting two such hybrid models, but they didn't yield the expected results. We also tried to expand our dataset by adding more data, but ran into issues due to differences in data structures and the extensive work needed for integration.

In the spirometry domain, our goal was to access larger datasets with demographic details. We wanted to integrate machine learning into spirometry for simpler assessments and the ability to predict lung conditions for diagnosis. Sadly, we faced challenges finding comprehensive datasets.

Unfortunately, we couldn't fully realize our research's potential due to time and resource constraints. Nonetheless, our work provides a foundation for future research in respiratory disease classification and lung health assessment. We hope our findings will inspire others to build upon them and advance this vital medical field.

Chapter 6

Conclusion and Future Works

6.1 Conclusion

In this study, we embarked on a comprehensive exploration of utilizing lung sounds and spirometry for improved respiratory health assessment. In the course of our investigation, we ventured into experimental evaluation, rigorously testing various classification models, including SVM, Decision Trees, CNNs, LSTM/GRU, Attention-based Models, and Stacked Denoising Autoencoders. These models were harnessed with different feature extraction techniques to classify respiratory diseases and lung sound patterns. Through a systematic assessment, we discerned their strengths and limitations, emphasizing the importance of metrics like accuracy, precision, recall, and F1-score in gauging model performance. Additionally, in our investigation of spirometry data, we emphasized the significance of reference values, Z-scores, and the lower limit of normal (LLN) to provide more precise assessments of lung function. This research underscores the potential of merging machine learning and medical data to enhance respiratory disease detection and lung health assessment, paving the way for more accurate diagnoses and personalized treatment strategies in the field of respiratory medicine.

6.2 Future Work

1. The unreliability and robustness of the classification model can be improved by gathering a larger and more varied dataset of lung sound and spirometry data that we are still searching for. It is possible to gain a more thorough understanding of respiratory diseases and how they manifest themselves by incorporating data from various populations, age groups, and demographic locations.

2. To evaluate the classification model's performance and efficacy in the real world, conduct rigorous validation studies and clinical trials. This entails working with healthcare organizations, selecting a patient population that is diverse, and comparing the model's performance to accepted diagnostic guidelines and professional judgment.

3. Examine the classification model's compatibility with electronic health record systems. This integration can support population-level studies on respiratory diseases, enable seamless information exchange, enable longitudinal analysis of patient data, and more.

4. Proposing cochleogram-based TF representation to improve the learning process of a CNN model in the classification of respiratory adventitious sounds which has not being applied in this context to the best of our knowledge.

5. Applying our classification models with different and multiple feature extraction to find the best suited one.

6. Utilizing a varied dataset that encompasses pediatric respiratory sound recordings.

Bibliography

[1] D. R. Holleman Jr and D. L. Simel, "Does the clinical examination predict airflow limitation?" en, *JAMA*, vol. 273, no. 4, pp. 313–319, 1995.

[2] N. Meslier, G. Charbonneau, and J. L. Racineux, "Wheezes," en, *Eur. Respir. J.*, vol. 8, no. 11, pp. 1942–1948, 1995.

[3] T. J. Barreiro and I. Perillo, "An approach to interpreting spirometry," *afp*, vol. 69, no. 5, pp. 1107–1115, 2004.

[4] M. Oud and E. J. W. Maarsingh, "Spirometry and forced oscillometry assisted optimal frequency band determination for the computerized analysis of tracheal lung sounds in asthma," *Physiol. Meas.*, vol. 25, no. 3, pp. 595–606, 2004.

[5] J. D. Leuppi, T. Dieterle, I. Wildeisen, *et al.*, "Can airway obstruction be estimated by lung auscultation in an emergency room setting?" en, *Respir. Med.*, vol. 100, no. 2, pp. 279–285, 2006.

[6] J.-C. Wang, J.-F. Wang, C.-B. Lin, K.-T. Jian, and W.-H. Kuok, "Content-based audio classification using support vector machines and independent component analysis," vol. 4, Jan. 2006, pp. 157–160. DOI: 10.1109/ICPR.2006.407.

[7] "COPD (chronic obstructive pulmonary disease)," en, *PrimaryCare*, vol. 7, no. 48, pp. 727–728, 2007.

[8] *NCI dictionary of cancer terms*, en, https://www.cancer.gov/publications/dictionaries/cancer-terms/def/respiratory-disease, Accessed: 2023-5-19, Feb. 2011.

[9] C. A. Vaz Fragoso, T. M. Gill, G. McAvay, P. H. Van Ness, H. K. Yaggi, and J. Concato, "Use of lambda-mu-sigma-derived Z score for evaluating respiratory impairment in middle-aged persons," en, *Respir. Care*, vol. 56, no. 11, pp. 1771–1777, 2011.

[10] Y. Nagasaka, "Lung sounds in bronchial asthma," en, *Allergol. Int.*, vol. 61, no. 3, pp. 353–363, 2012.

[11] P. H. Quanjer, S. Stanojevic, T. J. Cole, *et al.*, "Multi-ethnic reference values for spirometry for the 3–95-yr age range: The global lung function 2012 equations," en, *Eur. Respir. J.*, vol. 40, no. 6, pp. 1324–1343, 2012.

[12] N. Jaitly and G. E. Hinton, "Vocal tract length perturbation (vtlp) improves speech recognition," in *Proc. ICML Workshop on Deep Learning for Audio, Speech and Language*, vol. 117, 2013, p. 21.

[13] M. Mineshita, H. Kida, H. Handa, *et al.*, "The correlation between lung sound distribution and pulmonary function in COPD patients," en, *PLoS One*, vol. 9, no. 9, e107506, 2014.

[14] Q.-Y. Dai, C.-P. Zhang, and H. Wu, "Research of decision tree classification algorithm in data mining," *Int. J. Database Theory Appl.*, vol. 9, no. 5, pp. 1–8, 2016.

[15] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, IEEE, 2016.

[16] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," en, *Neurocomputing*, vol. 184, pp. 232–242, 2016.

[17] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," en, *PLoS One*, vol. 12, no. 5, e0177926, 2017.

[18] G. Chambres, P. Hanna, and M. Desainte-Catherine, "Automatic detection of patient with respiratory diseases using lung sound analysis," in *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, IEEE, 2018, pp. 1–6.

[19] N. Das, M. Topalovic, and W. Janssens, "Artificial intelligence in diagnosis of obstructive lung disease: Current status and future potential," en, *Curr. Opin. Pulm. Med.*, vol. 24, no. 2, pp. 117–123, 2018.

[20] J. Haynes, "Basic spirometry testing and interpretation for the primary care provider," en, *Can. J. Respir. Ther.*, vol. 54, no. 4, pp. 92–98, 2018.

[21] R. Kammoun, I. Ghannouchi, S. Rouatbi, and H. Ben Saad, "Defining and grading an obstructive ventilatory defect (OVD): 'FEV_1/FVC lower limit of normal (LLN) vs. Z-score' and 'FEV_1 percentage predicted (%pred) vs. Z-score'," en, *Libyan J. Med.*, vol. 13, no. 1, p. 1 487 751, 2018.

[22] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, and G. Li, "Computational intelligence enabling the development of efficient clinical decision support systems: Case study of heart failure," en, in *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*, Elsevier, 2018, pp. 123–133.

[23] G. Serbes, S. Ulukaya, and Y. P. Kahya, "An automated lung sound preprocessing and classification system based OnSpectral analysis methods," in *Precision Medicine Powered by pHealth and Connected Health*, Singapore: Springer Singapore, 2018, pp. 45–49.

[24] R. V. Sharan, U. R. Abeyratne, V. R. Swarnkar, S. Claxton, C. Hukins, and P. Porter, "Predicting spirometry readings using cough sound features and regression," *Physiol. Meas.*, vol. 39, no. 9, p. 095 001, 2018.

[25] V. Viswanath, J. Garrison, and S. Patel, "SpiroConfidence: Determining the validity of smartphone based spirometry using machine learning," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2018, pp. 5499–5502.

[26] M. Vukoja, A. Bokan, G. Vujasinovic, and I. Kopitovic, "The differences in spirometry predictive equations in classifying presence and degree of lung function impairment: Which suit fits the best?" en, *Lung*, vol. 196, no. 1, pp. 87–92, 2018.

[27] J. Zhang, X. Hu, X. Tian, and K.-F. Xu, "Global lung function initiative 2012 reference values for spirometry in asian americans," en, *BMC Pulm. Med.*, vol. 18, no. 1, 2018.

[28] *Chronic obstructive pulmonary disease (COPD): Overview*. Institute for Quality and Efficiency in Health Care (IQWiG), 2019.

[29] Z. Huang, F. Yang, F. xu, X. Song, and K.-L. Tsui, "Convolutional gated recurrent unit–recurrent neural network for state-of-charge estimation of lithium-ion batteries," *IEEE Access*, vol. PP, pp. 1–1, Jul. 2019. DOI: 10.1109/ACCESS.2019.2928037.

[30] N. A. Huprikar, A. B. Holley, A. J. Skabelund, *et al.*, "A comparison of global lung initiative 2012 with third national health and nutrition examination survey spirometry reference values. implications in defining obstruction," en, *Ann. Am. Thorac. Soc.*, vol. 16, no. 2, pp. 225–230, 2019.

[31] Y. Ma, X. Xu, Q. Yu, *et al.*, "LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-ResNet deep learning algorithm," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, IEEE, 2019, pp. 1–4.

[32] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano, and S. Kido, "Automatic classification of large-scale respiratory sound dataset based on convolutional neural network," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, IEEE, 2019, pp. 804–807.

[33] Phung and Rhee, "A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets," en, *Appl. Sci. (Basel)*, vol. 9, no. 21, p. 4500, 2019.

[34] *Respiratory sound database*, https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database, Jan. 2019.

[35] A. Tautan, M. Dogariu, and B. Ionescu, "Detection of epileptic seizures using unsupervised learning techniques for feature extraction," vol. 2019, Jul. 2019, pp. 2377–2381. DOI: 10.1109/EMBC.2019.8856315.

[36] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," en, *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, 2019.

[37] V. Basu and S. Rana, "Respiratory diseases recognition through respiratory sound with the help of deep neural network," in *2020 4th International Conference on Computational Intelligence and Networks (CINE)*, IEEE, 2020.

[38] S. Gairola, F. Tom, N. Kwatra, and M. Jain, "RespireNet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting," 2020.

[39] B. Gosmar, *Neural networks and speech recognition*, en, https://www.gosmar.eu/machinelearning/2020/05/25/neural-networks-and-speech-recognition/, Accessed: 2023-9-17, May 2020.

[40] M. K. Gurucharan, *Machine learning basics: Support vector machine (SVM) classification*, en, https://towardsdatascience.com/machine-learning-basics-support-vector-machine-svm-classification-205ecd28a09d, Accessed: 2023-9-17, Aug. 2020.

[41] G. Rudraraju, S. Palreddy, B. Mamidgi, *et al.*, "Cough sound analysis and objective correlation with spirometry and clinical diagnosis," en, *Inform. Med. Unlocked*, vol. 19, no. 100319, p. 100 319, 2020.

[42] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," *Applied Acoustics*, vol. 158, p. 107 020, 2020.

[43] J. Shrestha, S. Razavi Bazaz, H. Aboulkheyr Es, *et al.*, "Lung-on-a-chip: The future of respiratory disease models and pharmacological studies," en, *Crit. Rev. Biotechnol.*, vol. 40, no. 2, pp. 213–230, 2020.

[44] D. Singh, B. K. Singh, and A. K. Behera, "Comparative analysis of lung sound denoising technique," in *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, IEEE, 2020.

[45] N. Adaloglou, *How the vision transformer (ViT) works in 10 minutes: An image is worth 16x16 words*, en, https://theaisummer.com/vision-transformer/, Accessed: 2023-9-15, Jan. 2021.

[46] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," en, *Data Brief*, vol. 35, no. 106913, p. 106 913, 2021.

[47] U. Kiran, *MFCC technique for speech recognition*, en, https://www.analyticsvidhya.com/blog/2021/06/mfcc-technique-for-speech-recognition/, Accessed: 2023-9-17, Jun. 2021.

[48] L. Nanni, G. Maguolo, S. Brahnam, and M. Paci, "An ensemble of convolutional neural networks for audio classification," en, *Appl. Sci. (Basel)*, vol. 11, no. 13, p. 5796, 2021.

[49] T. Nguyen and F. Pernkopf, "Lung sound classification using co-tuning and stochastic normalization," 2021.

[50] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2595–2603, 2021.

[51] O. Calzone, *An intuitive explanation of LSTM - ottavio calzone*, en, https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c, Accessed: 2023-9-15, Feb. 2022.

[52] W.-B. Ma, X.-Y. Deng, Y. Yang, and W.-C. Fang, "An effective lung sound classification system for respiratory disease diagnosis using DenseNet CNN model with sound pre-processing engine," in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, IEEE, 2022.

[53] L. Marshall, *Spirometry predictive equations: NHANES III vs. GLI 12*, https://www.vitalflohealth.com/post/spirometry-predictive-equations-nhanes-iii-vs-gli-12, Accessed: 2023-9-19, Apr. 2022.

[54] M. R. Munawar, *What is data preprocessing, why do we need that?* en, https://medium.com/nerd-for-tech/what-is-data-preprocessing-why-we-need-that-2846b8b04bc4, Accessed: 2023-9-13, Apr. 2022.

[55] B. Zimmerman and D. Williams, *Lung Sounds.* StatPearls Publishing, 2022.

[56] S. Bae, J.-W. Kim, W.-Y. Cho, *et al.*, "Patch-mix contrastive learning with audio spectrogram transformer on respiratory sound classification," 2023.

[57] M. Falvo, D. Ibraheem, B. Samy, and J. Therkorn, "Raw Spirometry Dataset: Inspiratory and Expiratory," Feb. 2023. DOI: 10.6084/m9.figshare.19196543.v2. [Online]. Available: https://figshare.com/articles/dataset/Raw_Spirometry_Dataset_Inspiratory_and_Expiratory/19196543.

[58] R. Kundu, *Precision vs. recall: Differences, use cases & evaluation*, en, https://www.v7labs.com/blog/precision-vs-recall-guide, Accessed: 2023-9-13, May 2023.

[59] A. Lamberti, *From sound to sight: Using vision transformers for audio classification*, en, https://medium.com/artificialis/from-sound-to-sight-using-vision-transformers-for-audio-classification-aa6a0293914c, Accessed: 2023-9-17, Jan. 2023.

[60] Y. Li, J. Wang, X. Dai, *et al.*, "How does attention work in vision transformers? a visual analytics attempt," 2023.

[61] M. Riva, *Interpretation of loss and accuracy for a machine learning model*, https://www.baeldung.com/cs/ml-loss-accuracy, Accessed: 2023-9-13, Dec. 2023.

[62] *What are the symptoms of lung cancer?* en, https://www.cdc.gov/cancer/lung/basic_info/symptoms.htm, Accessed: 2023-9-13, Aug. 2023.

[63] *What is a Short-Time fourier transform?* en, https://www.collimator.ai/reference-guides/what-is-a-short-time-fourier-transform, Accessed: 2023-9-16, Jul. 2023.

[64] AAFA Community Services, *What happens in your airways when you have asthma?* en, https://community.aafa.org/blog/what-happens-in-your-airways-when-you-have-asthma, Accessed: 2023-9-13.

[65] American Lung Association, *Spirometry*, en, https://www.lung.org/lung-health-diseases/lung-procedures-and-tests/spirometry, Accessed: 2023-9-13.

[66] *Asthma*, en, https://www.who.int/news-room/fact-sheets/detail/asthma, Accessed: 2023-9-13.

[67] *Bronchiectasis*, en, https://www.nhs.uk/conditions/bronchiectasis/, Accessed: 2023-9-13.

[68] *Bronchiolitis*, en, https://www.nhs.uk/conditions/bronchiolitis/, Accessed: 2023-9-13.

[69] *Chronic obstructive pulmonary disease (COPD)*, en, https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd), Accessed: 2023-5-19.

[70] *Chronic respiratory diseases*, en, https://www.who.int/health-topics/chronic-respiratory-diseases, Accessed: 2023-9-13.

[71] C. C.-1. Demo, *CS231n convolutional neural networks for visual recognition*, https://cs231n.github.io/convolutional-networks/, Accessed: 2023-9-13.

[72] A. Gogna and A. Majumdar, *Discriminative autoencoder for feature extraction: Application to character recognition*, https://arxiv.org/ftp/arxiv/papers/1912/1912.12131.pdf, Accessed: 2023-9-17.

[73] *Lung sounds: What do they mean?* en, https://www.webmd.com/lung/lung-sounds, Accessed: 2023-9-13.

[74] *Up to 650 000 people die of respiratory diseases linked to seasonal flu each year*, en, https://www.who.int/news/item/13-12-2017-up-to-650-000-people-die-of-respiratory-diseases-linked-to-seasonal-flu-each-year, Accessed: 2023-9-13.

[75] *Wheezing*, en, https://my.clevelandclinic.org/health/symptoms/15203-wheezing, Accessed: 2023-9-13.