UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EFFECTS OF PREPROCESSING TECHNIQUES ON COUGH BASED MACHINE LEARNING DIAGNOSIS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

BY

TAHIR MAHMOOD
Norman, Oklahoma
2021

EFFECTS OF PREPROCESSING TECHNIQUES ON COUGH BASED MACHINE LEARNING DIAGNOSIS

A THESIS APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

Dr. Ali Imran, Chair

Dr. Samuel Cheng

Dr. Hazem Refai

Acknowledgments

I would like to express my deepest gratitude to my adviser Dr. Ali Imran who guided me throughout my Masters. His dedication and passion inspired me to appreciate the importance of research which led me in devising a practical thesis topic.

I would also like to thank the rest of the committee, Dr. Hazem Refai and Dr. Samuel Cheng for taking the time reviewing the manuscript and attending my thesis defense.

To all my friends and colleagues in AI4Networks Center who supported and helped me, I am very grateful and proud to be a part of this Center.

Lastly, I would like to dedicate this work to my beloved family who always believed in me, gave me strength and endless encouragement.

Table of Contents

List of Figures

List of Tables

Abstract

COVID-19 pandemic outbreak has taken the world by storm in the 18 months and the ramifications are by no means curtailing. The need of the hour with COVID-19 and other pulmonary diseases is a quick online diagnosis by handheld devices. In the light of these constraints, scientists are relying on audio based automated techniques since clinicians routinely use audio cues from the human body (e.g. vascular murmurs, respiration, pulse, bowel sounds etc) as markers for diagnoses of diseases or the development of ailments. Until recently, such signals have been commonly obtained during scheduled visits via manual auscultation. Research has also begun to use digital technologies to collect body sounds for cardiovascular or respiratory tests, e.g. from stethoscopes, which can then be used for automated artificial-intelligence-based analysis. An early study has promised to detect COVID-19 from cough and speech diagnostic signals. This research work describes how preprocessing techniques can enhance the performance of a methodology established over a large-scale crowd-sourced dataset of respiratory audios and in what ways preprocessing techniques ameliorate the performance of cough based diagnosis. Our findings demonstrate that a machine learning classifier will better distinguish a healthy individual from individual with cough due to bronchitis, pertussis or COVID-19 by applying preprocessing techniques. Robust results have been procured by user-based data split-up for the K-fold learning methodology. The results show a noticeable increase in the efficacy of the application of preprocessing techniques in an algorithmic pipeline. These results are rudimentary and only the tip of the iceberg of the potential of cough and audio-based machine learning. The research opens the door for enhancing the performance of lightweight machine algorithms to be comparable with their more complicated and resource-consuming counterparts. Such advancements can be of paramount significance in the practical field of application deployment.

CHAPTER 1

Introduction

1.1 Importance

Pulmonary maladies account for four of the top ten causes for death worldwide and are especially prevalent in low-income countries [1]. It is estimated that more than 1 billion people worldwide suffer from respiratory ailments [2]. Not only this but figure 1.1 gives brief statistics of deaths resulted from pulmonary diseases worldwide in 2008. Cough is a common symptom of many of these ailments, including (but not limited to) tuberculosis, lower respiratory infection, asthma, COPD, cystic fibrosis and over a hundred others [3]. Symptom tracking is an important part of the health care process at all stages including screening the general public to find new cases, assessing new patients and tracking long-term cases [4, 5].

1.2 Background

Audio sounds generated by the human body (e.g. lung, heart, bowel sounds, vascular murmurs etc) have been used by clinical researchers and physicians in identification and diagnosis of diseases. Howbeit, these signals were generally obtained during planned visits, until recently, by means of a manual auscultations. Current systems to track pulmonary ailments through cough sounds include patient self-reporting, manual cough counting and analysis, and automated cough frequency trackers [5, 6, 7]. Self-reporting of cough frequency and cough characteristics have been shown to lack the accuracy necessary for usage in clinical situations [8]. Manual cough counting, due to the unpredictable and intermittent nature of coughs, can be a very time-intensive process requiring a dedicated listener to record all

Deaths attributed to	Worldwide	WHO European Region
Ischaemic heart disease	7.3 million (12.8%)	2.40 million (24.7%)
Cerebrovascular disease	6.2 million (10.8%)	1.40 million (14.0%)
Lower respiratory infections	**3.5 million (6.1%)**	**0.23 million (2.3%)**
COPD	**3.3 million (5.8%)**	**0.25 million (2.5%)**
Diarrhoeal diseases	2.5 million (4.3%)	0.03 million (0.3%)
HIV/AIDS	1.8 million (3.1%)	0.08 million (0.8%)
Trachea/bronchus/lung cancer	**1.4 million (2.4%)**	**0.38 million (3.9%)**
Tuberculosis	**1.3 million (2.4%)**	**0.08 million (0.8%)**
Diabetes mellitus	1.3 million (2.2%)	0.17 million (1.7%)
Road traffic accidents	1.2 million (2.1%)	0.12 million (1.2%)

Fig. 1.1: The 10 most common causes of death in 2008. Source: World Health Organization (WHO) World Health Statistics 2011.

cough events over a time duration large enough to gather cough data for diagnostic purposes.

1.3 Related Studies

Research has utilized digital technology to acquire pulmonary sounds via digital stethoscopes and carry out automatic analysis on the data [9], such as for wheeze detection in asthma patients [10, 11]. Researchers have been exploiting the utility of human respiratory sounds to aid early diagnosis of several ailments such as Parkinson's disease, which is, associated with softness of speech resulting from plummeted coordination of the vocal muscles [12, 13]. Studies have shown that the detection of coronary artery disease is related to voice frequency; hardening of arteries affect voice generation [14] while pitch, vocal tone, rhythm, volume and rate correlate with maladies such as traumatic brain injury, post-traumatic stress disorder [15] and psychiatric conditions [16]. In Kosasih et al. [17], the frequencies having a range far beyond the human perception are leveraged along with wavelet analysis to pro-

Table 1.1: List of Acronyms

Acronym	Description
AI	Artificial Intelligence
NB	Naïve Bayes
SVM	Support Vector Machine
TD-DNN	Time Delay Deep Neural Network
DT	Decision Tree
LCM	Leicester Cough Monitor
MRI	Magnetic Resonance Imaging
NN	Neural Network
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
k-NN	k Nearest Neighbor
DA	Discriminant Analysis
LR	Logistic Regression Model
BGS	Bispectrum Score
LogE	Log Energy
Zcr	Zero Crossing
Kurt	Kurtosis
COPD	Chronic obstructive pulmonary disease
CNN	Convolutional Neural Network
RF	Random Forest
MLP	Multilayer Perceptrons
MFCC	Mel Frequency Cepstral Coefficents
NGS	Non-Gaussian score
FFT	Fast Fourier Transformation
HACC	Hull Automatic Cough Counter
PCA	Principal Component Analysis
CPNN	Constructive Probabilistic Neural Network
CHF	Congestive Heart Failure

cure information that is diagnostically significant for cough. DWT and CWT were applied to these samples in parallel with frequency time analysis after amplitude normalization. Various features are chosen from CWT and DWT. Lastly, regression analysis accompanying thresholding was used for classification purposes. Using two microphones, 90 or more cough sounds were explored along with demographic and clinical features from 4 patients. Recording information include Audition software with audio interface connection at a sampling rate of 192 KHz, running on a 16-bit, stereo, Windows XP laptop. Coefficients of determination of 77-82% at high frequencies ranging from 15 kHz to 90 kHz were acquired. An increase of coefficients of determination to 85-90% was observed by the combination of high and low frequencies below 15kHz. This paper is trying to procure useful information at high and low frequencies that can help in better diagnosing various diseases. Crook et al [18] was aiming at the objective cough monitoring via HACC and LCM software to detect the decrease in cough frequency during AE-COPD Convalescent. The fact that objective cough monitoring is sensitive to clinically meaningful change was demonstrated through cough frequency decrease during AE-COPD convalescence. Cough was monitored 24-hours via a hybrid system comprising of LCM software and HACC along with other tests and questionnaires. The aggregated specificity of cough counting through hybrid system was 98.2% and a sensitivity of 57.9% along with a positive predictive value of 80.9% while a negative predictive value of 94.6%.The research of objective cough monitoring through the LCM and HACC software indicates that the hybrid system has the potential to narrow down the disease, during AE-COPD convalescence. This study can be a predecessor of COPD patients home monitoring. Knocikova et al [19] aims at figuring out AB and COPD from non-infectious individuals based on the analysis of the properties of voluntary cough audio. After an initial study of tussiphonogram, wavelet transform was applied since the cough sounds were non-stationary. For better feature selection, backward selection algorithm was adopted to lessen cross-correlation among variables. Lastly, the classification among three classes were carried out by two linear functions. As far as the data set is concerned, a total of 65 cough audios were gathered from normal being while patients with AB and COPD provided 26 recordings. Among healthy subjects 15 were females and 11 were males with a median age 22 year whereas a total of 22 COPD patients involving 6 females and 16 males with median age 67 years were considered during the research work. Lastly, 17 patients were related to AB with 8 males, 9 females having a median age 32 years. The sampling frequency was kept to 11025 Hz. Cough sounds were classified using DA with a accuracy rate of about 85-90%. Such an approach of cough analysis provides an objective quantification of cough sounds with a fruitful diagnostic and prognostic value. There is a research study that aimed at using crowdsourced collection of COVID-19 related audios worldwide [20]. Most of the research techniques out there collect the data and allow the features to be extracted from these signals prior to the classification stage. A lion's share of the research work does not stress enough on the vitality of preprocessing techniques and how these techniques are corelated to the upcoming stages of feature extraction, feature engineering and classifier-based categorization within the big picture 4.1.

1.4 Contributions

To the best of our knowledge, this is the first endeavor to adopt a similar approach [20, 21] with the introduction of preprocessing techniques between the data collection and feature extraction stages considering the grand scheme of events. Not only this but how a combination of these resource light pre-processing effects lead to a better performance in stark contrast to their absence. On top of that, this research work describes our preliminary findings for the fact that how these pre-processing techniques can boost the efficiency of resource easy classifiers (those machine learning classifiers that require less time and memory to get trained and give out predictions) to their resource heavy (those machine learning classifiers that require much time and memory to get trained and give out predictions) counter parts by diminishing the effects of noise, amateur audio clipping/trimming for data generation, faulty samples and last but not the least, mic based incompatibilities. To better test the veracity of our results, user based cross validation is carried out such that no same user's cough samples are used in both training as well as testing which undermines overfitting, by any means.

We collected data from various sources. For further details, section 2.2 of Imran et al. [21] may be visited. Patients with three different types of diseases were taken into consideration including Pertussis, Bronchitis and COVID-19. A control group of normal individuals was also accounted for as the fourth category. For this research work, only cough sounds were collected and audios such as wheezing, snoring, sneezing etc were avoided altogether. More precisely, the main contributions of this research work are:

• Nullify the effect of distracting factors during sound recording that confuse the classifier.

• Initial finding of how a pre-processing technique can assist in enhancing efficacy metrices of a spectrum of classifiers from low resource dependence to high resource dependence.

• Illustration for the fact that how an average performing classifier be made comparable to better performing classifiers by the introduction of pre-processing effects for speech recognition in general and cough based diagnosis in particular.

• How a combination of pre-processing techniques (cascaded) can be more useful than the individual ones. Moreover, how their order is of any significance concerning the performance.

• Adding robustness to the results by applying the inconspicuous, yet cardinal, approach of user-based k-fold validation

• Discussion of the findings, their promise, and an illustration of some future directions in the COVID-19 progression-detection and pre-screening for our study and sound diagnosis.

1.5 Under preparation for submission

1. Preprocessing Techniques on Machine Learning Based Diagnosis via Cough
Tahir Mahmood, Usama Masood and Ali Imran

2. A Survey on Cough Sound Analysis Using Artificial Intelligence for Detection and Diagnosing Pulmonary Diseases
Aneeqa Ijaz, **Tahir Mahmood**, Muhammad Nabeel, Mayda Sajid Hashmi, Iryna Posokhova and Ali Imran

1.6 Organization

The rest of this thesis is organized as follows: Chapter 2 highlights the motivation to venture for such a research. It provides a clinical background on top of the related work in the domain of artificial intelligence. Chapter 3 is pertaining to the data collection and gathering protocols and standards. The dire demands of the present day are also addressed. In Chapter 4, the basic methodology is being discussed. The emphasis is on pre-processing techniques. Not only this, but the feature extraction methods are also taken into consideration. Moreover, classifiers are also addressed in that section.Chapter 5 deals with the preprocessing approaches that were taken into consideration including both, single techniques and cascaded ones. The validation procedure is also discussed in that chapter.Chapter 6 deal with the timing diagrams for training and prediction timings, how well the pre-processing techniques enhance the accuracy and some deep insights that can allow use to comprehend categorization from the classifier's prospect. Finally, chapter 7 gives the conclusions and future lines of research.

CHAPTER 2

Motivation and Related Work

Cough is produced as a response to irritants, trauma or diseases. This condition is characterized by gases gushing out of the mouth from the nasal and pulmonary track. Not all coughs are considered same, medically speaking, since every respiratory condition may effect the human body in a very specific manner and leads to a different type of cough such as wet cough, dry cough etc. Although a challenging task in its own right, the cough produced in response to a particular stimulus may hold the secrets to its identification and root causes. By the same token, cough sounds can play a crucial part in automated diagnosis of the diseases.

2.1 Motivation

Cough may come up as a common symptom for numerous respiratory ailments [22]. Once this symptom shows up, patients reach out to a physician or some sort of therapeutic care. Depending on its severity and nature, different diseases can be associated with the patient. In case of a severe cough, primarily caused by the upper respiratory tract infections, coughing lasts up to several weeks. More often than not, patients don't need special attention [23] but in some other cases such as that of a chronic cough, lasting more than 8 weeks, medical consultation is necessary. Early diagnosis of an unremitting cough can provide assistance in controlling the infection's outbreak at its initial stages. However, once the cough sets in, it becomes all the more painstaking to narrow down the disease at a domestic level, and its root cause, because of a long list of respiratory conditions related to the same apparent symptom [24]. Consequently, a visit to a health care center becomes all the more inevitable for accurate diagnosis of the disease. In some cases, clinical testing, with all its protocols, is conducted that take up much of the time for the diagnosis of the aliment. It may already be too late by the time patients are accurately diagnosed. Hence, a triage screening is the need of the hour so that the infected patients could take some rudimentary precautions prior to the mainstream clinical and medical diagnostic procedures.

2.2 Related Work

As a possible indicator of health and behaviour, researchers have long recognised the significance of cough. Previous studies have shown that unique latent features are associated with distinct respiratory syndromes. Such distinct features can be obtained by applying appropriate mathematical transformations and signal processing techniques over the cough sounds. These features can then be utilized to train a sophisticated AI engine for the preliminary diagnosis solely based on cough. To detect sounds from the lungs or heart via stethoscopes, purpose-built external microphone recorders have been used. Such devices needs a specialist to use and interpret although recent trends suggest that such devices are better at the hands of a common man as well due to the ease of the usage in technology. Albeit expansive yet techniques such as MRI and sonography are becoming easier to use making the disease analysis all the more swift. Moreover, trends in commodity devices, app development and automated audio interpretation has shown the potential to further the idea to offer sound as a less-expansive alternative that better deals with the diagnosis of diseases. Microphones on commodity devices including smartphones and wearables have used sound to perform better speech analysis techniques. In [25], the audio from the microphone was used to comprehend the user context (spoken words by people, objects recognized or text written on signs in the environment) and this information is accumulated to make up a view of the ambience of places around a city. In Emotionsense [26], Gaussian mixture models are utilized for detecting users' emotion in-the wild through the microphone acting as a sensor. In [27], authors analyze audios emitted while the user is asleep to determine sleep apnea episodes. Similar research works have also utilized sounds to distinguish wheezing and asthma [10, 11]. Machine learning methods have been devised to identify and diagnose respiratory ailments from audios [9] and more particularly cough. [28] uses CNNs to detect cough in the presence of background audio, and diagnose three potential maladies namely: bronchitis, pertussis and bronchiolitis based on their unique audio characteristics. Clinical work has focused on using voice analysis for specific maladies. For instance, in Parkinson's disease, laryngograph equipment and microphone have been used together to detect the softness of speech as a consequence from lack of coordination over the vocal muscles[12, 13]. Voice features are also utilized to diagnose bipolar disorder [16]. It also correlates pitch, tone, rhythm, volume and rate with signs of invisible conditions like post traumatic stress disorder, traumatic brain injury and depression[15]. Voice frequency has been associated with the coronary artery disease caused by the hardening of the arteries that may affect voice production [14]. Companies such as the Mayo Clinic and Israeli-based Beyond Verbal have expressed in press releases that they are piloting these methodologies. The outbreak of COVID-19 has caused researchers to see whether cough audio can be of any good in diagnosing the corona virus or not [29]. In [30], digital stethoscope data from lung auscultation is utilized as a diagnostic signal for COVID-19. In [21], a research work for the detection of cough audios related to COVID-19 is elucidated using a cohort of 48 COVID-19 patients that are tested positive versus other pathological causes of cough, on which an ensemble model is trained. In [31], speech recordings gathered from COVID-19 patients are analyzed to categorize the health state of patients automatically from four aspects namely sleep quality, severity of illness, anxiety and fatigue. Quatieri et al. [32] showed that variations in vocal patterns can be a potential bio-marker for COVID-19.[20]

used an entirely crowd-sourced dataset to distiguish COVID-19 patients from the rest of the lot, for which the ground truth is considered to be what the users state is in terms of symptoms and COVID-19 testing status. Not only this but there were challenges that they overcame for data coming from different microphones and possibly in very different environments. The closest to our work is [20], from which our approach differs in two significant ways. Firstly, their data is gathered from crowd sources whereas in comparison our data is collected in a controlled setting, which means that we have a more challenging task at our hands to come up with good performance metrices with lesser amount of data. Secondly, they used an end to end deep learning model on their data set consisting of around 100 samples; deep learning models often over-fit on such very small data-sets, so we chose a different strategy. We use simple machine learning models such as SVM with various features (handcrafted and obtained through transfer learning) and data augmentation to overcome such issues. Other crowd-sourced approaches of this kind are starting to emerge: in [33] a web form to gather sound data is presented, which collected about 570 samples but does not report any COVID-19 detection analysis.

In Swarnkar et al [34], an objective method of automated analysis of accessing acute asthma was developed via cough audios. Several features were extracted for each fragment, that is, the segmentation of each cough audio in three non-overlapping fragments. For every cough sound, a total of 76 features were extracted and 13 wavelet features were also procured. Leave-one-out cross-validation methodology was employed for training LR algorithm in a variety of parameter settings. Lastly, information extracted from the respiratory rate was merged with cough features. Along with this, RR was transformed into BI for LR models.

In [35], authors made the pioneering effort for the categorization of Wet and Dry cough using LR based model. The extracted features were LogE, Zcr, BGS, NGS, Kurt, formants frequencies and MFCC. The recording of the data was done by a low-noise microphone for 4-6 hours by a hypercardiod beam pattern. The distance varied from 40 cm to 70 cm between patient and the microphone as a consequence of movements by the patient while recording the data.

Parker et al [36] utilized various machine learning classifiers to categorize cough sounds into pertussis and non-pertussis. Cough samples were gathered belonging to patients with pertussis, croup and other similar maladies by a series of recordings. After manual segmentation, features such as MFCC were obtained to classify using three different classifiers including k-NN, RF with 200 trees and NN.

In Windmon et al [37] , TussisWatch, a smart-phone based system for early identification of CHF or COPD was proposed to process and record cough episodes. This approach can be broken down into five key steps :(1) filtering noise; (2) using domain expertise to chunk down every cough episode into multiple segments that are indicative of disease or of a non-infective being; (3) for each cough segment, narrowing down a limited number of audio features; (4) nullifying inherent biases as a consequence of sample size differences; and lastly, (5) designing a two-level classification scheme based on the idea of processing a recorded cough segment at two stages using classifiers such as RF,k-NN, SVM and NB.

Subirana et al [38] aimed at diagnosing COVID-19, using cough audio samples by applying transfer learning. As a consequence of scanty nature COVID-19 data, output layers of pre-trained ResNet5032 and DenseNet20131 architectures were trained on speech data set. These layers were taken as features to further train on the target domain of scarcely available COVID-19 audio. Evaluation was performaned on machine learning classification algorithms such as RF, SVM, k-NN and LR over a set of 5 cross validation test splits.Finally PCA was applied to come up with a visual differentiation between COVID-19 and non-infectious coughs after. This paper also focused on the exchanging notes and data from medical and engineering fields along with a worldwide template “Sigma”.

A systematic review that provides the detailed classification/characteristics of lung audios is elaborated in [39]. The authors also discuss the machine learning approaches being used for lung sound anomaly analysis. In [40], hybrid perceptual and cepstral feature set (PerCepD) is proposed for automatic breath sound detection. The research work accentuates the fact that respiratory diseases such as bronchitis, pneumonia, flu etc can be significantly understood by using breath sounds. The authors utilized ANN and SVM models for the categorization that showed high accuracy.

Based on the idea that obstruction in lung airways may cause sounds like crackles, wheezes, stridor etc, Bokov et al. [41] performed a study for the identification of wheezing sound in infants of 20 months. The data was recorded via a smartphone. A sensitivity of 71.4% and specificity of 88.9% was achieved in a two phase algorithm that performed signal analysis along with SVM training.

Specifically for the identification of auscultated sounds for the tracheal–bronchial breath audios, the authors performed a study in [42]. They compare the performance of CPNN against SVM and for the classification of a wide range of tracheal–bronchial breath sounds. CPNN achieved a 97.8% accuracy, while SVM attained 96.2% and MLP has 77.8% accuracy.

CHAPTER 3

Data Collection

In this section, we shall be discussing the data collection and preparation methods. Details with regards to data storage and utilization is also provided.

3.1 Quality Assurance

The data was gathered under controlled environments. The sources of data were online as well as applications on hand held devices. Once the samples were collected, the time duration of samples was set to two by trimming them. The software primarily used for that purpose was Audacity. On top of that, extraneous audios were minimized by removing the chunks from the clipped audios. All these protocols were taken into consideration for the fact that better accuracy and performance matrices were to be maximized under clinical conditions. The work done by Brown et al [20] had gone for the option of crowd sourcing data with less treated data with data augmentation. The data will be more in quantity but, in our approach, the quality of data is by no means questionable albeit less in quantity. Which leads to the fact that the results produce by this research are better in performance, robust and reliable.

3.2 Facts and Figures

The data was the similar to the one described in [21] gathered by the AI4Networks lab members at the University of Oklahoma, Tulsa related to the three diseases and one normal (non-infectious) category. For further details, section 2.2 of Imran et al. [21] may be visited.

3.3 Need of the Hour

Concerning the proper identification of disease, accuracy of the system is particularly crucial since false positives and true negatives have a major impact in controlling pandemic outbreaks. For these reasons, an approach other than crowd-sourced was utilized under controlled conditions to better ensure the robustness of the framework. We are using mobile app and web portal to collect data from our medical collaborators working from different parts of the world. Lastly, more than one samples are taken from every user to make sure that at least one sample is fulfilling all the requirements.

CHAPTER 4

Method

As a consequence of a mild sized dataset and the significance of performance efficacy under the public safety implications of our research, feature based machine learning techniques are employed. In this phase, we describe pre-processing methods that help in edification of the extracted features for the purpose of training our classifiers, considering particular idiosyncrasies of our audio database (e.g., longitudinal mobile customers and cross-validation). Not only we analyse special varieties of handcrafted features, but also the effects of pre-processsing techniques on these classifier. We examined classifiers inclusive of LR, Gradient Boosting, SVM, RF, k-NN and . The figure 4.1 shows our contribution to the knowledge value in the grand scheme of things. The pre-processing techniques step is of high significance with regards to the research work under consideration.

4.1 Preprocessing

In selecting the type of pre-processing techniques, its important to be mindful of the fact that the techniques should modify the audio in a roughly linear and time-invariant fashion, that is, the principle of superposition should hold true for these techniques. These techniques are relatively easy to deal with analytically as much theoretical research work is out there dealing with linear and time-invariant systems. Owing to this fact, our research work has a great potential of being carried forward and adopted well with existing machine learning cough diagnosis [21, 20]. Although a bunch of preprocesing techniques were applied, the most important ones are time stretching and pitch shifting. These two techniques complement each other in their

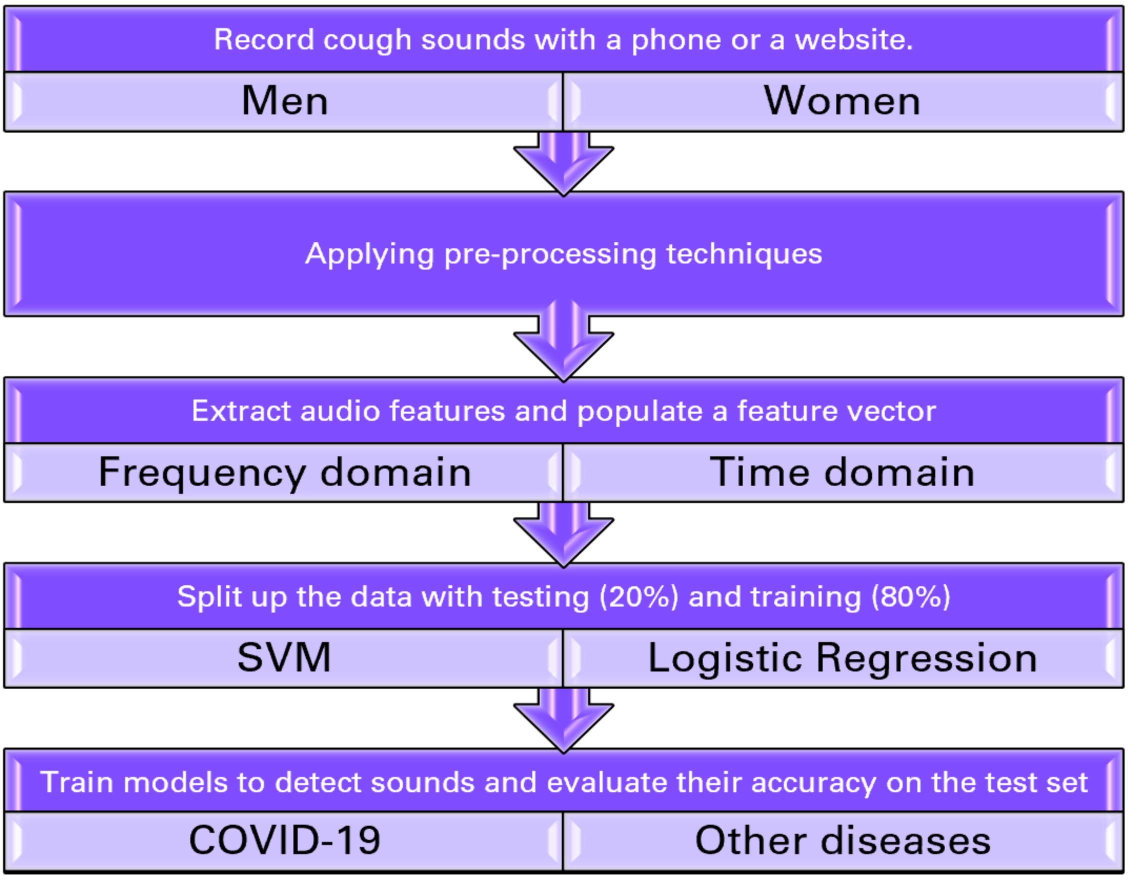

Fig. 4.1: Grand scheme of things for machine learning based diagnosis

working mechanism, that is, time stretching is the process of changing the duration of an audio signal without affecting its pitch. Pitch shifting is the opposite: the process of changing the pitch without affecting the speed. We used "librosa" library in python to produce these effects. There are four different formats of analysis that are being carried out namely: time shifting only, pitch shifting only, first time stretch then pitch and first pitch shift then time stretch. The details of these shall be provided in the next chapter.

Experimental results have shown that the order of the two techniques, taking the two pre-processing techniques in any order, have little to no effect on accuracy provided the appropriate values for shifting and stretching are known.

4.2 Feature Extraction

The raw audio waveform recorded is having a sampling rate of 44100 Hz. Librosa was used as our sound processing python library. Features were extracted based on the principles of signal processing developed by the librosa library. These features, after the application of statistical methods, are then engineered (combined) together in the form of a vector. This feature vector representing and encapsulating crucial properties, traits and other characteristics of the audio sample is referred to as a handcrafted feature for the sake of this research work. Several handcrafted features are extracted from the resampled audio. The features are extracted at the segment and frame level, covering structural-based, frequency, temporal and statistical attributes. A frame is a chunk (subset) of the whole sound data contained in a segment while a segment is the complete instance of one sound recording. Research shows that frequency-based, structural, statistical and temporal attributes can be used to differentiate one type of audio from the other [20]. The following features take all such attributes into consideration. A complete list is provided below:

Duration : The total time of the recording after trim leading and trailing silence.

Onset : The total number of pitch onsets (pseudo syllables) is calculated from the signals, by figuring out peaks from an onset strength envelope, which is calculated by summing each positive first-order difference across each Mel band .

Tempo : It is the rate of beats that occur at regular intervals temporally.

Period : The main frequency of the envelope of the signal.

RMS Energy : The power of the signal is the root-mean-square of the magnitude of a short-time Fourier transform.

Spectral Centroid : The average (centroid) extracted for one frame of the magnitude spectrogram.

Roll-off Frequency : The center frequency such that at least 85% of the energy of the spectrum in this frame for a spectrogram bin is found in this bin.

Zero Crossing : The rate at which sign-changes in a signal.

MFCC : On a nonlinear Mel scale, mel-Frequency Cepstral Coefficients are calculated from the power spectrum that is based on a linear cosine transform for the log power spectrum. We use the first 13 components.

-MFCC : The temporal differential (delta) of the MFCC.

2-MFCC : The differential of the delta of the MFCC (acceleration coefficients) .

For the rest of the chapters, handcrafted features are the feature vectors when the above mentioned temporal and latent features were applied with the 11 statistical approaches namely: mean, median, root-mean-square, maximum, minimum, 1st and 3rd quartile, interquartile range, standard deviation, skewness, and kurtosis. The total length of one handcrafted feature is 748. Given the high dimensionality of the features, we cannot present all distributions. Therefore, we focus only on the mean statistical feature of each feature family (e.g., Centroid is the values of Centroids per frame averaged together for one sample) for illustration purposes although the handcrafted features take into consideration all the aforementioned statistical methods. The box plots in 4.2 show that coughs for the four different classes.

Only for the sake of comprehension of box plots, it can be seen that for the case of silence removal, pertussis has the most number of outliers in the box plot. The box

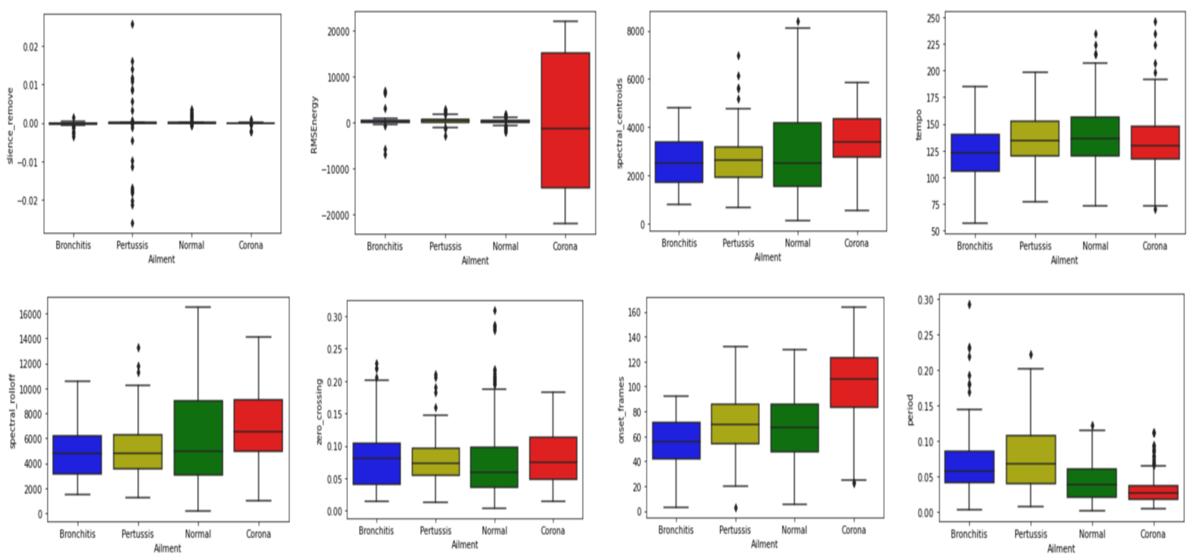

Fig. 4.2: Comparison of features extracted from samples of 4 categories

plots for tempo are similar for the four classes with almost matching ranges. The values are high and so is the mean for the box plot of COVID-19 samples for onset frame feature. High value of RMS can be seen in stark contrast for COVID-19 then the rest. The period of COVID-19 samples is less in magnitude and sparsity as well from the others. Other trends of spectral centroid, spectral roll off and zero-crossing can be seen for the four categories. This may also suggest that coughs are useful sounds for classifying users as COVID-19 or non-COVID-19.

4.3 Machine Learning Classifiers

A machine learning classifier uses training dataset to comprehend the association between the lables of a particular class and the data itself. As far as the scope of this study is concerned, the audio data of cough samples, after the grouped K-fold split, is used for training.

4.3.1 Types

Two types of classifier were given attention for our categorization task:

Lazy Learners

Lazy learners store the data for training and wait until a testing data sample appears. When it does, categorization is carried out based on the most pertinent data in the stored training data. Lazy learners have less training time but more time in predicting in stark contrast to eager learners. We are using k-NN and its prediction time is higher compared to other classifiers under consideration. It shall be viewed quantitatively in the later chapters.

Eager Learners

Eager learners construct a model based on the training data prior to receiving the data for classification. They commit to a single hypothesis that encompasses the entire instance space. Eager learners take a long time, because of model construction, to train and less time to predict. One example of such a classifier in our research work is DT.

4.3.2 Classifiers Used

Once the feature matrix is established, the feature vectors, representing audio samples, are setup for the training and testing of the classifiers. Six different machine learning classifiers namely: NB, RF, XGBoost, k-NN, SVM and LR are used. These classifiers are analysed and compared in terms of their performance. This comparative analysis can provide a blue print to better comprehend the utility of off the shelf classifiers for cough diagnosis in particular and speech recognition in general. The split between the training and testing samples was kept in the ratio of 4:1, that is, 80% of the data was utilized for training and 20% for testing. In our research work, group based k-fold cross-validation was carried out in which the data split up is based on the users. In other words, for one iteration, the samples from the same user can't be in both testing and training. If feature vectors of a particular user are a part of training in one iteration, then they could be in the testing part for the next iteration but not both in the same iteration. This approach curtails over fitting and all such trends that can cause the classifier to come up with a high value of accuracy without getting associated with incongruous traits. The details of experimental setup shall be explained in the upcoming chapter. The overall results are averaged out after several iterations to avoid extreme conclusions in a particular iteration.

The tuning of the hyperparameters of these six classifiers were carried out as well through hit and trial to set the best results for the base case, without pre-processing. A more extensive grid search can prove to be helpful for more robust performance.

CHAPTER 5

Evaluation

We now detail our evaluation of the classification of audio samples as for the four different categories namely: Bronchitis, Pertussis, Normal (non-infectious) and COVID-19 using features described in previous section. Findings and results are discussed later in this chapter.

As far as the categorization task is concerned for our base case, the methodology described in the previous section is carried out without the addition of pre-processing techniques for the six classifiers. Based on the data collection we focus on four approaches as discussed previously by applying pre-processing techniques to multi-class classification for three clinical diseases with the fourth as normal beings. Based on this road map four different tasks are carried out. Figure 5.1 and table 5.1 show the results procured by the six machine learning classifiers that were trained and validated utilizing the hand crafted features and then these results were averaged out over 100 iteration of k-fold grouped validation.

Table 5.1: Average efficiency matrices for the classifiers, over several iterations, for time stretching

	k-NN	**Naïve Bayes**	**Logistic Regression**	**Random Forest**	**XGBoost**	**SVM**
Accuracy	92.68%	84.19%	92.8%	92.82%	94.26%	91.33%
Precision	93.07%	85%	93.05%	93.45%	94.48%	91.88%
Recall	92.68%	84.19%	92.8%	92.82%	94.26%	91.26%
F1 score	92.73%	84.22%	92.76%	92.74%	94.21%	91.45%
Mean error rate	26.02%	40.28%	21.68%	17.8%	16.18%	30.03%

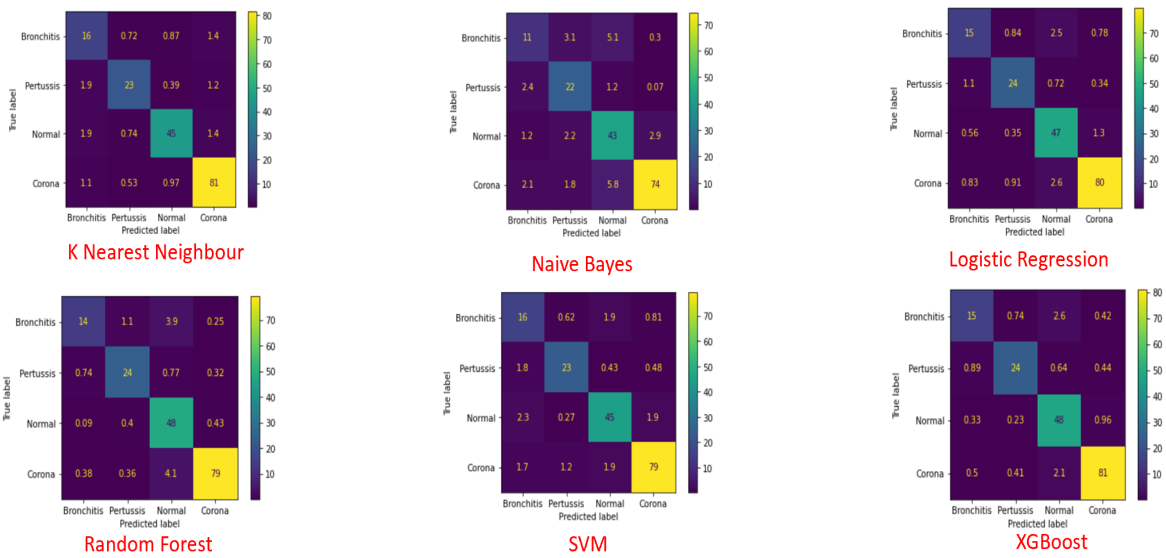

Fig. 5.1: Confusion matrices for six classifiers without pre-processing

5.1 Preprocessing Approaches

The following approaches were taken into consideration with regards to the pre-processing techniques:

5.1.1 Time Stretching only

In this approach, only time stretching is applied to the captured cough audio signals as a preprocessing technique.

Explanation As described previously, the cough audios are gathered and after conditioning, these sounds are applied to the pre-processing technique of time stretching only. The values of stretch are different depending on the type of classifiers to procure the best results at a sampling rate of 44100 Hz. The stretching or compressing of a cough audio can alter the state of its features extracted, especially those that are temporal in nature. Sounds possessing differences in temporal traits show contrasting values for the same feature and so time stretch plays a role in improving the overall classification process. The exact amount of stretching required to improve the performance in a noticeable contrast to its absence was done via trial and error methodology, that is to plug in a range of values for the stretching factor to see which one gives the best performance. Userbased cross-validation was applied which shall be discussed in the next section. Figure 5.2 shows the confusion matrices procured when samples are subjected to time-stretching for the six machine learning classifiers under consideration.

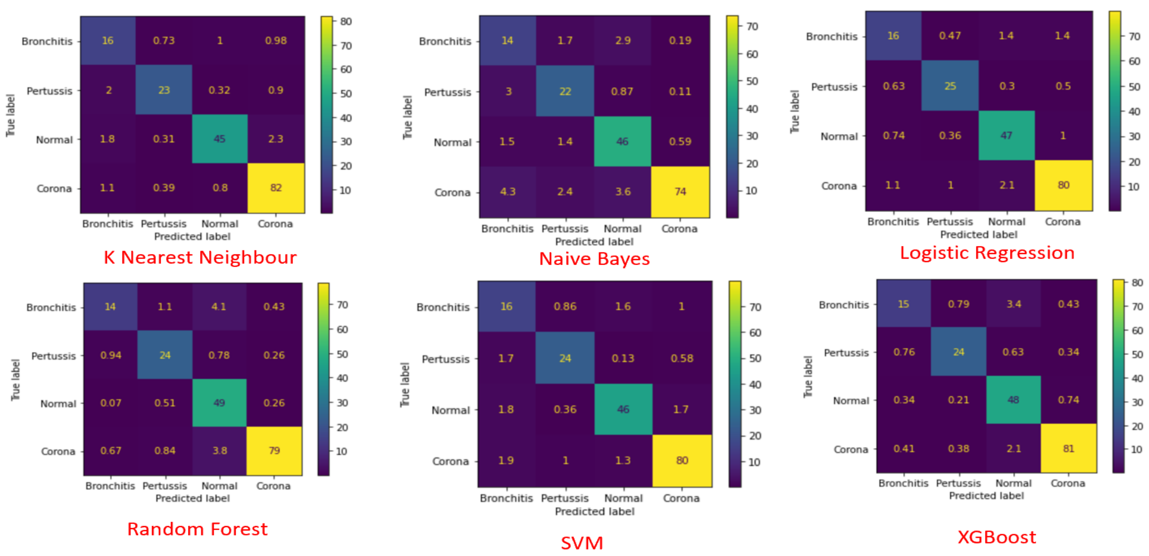

Fig. 5.2: Confusion matrices for six classifiers subjected to time stretch

Note that the confusion matrices are not normalized and are averaged out over several iterations of the evaluation process. For now this approach is carried out but theoretical analysis can be employed to come up with a formulation and statistical framework to determine best efficacy parameters. Table 5.2 shows the performance metrices for the aforementioned approach

5.1.2 Pitch Shifting only

In this approach, only pitch shift is applied to the captured cough audio signals as a preprocessing technique.

Table 5.2: Average efficiency matrices for the classifiers, over several iterations, for time stretching

	k-NN	**Naïve Bayes**	**Logistic Regression**	**Random Forest**	**XGBoost**	**SVM**
Accuracy	92.95%	87.35%	93.83%	92.34%	94.07%	92.18%
Precision	93.39%	88.92%	94.03%	92.99%	94.35%	92.59%
Recall	92.95%	87.35%	93.83%	92.34%	94.07%	92.18%
F1 score	93%	87.74%	93.81%	92.59%	93.99%	92.27%
Mean error rate	23.14%	44.34%	23.34%	21.42%	17.22%	29.42%

Explanation As previously discussed, the cough sounds are recorded and after trimming, these audios are applied to the pre-processing technique of pitch shift only. The values of shift vary depending on the type of classifier used. The shift in the pitch of a cough audio can alter the state of the features extracted, especially those that are strongly tied with the frequency response of a sound. Audios possessing differences in the properties of frequency response show contrasting values for the same feature and so pitch shift plays a part in improving the classification process. To obtain the best performance at a sampling rate of 44100 Hz in a noticeable contrast to its absence, shifting factors were figured out through trial and error methodology, that is to put in a range of different values for the pitch shift and observe the one's with best results. Userbased cross-validation was applied which shall be discussed in the next section. Figure 5.3 shows the confusion matrices procured when samples are subjected to time-stretching for the six machine learning classifiers under consideration.

Its important to note that the confusion matrices as shown in figure 5.3 are not normalized, but rather averaged out over 100 iterations of the validation process. The userbased validation process was carried out to nullify any over-fitting trends. For now this approach is carried out but theoretical analysis can be employed to come up with a formulation and statistical framework to determine best efficacy parameters, in our case the amount of pitch shift required for ideal results. Table 5.3 shows the performance metrics for the aforementioned approach.

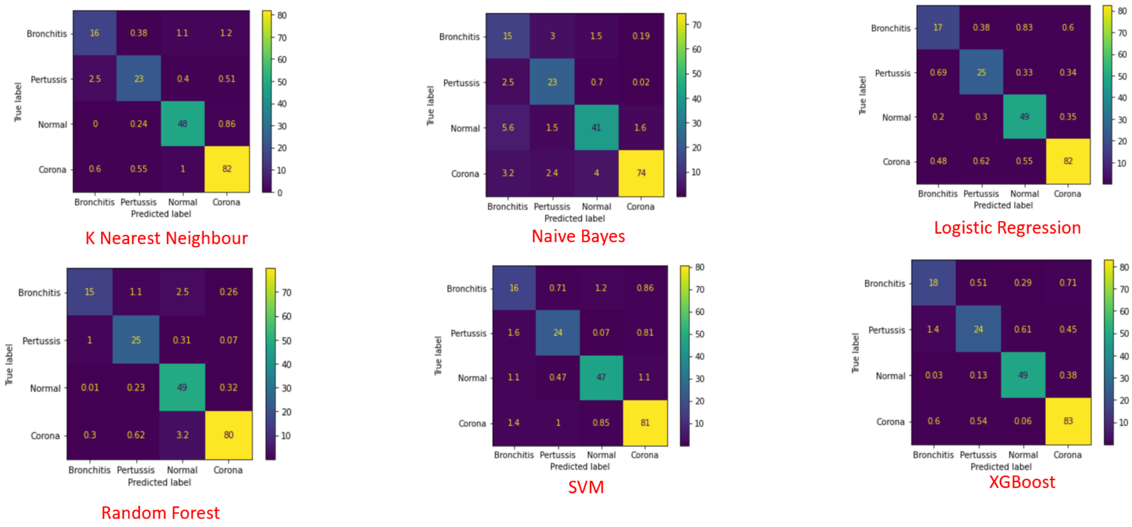

Fig. 5.3: Confusion matrices for six classifiers subjected to pitch shift

Table 5.3: Average efficiency matrices for the classifiers, over several iterations, for pitch shift

	k-NN	**Naïve Bayes**	**Logistic Regression**	**Random Forest**	**XGBoost**	**SVM**
Accuracy	94.72%	85.36%	96.83%	94.43%	96.82%	93.65%
Precision	94.93%	87.52%	96.96%	94.85%	96.92%	93.92%
Recall	94.72%	85.36%	96.83%	94.43%	96.82%	93.65%
F1 score	94.7%	85.99%	96.81%	94.4%	96.8%	93.7%
Mean error rate	17.09%	45.92%	11.35%	13.45%	11.25%	23.68%

5.1.3 Pitch Shift then Time Stretch

In this case, the audio sample is initially subjected to pitch shifting pre-processing technique and then applied to a time stretching approach in cascade.

Explanation As stated previously, the cough sounds are collected and after editing, these files are applied to the pre-processing techniques of pitch shifting first and then time stretching. These techniques help alter the audio's frequency components and temporal outlook that can divulge key aspects unique to one category of diseases over the others. The extracted features are then better off accentuating these aspects, assisting the classifier to make more accurate predictions. The values of stretch and shift are different depending on the type of classifier used to obtain the best performance at a sampling rate of 44100 Hz. It should be noted that the stretching and shifting factors for cascaded setup are different for the case of each technique applied individually. The exact amounts of pitch shift and time stretch required to enhance the performance in a noticeable contract to their absence was done through trial and error methodology. The way it works is by plugging in a bunch of values for the time stretch and pitch shift and monitoring those that come up with best results. Userbased cross-validation was applied which shall be discussed in the next section. Figure 5.4 shows the confusion matrices procured when samples are subjected to time stretch after pitch shift for the six machine learning classifiers under consideration.

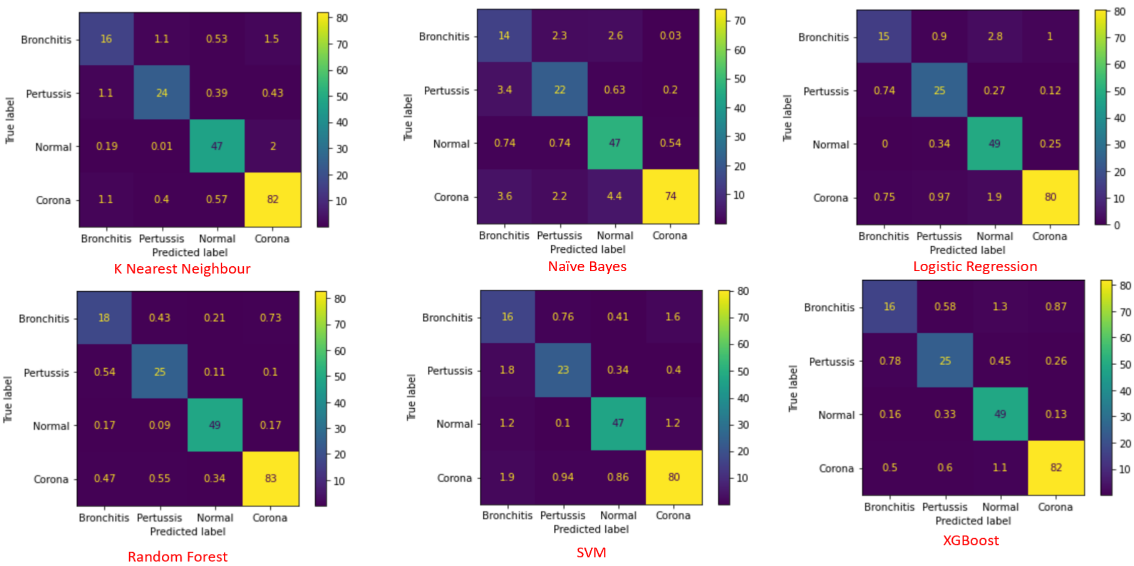

Fig. 5.4: Confusion matrices for six classifiers subjected to cascaded pitch shift followed by time stretch

Note that the confusion matrices are not normalized and are averaged out over several iterations of the evaluation process. For now this approach is carried out but theoretical analysis can be employed to come up with a formulation and statistical framework to determine best efficacy parameters. Table 5.4 shows the performance metrices for the aforementioned approach.

Table 5.4: Average efficiency matrices for the classifiers for cascaded preprocessing techniques, that is, first pitch shift then time stretching

	k-NN	**Naïve Bayes**	**Logistic Regression**	**Random Forest**	**XGBoost**	**SVM**
Accuracy	95.31%	88.24%	97.87%	94.37%	96.05%	93.37%
Precision	95.6%	90.1%	97.93%	94.66%	96.19%	93.74%
Recall	95.31%	88.24%	97.87%	94.37%	96.05%	93.37%
F1 score	95.35%	88.74%	97.87%	94.27%	96.01%	93.45%
Mean error rate	16.32%	45.23%	8.6%	19.98%	14.02%	26.62%

5.1.4 Time Stretch then Pitch Shift

In this approach, each audio sample is initially subjected to time stretching and then applied to a pitch shifting pre-processing technique in cascade.

Explanation As previously stated, the cough audios are recorded and once the audio files are edited, these sounds are applied to the pre-processing techniques of time stretching first and then pitch shifting. These pre-processing techniques vary the temporal characteristics and frequency components of the cough sound that can uncover key aspects unique to one class of ailments over the rest. The extracted features are then in a better position to highlight these aspects, assisting the classifier to make more accurate predictions. The values of stretching and shifting factors are different depending on the type of classifier used to procure the best performance at a sampling rate of 44100 Hz. It should be noted that the stretching and shifting factors for cascaded setup are different for the case of each technique applied individually. The exact amount of shifting required to better the performance in a noticeable contrast to its absence was done through trial and error methodology, that is to plug in a range of values for the shifting and stretching to see which give the best performance. Userbased cross-validation was applied which shall be discussed in the next section. Figure 5.5 shows the confusion matrices procured when samples are subjected to time-stretching prior to pitch shifting for the six machine learning classifiers under consideration.

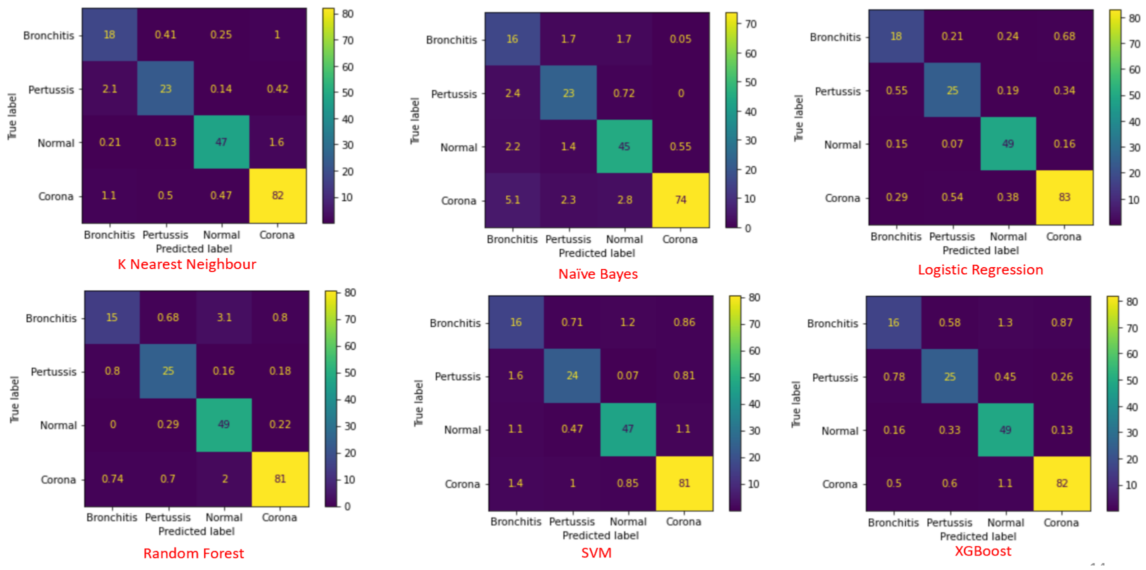

Fig. 5.5: Confusion matrices for six classifiers subjected to cascaded time stretch followed by pitch shift

Table 5.5: Average efficiency matrices for the classifiers for cascaded preprocessing techniques, that is, first time stretching then pitch shift

	k-NN	Naïve Bayes	Logistic Regression	Random Forest	XGBoost	SVM
Accuracy	94.77%	88.03%	97.81%	94.43%	96.05%	93.5%
Precision	94.86%	89.35%	97.88%	94.66%	96.18%	93.9%
Recall	94.77%	88.03%	97.81%	94.37%	96.05%	93.5%
F1 score	94.73%	88.32%	97.81%	94.27%	96.01%	93.58%
Mean error rate	19.61%	38.06%	9.29 %	13.45%	14.01%	27.49%

Note that the confusion matrices are not normalized and are averaged out over several iterations of the evaluation process. For now this approach is carried out but theoretical analysis can be employed to come up with a formulation and statistical framework to determine best efficacy parameters. Table 5.5 shows the performance metrices for the aforementioned approach.

5.2 User-based Cross Validation

We create training and test datasets from user splits that were disjoint, ensuring that the samples from the same user do not appear in both splits. Note that this does not lead to a perfectly balanced class splits. It is never easy to guarantee that a split chooses a representative test-set, so we performed a 5-fold-like cross validation using 20 different random seeds to select disjoint users (80%/20% split). In essence, this setup resembles a nested cross-validation [7]. For each iteration, the number of users are divided while keeping a track of the users included in this experimentation. We conduct extensive experiments over 100 iterations, that is, the split was 4:1 train:test to satisfy the 5-fold-like cross validation and 20 different random seed will give a total of 5*20 = 100 iterations. We selected several standard evaluation metrics such as the Accuracy, Precision, mean error rate, F1 score and Recall. In the following section we report the performance of our three tasks.

CHAPTER 6

Discussion

In this research work, hand-crafted features were developed to establish a pipeline based on cough audio samples. 6 classifiers were used and each had its own benchmark performance. The two pre-processing techniques in a varying combinations were applied that enhanced the overall accuracy. The technique of user based K-fold validation was carried out to come up with robust results. The following sections elucidate various approaches to gauge performance criteria.

6.1 Timing Diagrams

Choosing the best classifier, for a problem under consideration, is an important aspect of developing a classification pipeline. The selection can be made by different point of views. Generally, the obtained classification performance is the most important consideration. Howbeit, the No-Free-Lunch theorem [43] lets us know that there is no algorithm that can be considered the best, unanimously. Moreover, if the expected performance criteria of several algorithms are the same, the algorithm with a lesser run-time constraint is usually preferred. Not only this but, slight plummet in performance may be tolerated if the reduction in run-time is significantly less. From a practical view point, only a limited amount of time is available for the computation of the results. This aspect is specially true for industrially deployable apps that require run time computational assistance from the edge device. Furthermore, if the user has to pay for the computation time, he might not want to start a possibly time-consuming process without any idea about its duration.

Classifiers can be differentiated based on various properties associated with their accuracy, specificity, sensitivity, error rate, training time, prediction time and memory consumption. Figures 6.1 shows prediction timing diagrams for six different classifiers .

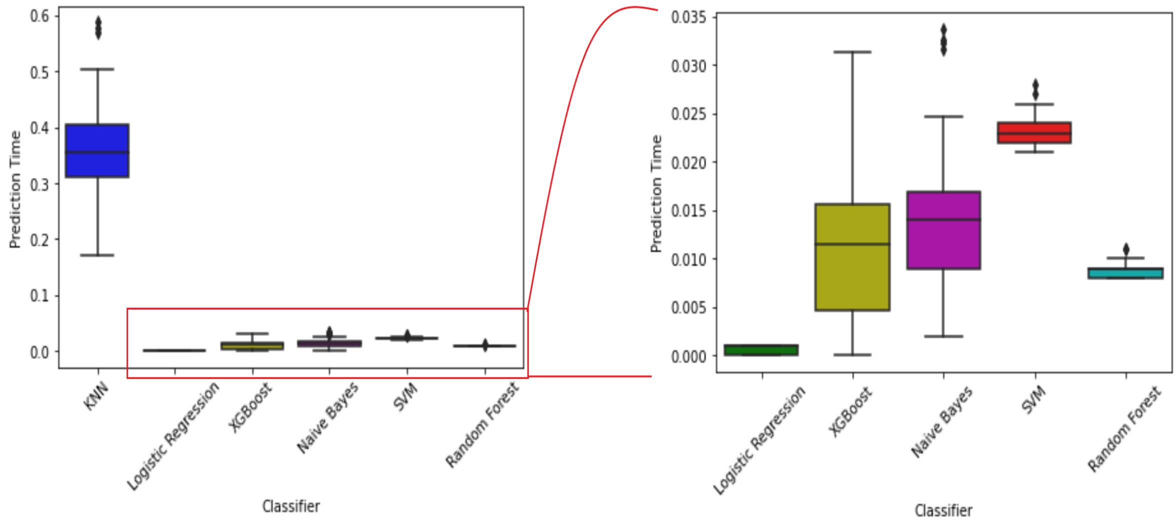

Fig. 6.1: Prediction time for the six different classifiers in seconds

Usually, an algorithm is elucidated by a general statement regarding its complexity. For example, MLP are expected to have a relatively high training time instead of a k-NN approach. Nonetheless, the actual run-time mostly depends on the dataset and the exact parameters of the algorithm. Furthermore, categorical time approximations such as "low" or "high" do not provide the users with the same amount of insights like actual time values leveraging real units. For example, a "high" run-time may mean "several hours" to "several days" to "several weeks" or even longer. Such nominal values are only useful for comparing multiple classifiers. In noticeable contrast to this aspect, more precision is associated while comparing the time elapsed by several classifiers when real numbers i.e. seconds, milliseconds, nanoseconds etc are involved. Additionally, actual time values and units make the approximation much more fruitful for the users. Theoretically, the computational complexity is also known for most algorithms. Since constant terms are ignored in computational complexity theory, the practical usefulness of these indications is limited as well. A method for determining the run-time of a classification algorithm was discussed in [44]. The figure 6.2 shows the comparison of training times of six different classifiers utilized in our study.

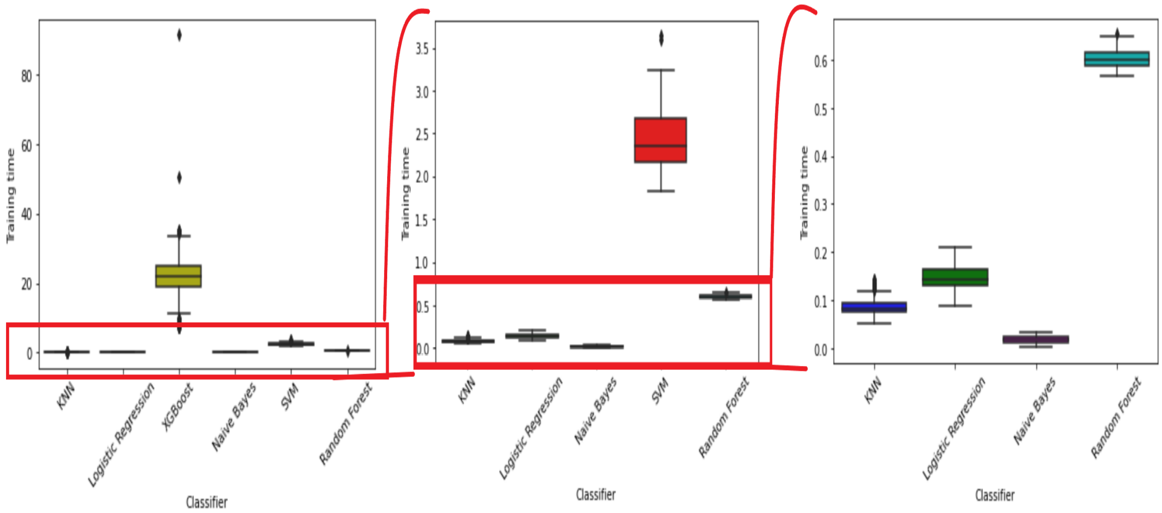

Fig. 6.2: Training time for the six different classifiers in seconds

Albeit their performances are much better than their lighter counterparts, it can be seen that the bulkier the classifier (those classifiers that consume more resources in terms of processing power and memory), the more the time it takes to predict and train. Figures 6.1 and 6.2 give a quantitative proof to this statement. This is where the real utility of preprocessing kicks in since the preprocessing techniques improve the accuracies in general, so then, some light weight classifiers can out perform their bulkier counterparts after the application of these techniques.

6.2 Results

This section shows the accuracy box plots for various approaches used in previous sections and what results and know-how we can derive from that.

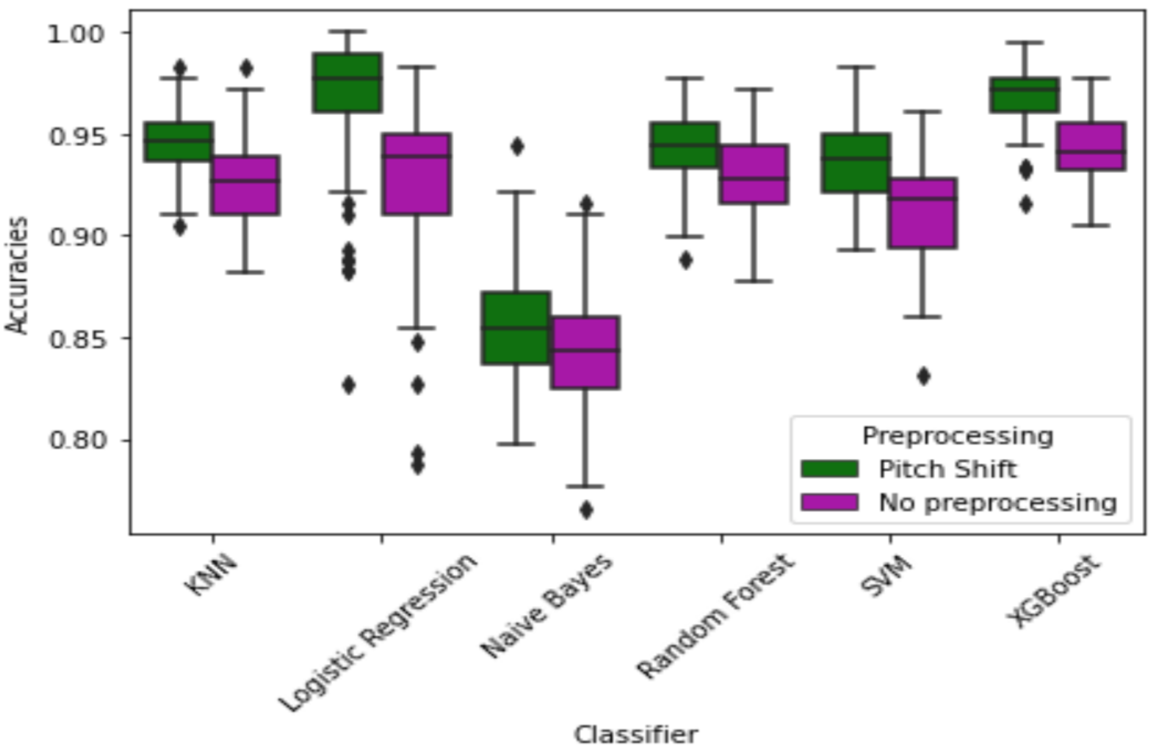

Fig. 6.3: Accuracy box plots for pitch shift

6.2.1 Performance Enhancement

It is evident for the figures 6.3 and 6.4 that preprocessing techniques increase the overall accuracies. The accuracy in these box plots are given in probabilities. This is due to real environment factors such as noise, erroneous samples, filtering and compression of audios that may introduce some detrimental effects on the sounds that confuse the classifiers. When pre-processing techniques are introduced, they act to nullify such effects and help to increase the accuracy. In our case, time stretching expands or compresses the audio. So if a temporal feature say period holds distinctive property of a disease from the rest, then time stretching can help increase or decrease the period of the audios and help the classifier, for example, LR to predict in an improved fashion.

Figure 6.5 shows how the periods of the audio samples are decreased when a time stretch of 9 was applied. In other words, the audio will be played-back at a speed 9 times faster. Similarly, used cases can be proposed for pitch shift as well. It is

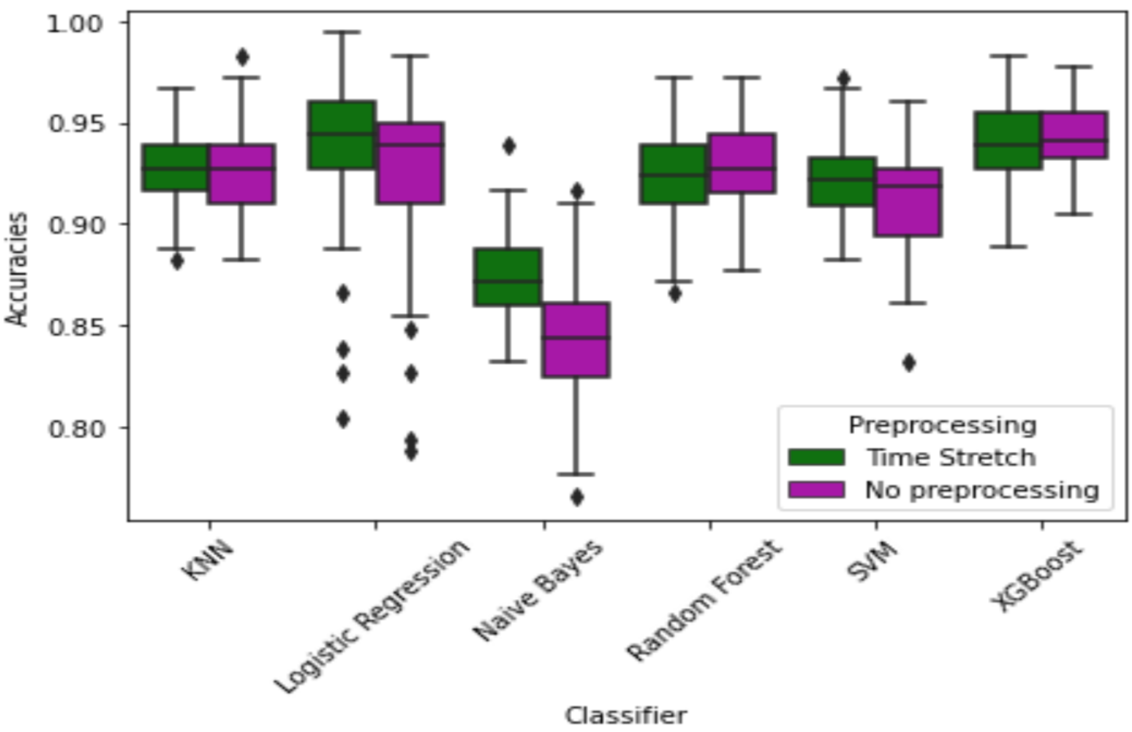

Fig. 6.4: Accuracy box plots for time stretch

because pitch shift can assist in focusing parts of the frequency response of the cough samples that hold pertinent information with regards to classification. Figure 6.6 shows the reduction in the values of spectral centroid and spectral rolloff values when pitch shift of -14 was applied. A comparison of figure 4.2, 6.5 and 6.6 shows how these techniques can vary the type of feature vector generated. These feature vectors have a better potential to assist in improved categorization of the disease. For now, only one dimensional features are shown in these box plots. Two dimensional features were also involved to generate feature vectors but that are not taken into consideration in these plots. Tables 6.2 and 6.1 will shed more light on why time shift was kept to 9 and pitch shift was equal to -14 in figures 6.5 and 6.6 respectively, especially when the classifier is LR. In fact, LR gave best results for pitch shift of -14 and time stretch of 9 in single pre-processing techniques. More on that can be observed in tables 5.2 and 5.3.

Its important to note that not every stretching and shifting factor can guarantee

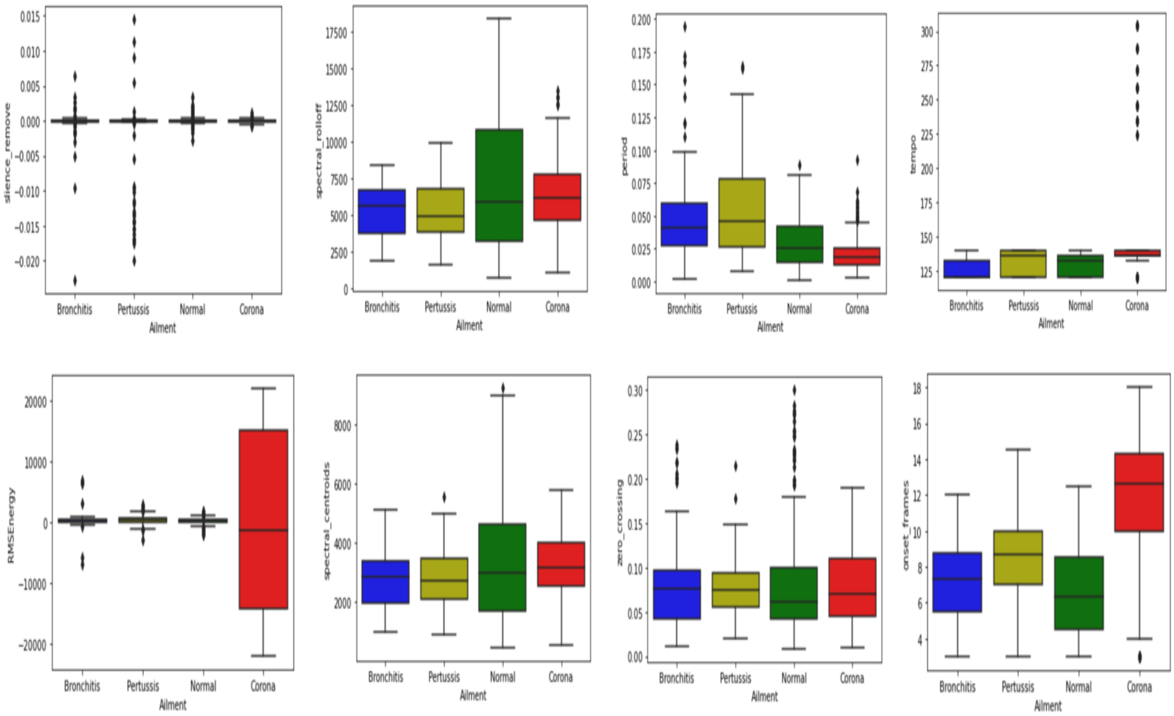

Fig. 6.5: Box plots of features when time stretch (stretching factor=9) was applied

Table 6.1: Performance metrices for various values of time stretch with Logistic Regression as the classifier

	time stretch=9	time stretch=5	time stretch=1/5	time stretch=1/9
Accuracy	93.83%	92.96%	91.81%	91.7%
Mean error rate	23.34%	26.39%	24.31%	24.94%

enhanced accuracies but only special parameters can help come up with efficency betterment. Tables 6.1 and 6.2 show efficiency of LR for a bunch of random values associated with pitch shift and time stretch. It can be seen that for value of pitch shift =14, the accuracy is worse then the base case, that is, without preprocessing. Most shifting and stretching factors reduce the accuracies. For now, trial and error methodology was carried out to come up with the appropriate values of shifting and

Table 6.2: Performance metrices for various values of pitch shift with Logistic Regression as the classifier

	pitch shift=14	pitch shift=2	pitch shift=-2	pitch shift=-14
Accuracy	87.41%	92.26%	94.5%	96.83%
Mean error rate	38.29%	23.88%	18.64%	11.35%

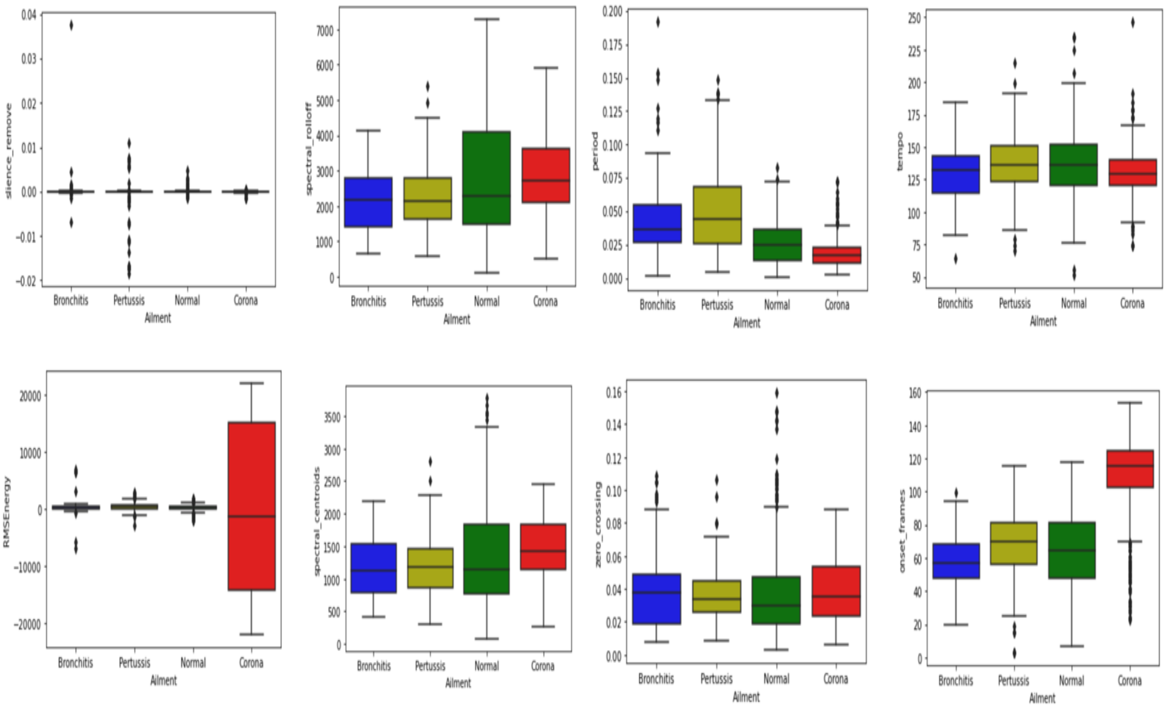

Fig. 6.6: Box plots of features when pitch shift (n-step=-14) was applied

stretching factors.

6.2.2 Cascaded Approaches are Generally Superior

The results of figures 6.7 and 6.8 illustrate that for most of the classifiers, the cascaded approach gives better results instead of a single pre-processing technique. The accuarcies in these box plots are given in probabilities. The logical reason behind is that as one technique is applied, the detrimental effects are taken care off in that dimension. The other techniques will better describe different aspects that were unaffected by the previous pre-processing technique. In our case, the time shift technique ameliorates the time dimension while pitch shift technique improves the frequency domain. By the same token, the more the techniques, the greater the potential to gain better accuracies.

An important thing that needs to be taken into consideration is that the parameters for a given technique may be adjusted according to the order in which the technique

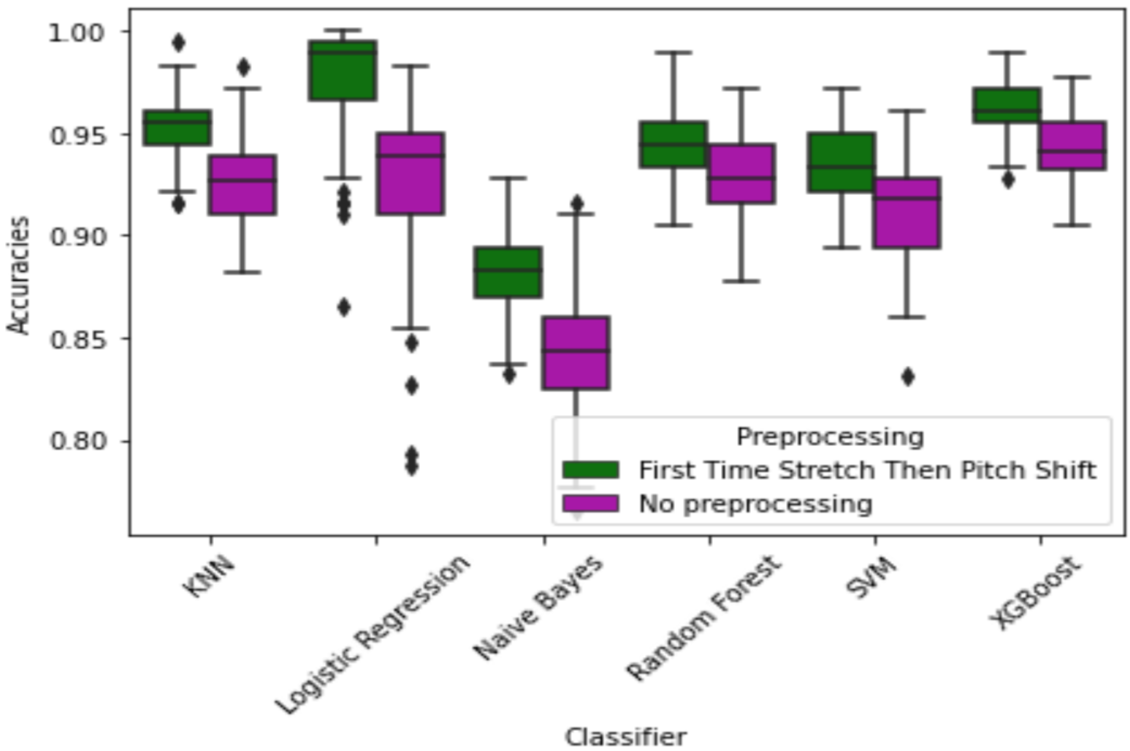

Fig. 6.7: Accuracy box plots regarding cascaded preprocessing techniques during iterations of validation procedure

was applied. For example, in the case of NB classifier, the best accuracy was obtained with time stretching first and then pitch shifting for half step value of -9 and stretching factor of 6. On the contrary, the best accuracy was obtained with pitch shifting first followed by time stretching for half step value of 2 and stretching factor of 10.

Its also important to note that, for now, the order of pre-processing techniques doesn't matter. Time stretch followed by pitch shift will have similar accuracy for pitch shift followed by time stretch given the classifier remains constant. The academic reasoning is that both theses pre-processing techniques are linear and time invariant, that is, they don't cause any non-linear change in the original signal. So, the principle of superposition holds true and cumulative effect of the two pre-processing techniques in cascade is independent of their order. This part of the discussing has several future avenues to it as shall be pointed out in the next chapter.

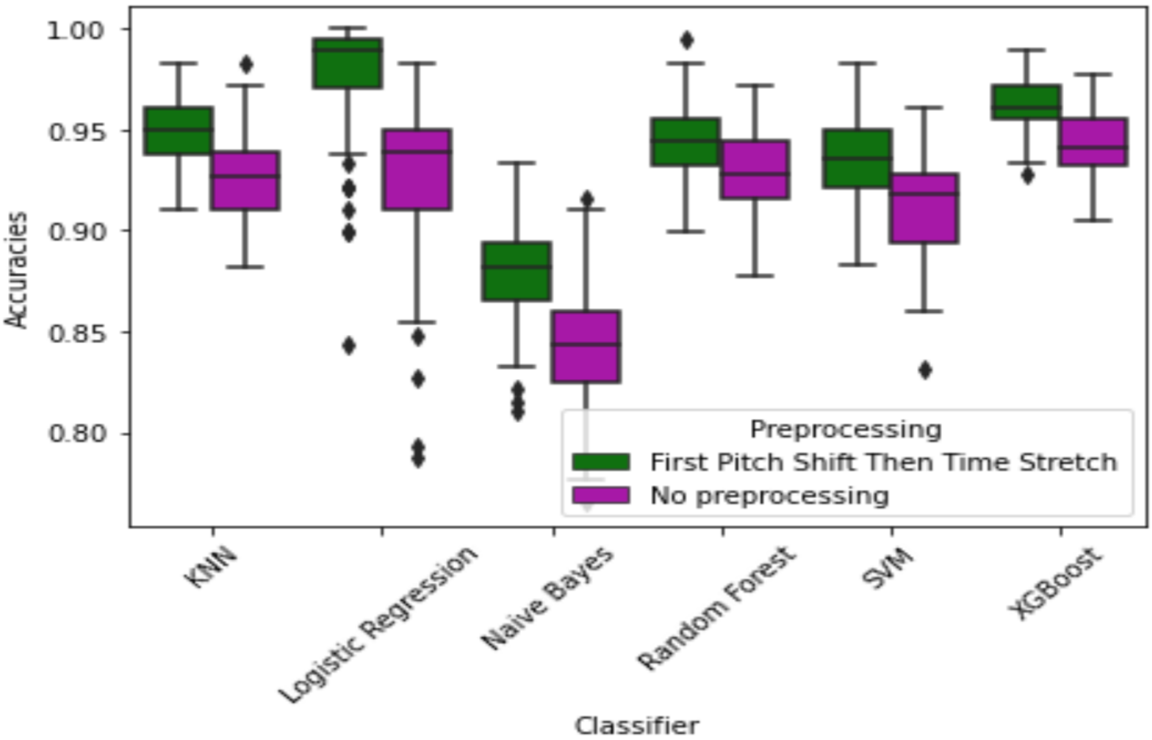

Fig. 6.8: Box plots for accuracies regarding cascaded preprocessing techniques during iterations of validation procedure

6.3 Insights

The experimentation and analysis at a physical level of the audio reveals that there are several physical factors that take part in misclassifying the samples. There can be a myriad of phenomenon going at the recording level that can confuse the classifier to misclassify. These causes can give researchers a starting point to address the problems arising at the time of recording so the software level signal processing techniques could be developed to mitigate the adverse effects of these factors that may help the classifier in its classification at a later stage. The encircled parts of the audio were determined by actually hearing the cough sound to figure out the aberrations. The following is by no means an exhaustive list of the detrimental aspects that can hamper the performance of the classifiers:

6.3.1 Background Noise

A very common reason of misclassification of cough samples is background noise. The common vindication behind this is that sound prevalent in the background can cause to tampering with the values associated with the handcrafted features, discussed in the previous chapters, that appear atypical to that of cough during the training stage. This causes misdiagnosis of one type of disease to another. Figure 6.9 shows the time and frequency domain response of COVID-19 sample mis-categorized as Pertussis. The pre-processing techniques that are adopted can help nullify these effects to better assist the classifiers during evaluation or testing phase. The encircled regions in figure 6.9 are determined by hearing the original audio.

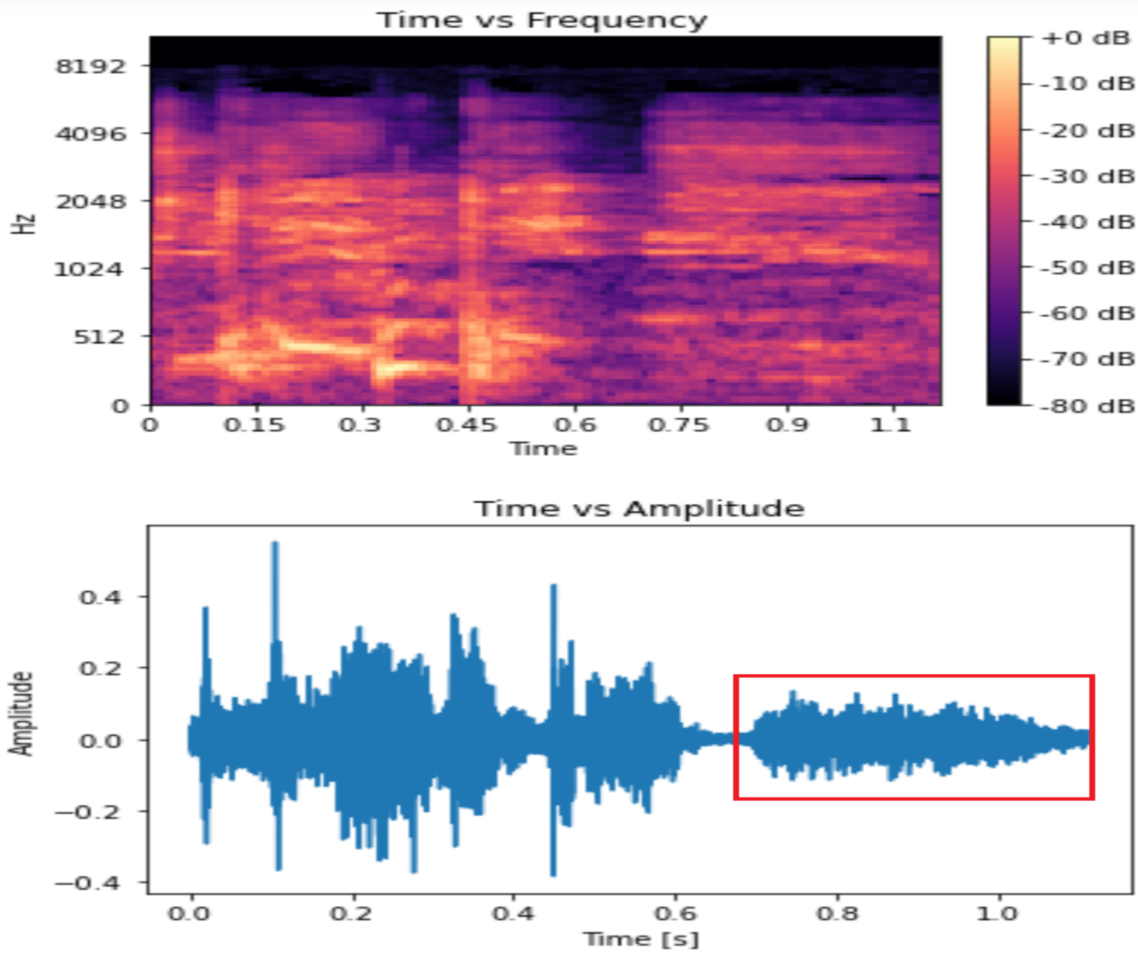

Fig. 6.9: Time domain and frequency response of a sample with background noise

6.3.2 Illtreated Samples

Audio samples that are gathered from masses, especially those that are not through the app, require treatment and standardization before they can be fit for the algorithm to be used. Generally these techniques are manual but some TD-DNN may also be used [45]. In our case, manual audio trimming was carried out for the classification purposes. More often than not, the relevant bouts of the cough get trimmed and so the information procured by the feature vector is somewhat unrepresentative from a more typical sample hence leading to a misclassification. Figure 6.10 below shows an audio cough sample of a normal individual classified as COVID-19. The encircled bout shows the fact that it was clipped at a time when its amplitude response was not fully manifested, that is, proper clipping would have helped express the full time domain cough signal. The encircled part of the amplitude plot accentuates untimely clipping of the cough bout.

6.3.3 Misleading Samples

Sometimes, slackness on part of the user can play a part in obtaining misleading samples. For reasons on part of the user namely distance, angle from the microphone, too fast audio recording, recklessness etc. can make an audio appear something that it is not. Even from the stand point of an unaided ear, the playback sound appear to be yelling, laughing, snoring etc in lieu of coughing. These erroneous samples can prove to be potential pitfalls for the classifiers. Most classifiers may not get these samples right during evaluation or training since the nature of audio appears different to what it actually should have been. Figure 6.11 shows the audio of a bronchitis patient that was falsely classified as pertussis patient by the classifier. The unaided hearing of the sample appears as if the patient is laughing.

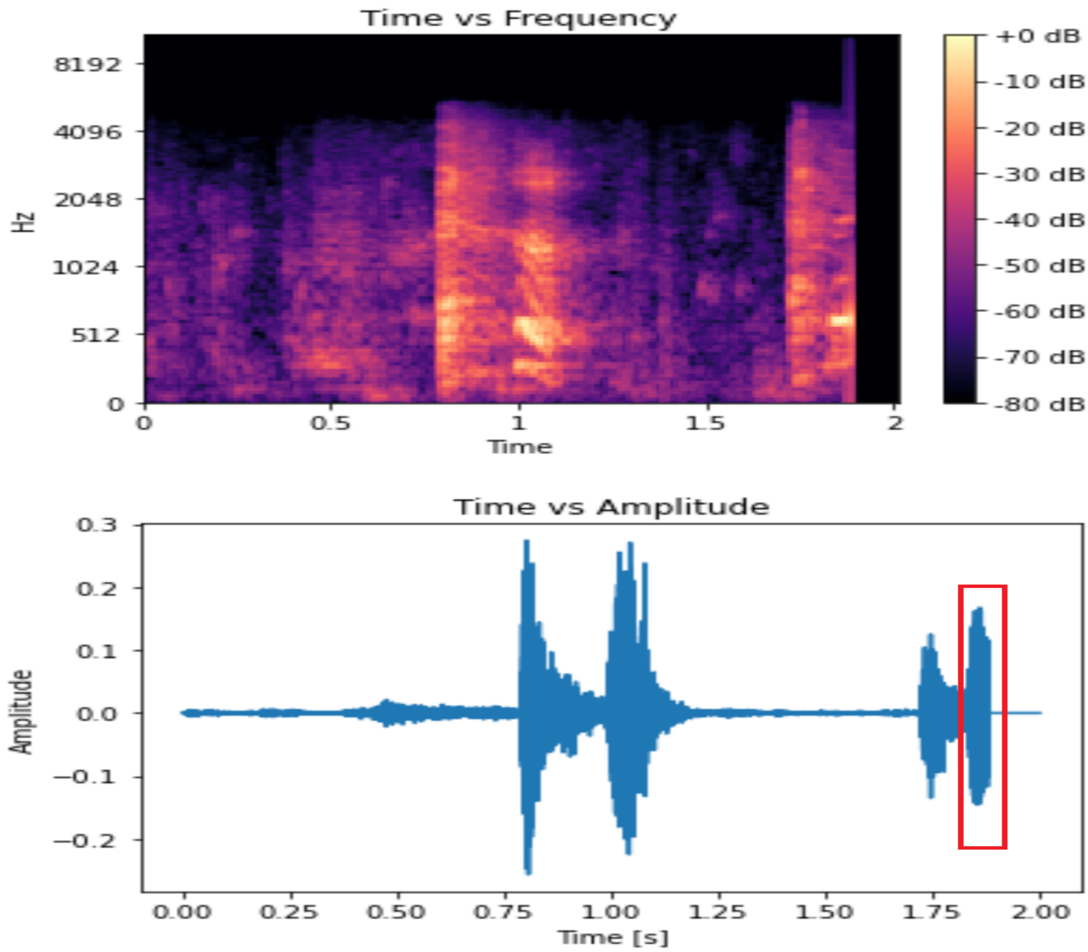

Fig. 6.10: The encircled bout shows an untimely audio clipping

6.3.4 Extraneous Sounds

Sometimes during the process of cough recording, sounds accompany recording that have nothing to do with the cough itself but are generated because of some random processes occurring in the background including door knocking, table slamming, random thud etc. The sources of these sounds are hard to pinpoint but improved controlled conditions can ameliorate the recordings. The features extracted from such an audio appear eccentric to those having regular audio cough. Improved trimming procedures can slice away the unwanted part in an audio for better performance. Figure 6.12 shows a misclassified sample of a normal (non-infectious) person's cough categorized as COVID-19 patient.

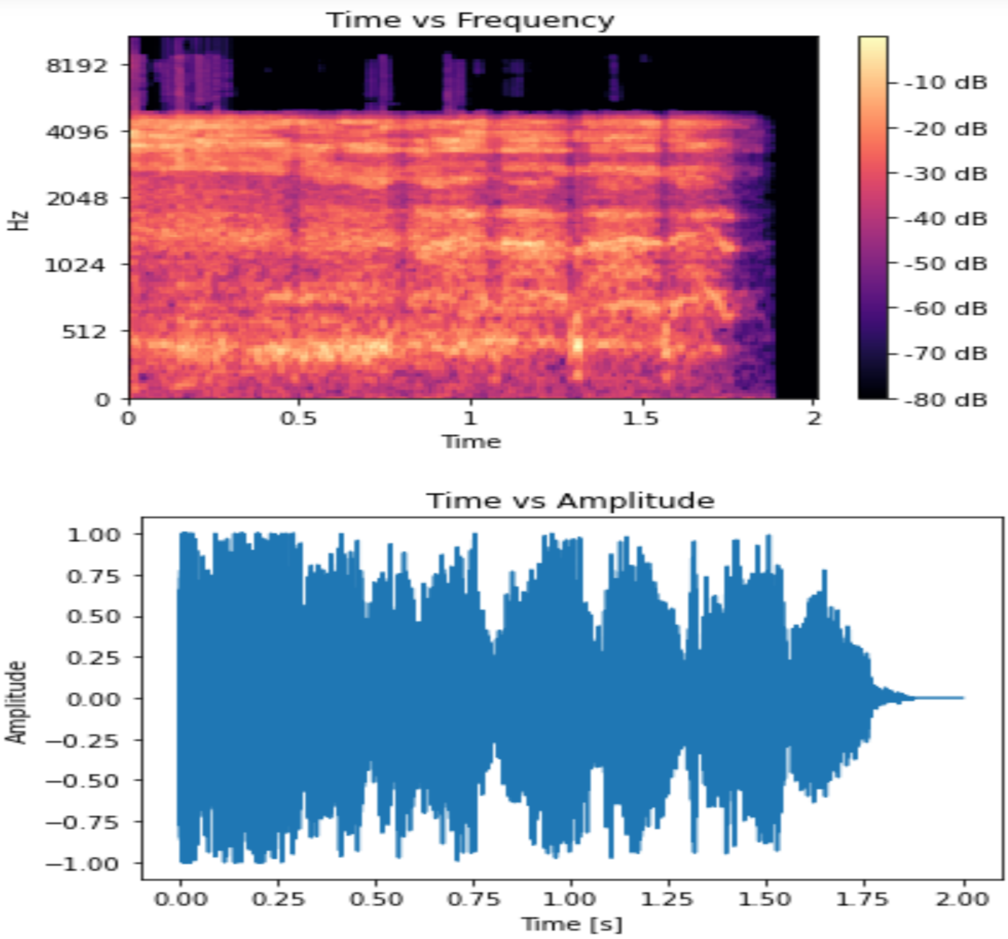

Fig. 6.11: misleading sample that doesn't resemble a typical cough

6.3.5 Unusual Cough

While collecting cough samples, the usual protocol is that users are told to cough in a non-spontaneous manner. Not all users have the ability to reproduce coughs that are typical of their sporadic counterparts. This can lead to cough samples that are more or less unusual. These cough samples appear unnatural to any ordinary listener and thus have an effect on the classifier as well. Classifier may get confused in grappling with these samples during testing or validating stages as a consequence of their aberrant nature. Although pre-processing techniques can come in handy to ameliorate the situation but better quality of samples on users' part can give the required results. Users can provide better samples via the app especially when they feel like coughing spontaneously. Figure 6.13 shows the sample of a COVID-19

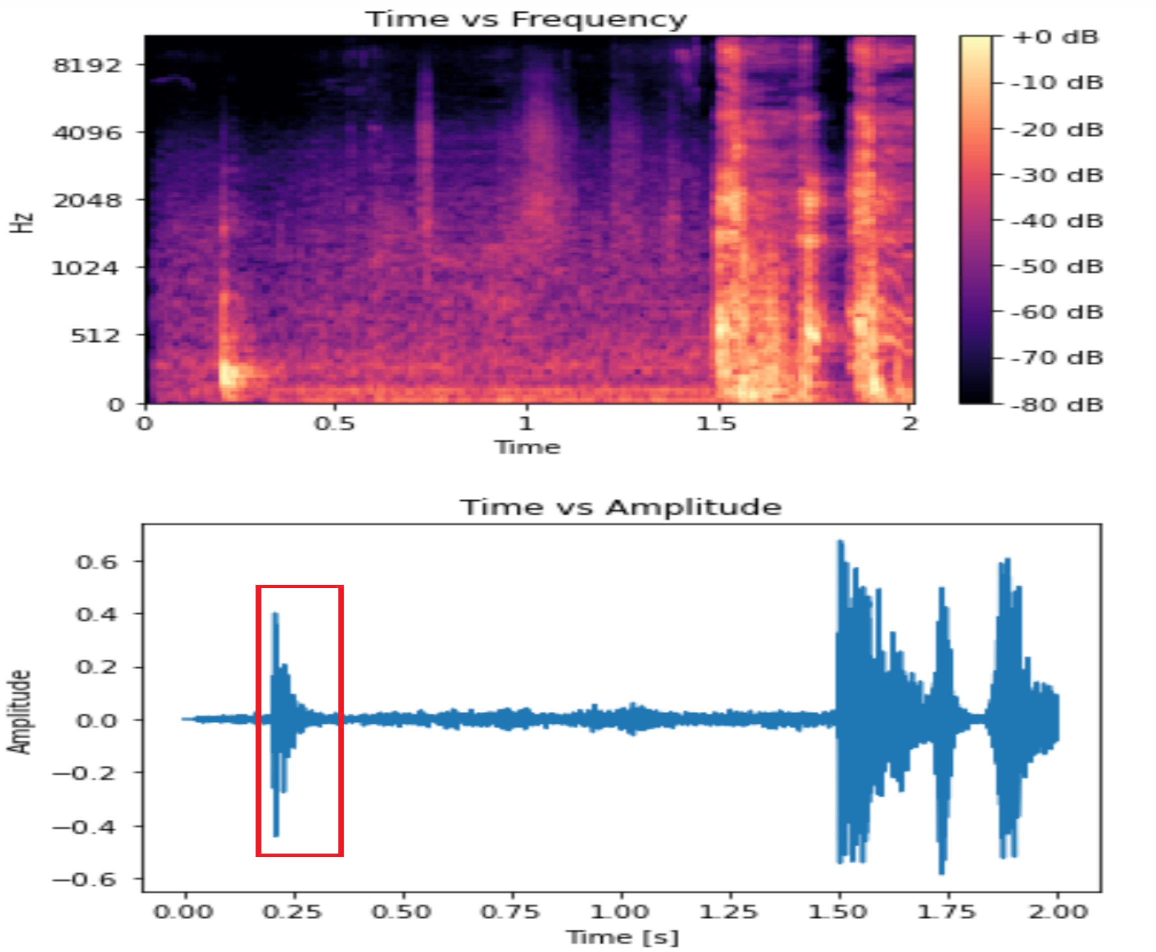

Fig. 6.12: Time domain and frequency response of a sample with extraneous audio effects patient miscategorized as a pertussis patient .

6.3.6 Too Loud for the Microphone

In the event of capturing non-spontaneous coughs from patients, some users leverage the devices that may not have the tolerance for loud voice signals. The microphone of the device may get saturated during the bouts of the cough and so useful information pertaining to cough may not get captured in the recording. Thus, leading to a misclassification on classifier's part. Popular mainstream approaches, such as normalization [46], can reduce the adverse effects caused by poor microphone recording, but the quality and quantity of data along with type of classifier has a major role to play in overcoming such pitfalls. Moreover, standardization of devices

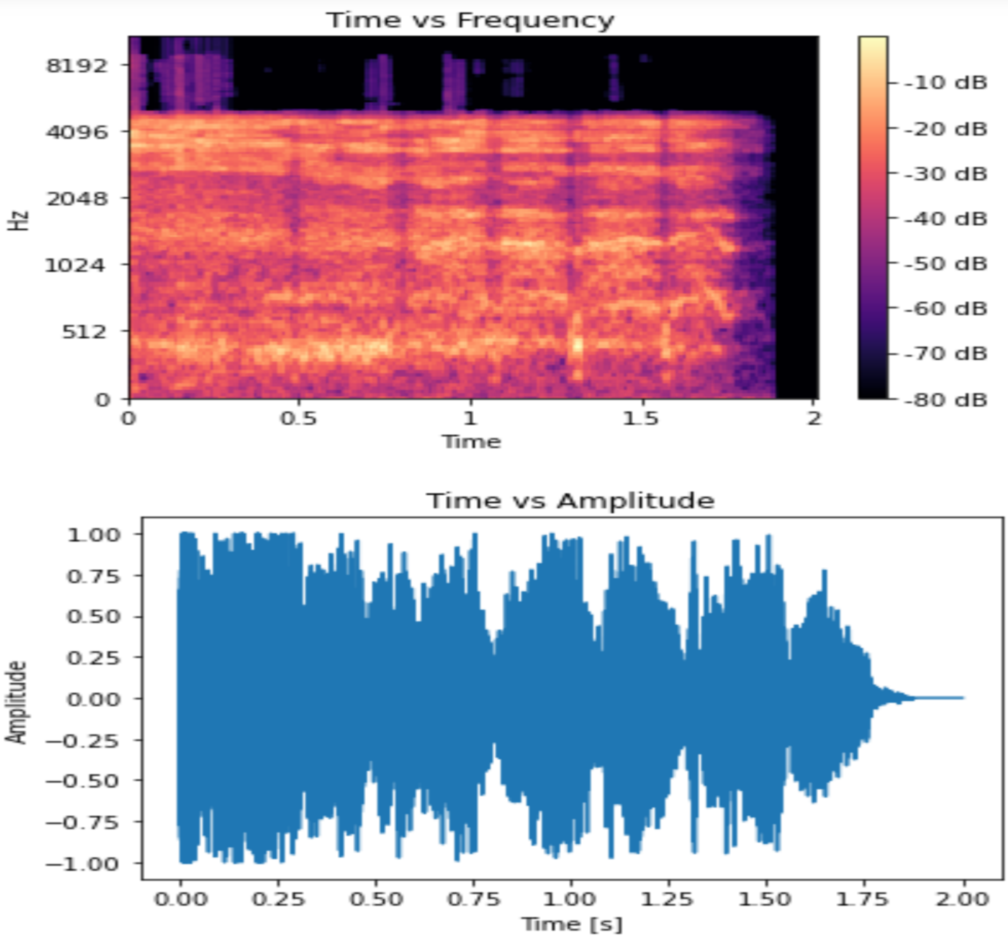

Fig. 6.13: Time domain and frequency response of a sample that isn't exactly a cough

utilized for gathering data can be of massive assistance to better the performance during clinical trials. Figure 6.14 shows the audio cough sample of a pertussis patient misclassified as normal being. The encircled regions in the amplitude response highlights magnitude saturation.

6.3.7 Too Many Bouts

Another issue associated with non-spontaneous cough recording is the number of bouts associated with one episode of cough. Users may produce many cough bouts that may not be the case for usual coughs and so these samples are anomalous to the rest of the lot. Classifiers may have a hard time classifying these samples during testing of validating stages because of the peculiar nature of these coughs. Generally,

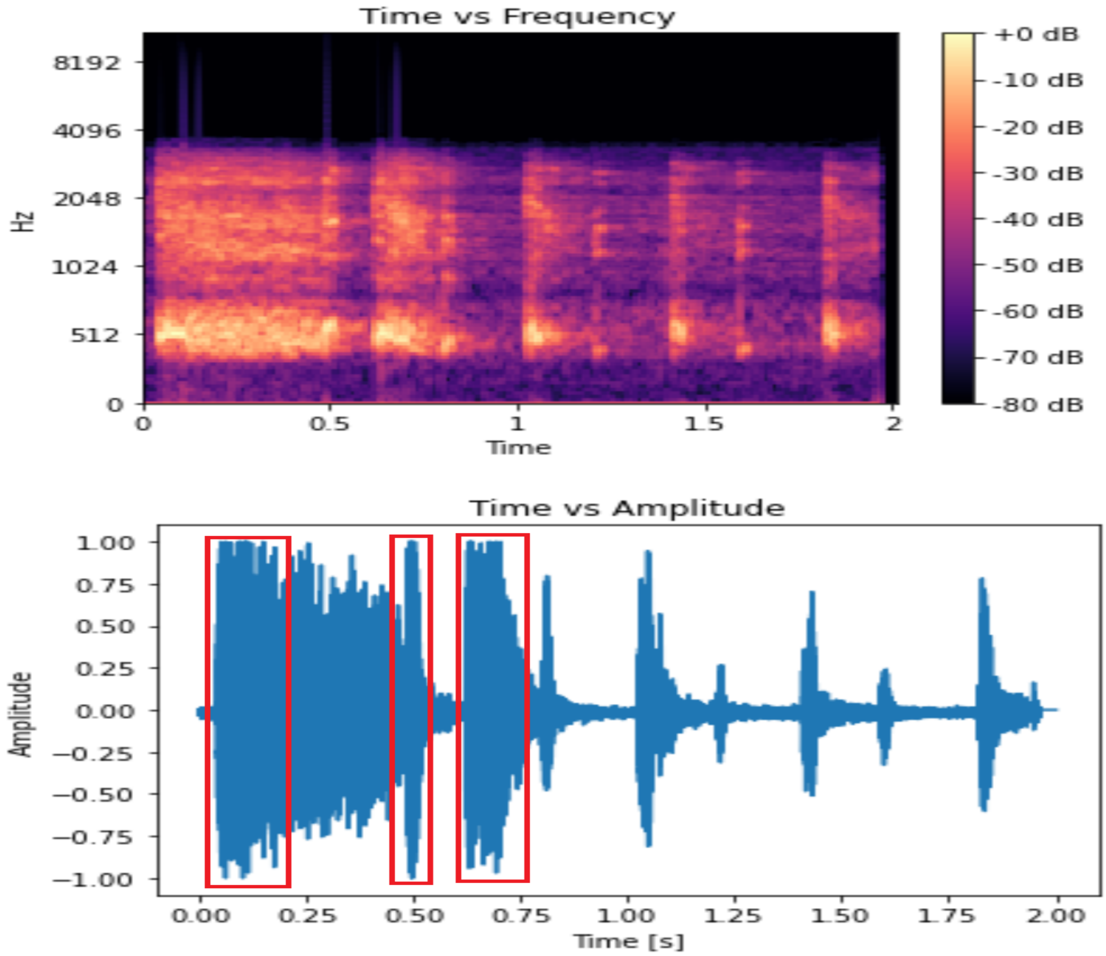

Fig. 6.14: Time domain and frequency response of a sample with amplitude saturation of device

there is a wide spectrum associated with the number of bouts during the coughing process. Normally, two to three bouts are there in a usual non-spontaneous cough but when this number goes beyond five then such cough moves to the abnormal side of the spectrum. Trimming techniques and sound conditioning can be useful to nullify these effects. Moreover, clinical approaches of initiating natural versions of cough via stimuli are highly useful to gather more consistent data. Figure 6.15 shows a cough sample of a normal being misclassified as pertussis patient.

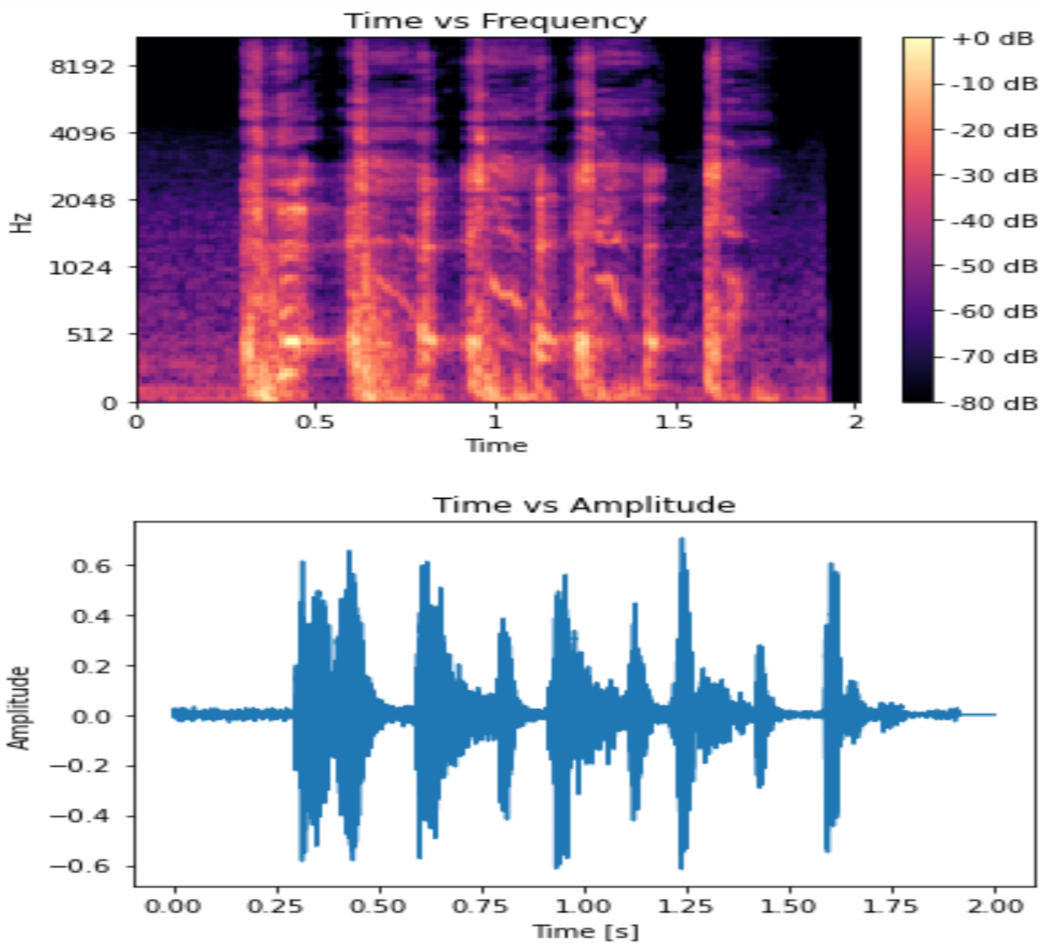

Fig. 6.15: Time domain and frequency response of a sample with too many bouts

6.3.8 White Noise

More often than not, patients are asked to record audio samples in controlled environment with as less noise as possible. Howbeit, these effects can never be abated altogether and some of these sounds make their way to the classification pipeline. Electronic devices and machines, of some other sort, operating in the background such as fan, generator etc. can produce the effect of white noise that can hamper the overall performance. There are several off the shelf signal processing techniques that can come up with the solution to solve the problem of white gaussian noise. Pre-processing techniques such as the application of a filter are typical solutions to these problems. Figure 6.16 shows white noise causing a COVID-19 audio cough sample to be misclassified as a normal (non-infectious) by NB classifier.

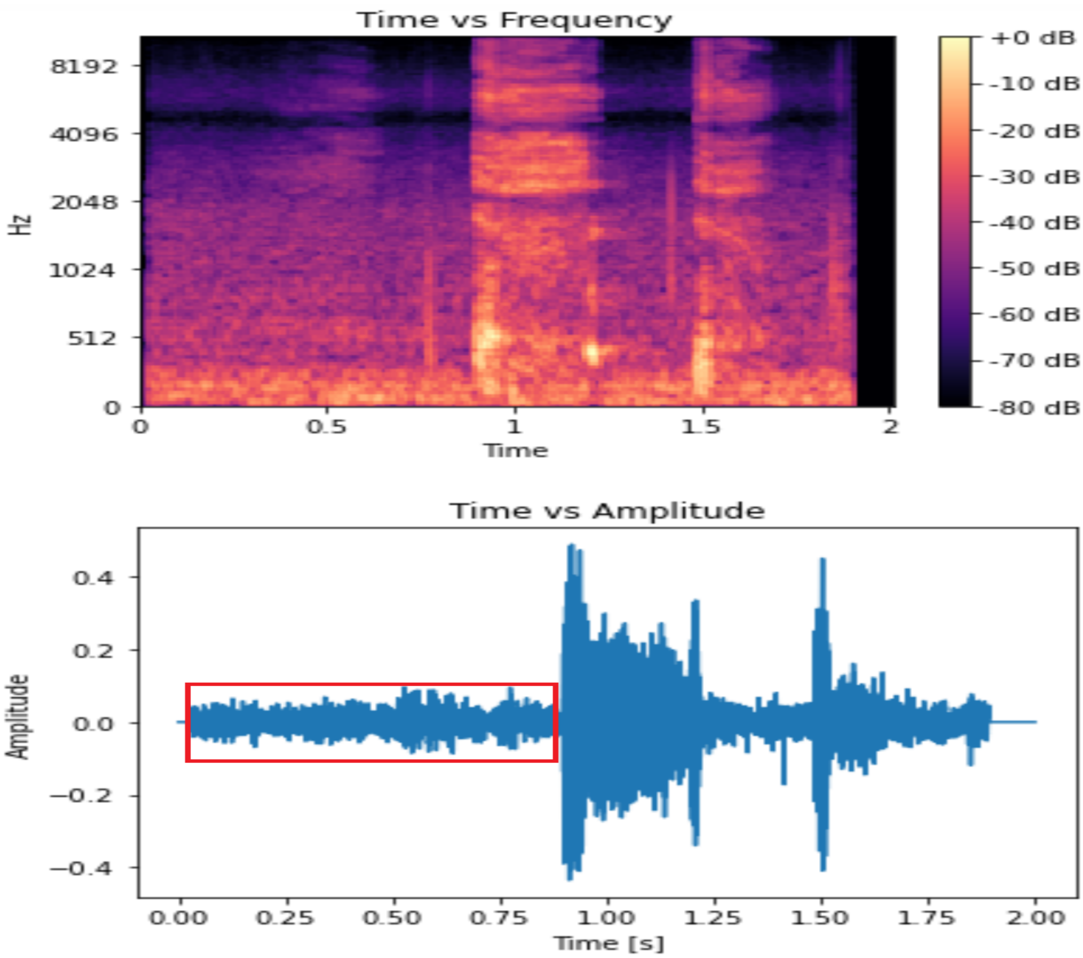

Fig. 6.16: Time domain and frequency response of a sample with white noise

6.4 Negative Findings

In the previous sections, we have seen that the techniques of time stretch and pitch shift has helped us in achieving performance unprecedented in their absence. Finding these two pre-processing techniques was never a stroke of luck but several pre-processing techniques were chanced and finally the successful ones were adopted. These are not the only two techniques out there that showed promise. Various other techniques including filtering, envelope detection, precursive effects, harmonic effects, pre-emphasis, normalization and several others. Moreover, customized coded effects using the classical techniques of signal processing can also be employed to come up with a solution to solve the problem of a pre-processing technique that improves the performance.

The scope of this research work was not limited to the aforementioned frequency and time domain techniques, but the following pre-processing were also taken into consideration and has the potential to deliver the results but for now, their performance isn't up to the mark.

6.4.1 Off the Shelf

The pre-processing techniques that were procured from the pre-built libraries belong to this section.

Filter

Once the audio is loaded, butter worth band pass filter was applied to reduce the noise and other unwanted effects in the cough audios. The big deal was in finding the appropriate values of lower cut-off and higher cut-off values that can promise better performance.

Phase Vocoder

A phase vocoder analyzes the input signal via the FFT, that decomposes the signal into its frequency components. This technique is also a part of librosa library, but it ultimately gives you the stretched signal in which time stretching has already been done. This is one reason why the performances of time stretch and phase vocoder are same.

6.4.2 Customized Development

This subsection deals with the preprocessing techniques that were developed using the principles of signal processing.

Table 6.3: Performance metrices for promising pre-processing techniques with Logistic Regression as the classifier

	Filter	Phase Vocoder	Imitate Filter	Envelope Detection	Thresholding
Accuracy	91.31%	93.87%	93.11%	85.81%	90.82%
Mean error rate	24.08%	23.58%	22.21%	39.58%	28.63%

Imitate Filter

A windowing mechanism of zeroing out the higher frequency components manually (by multiplying frequency response with a masking sequence of zeros and ones). The performance showed improvements with a few classifiers, but the accuracies dropped in other cases as well.

Envelope Detection

In this approach, the signal used for the machine learning pipeline ahead was not the signal itself but its envelope. The magnitude response was calculated by taking the absolute value of Hilbert transform. The accuracy wasn't as expected.

Thresholding

For a frequency response of a cough signal, only those frequencies were kept that had magnitudes greater than a certain threshold. The accuracy wasn't as expected.

Table 6.3 shows the accuracies and mean error rate for the aforementioned promising pre-processing techniques.

CHAPTER 7

Conclusion and Future Work

In this research work, we leveraged cough audios to diagnose three different diseases, accompanied by the fourth normal (non-infectious) category, and apply a couple of pre-processing techniques along with their combinations that can enhance the performance of a pre-existing machine learning classification pipeline. Although audio based diagnosis techniques are prevalent but to the best of our knowledge, this is the first attempt that aimed at coming up with quantitative and qualitative aspects of pre-processing techniques prior to a machine learning framework. Not only does these pre-processing techniques provide edification with regards to the efficacy parameters but also improve the performance of low resource dependent classifiers to their high resource dependent counterparts. This approach has the potential of a prototype or blue-print study for major breakthroughs in the industrial aspects of application deployment.

This study has the potential to give fruitful results for non-cough audios such as breathing, snoring etc and a combination of them can prove to be useful in better diagnostic performance. Dimensionality reduction techniques to trammel the length of feature vector can enhance the training and prediction time for the classifies. On top of that, better performance can also be expected from it. A cascade of more pre-processing techniques and their combination can also enhance the performance. A theoretical framework can be developed that can assist in recognizing the parameters of the pre-processing techniques that can further the performance in lieu of trial and error methodology. Mic based K-fold cross validation is also applied for better analysis perspective. Novel pre-processing techniques can also be established, using the principles of signal processing, that can better countermand the effects of the environment. Hyper parameter tuning of machine learning classifiers can play a vital role in the future of this study. Last but not the least, other preprocessing techniques such as normalization, filtering etc should also be utilized to improve the accuracy metrices.

Bibliography

[1] W. H. Organization *et al.*, "The top 10 causes of death. january 2017," 2017.

[2] D. Marciniuk, T. Ferkol, A. Nana, M. M. de Oca, K. Rabe, N. Billo, and H. Zar, "Respiratory diseases in the world. realities of today–opportunities for tomorrow," *African Journal of Respiratory Medicine Vol*, vol. 9, no. 1, 2014.

[3] J. Korpáš, J. Sadloňová, and M. Vrabec, "Analysis of the cough sound: an overview," *Pulmonary pharmacology*, vol. 9, no. 5-6, pp. 261–268, 1996.

[4] R. G. Loudon and L. C. Brown, "Cough frequency in patients with respiratory disease," *American Review of Respiratory Disease*, vol. 96, no. 6, pp. 1137–1143, 1967.

[5] J. Hsu, R. Stone, R. Logan-Sinclair, M. Worsdell, C. Busst, and K. Chung, "Coughing frequency in patients with persistent cough: assessment using a 24 hour ambulatory recorder," *European Respiratory Journal*, vol. 7, no. 7, pp. 1246–1253, 1994.

[6] A. Spinou and S. S. Birring, "An update on measurement and monitoring of cough: what are the important study endpoints?" *Journal of thoracic disease*, vol. 6, no. Suppl 7, p. S728, 2014.

[7] R. S. Irwin, M. H. Baumann, D. C. Bolser, L.-P. Boulet, S. S. Braman, C. E. Brightling, K. K. Brown, B. J. Canning, A. B. Chang, P. V. Dicpinigaitis *et al.*, "Diagnosis and management of cough executive summary: Accp evidence-based clinical practice guidelines," *Chest*, vol. 129, no. 1, pp. 1S–23S, 2006.

[8] S. J. Barry, A. D. Dane, A. H. Morice, and A. D. Walmsley, "The automatic recognition and counting of cough," *Cough*, vol. 2, no. 1, pp. 1–9, 2006.

[9] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," *PloS one*, vol. 12, no. 5, p. e0177926, 2017.

[10] S.-H. Li, B.-S. Lin, C.-H. Tsai, C.-T. Yang, and B.-S. Lin, "Design of wearable breathing sound monitoring system for real-time wheeze detection," *Sensors*, vol. 17, no. 1, p. 171, 2017.

[11] D. Oletic and V. Bilas, "Energy-efficient respiratory sounds sensing for personal mobile asthma monitoring," *Ieee sensors journal*, vol. 16, no. 23, pp. 8295–8303, 2016.

[12] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova, "Speech disorders in parkinson's disease: early diagnostics and effects of medication and brain stimulation," *Journal of neural transmission*, vol. 124, no. 3, pp. 303–334, 2017.

[13] B. Erdogdu Sakar, G. Serbes, and C. O. Sakar, "Analyzing the effectiveness of vocal features in early telediagnosis of parkinson's disease," *PloS one*, vol. 12, no. 8, p. e0182428, 2017.

[14] E. Maor, J. D. Sara, D. M. Orbelo, L. O. Lerman, Y. Levanon, and A. Lerman, "Voice signal characteristics are independently associated with coronary artery disease," in *Mayo Clinic Proceedings*, vol. 93, no. 7. Elsevier, 2018, pp. 840–847.

[15] D. Banerjee, K. Islam, K. Xue, G. Mei, L. Xiao, G. Zhang, R. Xu, C. Lei, S. Ji, and J. Li, "A deep transfer learning approach for improved post-traumatic stress disorder diagnosis," *Knowledge and Information Systems*, vol. 60, no. 3, pp. 1693–1724, 2019.

[16] M. Faurholt-Jepsen, J. Busk, M. Frost, M. Vinberg, E. M. Christensen, O. Winther, J. E. Bardram, and L. V. Kessing, "Voice analysis as an objective state marker in bipolar disorder," *Translational psychiatry*, vol. 6, no. 7, pp. e856–e856, 2016.

[17] K. Kosasih, U. R. Abeyratne, and V. Swarnkar, "High frequency analysis of cough sounds in pediatric patients with respiratory diseases," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 5654–5657.

[18] M. G. Crooks, Y. Hayman, A. Innes, J. Williamson, C. E. Wright, and A. H. Morice, "Objective measurement of cough frequency during copd exacerbation convalescence," *Lung*, vol. 194, no. 1, pp. 117–120, 2016.

[19] J. Knocikova, J. Korpas, M. Vrabec, and M. Javorka, "Wavelet analysis of voluntary cough sound in patients with respiratory diseases," *J Physiol Pharmacol*, vol. 59, no. Suppl 6, pp. 331–40, 2008.

[20] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, and C. Mascolo, "Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3474–3484.

[21] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, M. S. Riaz, K. Ali, C. N. John, M. I. Hussain, and M. Nabeel, "Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via an app," *Informatics in Medicine Unlocked*, vol. 20, p. 100378, 2020.

[22] R. S. Irwin, C. L. French, A. B. Chang, K. W. Altman, T. M. Adams, E. Azoulay, A. F. Barker, S. S. Birring, F. Blackhall, D. C. Bolser *et al.*, "Classification of cough as a symptom in adults and management algorithms: Chest guideline and expert panel report," *Chest*, vol. 153, no. 1, pp. 196–209, 2018.

[23] A. Morice, L. McGarvey, and I. Pavord, "Recommendations for the management of cough in adults," *Thorax*, vol. 61, no. suppl 1, pp. i1–i24, 2006.

[24] P. G. Gibson, A. B. Chang, N. J. Glasgow, P. W. Holmes, A. S. Kemp, P. Katelaris, L. I. Landau, S. Mazzone, P. Newcombe, P. Van Asperen *et al.*, "Cicada: Cough in children and adults: Diagnosis and assessment. australian cough guidelines summary statement," *Medical Journal of Australia*, vol. 192, no. 5, pp. 265–271, 2010.

[25] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao, "Automatically characterizing places with opportunistic crowdsensing using smartphones," in *Proceedings of the 2012 ACM conference on ubiquitous computing*, 2012, pp. 481–490.

[26] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, "Emotionsense: a mobile phones based adaptive platform for experimental social psychology research," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 2010, pp. 281–290.

[27] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, 2015, pp. 45–57.

[28] C. Bales, M. Nabeel, C. N. John, U. Masood, H. N. Qureshi, H. Farooq, I. Posokhova, and A. Imran, "Can machine learning be used to recognize and

diagnose coughs?" in *2020 International Conference on e-Health and Bioengineering (EHB).* IEEE, 2020, pp. 1–4.

[29] G. Deshpande and B. Schuller, "An overview on audio, signal, speech, & language processing for covid-19," *arXiv preprint arXiv:2005.08579*, 2020.

[30] Y. Huang, S. Meng, Y. Zhang, S. Wu, Y. Zhang, Y. Zhang, Y. Ye, Q. Wei, N. Zhao, J. Jiang *et al.*, "The respiratory sound features of covid-19 patients fill gaps between clinical data and screening methods," *medRxiv*, 2020.

[31] J. Han, K. Qian, M. Song, Z. Yang, Z. Ren, S. Liu, J. Liu, H. Zheng, W. Ji, T. Koike *et al.*, "An early study on intelligent analysis of speech under covid-19: Severity, sleep quality, fatigue, and anxiety," *arXiv preprint arXiv:2005.00096*, 2020.

[32] T. F. Quatieri, T. Talkar, and J. S. Palmer, "A framework for biomarkers of covid-19 based on coordination of speech-production subsystems," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 1, pp. 203–206, 2020.

[33] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S. R. Chetupalli, P. K. Ghosh, S. Ganapathy *et al.*, "Coswara–a database of breathing, cough, and voice sounds for covid-19 diagnosis," *arXiv preprint arXiv:2005.10548*, 2020.

[34] V. Swarnkar, U. Abeyratne, J. Tan, T. W. Ng, J. M. Brisbane, J. Choveaux, and P. Porter, "Stratifying asthma severity in children using cough sound analytic technology," *Journal of Asthma*, pp. 1–10, 2019.

[35] V. Swarnkar, U. R. Abeyratne, Y. A. Amrulloh, and A. Chang, "Automated algorithm for wet/dry cough sounds classification," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, 2012, pp. 3147–3150.

[36] D. Parker, J. Picone, A. Harati, S. Lu, M. H. Jenkyns, and P. M. Polgreen, "Detecting paroxysmal coughing from pertussis cases using voice recognition technology," *PloS one*, vol. 8, no. 12, 2013.

[37] A. Windmon, M. Minakshi, P. Bharti, S. Chellappan, M. Johansson, B. A. Jenkins, and P. R. Athilingam, "Tussiswatch: A smart-phone system to identify cough episodes as early symptoms of chronic obstructive pulmonary disease and

congestive heart failure," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1566–1573, 2018.

[38] B. Subirana, F. Hueto, P. Rajasekaran, J. Laguarta, S. Puig, J. Malvehy, O. Mitja, A. Trilla, C. I. Moreno, J. F. M. Valle *et al.*, "Hi sigma, do i have the coronavirus?: Call for a new artificial intelligence approach to support health care professionals dealing with the covid-19 pandemic," *arXiv preprint arXiv:2004.06510*, 2020.

[39] R. Palaniappan, K. Sundaraj, and N. U. Ahamed, "Machine learning in lung sound analysis: a systematic review," *Biocybernetics and Biomedical Engineering*, vol. 33, no. 3, pp. 129–135, 2013.

[40] B. Lei, S. A. Rahman, and I. Song, "Content-based classification of breath sound with enhanced features," *Neurocomputing*, vol. 141, pp. 139–147, 2014.

[41] P. Bokov, B. Mahut, P. Flaud, and C. Delclaux, "Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population," *Computers in biology and medicine*, vol. 70, pp. 40–50, 2016.

[42] R. Folland, E. Hines, R. Dutta, P. Boilot, and D. Morgan, "Comparison of neural network predictors in the classification of tracheal–bronchial breath sounds by respiratory auscultation," *Artificial intelligence in medicine*, vol. 31, no. 3, pp. 211–220, 2004.

[43] D. H. Wolpert, "The lack of a priori distinctions between learning algorithms," *Neural computation*, vol. 8, no. 7, pp. 1341–1390, 1996.

[44] M. Reif, F. Shafait, and A. Dengel, "Prediction of classifier training time including parameter optimization," in *Annual Conference on Artificial Intelligence.* Springer, 2011, pp. 260–271.

[45] R. V. Sharan, U. R. Abeyratne, V. R. Swarnkar, and P. Porter, "Automatic croup diagnosis using cough sound recognition," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 2, pp. 485–495, 2018.

[46] C. Infante, D. Chamberlain, R. Fletcher, Y. Thorat, and R. Kodgule, "Use of cough sounds for diagnosis and screening of pulmonary disease," in *2017 IEEE Global Humanitarian Technology Conference (GHTC).* IEEE, 2017, pp. 1–10.