Automated Diagnosis of Respiratory Disorders Using Convolutional Neural Networks

Sadashiv Dalvi

Supervised by Dr. Mahdi Maktab Dar

A thesis submitted in fulfillment of the requirements
for the degree MSc Applied Data Science

School of Computing and Information Science
Anglia Ruskin University

09/09/2024

Declaration

I, **Sadashiv Dalvi**, declare that the work in this dissertation titled *"Automated Diagnosis of Respiratory Disorders Using Convolutional Neural Networks"* is carried out by me. This work has not been submitted to Anglia Ruskin University or any other educational institution for the award of a degree or educational qualification. I also declare that the information published in this dissertation has been obtained and presented according to academic rules and ethical conduct. Any information obtained from other sources has been properly referenced.

Acknowledgement

I'd like to express my heartfelt gratitude to everyone who has helped me on this journey.

To my supervisor, Dr. Mahdi Maktab Dar, I appreciate your invaluable advice, insightful feedback, and unwavering support. Your guidance and encouragement were critical in shaping this report and my academic progress.

My family has provided me with unwavering support and understanding. Your patience and encouragement have helped me achieve my goals.

Thank you for believing in me and helping me achieve this goal.

Table of Contents

List of Figures

List of Tables

Abstract

The research focuses on the benefits of using automation in the diagnosis of lung disorders. The mortality rates for lung diseases have increased over the years. After the COVID-19 pandemic, It has become extremely important that lung disorders are diagnosed at an early stage for better chances of recovery. Researchers have developed various Machine Learning and Deep Learning models for accurately predicting lung diseases. Various techniques such as CNN Models, LSTM models, and SVM models have shown promising results. The main aim of the research is to investigate if a hybrid model consisting of CNN, LSTM, and Attention layers would improve the performance of the model significantly. MFCCs are computed for all the recordings after data augmentation and then fed to each model. The models were then compared and evaluated on different metrics. The CNN model with LSTM performed better than the baseline CNN Model and the hybrid attention model because of its ability to capture temporal dependencies. Risk assessment was implemented using the softmax function of the CNN+LSTM Model. Risk assessment furnishes healthcare establishments with valuable insights to enable them to distinguish between patients with varying risk profiles. Research in this domain should be encouraged Because automation improves diagnostic accuracy and consistency through objective data analysis and sophisticated algorithms. Automation also helps medical professionals by lessening their workload and providing more access to cutting-edge diagnostic tools, especially in underprivileged areas.

Chapter 1

Introduction

1.1 Overview

According to studies done by NHS England, 1 out of 5 people in England are diagnosed with some form of respiratory disease. It is responsible for 3rd largest number of deaths in England trailing only Cancer and Cardiovascular diseases. The most common respiratory diseases are Asthma, Chronic obstructive pulmonary disease (COPD), and occupational lung disease. Annually, The economic load for all lung conditions accounts for GBP 11 billion to the NHS UK. Worldwide, Respiratory diseases claim around 4 million lives yearly, per the National Institute of Health (NIH). These numbers have increased significantly in the aftermath of the COVID-19 pandemic. This implies that lung diseases cause tremendous health and social and economic impacts on society. Hence, it becomes very imperative that these diseases are diagnosed at an early stage due to their extreme consequences.

Some of the methods used nowadays include bronchoscopy, CT scan, Chest X-ray, Echocardiogram, etc. While Hospitals and healthcare institutions offer precise diagnoses, their efficiency is limited by professional expertise and personalized interpretations. Also, these methods are costly and the patient undergoing CT scans are at risk of developing cancer due to their exposure to radioactive particles. Recent developments in electronic stethoscopes have kindled curiosity in contact-free healthcare and automated lung disease prediction.

1.2 Problem Background

In recent years, Machine learning and Deep Learning algorithms have been used in various healthcare domains such as predicting heart anomalies, diagnosing skin cancer, and even predicting lung disease. However, there is still scope for additional research and examination when it comes to lung disease classification because of

the nature of the data. Studies have developed neural network models for better classification of breathing sounds which are extremely crucial for the prediction of the lung disorder. Crackles in the sound signify lung conditions such as COPD, Pneumonia, and lung fibrosis whereas wheezes are high-pitched sounds linked with asthma and COPD.

Various studies have been conducted across this domain. Different Machine learning models were used to make accurate lung disease prediction. Because of the complexity of the dataset, various deep learning methods were also applied. Deep learning model have an competitive edge over Machine learning models because of their ability to perform better on complex data and large datasets. Research's like feeding MFCCs to SVM model and feeding spectrograms to CNN models have also showed promising results. Models are also built upon transfer learning techniques.

1.3 Research Aim

The aim of the research is to build a powerful and robust deep learning model that can precisely predict respiratory illness such as COPD, Bronchiolitis, URTI and Pneumonia. Also, the aim is to perform a risk assessment task to predict if a healthy patient is at risk of developing a lung disease. The research seeks to improve diagnostic precision and early identification, aid in assessing risk, and support customized treatment strategies, ultimately enhancing patient results and lessening the strain on healthcare systems.

1.4 Research Objectives

- To investigate the benefit of various data augmentation techniques to mitigate the class imbalance problem.
- To investigate if Hybrid models such as CNN + LSTM and CNN+ LSTM + Attention mechanism improve the model's performance.
- To perform risk assessment on the dataset

1.5 Research Scope

In-scope: This study aims to create a deep learning model to automatically diagnose and classify the risk of respiratory diseases such as COPD, Bronchiolitis, Pneumonia, and URTI, using the respiratory sound dataset from the ICBHI conference. This study will include preparing respiratory sound recordings, extracting

important features, and creating and assessing a deep learning model that can accurately classify respiratory conditions.

The study will focus on improving current diagnostic models, which can be inaccurate because of the complicated nature of respiratory sounds and the diverse conditions of patients. The goal of the research is to address these difficulties by combining attention mechanisms and cutting-edge deep learning methods, providing a more accurate and dependable diagnostic tool. The assessment criteria will consist of precision, recall, F1-score, and overall accuracy, aiming to elevate the diagnostic capabilities for each individual condition. The study will also investigate how this model could help with identifying illnesses early and creating personalized treatment plans, which can lead to improved outcomes for patients.

Out of scope: Although the goal of this study is to enhance the precision and dependability of diagnosing respiratory conditions through deep learning, there are certain elements that fall outside the scope of the research. The research will not focus on implementing the model directly in clinical settings, including integrating it into current healthcare systems or using it in real-time clinical settings. In addition, the study will not discuss diagnosing respiratory illnesses other than COPD, Bronchiolitis, Pneumonia, and URTI. The research will not include investigating other data sources or different types of information, like patient demographics or imaging results, and will only concentrate on the respiratory sound recordings from the ICBHI dataset.

Chapter 2

Literature Review

2.1 Approach 1

According to a study by Bardou et al (2018), in the field of bioinformatics, the researcher found that the lung sound emanates crucial information relating to pulmonary disorders and physicians evaluating patients with pulmonary conditions obtain this information with the help of traditional techniques of auscultation. But this technique is basically very basic and limited. Wrong diagnosis is highly probable in case doctors and physicians are not well trained to perform this procedure. This may lead to wrong treatments of the patient and that would lead to physical harm of the patient. What makes the task of the physicians more difficult is the fact that lung sounds are non-stationary, which naturally increases the complications of the analytical procedure, which involves recognition and distinction of the lung sounds. As a result, it is very important to develop an automatic recognition system which can help to overcome the limitations of the traditional method of auscultation. Bardou et al. (2018)

2.2 Approach 2

Pramono et al (2019), in their study have evaluated various features for classifying wheezes and normal sound of the respiration. In this study, they found that diseases such as Asthma which can be categorised as Chronic Respiratory Disease and Chronic Obstructive Pulmonary Disease can be categorised as health-related problems leading to a huge number of deaths globally. It is only possible to prevent deterioration of these diseases through actively monitoring the symptoms, however, these are not curable. Wheezing sounds in the breathing are key symptoms and indicators which must be monitored at an early stage of the disease to ensure the disease does not aggravate beyond a point where it is no longer possible to man-

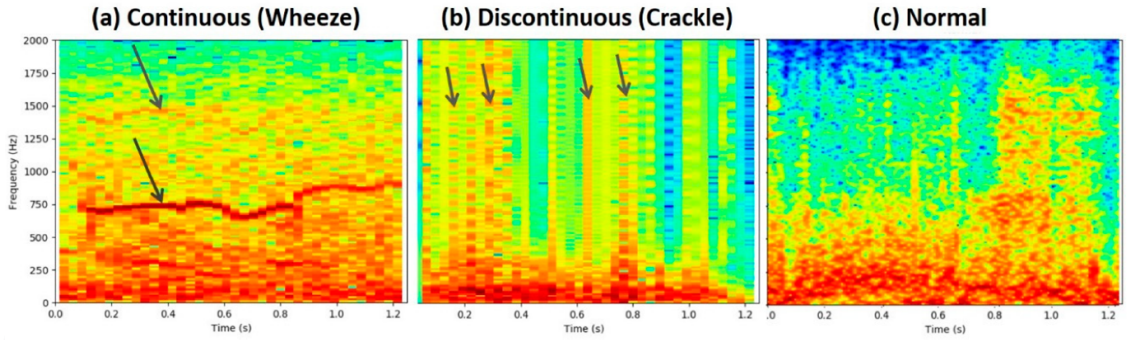


Figure 2.1: Classification of Lung Sounds

age it. Wheezing can occur anytime without any prior indication, which means an automatic wheezing detector which would be able to continuously monitor wheezing sounds automatically would be very useful in the process of managing these respiratory diseases.

This research aims to evaluate the different types of features and distinguishing ability associated with wheezing sounds mentioned in the previous studies. A total of 105 features are evaluated which are focused on automatic detection of wheezing sounds during respiration. After a comprehensive evaluation of various discriminatory abilities of different types of wavelet, time, spectral and cepstral features which includes a total size of 105 for automatic identification of wheezing sounds and pattern of breathing. The study found that specific individualistic features such as tonality index and MFCC can more precisely detect wheezing sounds. But their computational requirements are more intricate and higher compared to easier features related to time domain. It has been further shown in the study that, even though usage of multiple features enhances accuracy of classification in a few cases, the improvement of performance reduces after a specific number of features. The present research used a very simple classifier, using other more complex classifiers such as artificial neural network or vector machines may further increase the performance of classification. Therefore, this research concludes that it is essential to take note of all the competing requirements while selecting a wheeze detection feature for a wide range of applicability.Pramono et al. (2019)

2.3 Approach 3

Pham et al. (2020), conducted research on inception-based network and multi-spectrogram ensemble which is used for making predictions on respiratory and lung diseases and anomalies. Through respiratory sound input, the present study proposes an inception based neural network for identifying lung diseases. Firstly, the

respiratory sounds of the patients are recorded. These sounds are transformed into spectrograms where spectral and temporal information are clearly given. This is remitted to as a front-end feature extraction. After that, the transformed spectrograms are installed into the suggested network which is identified as back end categorisation in order to detect if the patients are affected by lung diseases or respiratory disorders. The present research has been conducted under the guidelines of ICBHI benchmark respiratory sound data set. The present research has also compared the multiple spectrogram ensemble and inception-based network with state-of-the-art methods of detecting respiratory illness. The outcome of the comparison suggested that the proposed system is highly competitive with the state-of-the-art system. In some scenarios, the ensemble system also had outperformed the state-of-the-art system. As a result of the research exploration and extensive experiment on the metadata set of ICBHI, the best model proposed uses inception 01 architecture and the state-of-the-art systems are conquered by Gammatone gram and Scalogram in both tasks which establishes and validates the efficiency of deep learning in order to diagnose respiratory diseases at an early stage.

2.4 Approach 4

Aykanat et al. (2017) opines that in the medical field, due to technological development and introduction of the computer system, diagnosis and analysis of various diseases has become much easier. Technology has especially facilitated non-invasive diagnostic procedures which can help to identify a wide range of health conditions which previously needed complicated diagnosis, intervention or treatment. The present research papers investigate various non-invasive procedures of categorizing breathing sounds which are recorded with the help of an electronic stethoscope and software of audio recording by deploying numerous Machine learning algorithms. A cost effective and simple to use electronic stethoscope has been developed which facilitates the storage and recording of respiratory sounds captured on a computer system. 17,930 lungs were recorded from 1630 study participants with the help of this device. Two types of machine learning algorithms were used which are Mel frequency cepstral coefficient also called MFCC that features in a Support vector machine or SVM and images of the spectrograms in the CNN.

Because the usage of MFCC features with the Support vector machine algorithm is a commonly accepted categorisation procedure for audio. Its results were utilized for setting a benchmark for the Convolutional Neural Network algorithm. In this research study four data sets were categorized in order to make classification of respiratory audio. The first category is pathological compared with healthy, the

second is classification between rale, rhonchus and normal sound. The third category is classification of singular respiratory sound and the fourth one is classification of audio with all sound types. The accuracy results of all the above-mentioned classification are 86

2.5 Approach 5

Zhang et al. (2023) conducted research for detection and classification of pulmonary disease with the help of respiratory audio files by using long short-term memory neural networks. The research was conducted with the aim of improving diagnostic accuracy for respiratory diseases. This research proposes a novel methodology to increase precision of diagnosis with the use of respiratory audio recordings to treat and improve conditions like chronic obstructive pulmonary disease, pneumonia, bronchiolitis, upper respiratory tract infection etc. The methodology of this research study involved training of four different machine learning algorithms on a data set composed of 920 respiratory audio files. Digital stethoscopes were used to record these lung sounds which created the Respiratory Sound Database.

Convolutional Neural Network, CNN with unidirectional long short term memory, CNN with bi-directional long short term memory, and Long Short-Term Memory were the models that helped accomplish the study. Precision, accuracy, recall, and F-1 score were among the metrics used to evaluate the aforementioned models. Long short term memory, with an accuracy rate of 98.82%, was the algorithm that performed the best. F-1 received a score of minus 97 points. As a consequence of its propensity to forecast the future state sequence using audio signals, the LSTM algorithm outperformed the other algorithms examined in terms of predictive accuracy, according to the study's findings. In summary, the study demonstrates that the LSTM model is able to precisely interpret the patient's lung sounds and makes timely predictions regarding the onset and severity of lung disease.

2.6 Approach 6

Cozzatti et al. (2022), conducted a research where a proposal was given on a poorly-supervised approach based on a machine learning tool which would help to alert patients about potential respiratory problems. The study suggests that a wide range of pathologies has the potential to impact the respiratory system in the human body, which generally leads to the development of more severe illnesses and can even lead to death in some cases. It is essential to study and find useful preventive practices which can provide essential assistance in improving the health condition of the patient.

The research aims to propose a tool which will be able to automatically detect respiratory illness and which will be user friendly and easily accessible. The method proposed will take advantage of Variational Autoencoder architecture which permits the employment of training pipelines of measurable complexity and potentially small size of dataset. The results of the study concluded that, the proposed tool has an accuracy of 57 Percent which is at par with the contemporarily existing robustly supervised approaches. The present research proposes a framework by modelling the MFCC taken from sounds of healthy respiratory systems by deploying a complete Convolutional Variational Autoencoder.

2.7 Approach 7

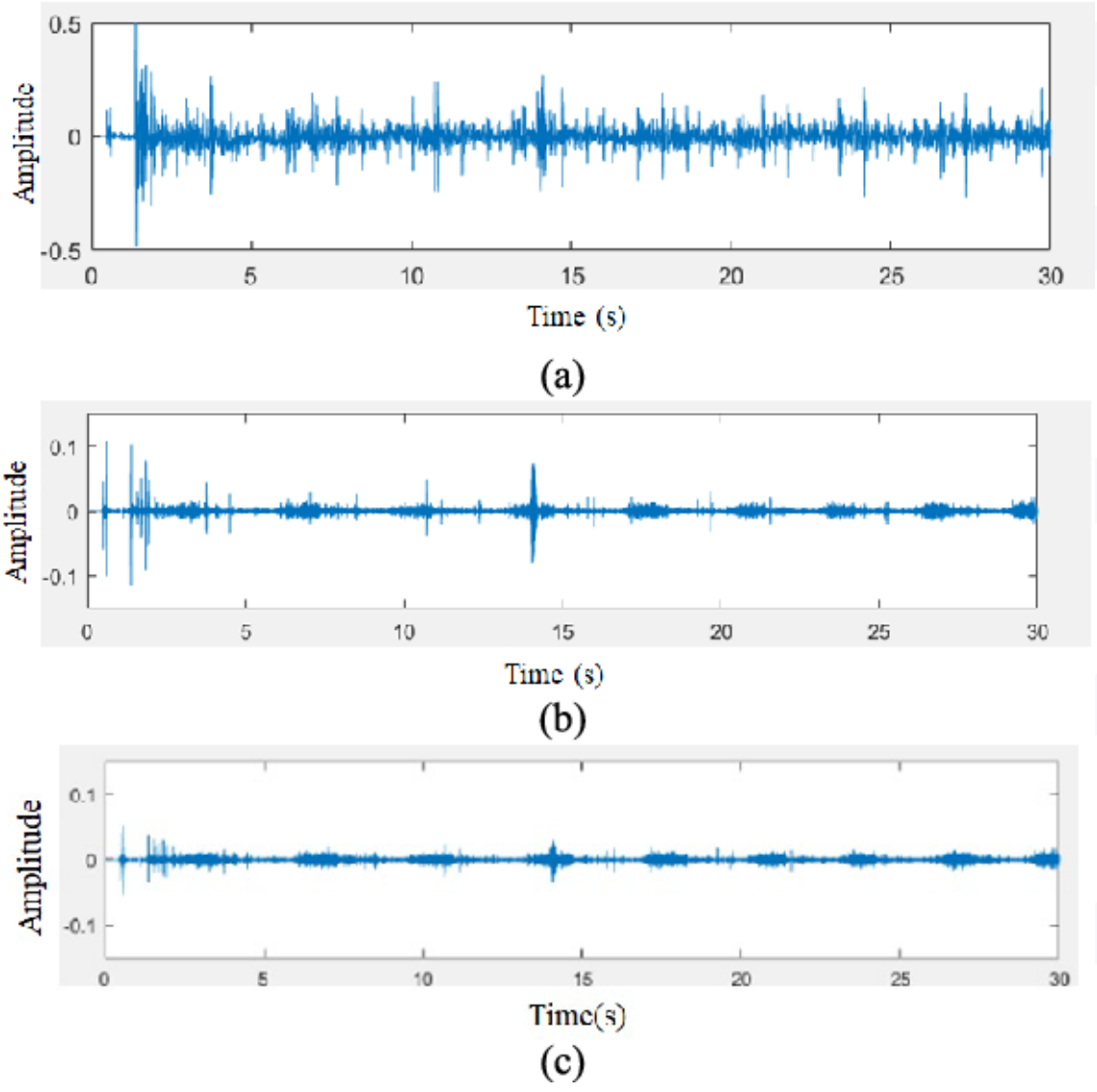


Figure 2.2: lung sound recognition algorithm based on VGGish-stacked BiGRU

Lal (2023), respiratory disease is one of the major causes behind deaths in the world. Therefore, this paper analyses the utilisation of artificial intelligence for classifying lung sounds and respiratory sounds using transfer learning. Sometimes there are different cases of missed diagnosis and lack of proper treatment for the respiratory symptoms, because the root cause is not analysed properly. Traditionally Lal, (2023), argues that the convolutional neural network (CNN) is not able to extract the temporal features of the different respiratory sounds. As a result, one of the major solutions to the problem is using an algorithm called lung sound recognition algorithm based on VGGish-stacked BiGRU, that combines the stacked bi-directional gated recurrent unit neural network. Therefore, this respiratory sound algorithm uses a feature extracted with a pretrained model using transfer learning.

In this paper Lal (2023), analyses the utilisation of transfer learning, and compares the work with utilizing the new method of Mel frequency Cepstrum. Lal, (2023), critically analyses the utilisation of transfer learning and convolutional network, for utilising the knowledge to get better understanding about Deep learning activities for recognition of respiratory sounds and patterns. Data preprocessing is used by the nonlinear time series for extracting the effective information throughout observation. Input processing of the lung sound data is also analysed utilising audio samples and the model parameters are analyse utilising the transfer learning approach. The experimental result is able to display that respiratory sound database after compiling with the scientific and providing a parameter setting, able to utilize proper classification effect by introducing the two-layer bidirectional gated recurrent with VGGish. However, the Lal, (2023) critiques about the complexity of the classification process using this network, as small datasets will not be able to provide accuracy of the model. Therefore, Lal, (2023), suggest that it is important to study how to improve accuracy of classification and reduce the over fitting, using larger databases of CT Scan images of the different patients.

2.8 Approach 8

The paper conducted by Wanasinghe et al. (2024), critically utilises a multi-featured integration for analysing respiratory sounds using extraction and classification techniques. The author mentions that detecting any kind of lung disease is important, considering the different types of respiratory elements present across the world and during the pandemic. As a result, the initial stage of lung disease detection mostly includes auscultation by the specialist. The author suggests utilising automation within the auscultation process in order to detect the lung disease with improved efficiency, using artificial intelligence to increase the accuracy of classification of

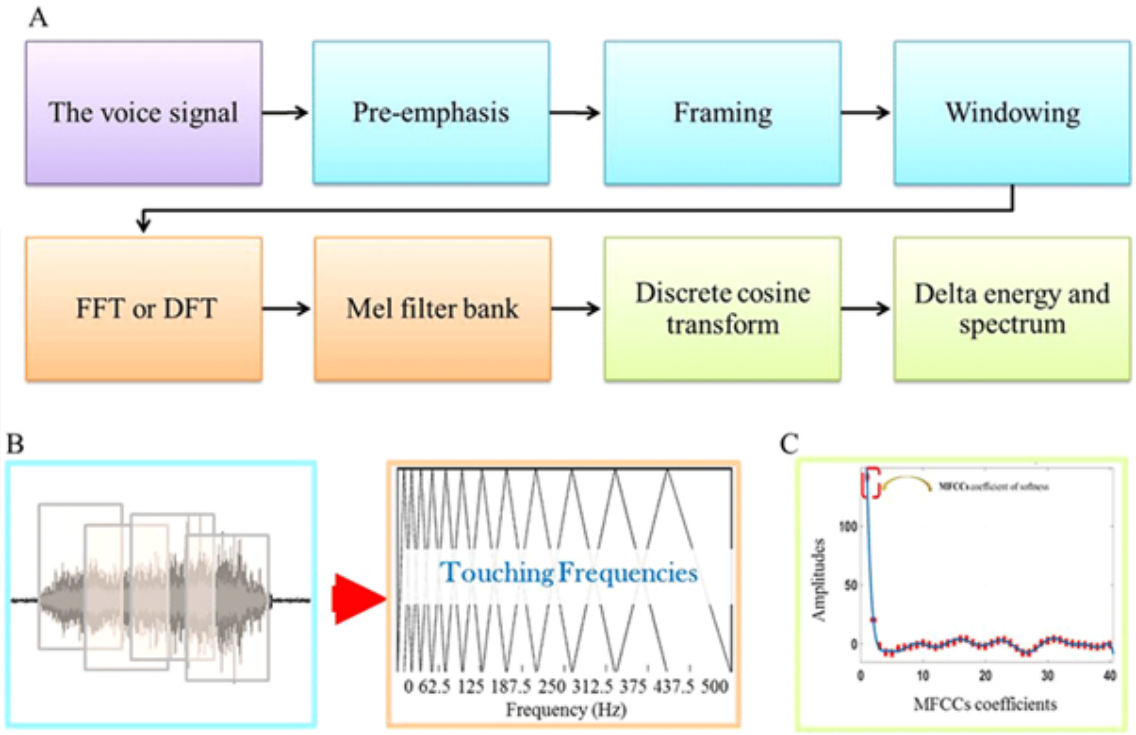


Figure 2.3: lung sound recognition algorithm Mel Frequency Cepstral Coefficient (MFCCs)

lung disease, based on the extraction features from the lung sounds. Classification and extraction used by automated devices are able to create a relationship between the different features among pulmonary disease. As a result, the paper has utilised two most important respiratory sound recordings using the lung sound dataset at Mendeley Data as well as the data et from ICBH 2017. After detailed exposition has been carried out employing the convolutional neural network, that utilises Mel spectrograms, chromogram, along with Mel Frequency Cepstral Coefficient (MFCCs), feature extraction and classification was utilised.

Wanasinghe et al. (2024)records that the highest accuracy has been able to beloved at 91.04 Percent classification for 10 classes. Therefore, the paper has been able to elaborate the explanation behind classification model prediction by utilising the explainable artificial intelligence (XAI). The paper has been able to study that the convolutional neural network has been able to classify and extract the respiratory sounds into 10 different classes by combining various types of audio specific features for enhancing the efficiency of the classification. However, Wanasinghe et al., (2024) argues about one of the major limitations of the study is that there is a lack of proper dataset and advanced featured extraction. Therefore, in the future the study will try with higher number of datasets for enhancing the efficiency of the automatic devices for classifying and extracting the respiratory sounds.

2.9 Approach 9

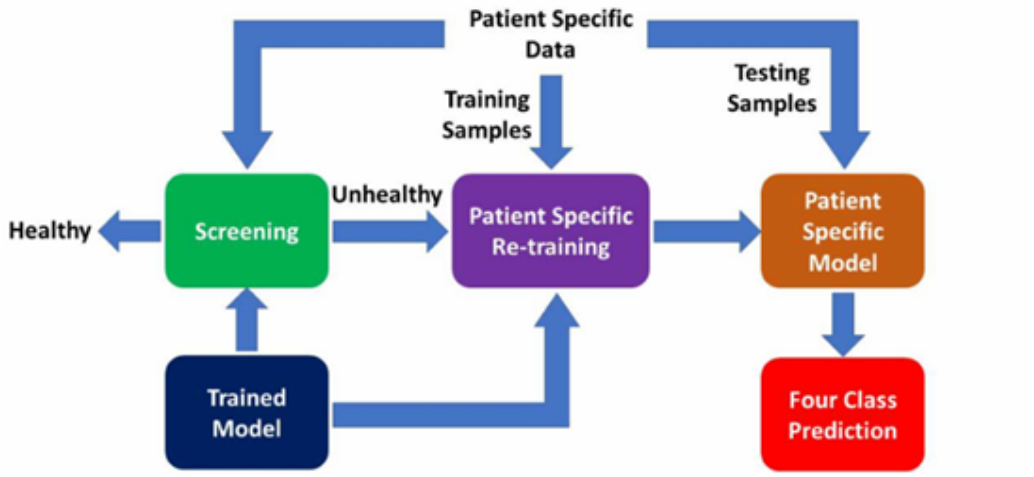


Fig. 3. Screen and model tuning strategy: First the patients are screened into healthy and unhealthy based on % of breathing cycles predicted as unhealthy. For patients predicted to be unhealthy, trained model is re-trained on patient specific data to produce patient specific model which then performs the four class prediction on breathing cycles.

Figure 2.4: CNN-RNN model strategy for lung sound classification)

According to Acharya & Basu (2020), in order to create a classification strategy and model for identifying the different respiratory sound anomalies, automatic diagnosis of pulmonary disease is necessary. Therefore, this paper proposes utilising a deep neural networking, under the convoluted neural network and RNN model classifying the different types of respiratory sounds based on the spectrograms according to the Mel frequency cepstral coefficient (MFCC). The study has implemented a patient specific model tuning strategy therefore for screening the respiratory patients and then building anomaly detection through the reliable CNN RNN model. The paper has also created a local log quantization strategy in order to reduce any kind of memory footprint within the wearable device. Hybrid deep neural network model has shown to achieve a code of 66.3% on the 4-class classification cycle as per the ICBHI scientific challenge respiratory sound database.

Upon selecting specific data from patients, the model is retrained and create the score of 71.81%. In addition, the weight quantization technique is able to achieve four times reduction in the memory cost without any kind of performance loss. Therefore, the paper has tried to propose a model which is able to provide a state-of-the-art classification system based on hybrid deep neural network as per the ICBHI dataset. The paper has also been able to prove that deep learning models are successfully important domain specific knowledge platforms for pre training breathing data and significantly is able to provide superior performance in comparison to other traditional methods. Lastly Acharya and Basu, (2020) has been able to prove that local log quantization of trained weight is able to reduce the memory requirement

in the variable device in a significant manner. Therefore, patient specific Retraining strategy is not only useful for creating and automated patient monitoring system what is able to also reduce the cost of Health Care services. However, the major limitation of this paper, critiques by Acharya and Basu, (2020) is the lack of availability of Pretrained models within audio which has prohibited the lung training time and has also prevented the author from verifying the hypothesis of the work. Therefore, in future the author plans to explore the transfer learning performance within image and audio dataset.

2.10 Approach 10

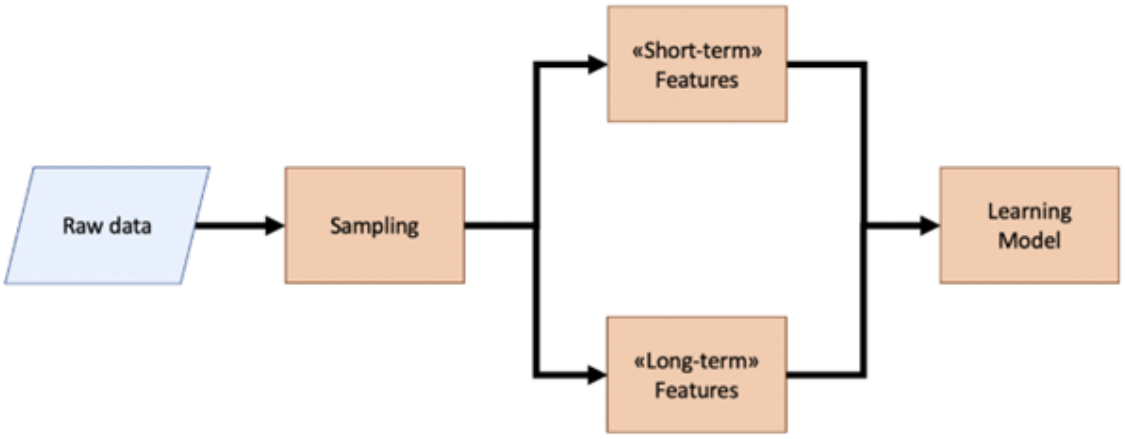


Figure 2.5: multi time scale feature for classifying respiratory sounds)

Monaco et al. (2020) in their paper has utilised multi time scale feature for classifying respiratory sounds in high accuracy. Monaco et al., (2020) has applied automatic classification of lung sounds, because it has gained attention since last few years after the pandemic has highlighted the urgent need for creating a development. Weights for the algorithm has been applied for classifying different types of respiratory sounds including wheezes, crackles and combination in a documented fashion using the ICBHI dataset. Monaco et al., (2020) suggest that the pandemic has amplified the need for developing computer assisted medicine for respiratory illness diagnosis and assessment. Therefore, the need in the scientific community has increased the ICBHI requires implementation of proper algorithms for respiratory sound classification.

Therefore, in this paper Monaco et al., (2020) applied a Framework for classification has been created utilizing short term features summarizing the properties of sounds on a time scale of 10th of a second. Secondly the paper also utilise is long term features assessing the properties of sound on the time scale of second.

The available dataset from ICBHI was cross validated using a neural network model having 126 subjects and 6895 respiratory cycles. It has reached an accuracy of 85% Precision of 80% in comparison to other models when compared through the body of literature. Therefore, this paper has also predicted the robustness of the model used by comparing different types of machine learning tools including random forest deep neural network and support vector machine. The model therefore has been suitable for not only small-scale application but also large-scale application in clinical practices, and for both multi time scale features, accurate classification is necessary for clinical interpretation. However, there is a limitation of the study, critiques by Monaco et al., (2020) is because it is a featured engineering process, applied hypothesis was created about the significant aspects of the different signals which was not utilised. Therefore, the classification can be enhanced by using general deep neural planning approach like ResNet or LSTM, although it would require future investigation.

2.11 Approach 11

Chen et al. (2019) in their paper has critically examined classification of respiratory sounds using triple classification by optimised S transform and deep residual network. They paper has been able to provide valuable data about smart diagnosis and telemedicine using non-invasive manners of detecting and diagnosing Pathology in the lungs. Wheezing is one of the most common coordinated sounds associated with COPD and asthma. On the other hand, discontinuous adventitious crackle is another most important clinically coordinated sound, which often occurs in Bronchitis or pneumonia patient. In the past the conventional systems had constraint feature extraction methods as well as contained artefacts, for which accuracy and reliability of the classification related to the normal sound crackling or wheezing was not accurate. As a result of which are novel method has been identified in this paper by Chen et al., (2019) which utilise is an optimised S-transform (OST) and deep residual network (ResNets) using system identification. The spectrogram of OST is first processed and then it is rescaled using the ResNets. Results by Chen et al., (2019) critically show that utilizing ResNets along with OST as a multi classification for respiratory sound is able to provide high sensitivity accuracy and specificity up to 96.27% 98.79% and 100% respectively. As a result, triple classification of respiratory sound out performs other deep learning based assembling CNN by almost 3.23% and artificial neural network by 4.63%.

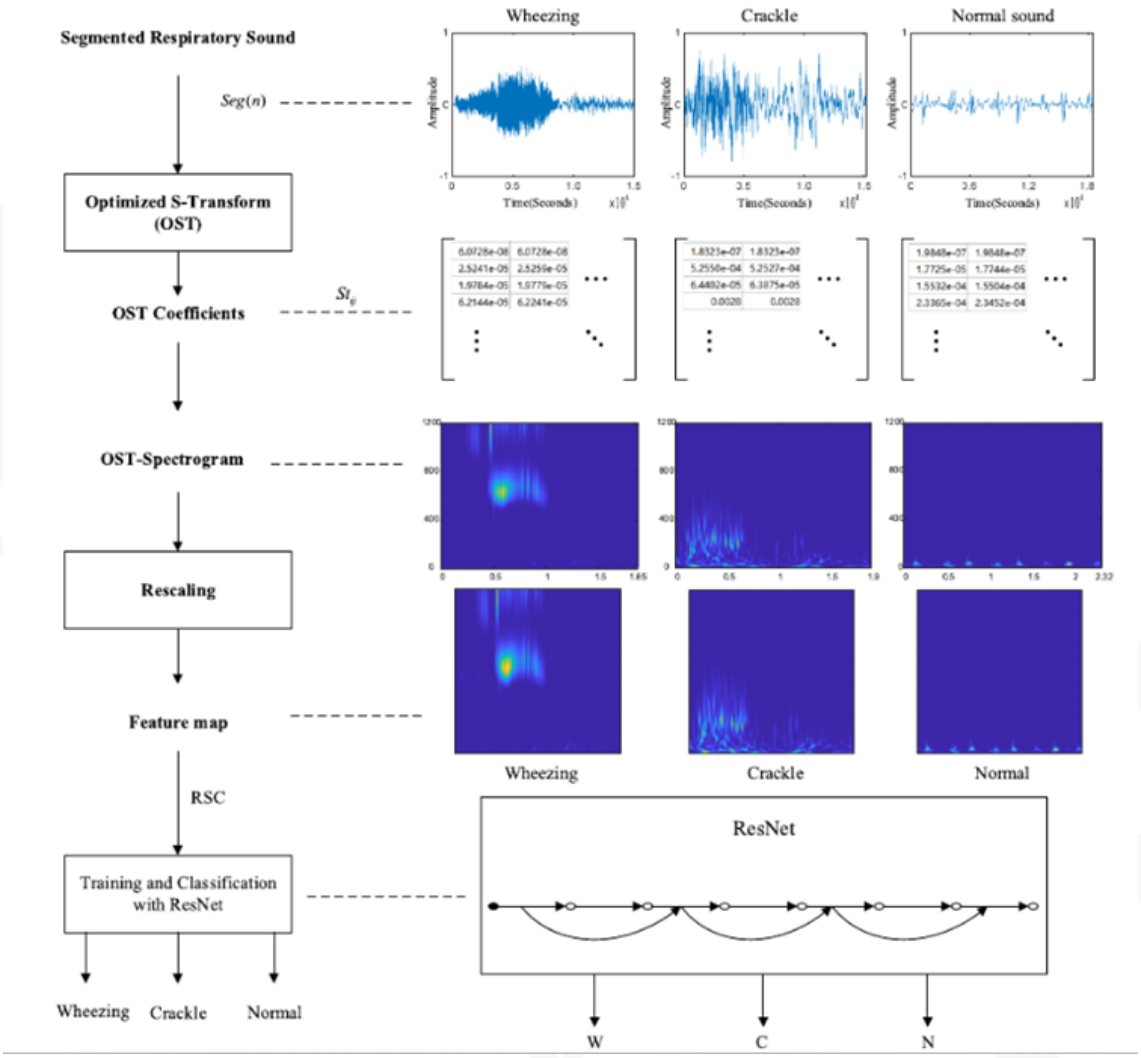


Figure 2.6: Flowchart of OST ResNets sound classification

2.12 Literature Review Summary

Name	Methodology	Strength	Limitations
Bardou et al (2018)	An automatic recognition system has been developed for the reduction of the problem associated with the auscultation method. Convolutional Neural Network design (CNN) also applied for the statistical calculations associated with this research.	The generation of automatic recognition systems in this research is effective for the reduction of limited areas of auscultation technique.	The auscultation method is essentially very basic and restricted. If doctors and physicians are not properly trained to perform this procedure, they will likely make the wrong diagnosis. This might prompt the wrong medicines of the patient and that would prompt actual mischief of the patient.
Pramono et al (2019)	Tonality index and MFCC, two individualistic features, were found to be more accurate at detecting wheezing sounds in the study. There are 105 features evaluated, all of which are focused on automatically detecting wheezing sounds made by breathing.	Managing these respiratory conditions would greatly benefit from the use of an automatic wheezing detector that could automatically monitor wheezing sounds all the time. The performance of classification may also be improved by employing more complex classifiers like vector machines or artificial neural networks	When compared to simpler time domain features, the computational requirements are more complex and higher.
Pham et al (2020)	Investigation and broad examination on the metadata set of ICBHI, the best model proposed utilizes commencement 01 engineering and the cutting edge frameworks are vanquished by Gammatone gram and Scalogram in the two undertakings which lays out and approves the effectiveness of profound learning to analyze respiratory illnesses at a beginning phase.	A multi-spectrogram ensemble and an inception-based network that are used to predict respiratory and lung diseases and anomalies. The current study proposes an inception-based neural network for diagnosing lung diseases based on respiratory sound input.	This is remitted as a front-end highlight extraction. From that point forward, the changed spectrograms are introduced into the proposed network which is distinguished as back-end categorization to recognize assuming that the patients are impacted by lung illnesses or respiratory problems.

Name	Methodology	Strength	Limitations
Aykanat et al (2017)	A low-cost, easy-to-use electronic stethoscope has been developed that makes it easier to store and record respiratory sounds recorded on a computer. This device was used to record 17,930 lungs from 1630 study participants.	Using a variety of machine learning algorithms, this study investigates various non-invasive methods for classifying breathing sounds recorded with an electronic stethoscope and audio recording software.	The use of MFCC features in conjunction with the Support vector machine algorithm is a complex method for categorizing audio that is generally accepted. The Convolutional Neural Network algorithm's benchmark was established using its findings.
Zhang et al (2023)	Four distinct machine learning algorithms were trained on a data set consisting of 920 respiratory audio files as part of the study's methodology. Computerized stethoscopes were utilized to record these lung sounds which made the Respiratory Sound Information base.	A new method has been used in this research to treat and improve conditions like chronic obstructive pulmonary disease, pneumonia, bronchiolitis, and upper respiratory tract infections by using respiratory audio recordings to improve diagnosis precision.	Long Short-Term Memory, Convolutional Neural Network, CNN ensembles with unidirectional long-term memory, and Convolutional Neural Network ensembles with bidirectional long-term memory
Cozzatti et al (2022)	The method relies on a machine learning tool and is poorly supervised and would assist in advising patients of potential respiratory issues. The proposed approach will make use of the Variational Autoencoder architecture, which makes it possible to use training pipelines with a measurable amount of complexity and possibly a small dataset.	The proposed tool matches the current robustly supervised approaches in terms of accuracy, which is 57%. It is vital for study and find helpful preventive practices which can give fundamental help with further developing the ailment of the patient.	The proposed model in the current review has had the option to give advancement brings about quiet autonomous trial convention in spite of the fact that it is ineffectively directed.

Name	Methodology	Strength	Limitations
Lal, (2023)	Data preprocessing using lung sound recognition algorithm based on VGGish-stacked BiGRU	Experimental result is able to display that respiratory sound database after compiling with the scientific and providing a parameter setting	Complexity of the classification process using this network, as small datasets will not be able to provide accuracy of the model
Wanasinghe et al., (2024)	Multi-featured integration for analysing respiratory sounds using extraction and classification techniques by Mel Frequency Cepstral Coefficient (MFCCs), feature extraction and classification was utilised.	The paper has been able to study that the convolutional neural network has been able to classify and extract the respiratory sounds into 10 different classes by combining various types of audio specific features for enhancing the efficiency of the classification.	One of the major limitations of the study is that there is a lack of proper dataset and advanced featured extraction.
Acharya and Basu, (2020)	The hybrid deep neural network, RNN model, and convoluted neural network are the three models of deep neural networks that are proposed to be used in this paper.	According to the study, there is a significant reduction in the memory requirement in the variable device when local log quantization of trained weight is used.	The author has been unable to verify the work's hypothesis and has prohibited lung training time due to the unavailability of pretrained models within audio.

Name	Methodology	Strength	Limitations
Monaco et al., (2020)	Multi time scale feature for classifying respiratory sounds in high accuracy automatic classification of lung sounds	Framework for classification has been created utilizing short term features summarizing the properties of sounds on a time scale of 10th of a second.	Applied hypothesis was created about the significant aspects of the different signals which was not utilised.
Chen et al., (2019)	Classification of respiratory sounds using triple classification by optimised S transform and deep residual network	Utilises an optimised S-transform (OST) and deep residual network (ResNets) using system identification	Utilizing ResNets along with OST as a multi classification for respiratory sound is able to provide high sensitivity accuracy and specificity up to 96.27% 98.79% and 100% respectively.

Table 2.1: Summary of Methods, Strengths, and Limitations

Chapter 3

Methodology

3.1 Data collection

The ICBHI (International Conference on Biomedical and Health Informatics) 2017 dataset is a collection of respiratory sound recordings intended to facilitate research on automated lung disease analysis. The dataset includes sound recordings from 126 patients of various genders, ages, and different respiratory conditions. The breathing sounds were collected using various electronic stethoscopes which are useful to capture the slightest details from these breathing sounds. The dataset consists of 920 audio recordings with a mixture of normal and abnormal respiratory sounds. Each recording differs in length but the total duration of all the recordings is approximately 5.5 hours. The abnormal sounds include wheezes which are high-pitched sounds and crackles which are popping sounds linked with different respiratory conditions. The dataset is a multi-class classification problem with classes such as Healthy, COPD, Pneumonia, Bronchitis, URTI, and Asthma.

3.2 Exploratory Data Analysis

The First step would be to import the CSV file containing the patient ID and their diagnosis with the help of the panda's library. Then using the unique function, it was found that there were 8 unique labels in the dataset which are Healthy, URTI, Asthma, COPD, LRTI, Bronchiectasis, Pneumonia, and Bronchiolitis.

The distribution of all the classes in the dataset is highly unbalanced as can be seen in the above figures. More than half the classes belong to the COPD class which would lead to extreme bias while training the Deep Learning model. This happens because the model comes across more samples from the majority class during the training stage which would lead to the model making predictions favoring the majority class due to insufficient samples from the minority class. The model

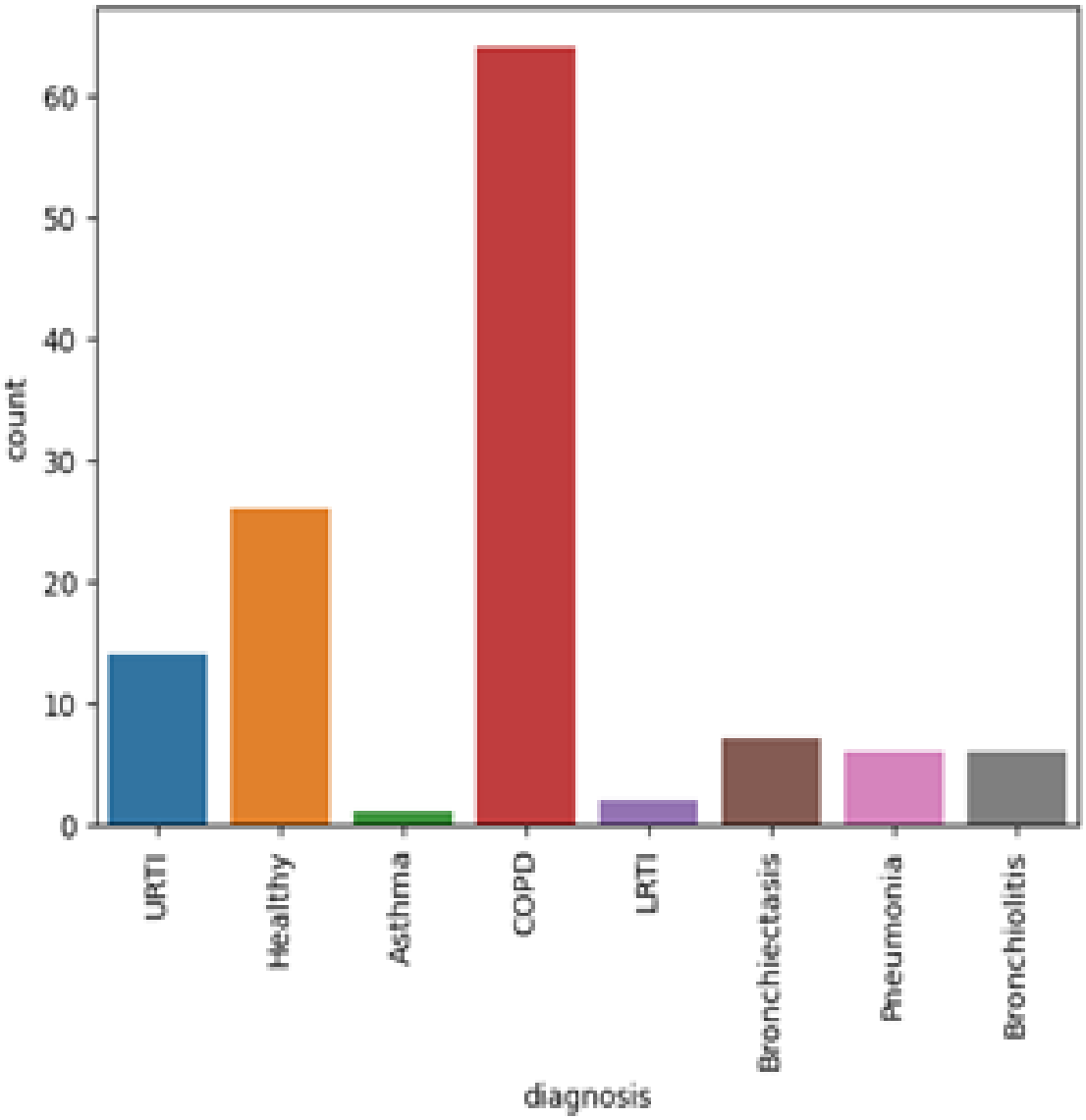


Figure 3.1: Distribution of Data

will not be able to learn patterns from the minority classes which would again lead to poor performance and misleading accuracy. Hence, all the samples in the dataset must be balanced for a robust model. Also, it was found that LRTI had 2 samples and Asthma had just 1 sample. The best course of action would be to drop these classes as it is often recommended to drop classes with a very tiny number of samples because it leads to severe class imbalance.

3.3 Data Augmentation

Data augmentation is an approach to increase the number of samples in a dataset by applying various transformations to the existing data. This is usually done when

the database is highly unbalanced. Data augmentation methods are applied for better generalization and a better-performing model. The type of data augmentation technique used differs from the nature of the dataset. As the ICBHI dataset consists of audio recordings data augmentation techniques suitable for audio files are used. The methods used are:

- Adding noise to existing data: Random noise is added to the existing data. It is done using the random function in the NumPy library. Noise is added to the existing sound recording so that the model would be able to discriminate between real signals and disturbances that might happen in real-life scenarios.

- Shifting the existing data: The existing audio data is shifted further or backward in time using the roll function from the NumPy library. A positive value is entered in the parameter of the roll function to move the audio sample further and a negative value is entered to move the sample earlier in time.

- Stretching the existing data: The duration of the audio samples is changed by slowing them down or speeding them up. This has to be done without changing the pitch of the sound. This can be done with the help time_stretch function from the librosa library which is a powerful toolkit used for audio data. The audio samples can be stretched by passing a rate bigger than 1 in the parameter section of the time_stretch function while a rate lesser than 1 will compress the audio sample.

- Shifting the pitch of the existing data: The pitch_shift function from the librosa library is used to carry out this kind of augmentation on the audio samples. The objective of this function is to change the pitch of the audio sample by a specified number of semitones. It is done by altering the frequency element of the audio signal.

3.4 Feature Extraction

Feature extraction is a method of converting raw data to features that can be used by Machine Learning Algorithms to perform tasks like Regression, Classification, or clustering. Raw data like Audio recordings consist of a lot of information that isn't useful to Machine Learning models directly. Feature extraction reduces the amount of data that is fed to the model by limiting it to its most informative parts. This helps in increasing the accuracy and performance of the model.

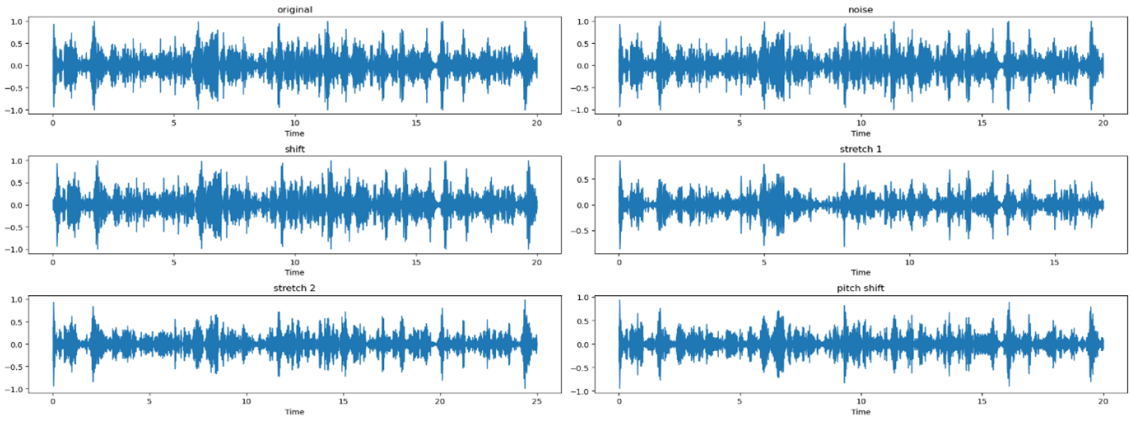
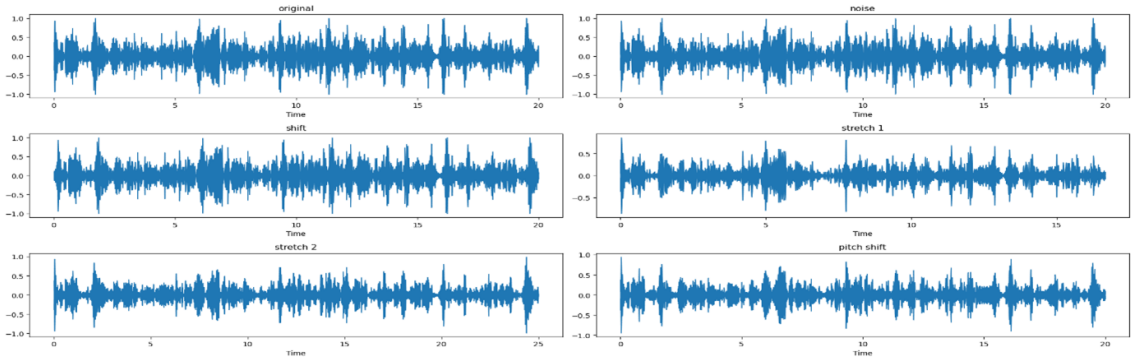


Figure 3.2: Audio Signals after Data augmentation

The first objective of feature extraction is to transform the audio samples into Mel-frequency cepstral coefficients (MFCCs) which are a type of feature when dealing with audio data. Essentially MFCCs are a set of coefficients that denote the short-term power spectrum of an audio recording on the Mel scale. A Cepstrum is calculated by applying inverse Fourier transformation to a log amplitude spectrum. MFCCs are calculated in a series of steps:

- Firstly, The audio signals are segregated into small overlapping frames. The length of these frames varies from 20-40 milliseconds.
- To reduce spectral leakage during a fast Fourier transform, a window function is multiplied with every frame to shrink the discontinuities at the edge of the frame.
- A Fast Fourier Transform is implemented on each frame to convert the time domain signal to a frequency domain signal.
- The power spectrum is passed through a series of filters as per the Mel scale. This is crucial as it highlights frequencies imperative for human hearing.
- After filtering, the logarithm of the filter bank energies is calculated.
- Lastly, The Discrete Fourier Transformation is implemented on log filter bank energies which provides us with Mel Frequency Cepstral Coefficients (MFCCs)

The Discrete Fourier transformation is used over the Inverse fast Fourier transformation while calculating MFCCs because It gives a better description of the audio signals. Most of the information is usually in the first few coefficients approximately (12-13) and the higher coefficients might represent noise. This helps in capturing important characteristics from the data.



The above picture displays the waveforms of the original audio and transformed audio after using data augmentation on the audio sample. This is done with the help of the waveshow function from Librosa library Populate accordingly. Explain all the necessary details, parameters, experimental setups, and components in detail.

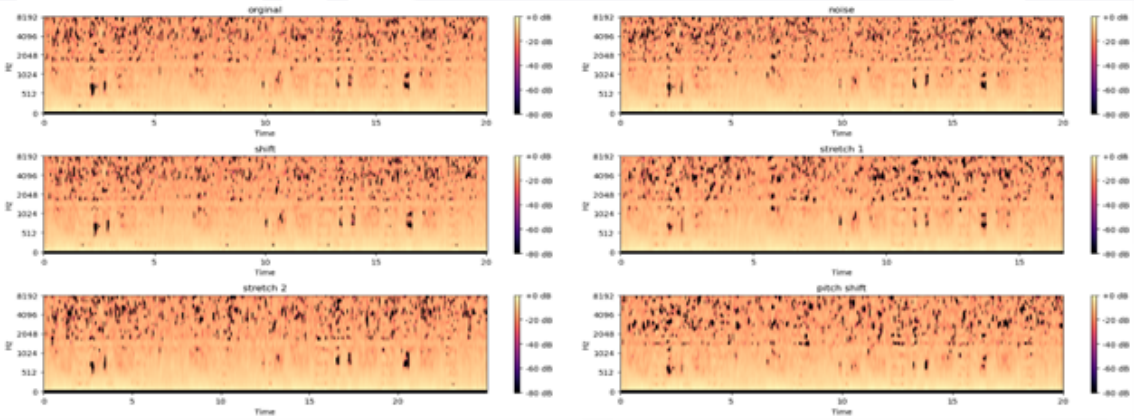


Figure 3.3: MFCCs using specshow

In the above figure, the Specshow function is used from the librosa library to display MFCCs after implementing various data augmentation techniques.

To calculate the MFCCs for all the audio samples, A function was created for simplicity so that iterates over all the audio samples one after the other. As the dataset comprised nearly 50% of COPD cases a limitation was set for 2 samples for each COPD patient and No data augmentation was implemented on COPD samples because that would again lead to an unbalanced dataset. Also, 3 Records were dropped which belonged to LRTI and Asthma as there weren't enough samples for the model to learn patterns for these classes. Data augmentation methods such as adding noise, pitch shift, and stretching were applied to all the samples for the rest of the classes. This would ensure that the dataset becomes populated with enough samples from all the classes which would then lead to a balanced dataset.

The number of features to be extracted from the MFCCs is completely dependent

on the nature of the dataset. Here,52 features are being extracted from MFCC to extract minute details from the audio samples, especially in the lung disease classification task where the smallest frequency change can be important in making an accurate decision. The calculated MFCCs and their corresponding labels are then converted to NumPy arrays to make them compatible with Deep learning models, improve speed, and for the simplicity of manipulation.

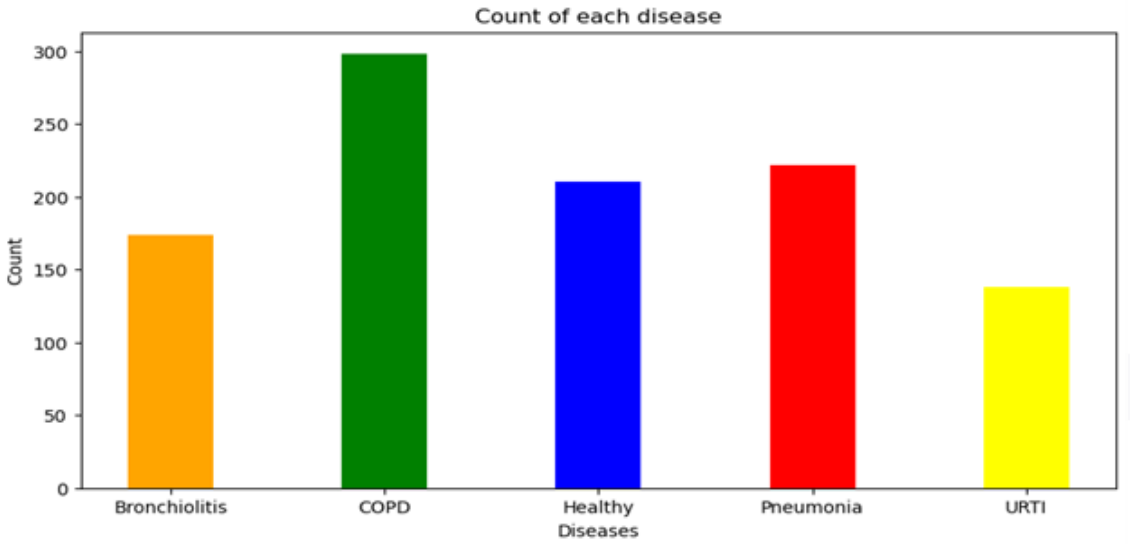


Figure 3.4: Distribution of classes after data augmentation

In the above figure, it can be seen that the dataset is more balanced than it was due to the data augmentation techniques used. The dataset is still not perfectly balanced but the Stratify method will be used while splitting the data into train and test.

3.5 One hot encoding

One hot encoding is a technique used in Machine learning that involves converting categorical labels to binary matrix representation. This is done to convert the categorical labels to a format compatible with ML/DL algorithms. The binary matrix representation of these categorical variables would be as follows:

- COPD- [1,0,0,0,0]
- Bronchiolitis- [0,1,0,0,0]
- Pneumonia - [0,0,1,0,0]
- URTI- [0,0,0,1,0]
- Healthy- [0,0,0,0,1]

3.6 Train-Validation-Test split

This is the process of dividing the dataset into 3 subsets (Train, Validation, and Test). The ML/DL model is trained on test data while the validation set checks the performance of the model on the data the model hasn't seen yet. The test set checks the final performance of the model after training and validation are completed. The model mustn't be just learning patterns from the training data but should also be able to generalize to unseen data to ensure that the model is ready to be deployed in the real world and can accurately make predictions.

The dataset is divided as follows:

- Training set – Approximately 75% of the original data
- Validation set – Approximately 17.5% of the original data
- Test set – Approximately 7.5% of the original data

The Stratify method makes sure that the division of classes among all 3 subsets is almost identical to the class distribution of the original data. This is done to ensure that all the classes are represented equally without bias. This will lead to the model being robust and better performance on the dataset.

3.7 Model building

We have built 3 different models i.e. CNN Model, the CNN with LSTM Model, and A hybrid CNN model with LSTM and Attention layer.

The convolutional Neural Network (CNN) model is built with the input shape (mfccs_features.shape[1], 1) where the first parameter takes mfcc feature extracted from each audio sample which is 52 in our case and 1 signifies each feature is a single scalar.

The below image represents the architecture of the CNN model. It can be seen that the input through three Conv1D Layer with a filter size of 64,128 and 256 respectively, three Maxpooling 1D layers with a pool size of 2, flatten layer, dense layer, dropout layer, and then the final output dense layer. The max pooling layer is used to reduce dimensionality. The activation of all the Conv1D layer and 1st Dense layer is Relu whereas the activation for the output dense layer is SoftMax as it is a multi-class classification problem. During Model compilation, Adam is chosen as the optimizer as it adjusts the learning rate and gradients to enhance the model training. The loss function is taken as categorical cross entropy as the labels are one hot encoded and the problem is a multi-class classification task.

Figure 3.5: Architecture of CNN Model

CNN model with LSTM LAYER

A combination of the CNN Model with the LSTM(Long Short Term Memory) layer is a very strong approach when working with sequential data. Sequential data refers to the data where data points are accumulated at regular intervals and the order of these data points contains useful information. Temporal dependencies are not a rare sight when working with sequential data. These dependencies refer to the event that a data point might have some connection with the previous data point in the sequence. LSTM layers can capture the temporal dependencies usually found in sequential data. Before building the model, the dimensions of the dataset are increased to a 3D array because LSTM expects a 3d array as input. The new shape would be (num_samples, num_features, 1)

The flowchart below represents the architecture of the CNN + LSTM Models. The model consists of Three Conv1D layers with a filter size of 2048,1024 and 512 respectively, Three max-pooling layers with pool sizes and stride of 2, Three layers of batch normalization, Two LSTM layers with 256 and 128 units respectively, two dense layers with 64 and 32 units respectively with activation as relu, Two dropout layers and finally a dense layer with activation as SoftMax. Batch normalization is used to normalize the output calculated by the previous layer to improve stability

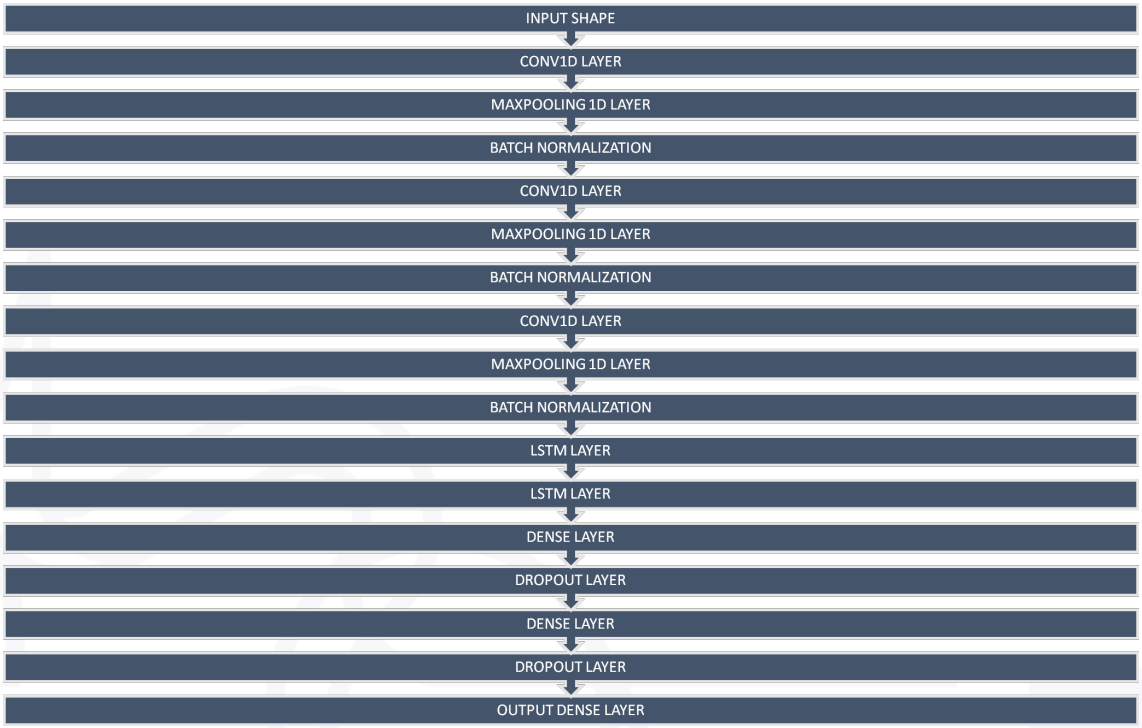


Figure 3.6: Architecture of CNN+LSTM Model

and increase speed during the model's training. Adam is chosen as the optimizer during model compilation and categorical cross entropy is taken as the loss function. Also, Early stopping is applied to stop the model from training when the validation accuracy does not improve over 20 epochs.

CNN + LSTM + Attention mechanism

The attention layer is a layer that focuses on a certain part of the sequence that is the most significant for making accurate predictions. The attention mechanism used for the model is of additive type. The attention mechanism determines a score for each data point in the sequence representing the relevance of that information. The sum of weighted input and bias is calculated before using the Tanh activation function. Hence it is called as an additive attention mechanism because it calculates the sum. Using the SoftMax function, The calculated scores are then normalized to create the attention weights which sum to 1. Attention layers are really useful for lung disease classification as they focus on the most important part of the sequence thus increasing the performance of the model.

The picture below represents the architecture of the CNN + LSTM + Attention layer model. The model has three Conv 1D Layers with a filter size of 2048,1024 and 512 respectively, Three max-pooling layers with pool size and stride of 2, Three batch normalization layers, Two LSTM Layers with 256 and 128 units, One attention layer, Two Dense layer, Two Dropout layer and lastly a dense layer with SoftMax activation. Adam is used as the optimizer while compiling the model.

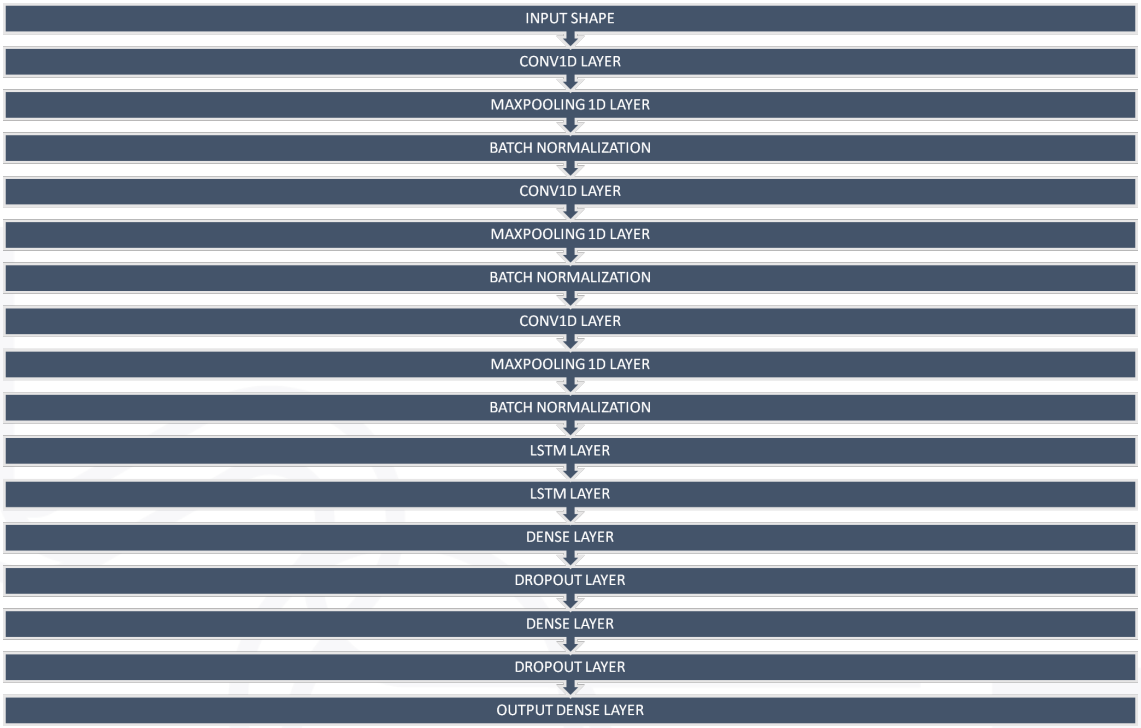


Figure 3.7: Architecture of CNN+LSTM+Attention Model

3.7.1 Evaluation metrics

- **Accuracy**: Several Evaluation metrics are considered to check the performance of each model. Accuracy is one of the important metrics that can provide a good description of the performance of the model. Accuracy is calculated as:

  (Number of correct predictions/Total number of predictions)

  However, Accuracy might not provide a good evaluation at all times. In the medical and healthcare sector, False Negatives could be very dangerous when compared to false positives. This could potentially lead to a life-or-death situation as the patient with the false negative diagnosis didn't get the healthcare that was needed at that time. Also, supposing the dataset consists of 90 records of class A and 10 records of class B and the model is predicting all the 100 records as class A. Here, The Accuracy would be 90% but the model is not good as it hasn't learned any patterns from class B

- **Recall**: Recall is also known as Sensitivity or True Positive Rate and is calculated as

  True Positives/(True Positives + False Negatives)

  Recall calculates the percentage of positive cases correctly identified by the model out of all the positive cases. A high recall is desired while dealing with

lung disease classification as they must predict all the true positives due to the nature of the classification task.

- **Precision**: Precision calculates the percentage of correctly predicted positive cases out of all the total positive predictions made by the model. Its calculated as

  True Positives/(True Positives + False Positives)

- **F1 Score**: F1 Score is used when both Precision and recall is importation in determining the performance of the model.F1 Score is the harmonic mean of Recall and Precision and It ranges between 0 to 1.It is Calculated as

  F1 Score=2×(Precision*Recall)/(Precision+Recall)

Chapter 4

Results

4.1 CNN Model results

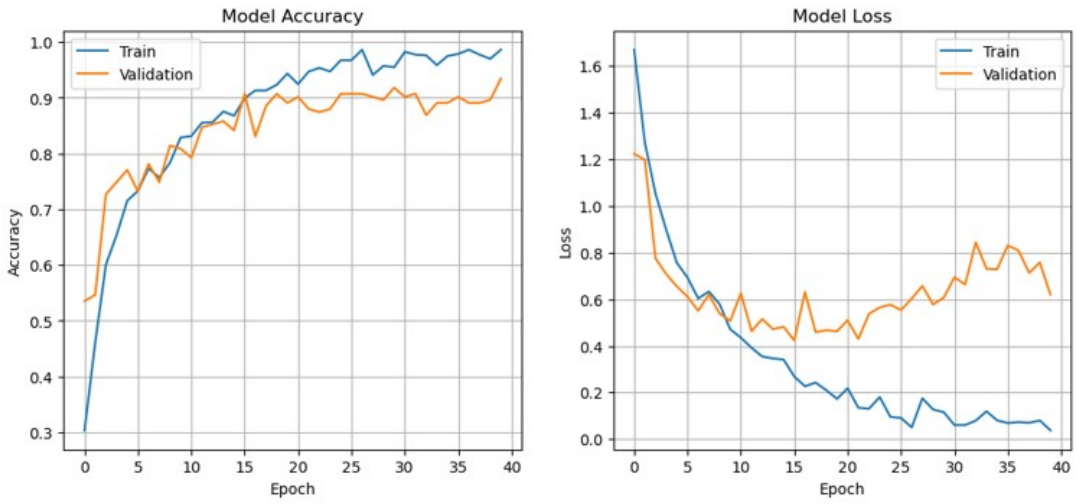


Figure 4.1: Accuracy and Loss of CNN Model

The CNN model accuracy on Training set is 98% whereas accuracy on test is 89%. The weighted average for Recall and Precision is 0.89 and 0.91 respectively and the F1 score is 0.89.

4.2 CNN+LSTM Model results

The CNN+LSTM Models accuracy on train is approximately 96% and accuracy on test is 95%. The weighted average for Recall and Precision is 0.95 each whereas F1 Score is 0.95 as well.

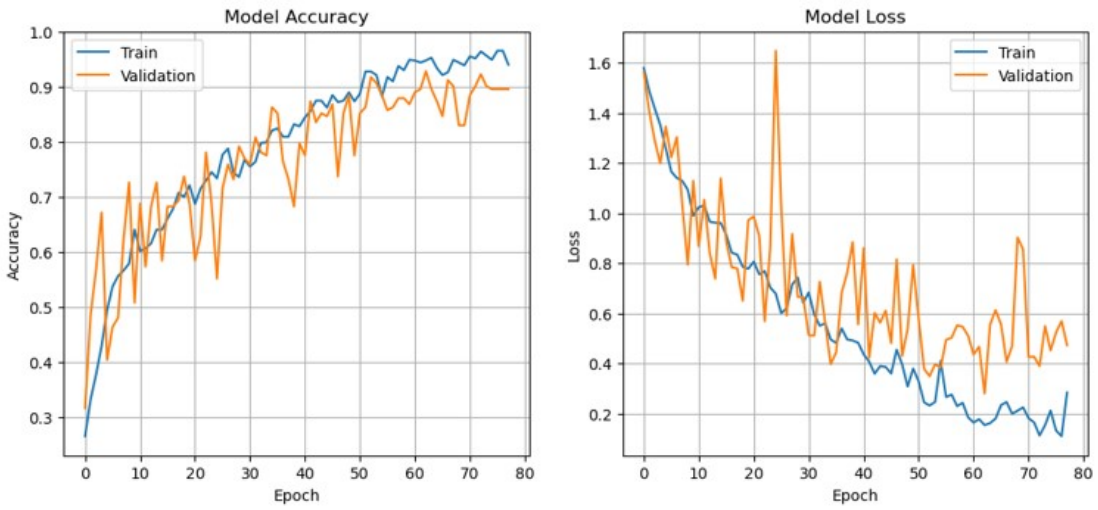


Figure 4.2: Accuracy and loss of CNN+LSTM Model

4.3 CNN+LSTM+ATTENTION Model Results

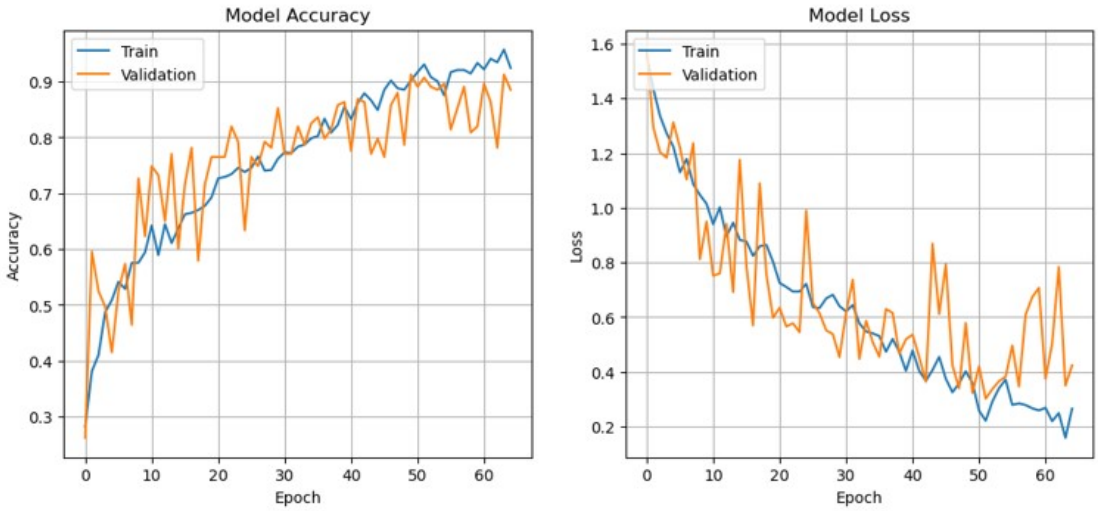


Figure 4.3: Accuracy and loss of CNN+LSTM+ATTENTION Model

The accuracy on the hybrid model on training set is around 93% whereas accuracy on test set is 90%. The weighted average for precision and recall is 0.92 and 0.91 respectively and the F1 score is 0.90.

4.4 CONFUSION MATRIX

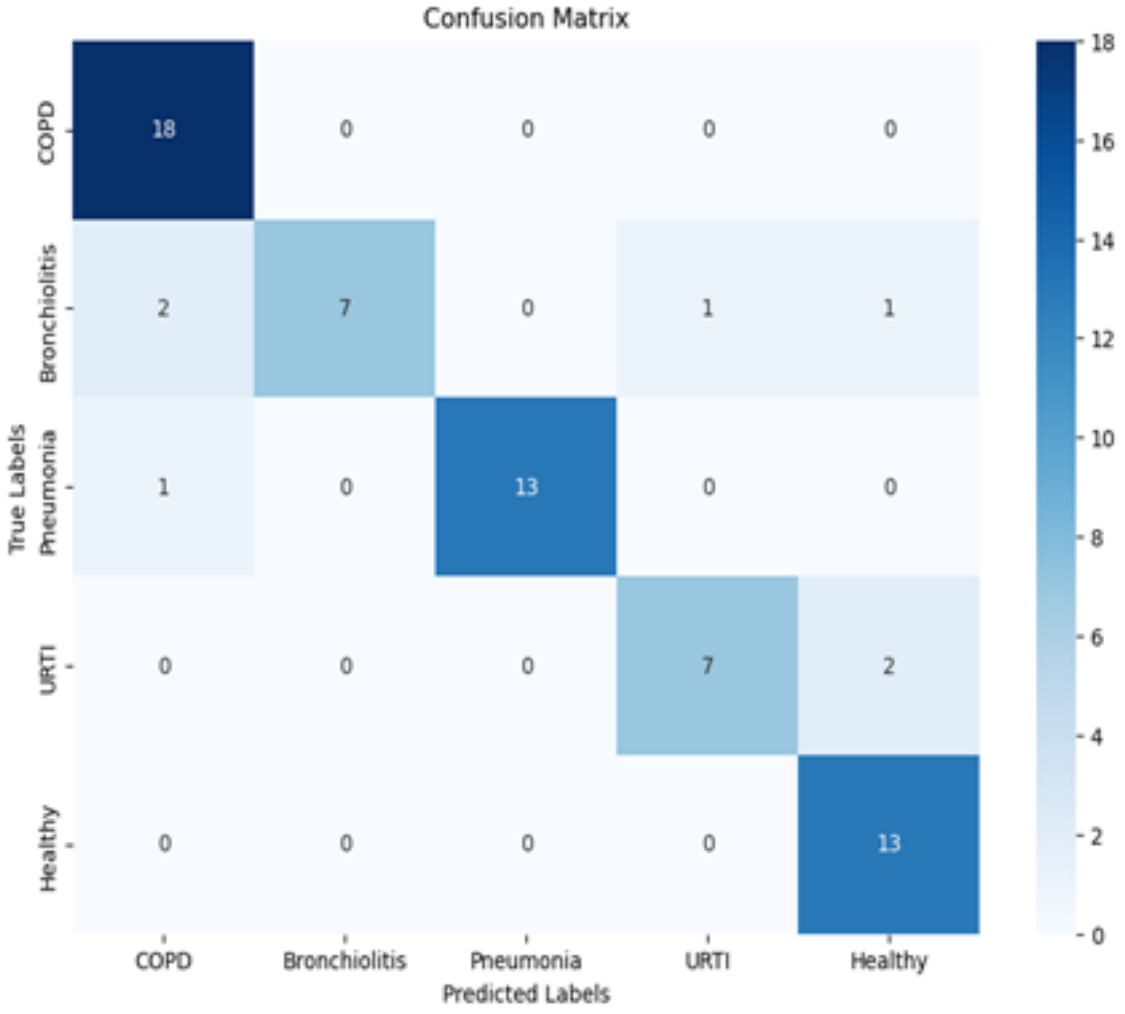


Figure 4.4: CONFUSION MATRIX OF CNN MODEL

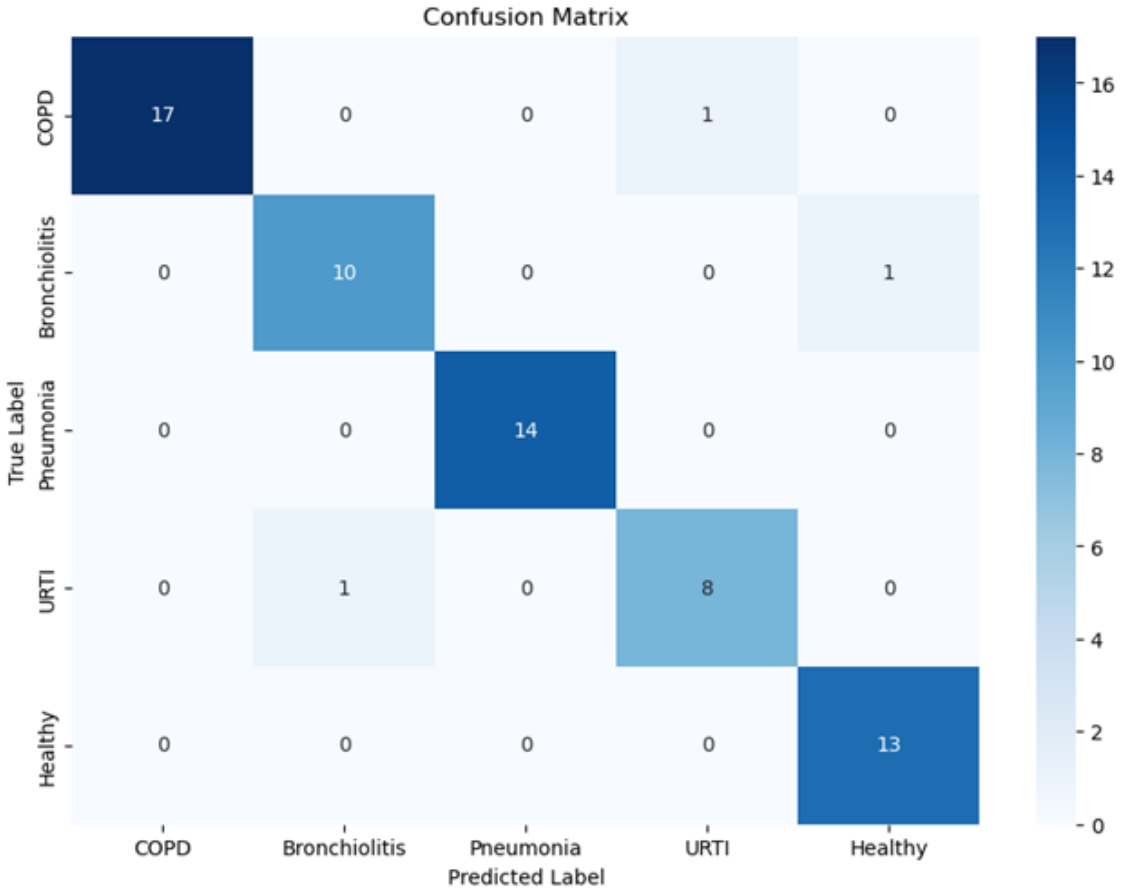


Figure 4.5: CONFUSION MATRIX OF CNN+LSTM MODEL

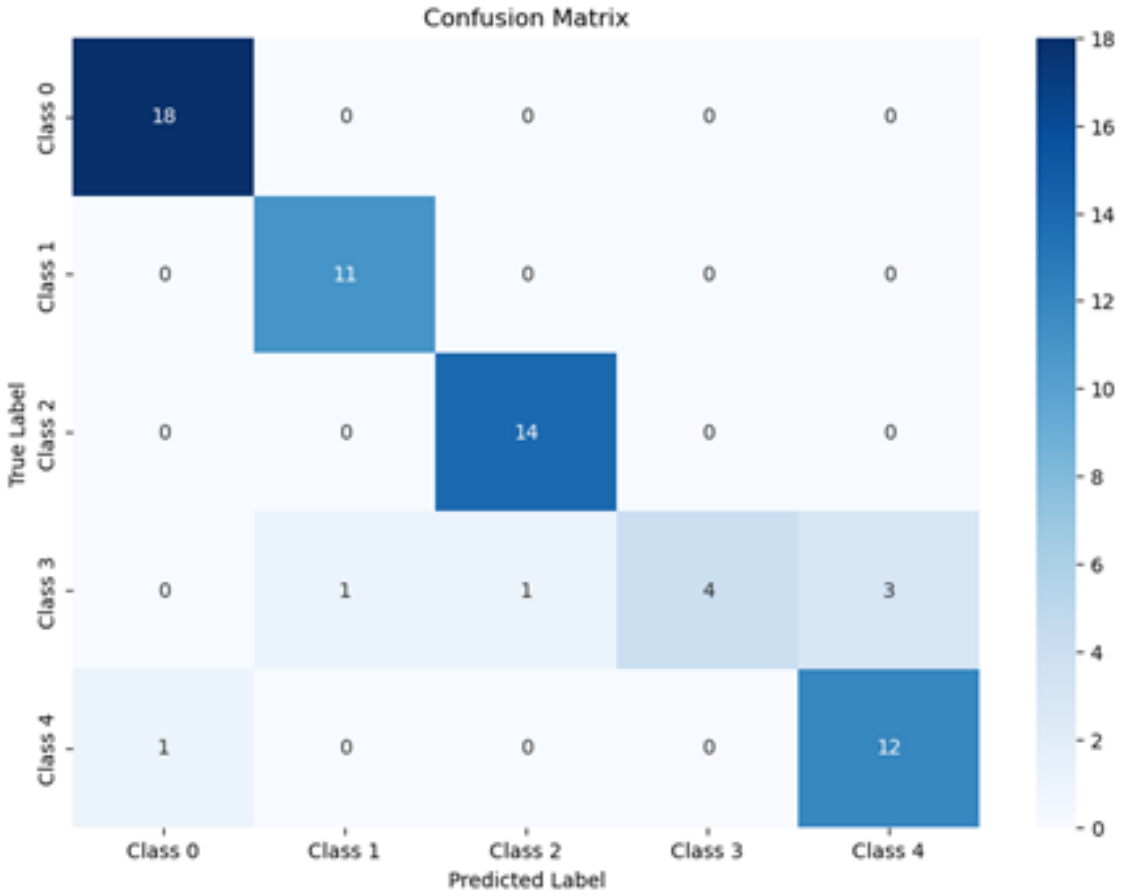


Figure 4.6: CONFUSION MATRIX OF CNN+LSTM+ATTENTION MODEL

Chapter 5

Discussion and Conclusion

The Study compares 3 different models which are CNN, CNN+LSTM and a hybrid model of CNN + LSTM + Attention mechanism. MFCCs were used as feature input for the models and the performance of these models were determines using various metrics such as accuracy, recall, precision and F1 score. The baseline CNN model had a test accuracy of 89 Percent after experimenting with different hyperparameters. The CNN + LSTM was the most promising model of the three models with a 95% test accuracy and F1 score of 0.95. The LSTM layers used are capable of capturing temporal dependencies in the data and works best on sequential data and hence perform far better than the other 2 models. The Hybrid model with attention mechanism produced results which were satisfactory with a test accuracy of 90% and a F1 Score of 0.90.The study highlights the effectiveness of integrating CNNs with LSTMs for sequential data tasks and paves the way for future research in this area.

The attention mechanism could work better if a different kind of attention mechanism is used. Additive kind of attention mechanism is used for this study with tanh activation. Perhaps, using a different kind of attention mechanism and changing the number of convolution layers and filter size could provide better results.

Risk prediction is also carried out using the SoftMax function. It calculates the probabilities for all the classes which sum up to 1. The 2nd largest probability is used for risk prediction. Supposing, A patient is deemed healthy by the model with a probability of 0.80% but he still falls under the risk of developing COPD with a probability of 0.20%. Risk assessment provides healthcare institutions with useful information so that they can manage to differentiate patients with different risk profiles.

This study also proves that using AI or Automation for medical technologies can positively impact the field of healthcare. Furthermore, it enables choosing a cutting-edge method for diagnosing and monitoring treatment processes. It also allows for

a technologically advance approach for diagnosis and medical treatments following that. It can be said that technology holds promise in this domain and Studies in this area should be highly encouraged.

Bibliography

Acharya, J. & Basu, A. (2020), 'Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning', *IEEE transactions on biomedical circuits and systems* **14**(3), 535–544.

Aykanat, M., Kılıç, Ö., Kurt, B. & Saryal, S. (2017), 'Classification of lung sounds using convolutional neural networks', *EURASIP Journal on Image and Video Processing* **2017**, 1–9.

Bardou, D., Zhang, K. & Ahmad, S. M. (2018), 'Lung sounds classification using convolutional neural networks', *Artificial intelligence in medicine* **88**, 58–69.

Chen, H., Yuan, X., Pei, Z., Li, M. & Li, J. (2019), 'Triple-classification of respiratory sounds using optimized s-transform and deep residual networks', *IEEE Access* **7**, 32845–32852.

Cozzatti, M., Simonetta, F. & Ntalampiras, S. (2022), Variational autoencoders for anomaly detection in respiratory sounds, *in* 'International Conference on Artificial Neural Networks', Springer, pp. 333–345.

Lal, K. N. (2023), 'A lung sound recognition model to diagnoses the respiratory diseases by using transfer learning', *Multimedia Tools and Applications* **82**(23), 36615–36631.

Monaco, A., Amoroso, N., Bellantuono, L., Pantaleo, E., Tangaro, S. & Bellotti, R. (2020), 'Multi-time-scale features for accurate respiratory sound classification', *Applied Sciences* **10**(23), 8606.

Pham, L., Phan, H., King, R., Mertins, A. & McLoughlin, I. (2020), 'Inception-based network and multi-spectrogram ensemble applied for predicting respiratory anomalies and lung diseases', *arXiv preprint arXiv:2012.13699* .

Pramono, R. X. A., Imtiaz, S. A. & Rodriguez-Villegas, E. (2019), 'Evaluation of features for classification of wheezes and normal respiratory sounds', *PloS one* **14**(3), e0213659.

Wanasinghe, T., Bandara, S., Madusanka, S., Meedeniya, D., Bandara, M. & de la Torre Díez, I. (2024), 'Lung sound classification with multi-feature integration utilizing lightweight cnn model', *IEEE Access* .

Zhang, P., Swaminathan, A. & Uddin, A. A. (2023), 'Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks', *Frontiers in Medicine* **10**, 1269784.